

Digitaalinen tiedonhaku kuurojen yhteisön ja oralismin kohtaamisesta.

Suomen kuurojen yhteisön kokemukset oralismista 1890–1922 ja niiden
jäljittäminen korpusaineistosta

Johan Wahlsten

Pro gradu -tutkielma

Arkistoalan ja asiakirjahallinnan maisteriohjelma, kulttuurihistoria

Historian, kulttuurin ja taiteiden tutkimuksen laitos

Humanistinen tiedekunta

Turun yliopisto

Toukokuu 2026

TURUN YLIOPISTO

Pro gradu -tutkielma

Arkistoalan ja asiakirjahallinnan maisteriohjelma, kulttuurihistoria

Historian ja arkeologian tutkinto-ohjelma, kulttuurihistoria.

Johan Wahlsten

Digitaalinen tiedonhaku kuurojen yhteisön ja oralismin kohtaamisesta.

Suomen kuurojen yhteisön kokemukset oralismista 1890–1922 ja niiden jäljittäminen korpusaineistosta

Sivumäärät: Tutkielman sivumäärä: 65. Liitteiden sivumäärä: 9.

Tutkielmassa tarkastellaan oralismin historiaa digitaalisen historiankirjoituksen ja tiedonhaun menetelmien näkökulmasta. Oralismi on aate, joka kiistää viittomakielen merkityksen kuurojen elämässä ja painottaa puheen ymmärtämisen ja tuottamisen merkitystä oppimisessa. Oralismin leviäminen koulujen kautta yhteiskuntaan heikensi merkittävästi Suomen kuurojen yhteisön asemaa ja aiheutti kuuroille vääryyden kokemuksia, joita on käsitelty muun muassa 2021 valtioneuvoston selonteossa *Viitotut muistot*. Tässä tutkimuksessa tarkastellaan tietokoneavusteisin menetelmin historian tutkimuksen ja arkistotieteiden näkökulmasta vääryyksien havainnoimista sanoma- ja aikakauslehdistä muodostetussa korpusaineistossa.

Tutkimuksen keskeinen menetelmä on kollokaatioanalyysi, jota sovelletaan tiedonhaun välineenä laajaan korpusaineistoon. Lisäksi tutkimus hyödyntää aikaisempia kuurojen historiaa käsitteleviä yleisteoksia, tutkimuksia ja *Viitotut muistot*-selontekoa vääryyksien määrittelemiseksi sekä paikantamiseksi korpusaineistosta. Tutkimuksessa pohditaan digitoitujen aineistojen erityispiirteitä, mahdollisuuksia ja rajoitteita historian tutkimuksessa. Samalla tarkastellaan, millä keinoin laajasta aineistosta voidaan paikantaa tutkimuksen kannalta merkityksellistä tietoa ja mitä tietokoneavusteisuus tarkoittaa arkistojen tiedonhaun kannalta.

Tutkielma osoittaa, että tietokoneavusteiset menetelmät mahdollistavat uusien näkökulmien muodostamisen tutkimusaineistosta ja täydentävät aiemman tutkimuksen tuloksia. Nyt löydetyt havainnot oralismin aiheuttamista vääryyksistä ovat pitkälti linjassa aiemman tutkimustiedon kanssa. Samalla tutkimusprosessissa korostuu korpusaineiston kattavien metatietojen vaade, jotta aineistoa voidaan esikäsitellä riittävästi tiedonhaun toteuttamista varten.

Avainsanat: Pro gradu -tutkielmat, kuurojen yhteisö, historian tutkimus, digitaalinen historian tutkimus, digitaaliset ihmistieteet, digitaaliset tutkimusmenetelmät, oralismi, sähköiset arkistot.

Sisällysluettelo

1	Johdanto	4
1.1	Tutkimuskysymys	4
1.2	Tutkimusaiheen taustaa ja metodien viitoittamalla tiellä	5
1.3	Tutkimusaiheen muotoutuminen osana tutkimusprosessia	8
1.4	Käsitteet ja tutkimustilanne	12
2	Kuurojen opetuksen historiankuva kollokaatioiden valossa	17
2.1	Suomen Kuurojen yhteisön opetuksen historiaa	17
2.2	Lähdeaineiston tarkastelu	20
2.3	Kollokaatiot digitaalisissa ihmistieteissä ja metodin tulokset	23
2.4	Kollokaatioiden tarkastelu	36
2.4.1	Döfstumma-andesvaga – kollokaatio ja <i>Suomen aistivälliskoulujen lehti</i>	36
2.4.2	Kustaa Killinen	42
3	Tutkimusprosessin tulokset	47
3.1	Tapaustutkimuksen analyysi ja kerätyn tiedon luonne	48
3.2	Sanomalehdet digitaalisina ilmentyminä	53
3.3	Tiedonhaku ja arkistojen hiljaisuudet	57
4	Lopuksi	66
	Lähteet	70
	Liitteet	75
	Liite 1. R-ohjelman koodi	75
	Liite 2. Matemaattiset kaavat	83
	Liite 3. Kollokaatioiden visualisoinnit ja tiedonhaun indeksit	84

1 Johdanto

1800-luvun jälkipuolella Suomeen rantautunut oralismin aate merkitsi jyrkkää muutosta Suomen kuurojen asemaan ja toimijuuteen.¹ Oralismi on määritelty aatteeksi, joka ei tunnista viittomakielen merkitystä kuurojen elämässä ja painottaa kaiken oppimisen perustuvan puheen ymmärtämiseen ja tuottamiseen.² 1800-luvun alkua ennen oralismin saapumista on pidetty Suomen kuurojen yhteisön historiassa kultakautena, jolloin viittomakieli oli kuurojen pääasiallinen kieli kouluissa ja kotona. Viittomakielen vahva asema yhdisti kuuroja, mikä näkyi muun muassa aktiivisena järjestötoimintana.³ Kulta-aikaa seurasi pitkä 1900-luvun vääryyksien ja kaltoinkohtelun kausi.⁴ Kuurojen historiaa käsittelevissä teoksissa on tutkittu aiemmin, miten koulun penkiltä koteihin ja yhteiskunnallisiin rakenteisiin leviävä oralismi aiheutti vääryyden ja kärsimysten kokemuksia kuurojen keskuudessa. Tässä tutkielmassa tarkastellaan, mitä asioita ja millä tavalla oralismin historiaa voidaan ymmärtää käyttämällä lähdeaineistona digitaalisia sanoma- ja aikakauslehtiä. Samalla tutkielmassa pohditaan digitoidun lähdeaineiston mahdollisuuksia ja rajoitteita tekemällä tapaustutkimus digitaalisen historian tutkimuksen menetelmin ja pohtimalla, millaista tapaustutkimuksen tieto on verrattuna aikaisempaan tutkimustietoon. Koska digitaalisessa muodossa olevia aineistoja on 2020-luvulla käytössä laajoissa volyymeissa, historian tutkimuksen olisi syytä huomioida digitaalisessa muodossa olevien aineistojen erityispiirteet ja keinot, millä näitä aineistoja voidaan hyödyntää tutkimuksessa. Suurin osa aiemmasta tutkimuksesta perustuu lähilukuun ja aineistojen analogisen kontekstin analyysiin, joten tämän tutkielman tutkimusprosessi on pyritty avaamaan mahdollisimman yksityiskohtaisesti, jotta vastaavanlaisten laajojen digitaalisten aineistojen analyysille ja käytölle tulisi hahmoteltua käytännön keinoja.

1.1 Tutkimuskysymys

Tässä tutkimuksessa tarkastellaan, voiko digitoiduista sanoma- ja aikakauslehtiaineistoista muodostaa lähdetietoja kuurojen oralismin liittyvistä vääryyksien kokemuksista ja millä keinoin ne ovat löydettävissä. Samalla tutkielmassa pohditaan, miten löydetty tiedot vertautuvat kuurojen historian yleisteoksiin ja aiempaan tutkimukseen sekä vääryyksien

¹ Katsui et al. 2021, 36–37.

² Laakso & Salmi, 2006, 28.

³ Katsui et al. 2021, 35.

⁴ Katsui et al. 2021, 17.

kokemuksesta tehtyyn selontekoon. Keskeisenä arkistotieteellisenä teemana on pohtia, minkälaisia tiedonhakuun liittyviä kysymyksiä tietokoneavusteisesti käsiteltyihin aineistoihin liittyy vertailtaessa aiempaan tutkimukseen.

Tutkielmassani tarkastellaan aikaa, jolloin oralismi vakiintui pedagogisena ja ideologisena suuntauksena Suomessa. Aineistonani ovat kaikki saatavilla olevat suomen- ja ruotsinkieliset digitoidut sanoma- ja aikakauslehdet vuosilta 1890–1922, sekä 2021 Suomen valtioneuvostolle tehty selvitys *Viitotut muistot* kuurojen yhteisöön koostuneista vääryyksistä. Aikarajaus on valikoitunut selvityksen perusteella, jonka mukaan 1890-luvun jälkeen asenteet kuuroja kohtaan alkoivat muuttua kielteisemmiksi.⁵ Tarkastelen tutkimuksessani, millainen historian kuva *Viitottujen muistojen* ja digitoitujen sanoma- ja aikakauslehtien tarkastelulla on mahdollista muodostaa, miten kuuroihin kohdistuneita vääryyksiä⁶ voidaan havaita 1890–1922 digitoiduista sanoma- ja aikakauslehdistä ja millaisia haasteita laajan digitaalisessa muodossa olevan korpusaineiston tutkimuskäytössä on. Tarkoituksena ei ole kumota aiempien tutkimuksien tuloksia, vaan pohtia korpusaineiston substanssia, tutkimukseen tarvittavia teknisiä ratkaisuja sekä mahdollisia tiedon saavutettavuuteen, käytettävyyteen ja ymmärrettävyyteen liittyviä aiheita.

Keskeiseksi kysymykseksi nousee tietokoneavusteinen aineiston analyysi tiedonhaun välineenä mahdollisuuksineen ja rajoitteineen. Koska Suomen kuuroista ja kuurojen yhteisöstä on tuotettu 2020-luvulla useita erilaisia tutkimuksia, tässä tutkielmassa pyritään ensisijaisesti pohtimaan, minkälaista lisätietoa tietokoneavusteisten menetelmien avulla on mahdollista tuottaa, eli minkälaisia tiedonhakumenetelmiä korpusaineistoon on mahdollista soveltaa ja miten aineisto tulee valmistella, ennen kuin tiedonhakua voidaan hyödyntää. Samalla tutkielmassa havainnoidaan, millaisia rajoitteita tietokoneavusteisuus tuottaa tiedonhaun kannalta: Mitä tietoja jää hahmottamatta aineiston käsittelyn seurauksena ja miten käyttökelpoista laaja kuurojen yhteisöä käsittelevä digitoitu materiaalia on.⁷

1.2 Tutkimusaiheen taustaa ja metodien viitoittamalla tiellä

Kuurojen yhteisössä 1800-lukua kutsuttiin kulta-ajaksi, jolloin kuurojen mahdollisuus itsenäiseen toimijuuteen sekä itseilmaisuuksiin oli huipussaan.⁸ Keskeisenä tekijänä aseman

⁵ Katsui et al. 2021, 35.

⁶ Katsui et al. 2021, 8–9.

⁷ Carbajal 2021, 1108.

⁸ Katsui et al. 2021, 35.

saavuttamiseksi oli C.O. Malmin 1846 Porvooseen perustama koulu,⁹ jonka aikuistunut oppilaskunta perusti 1880-luvulla kuurojen yhdistyksiä.¹⁰ *Tidskrift för döfstumma* alkoi ilmestyä 1887 ja sen sisarlehti *Kuuromykkäin lehti* 1896.¹¹ Malmin koulun opetus tapahtui viittomakielellä ja opetettavien aineiden paletti sisälsi esimerkiksi kirjoitetun ruotsin, matematiikan ja maantiedon.¹² Opetuksen ansiosta viittomakieli oli 1800-luvun jälkipuolella kuurojen virallinen kieli sekä tärkeä osa yhteisön identiteettiä.

Kuurojen oloja merkittävästi heikentänyt oralismi perustuu saksalaisen opettajan Samuel Heinicken (1727–1790) opetusmetodeihin. Menetelmän mukaan puhe ja ajattelu liittyivät toisiinsa, ilman puhetta ei ollut myöskään todellista ajattelua. Oralismi kietoutui 1800-luvulla evoluutioteorioihin ja rotuoppeihin. Viittomakieli oli oralismin mukaan ”alempi kielimuoto”, joka tulisi korvata kehittyneemmällä puhekielellä.¹³ Oralismi rantautui Suomeen koululaitoksen uudistajien mukana. Esimerkiksi kouluhallituksen ylitarkastajana toiminut Uno Cygnaeus vaati puhemenetelmää ainoaksi kuurojen opetusmetodiksi. On huomattava, että Suomeen on virrannut aina 1700-luvulta lähtien pedagogisia ja didaktisia vaikutteita juuri saksankielisistä maista. Siksi oralismi tuli lähes varkain osaksi suomalaisten opettajien pedagogiikkaa 1800-luvun aikana.¹⁴ Suomen senaatti virallisti aistivaliokoulujen tarkastajan Valter Forsiuksen suositusten mukaisesti oralistisen opetusasetuksen 1890, minkä seurauksena viittomakielen opetus käytännössä loppui Suomen valtiollisissa oppilaitoksissa.¹⁵ Kuitenkin joissakin oppilaitoksissa esimerkiksi Kurikan yksityiskoulussa viittomakielen opetusta jatkettiin aina vuoteen 1919 asti.¹⁶

Tultaessa 1920-luvulle kuuroihin alettiin soveltaa rotuhygienian mukaisia toimia, joista etenkin 1929–1969 voimassa ollut avioliittolaki herätti runsaasti vastustusta kuurojen yhteisössä jo laatimisvaiheessaan.¹⁷ Avioliittolaki koski useaa eri vammaisryhmää, joskin kuurojen kohdalla kielto koski pelkästään synnynnäisesti kuuroja.¹⁸ Avioliittolain kumoamista oli esitetty muutettavaksi valtioneuvoston toimesta 1961, mutta vasta 1969 kaikki kuurojen väliset avioliitot säädettiin taas sallituiksi.¹⁹ Oralismin voidaan sanoa hiipuneen vaiheittain

⁹ Laakso & Salmi 2006, 14–16.

¹⁰ Laakso & Salmi 2006, 62, Ensimmäinen virallinen yhdistys perustettiin Turussa 1886.

¹¹ Laakso & Salmi 2006, 69–71.

¹² Laakso & Salmi 2006 41–43.

¹³ Laakso & Salmi 2006, 144.

¹⁴ Laakso & Salmi 2006, 145, 150–152.

¹⁵ Laakso & Salmi 2006, 154–158.

¹⁶ *Kuuromykkäin lehti* 1.2.1922, nro 5, s.20. Myös Pietarsaaren koulussa opetus jatkui aina vuoteen 1932.

¹⁷ Laakso & Salmi 2006, 204.

¹⁸ Katsui et al. 2021, 59–60.

¹⁹ Laakso & Salmi 2006, 208–209.

1950-luvun jälkipuolella. Viittomakieli hyväksyttiin lailla yhdeksi Suomen virallisista kielistä vasta 1995.²⁰ *Viitottujen muistojen* mukaan oralismin kuurojen yhteisöön jättämät jäljet ovat edelleen ajankohtaisia.²¹ Selvityksen mukaan kuurot joutuvat kohtaamaan syrjiviä ja loukkaavia asenteita edelleen 2020-luvulla.²²

Tutkielmani lähdeaineistossa rotuhygieniä näyttäytyy viittomakielen vastustuksena, joskin tultaessa 1920-luvulle kuurojen itsensä tuottamassa aineistossa on merkkejä siitä, että kuurous ymmärrettiin negatiivisena periytyvänä ominaisuutena.²³ Viittomakielen vastustaminen vaikuttaa olleen voimakasta 1800- ja 1900-lukujen taitteessa opetusvaikuttajien sekä kuuroja opettaneiden keskuudessa.²⁴

Pääasiallinen tietokoneavusteinen menetelmäni on R-ohjelmointikielellä tehty sovellus kollokaatioanalyysistä,²⁵ joka on yksi digitaalisen kielentutkimuksen ja digitaalisten ihmistieteiden käytetyimmistä metodeista. Historiantutkimuksen ja arkistotieteiden kohdalla kollokaatioanalyysit ovat harvinaisempia.²⁶ Käyttämäni analyysi perustuu tutkijoiden Brezina Vaclav, Tony McEnery ja Stephen Wattam esittämään ajatukseen kollokaatioiden verkostoimisesta, minkä avulla pyritään hahmottamaan koko korpuksen kokoisia kokonaisuuksia ja millaisia yhteyksiä sekä konteksteja niistä olisi mahdollista mallintaa sanojen tasolla, kun taas perinteinen kollokaatioanalyysi on perehtynyt yleensä kielen rakenteellisiin ja kieliopillisiin piirteisiin.²⁷ Koska analyysin tarkoituksena on löytää uusia näkökulmia, en ole ottanut huomioon menetelmän keskeisen log-dice-pisteytyksen²⁸ lisäksi muita tilastollisia määreitä. Edellä mainittu analyysi on mahdollista tehdä valmiilla ohjelmistolla, mutta pyrin toteuttamaan analyysit ohjelmoimalla, jotta pystyn määrittämään tarkasti käytettävät algoritmit ja datan visualisoinnit. Lisäksi valmisohjelmistossa ei ole tietääkseni sovellusta, joka pystyisi listaamaan kollokaatioille lähdetekstejä vastaavat metatiedot. Kollokaatioanalyysia ei ole tietääkseni sovellettu eksploraatiivisesti historiantutkimuksen saralla, joten tuloksien tulkinta on luonteeltaan heuristista.

²⁰ Laakso & Salmi 2006, 327–328.

²¹ Esim. Katsui et al. 2021, 136–137.

²² Katsui et al. 2021, 140.

²³ *Kuuromykkäin lehti* 1.6.1920, nro 6–7, s. 52.

²⁴ Laakso & Salmi 2006, 195–196.

²⁵ Alaluku 2.3.

²⁶ Vaclav & Platt 2024, 142.

²⁷ Kyröläinen & Laippala 2020, 496–497.

²⁸ Gablasova et al. 2017, 164–166 & Öksüz, 2021 61.

Tutkimus keskittyy analysoimaan kuurojen kärsimyksen havaitsemista digitaalisista lähdeaineistoista, jolloin *Viitotut muistot* sekä *Maahan lämpimään* toimivat toisen käden lähteinä kuurojen yhteisön kokemien epäkohtien todentamisessa. Digitaalisten tokenisoitujen ja kollokaatioanalyysillä järjestettyjen sanoma- ja aikakauslehtikorpusten aineistoon perustuvia havaintoja vertaillaan toisen käden lähteillä. Itse havaintojen, eli lehtitekstien sisällön tarkastelussa sovelletaan lähilukua vertailtaessa niitä *Viitottujen Muistojen* kuvauksiin sorrosta.

Maahan lämpimään mukaan viittomakieltä ei voitu enää pitää kuurojen virallisena opetuskielenä 1898, mutta tästä näkökulmasta huolimatta olen rajannut tutkielmani ajallisesti *Viitottujen muistojen* perusteella. Selonteon mukaan kuuroihin suhtauduttiin myönteisesti 1890-luvun alkuun asti.²⁹ Takarajaus perustuu suomalaisen puhemenetelmän merkittävän pioneerin Kustaa Killisen kuolemaan 1922. Takaraja selkeytyi korpuksen kollokaatioanalyysin tulosten tutkimiseen jälkeen. Alun perin aineisto oli rajattu vuoteen 1921, jolloin oppivelvollisuus astui voimaan.

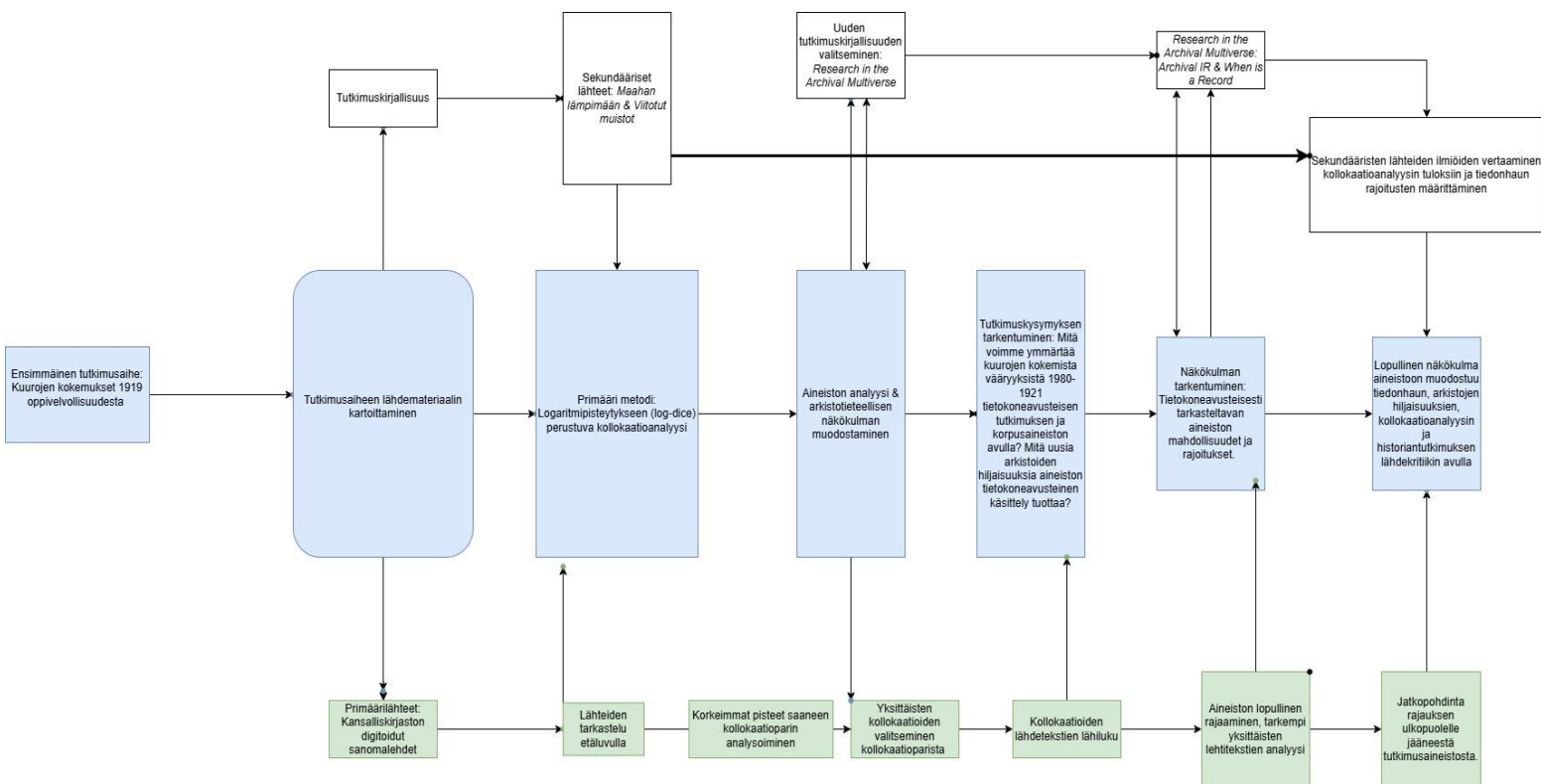
Olen myös etsinyt aineistosta mahdollisia kollektiivista muistia edustavia lähteitä, sillä kollokaatioanalyysin perusteella sanoma- ja aikakauslehtien artikkeleista ei voi löytää kuurojen omaa keskustelua oralismiin liittyen. Tästä syystä muistitietoaineiston kohdalla tutkimuksessa on hyödynnetty lähes yksinomaan *Viitottujen muistojen* tietoja kuurojen aikalaiskokemusten hahmottamiseksi. Tästä esimerkkinä termi kuuromykkä, jonka alkuperä vaikuttaa hämärtyneen aikalaisten keskuudessa jo tultaessa 1900-luvulle.³⁰ Tutkimusprosessin aikana kerättyjen tietojen perusteella Suomen kuuroja ei voi pitää yksiselitteisesti valtavirran sortamana erityisryhmänä.³¹ *Kuuromykkäin lehden* tekstien arviointia vaikeuttaa, että tekstien kirjoittajia ei ole aina mainittu edes nimimerkillä.

1.3 Tutkimusaiheen muotoutuminen osana tutkimusprosessia

²⁹ Katsui et al. 2021, 35.

³⁰ C.O. Malmin oma nimitys viittoville, ei puhuville, Katsui et al. 2021, 253–256.

³¹ Katsui et al. 2021, 255–256.



Tutkimusprosessikaavio

Tutkimusprosessi perustuu tässä pro gradu -tutkielmassa aineiston analyysiin tietokoneavusteisin tutkimusmenetelmin ja vertailemalla tuloksia toisen käden lähteisiin. Kyseessä on pitkälti tapaustutkimus tietäntyyppiseen aineistoon, jonka sisältö on tarkentunut tutkimusprosessin edetessä. Tästä eteenpäin yllä esitettyyn kaavioon viitataan tutkimusprosessikaaviona. Kaaviota luetaan vasemmalta oikealle. Nuolet kuvaavat laatikoiden, prosessien etenemistä kolmeen pääasialliseen osaprosessiin. Siniset suorakulmamalliset laatikot kuvastavat tutkimuksen menetelmällisiä vaiheita, niissä on aktiivisesti käytetty metodologista lähestymistapaa kuvattuihin aineistoihin. Soikea

sinipohjainen prosessi kuvaa lähdeaineiston hankkimista. Sininen prosessi voidaan ymmärtää lähimmäksi ”perinteistä” historian ja arkistotieteiden tutkimusta. Valkopohjaiset prosessit kuvaavat tutkimuskirjallisuuden kartoittamista, teoreettisen näkökulman muodostumista ja osan tutkimuskirjallisuuden käyttöä ristiin luettavana toisen käden lähdekirjallisuutena. Vihreäpohjaiset prosessit taas kuvaavat tietokoneavusteisen aineiston muodostamista ja käyttöä. Koska tietokoneavusteinen menetelmä on käytännössä myös ohjelmointikoodin kirjoittamista ja korjaamista sen mukaan, millaisia tuloksia aineistosta on saatu, voidaan se ymmärtää omana erillisenä prosessinaan, joka toisaalta mahdollistaa koko tapaustutkimuksen, mutta on samalla riippuvainen keskeisestä kriittisestä menetelmätutkimuksesta.

Digitoituna lähdeaineistonani ovat suomen- ja ruotsinkieliset korpuksat, joista toinen on koostettu sanahauulla ”kuuro* AND mykk*”. Toisen kohdalla yhdistin kaksi korpusa sanahauista ”döfstum*” ja ”dövstum*”, sillä jälkimmäinen kirjoitusasu näyttää vakiintuneen lehdissä tullessa 1910-luvulle. Vaihtoehtoisina sanahakuina olisi voinut käyttää laajempia ”kuuro” tai ”döf” -sanoja, mutta päädyin tarkempiin hakusanoihin aineiston rajaamiseksi. Käytännössä sanahaun takia toinen korpuksista kattaa kaikki ajanjakson aikana julkaistut suomenkieliset lehdet ja toinen vastaavasti kaikki ruotsinkieliset. Tavoitteeni on tutkia keräämäni aineiston ominaisuuksia sekä millaista tietoa laajasta aineistosta on ylipäätään mahdollista löytää sanahauilla ja onko 2021 julkaistun selvityksen esittämiä kehityskulkuja mahdollisuutta havaita korpusaineistosta. Tutkimuksen tämä vaihe on prosessikaavion kohta ”tutkimusaiheen lähdemateriaalin kartoittaminen” ja siitä haarautuvat prosessit.

Käsittelmäni korpusaineisto koostuu 76 943 tekstistä, joita tarkastellaan tietokoneavusteisin menetelmin. Käytännössä sovellan digitaalisessa kielentutkimuksessa käytettyä kollokaatioanalyysia³² eli sanojen tilastollisesti määriteltyä taipumusta esiintyä yhdessä. Menetelmää sovelletaan yhdessä etä- ja lähilukutapojen yhdistävässä analyysissa.³³ Analyysini on eksploraatiivinen, eli tavoitteenani ei ole löytää laajemmin yleistettäviä tilastollisia havaintoja, vaan nostaa hakutuloksista esille uusia näkökulmia.³⁴ Samalla tutkielmassani käsitellään arkistollista tiedonhaku³⁵ osana gradun tutkimusprosessia. Tiedonhaku tapahtuu sekä kansalliskirjaston digitoitujen sanoma- ja aikakauslehtikorpuksien muodostamisessa käytetyssä sanahaussa että korpuksille tehtävässä kollokaatioanalyysissa.

³² Alaluku 2.3.

³³ Hakkarainen et al. 2022, 116–117, 119.

³⁴ Pivovarova 2020, 1.

³⁵ Archival IR, Furner & Gilliland 2017, 581–631.

Koska kollokaatioita on erittäin paljon, tässä tutkielmassa on keskitytty vain kahden kollokaatioketjun analysoimiseen.³⁶ Ensimmäinen kollokaatioketju valikoitui tilastollisen testin tuloksena ja toinen korpusaineiston lähiluvun perusteella osana tutkimusprosessin vaihetta, jossa tutkimuskysymys on tarkentunut.³⁷

Jokainen kollokaatio saa pisteytyksen sen mukaan, miten uniikkeja sanojen vierekkäisyys on. Ensimmäinen tutkielmassa tarkemmin tarkasteltavista kollokaatioista ”döfstumma-blinda_och – och_döfstumma-andesvaga” on korkeimman pisteytyksen saanut sanayhdistelmä. Tämä tarkoittaa, että och-sanan oikealla ja vasemmalla puolella esiintyvät ”döfstumma-blinda” ja ”döfstumma-andesvaga” esiintyvät todennäköisesti juuri toistensa läheisyydessä.

Vastakohtana ”döfstumma_och_andesvaga” ei ole lainkaan korkeimman 50 log-dice-pisteytyksen saaneen kollokaation joukossa, koska kyseiset sanat esiintyvät toistensa kanssa määrällisesti useasti. Historiantutkimuksessa edellä mainittujen uniikkien sanaparien löytäminen voi liittyä poikkeuksellisten tietojen löytämiseen, mahdollisten entiteettien, kuten yksittäisten henkilöihin liittyvien titteleiden havaitsemiseen tai muihin sananhaun sanojen tasolla havaittavien, uniikkien sananläheisyyksien löytämiseen. Kollokaatiot tuottavat uutta tietoa ja uusia näkökulmia jo olemassa olevaan tietoon.

Kollokaatioanalyysin perusteella tiettyjä digitoituja sanoma- ja aikakauslehtiä on tarkasteltu tarkemmin sisällön osalta. Lehtien tekstejä ja toisen käden lähteitä ristiin lukemalla voidaan todeta oralismin olleen käänteentekevä ilmiö Suomen kuurojen yhteisön historiassa.

Digitaalisissa aineistoissa oralismiin ja puhemetodiin liittyvät keskustelut ovat yliedustettuina *Kuuromykkäin lehdessä* ja *Suomen aistivalliskoulujen lehdessä*.³⁸ Tarkastelemani ajanjakson muissa lehdissä oralismia ei mainita suoraan ja kuuroihin liittyvistä maininnoista ei voi vetää selviä johtopäätöksiä siitä, miten kuuroihin on ylipäätään suhtauduttu. Pikemminkin voidaan todeta, että sanoma- ja aikakauslehtiaineiston yleiskuva on ambivalentti tutkimushypoteesini osalta: Kuuroja kohtaan tunnettiin toisaalta myötätuntoa, mutta samanaikaisesti esimerkiksi kollokaatioanalyysien tulokset visualisoimalla yleisvaikutelma kuuroja kohtaan vaikuttaa kielteiseltä. *Viitottujen Muistojen* muistitietoaineisto sekä Laakson ja Salmen kuurojen historiaa käsittelevä *Maahan lämpimään* osoittaa, että tultaessa 1910-luvulle oralistisen opetuksen takia viittomakielestä tuli kuurojen yhteisön sisäinen kieli.³⁹ Samalla yhteiskunta

³⁶ Tästä lisää alaluvussa 2.3.

³⁷ S. 10, kuva: Tutkimusprosessikaavio.

³⁸ Suomen aistivalliskoulujen lehti 1.8.1896, nro 8–9, s.1.

³⁹ Laakso & Salmi 2006, 194.

alkoi vähitellen niputtaa kuurot yhteen muun muassa tylsämielisten ja mielisairaiden kanssa.⁴⁰ Rotuhygienian periaatteiden mukaisen avioliittolain⁴¹ valmistelu alkoi vuonna 1917, jolloin lainvalmistelukunta esitti synnynnäisesti kuurojen avioliittoa luvanvaraiseksi.⁴² Lähiluvun soveltaminen on tässä tutkielmassa rajattu niihin digitoituihin sanoma- ja aikakauslehtiteksteihin, joiden sisältö on löytynyt kollokaationalayysin avulla. Rajaus on tehty kollokaatioiden avulla löytyneiden tekstien suuresta määrästä johtuen. Samalla lähiluku on metodologinen lähestymistapa arvioida tiedonhaun näkökulmasta kollokaatioanalyysin tuloksia, jossa lähiluettavia tekstejä verrataan aikaisemman tutkimuksen tuloksiin.

Kolmas lähestymistapa aineistoon on eksploratiivinen. Tässä vaiheessa ensin muodostettua kollokaatioketjua on tarkasteltu sekä etäluvun että osan tekstien osalta lähiluvun avulla. Näistä lähdeteksteistä saatujen tulosten ja toisen käden lähteiden ristiin luvun perusteella valikoitui toinen tarkemmin tarkasteltava kollokaatioketju. Tämä toinen ketju liittyi ”döfstumma-andesvaga”- kollokaatioiden lähdeteksteissä esiintyneeseen Kustaa Killiseen. Kolmanteen vaiheeseen liittyy myös menetelmien arviointi tiedon käytettävyyden osalta, eli minkälaisia tiedonhaun mahdollisuuksia käsitellyllä aineistolla on. Tämä vaihe on kuvattu tutkimusprosessikaavan kohdassa ”Näkökulman tarkentuminen: Tietokoneavusteisesti tarkasteltavan aineiston mahdollisuudet ja rajoitukset” ja siihen liittyvissä prosesseissa.

1.4 Käsitteet ja tutkimustilanne

Kuurojen keskuudessa 1800-luvun loppupuolella vakiintunut nimitys ”kuuromykkä” on yksi keskeisimmistä käsitteistä avattavaksi. Kuuromykällä (döfstumma) tarkoitettiin 1900-luvulla henkilöä, joka kommunikoi pelkästään viittomalla.⁴³ Oralistisessa opetuksessa kuuromykkyys oli sanana negatiivinen, sillä se viittasi puhetaidon osaamattomuuteen,⁴⁴ kun taas Malmin koulussa kuuromykkä tarkoitti käytännössä viittomaa.⁴⁵ Malmin koulun opetuksen myötä kuuromykkä oli identifioiva sana esimerkiksi kuuromykkäin lehden kohdalla. Termistä käytiin jo tuolloin keskustelua ja nykyään se koetaan kuurojen keskuudessa erittäin loukkaavaksi. Tähän liittyen myös aistiviallinen on tärkeä termi hahmotettaessa, miten erilaisia

⁴⁰ Laakso & Salmi 2006, 203, alaluku 2.4.1, Hufvudstadsbladet 22.5.1908, nro 139, s.4. Ruotsiksi ”andesvaga.”

⁴¹ Katsui et al. 2021, 61.

⁴² Laakso & Salmi 2006, 203.

⁴³ Katsui et al. 2021, 255.

⁴⁴ Katsui et al. 2021, 38.

⁴⁵ Katsui et al. 2021, 254.

erityisryhmiä ei ole juurikaan käsitelty omista erityisistä lähtökohdistaan, vaan heidät on niputettu yhdeksi massaksi.

Kuurojen yhteisöä käsiteltäessä on avattava myös oralismin käsite. Ilmiönä oralismi on koskettanut kuuroja niin laajasti ja merkityksellisesti, että sen vaikutuksia yksilöiden sekä yhteisöiden tasolla käydään edelleen läpi.⁴⁶ Oralismi on myös tutkimuskysymystä selittävä käsite: Se auttaa hahmottamaan Suomen opettamisen historian siteet saksankielisiin maihin ja monien pedagogisten ja didaktisten vaikutteiden omaksumisen juuri 1800-luvun lopun Saksasta. Termi myös liittyy Suomen tarpeeseen pyrkiä kohti länttä omaksumalla samalla esimerkiksi rotuhygieenisiä ajatuksia.⁴⁷

Tiedonhaku ”Information retrieval” on arkistotieteiden tutkimuskentällä liitetty tietokoneavusteisiin menetelmiin, kuten Internetin hakukoneisiin.⁴⁸ Tultaessa 2020-luvulle arkistoaineistojen digitalisoituminen ja sähkösyntyisten aineistojen määrän kasvaminen kuitenkin tarkoittaa, että tiedonhakua ei voi tarkastella pelkästään tapana hakea informaatiota, vaan kokonaisuutena menetelmiä, joilla paikannetaan *digitaalisia objekteja*.⁴⁹

Kollokaatioanalyysi toimii käytännössä korpusaineiston tietojenhakumenetelmänä ja tähän liittyen tutkielmassani tarkastellaan kriittisesti sanahaun ominaisuuksia. Väitän, että sanahauulla koottu korpusaineisto ei ole pelkästään leikattu osakokonaisuus alkujaan yksittäisten lehtien teksteistä muodostuvaa arkistoaineistoa,⁵⁰ vaan haettu sana aiheuttaa aineistoon sidoksisuutta, joka paljastuu tarkastelemalla kollokaatioanalyysin tuloksia.

Käyttämistäni tutkimuskirjoista *Maahan lämpimään* on lähinnä yleisesitystä Suomen kuurojen yhteisön historiasta ja sivuaa myös eniten omaa aihettani. Viimeisimmistä tutkimuksista Oulun yliopistossa toteutettu monitieteinen Hiljaisen kansan äänet - tutkimusohjelma sivuaa aihettani kuurojen musiikkikasvatuksen historian osalta. Hisayo Katsuin ja kumppanien laaja katsaus *Viitotut muistot* sisältää kattavan kuvauksen kuuroihin kohdistuneista väärinkäytöksistä, jotka ovat relevantteja graduni kannalta. En ole onnistunut löytämään historian alalta tutkimusta, joka käsittelee tarkemmin kuurojen kokemuksia vuosina 1890–1922. Sen sijaan viittomakielen ja oralismin yhteistä historiaa on tutkittu jonkin verran muissa Pohjoismaissa.⁵¹

⁴⁶ Katsui et al. 2021, 285.

⁴⁷ Laakso & Salmi 2006, 152.

⁴⁸ Furner & Gilliland 2017, 582–583.

⁴⁹ Rozenberg 2021, 211. Myös alaluvut 3.2. & 3.3.

⁵⁰ Furner & Gilliland 2017, 588–600.

⁵¹ <https://www.stockholmsdf.se/event/dovas-historia-och-kultur-i-skolan/>

Käyttämäni metodikirjallisuus sisältää runsaasti korpus- ja kieliteknologian oppaita, jotka tukevat käyttämiäni tietokoneavusteisia analyysimenetelmiä. Historian metodioppaiden joukosta löytyy hyödyllisiä historiografisia sekä tutkimusteoreettisia oppaita digitaalisen historian tutkimuksen suhteen, mutta oman tutkimukseni kohdalla kielentutkimuksen, digitaalisten ihmistieteiden ja arkistotieteiden metodit ovat paremmin sovellettavissa lähdeaineistooni. Tietokoneavusteisia eksploratiivisia historian tutkimuksia on viime aikoina tehty etenkin koneoppimisen menetelmien avulla⁵² ja aihepiiriin liittyen on myös julkaistu huomattava määrä ohjemateriaalia.⁵³ Digitaalisissa ihmistieteissä käytetään 2020-luvulla luonnollisen kielen prosessointiin ja koneoppimisen mahdollisuuksia. Perinteisempiä algoritmisia menetelmiä on hyödynnetty muun muassa Turun yliopistossa tehdyssä sanoma- ja aikakauslehtitekstien uudelleenkäyttöä tarkastelevassa tutkimuksessa.⁵⁴ Suomen koulujärjestelmästä on olemassa runsaasti tutkimuksia, joten niiden osalta lähdeluetteloon tulee vielä runsaasti lisäyksiä. Kuurojen asemaa 1920-luvun koulutusjärjestelmässä on edellä mainittujen tutkimushankkeiden lisäksi sivuttu useissa kasvatustieteen, erityispedagogian ja kasvatussosiologian tutkimuksissa. Minna Muiluvuoren 2025 kirjoittama Turun yliopiston etnologian pro gradu -tutkielma *Kuurot ja arki Vammaisuuden representaatioita Kuuromykkäin lehdessä vuosina 1896–1912* käsittelee yhdeksässä lehdessä esille nostettuja diskursseja kuurojen ja kuulevien valtasuhteiden välillä.

Pyrin huomioimaan tarkasti eettiset seikat liittyen kuurojen 1900-luvun alun asemaan. Lähtökohtani on, että 1900-luvun alun kuurot eivät olleet yksiselitteisen sorretti ihmisryhmä, vaan esimerkiksi *Kuuromykkäin lehdessä* käytiin kiivasta keskustelua kuurojen oikeuksien puolesta esimerkiksi oralismin suhteen. Kuurojen yhteisö on ja on ollut hyvin moninainen, mihin tulen kiinnittämään erityistä huomiota.⁵⁵ En kuitenkaan sivuuta tosiasiaa, että kuurot, kuten myös muut erityisvähemmistöt, ovat joutuneet Suomessa 1900-luvun aikana erinäisten sortotoimenpiteiden ja ennakkoluulojen kohteeksi.⁵⁶ Tästä syystä kuurojen kokemat oralismin aiheuttamat vääryydet rinnastavat yhteisön 1900-luvun alun aseman muihin kielivähemmistöryhmiin, kuten saamelaisiin.⁵⁷

⁵² Pivovarova 2020, 3–4.

⁵³ Esim. Wilkinson 2023.

⁵⁴ Hakkarainen, Paju & Salmi. 2023. "Towards an Ontology and Epistemology of Text Reuse".

⁵⁵ Katsui et al. 2021, 33–34.

⁵⁶ Katsui et al. 2021, 52–53.

⁵⁷ Nyysönen 2014, 66–72.

Käytyäni määrällisesti läpi digitaalisia aineistoja olen löytänyt muutamia kiinnostavia tapauksia, joita käytän esimerkkinä lähi- ja etäluvun yhdistämisestä. Toisessa esimerkkitapauksessa käydään läpi puhemenetelmän pioneerin Kustaa Killisen merkityksiä painottaen lähiluvun merkitystä korpusaineistossa. Toisessa taas tarkastellaan etälukupainotteisesti ruotsinkielisen lehdistön käyttämiä ilmaisia kuuromykyistä tarkastelemalla ”döfstumma-andesvaga” ja suomenkielisen lehdistön käyttämää ”johtaja-Killinen” -kollokaatiota. Esimerkkitaupaukset ovat valikoituneet tietokoneavusteisen analyysin visualisointien tarkastelun, log-dice-pisteytyksien sekä lähiluvun perusteella. Lisäksi tapauksien välillä on olemassa yhteys, joka havainnollistaa digitaalisen historian tutkimuksen mahdollisuuksia laajojen aineistojen piilevien yhteyksien löytämisessä.

Tiedonhaku vaikuttaa olevan edelleen melko vähäisesti käsitelty aihe niin arkistotieteen kuin historian tutkimuksen aloilla. Tämä luultavasti johtuu tiedonhaun teknisten toteutusten järjestelmäkohtaisuudesta. Esimerkiksi käyttämäni kollokaatioanalyysi vaatii Kansalliskirjaston käyttöliittymän tuottaman sanoma- ja aikakauslehtikorpuksen toimiakseen, jolloin sitä ei voida soveltaa muihin korpusaineistoihin ilman koodin muokkaamista. Tarkoitukseni on tarkastella tiedonhallintaa käyttäjän näkökulmasta, jolloin fokuksessa on haetun ja käsitellyn tiedon substanssi.⁵⁸ 2017 julkaistu artikkelikokoelma *Research in the Archival Multiverse* esittelee arkistotieteellisen näkökulman tiedonhakuun, joka kuitenkin sisältää osittain vanhentunutta tietoa juuri hakualgoritmeihin liittyvän teknologian osalta.

Tutkielman rakenne vastaa tutkimusprosessikaavion rakennetta. Aluksi esitellään aiheen historiallinen konteksti ja siihen liittyvä tutkimusmateriaali mahdollisimman kompaktisti. Tähän materiaaliin kuuluvat Suomen kuurojen opetuksen murroskohdat, eli 1850-luvulta C. O. Malmin koulun perustumisesta alkaneen viittomakielen kauden ja vuosisadan jälkipuolella alkaneen oralismin ajan.⁵⁹ Olen rajannut tutkielmani 1890–1922 väliselle aikajaksolle, mutta kontekstoinnin osalta tutkimuksessa on avattava myös 1800-luvun alun kuurojen yhteisön vaihteita. Tämän jälkeen tarkastellaan, miten sanoma- ja aikakauslehtikorpusta pystyy etälukemaan ja miten kehittämäni kollokaatioanalyysi toimii käytännössä historian tutkimuksen etäluvun tiedonhankintaa kehittäväenä välineenä. Tutkielmassa sivutaan kuurojen yhteisön kokemia vääryyksiä 1900-luvun alussa 2021 selonteon materiaalin valossa,

⁵⁸ Furner & Gilliland 2017, 581.

⁵⁹ Katsui et al. 2021, 71–73.

mutta keskeisenä näkökulmana on pohtia, mitä sanoma- ja aikakauslehtikorpus voi kertoa tutkimusaiheen historiasta ja miten.

Toinen käsittelyluku kattaa korpusaineiston kollokaatioanalyysin tulokset, tarkemman katsauksen sanoma- ja aikakauslehtikorpuksista ja sen, millaisia lähteitä digitoidut sanoma- ja aikakauslehdet oikeastaan ovat. Samalla pohditaan, millaisia haasteita digitaalisten aineistokokonaisuuksien tiedonhakuun liittyy. Keskeisimpiä teemoja menetelmien osalta on sanoma- ja aikakauslehtikorpuksen olemus datakehikkoon tokenisoituna teksteinä.⁶⁰

Tokenisointi tarkoittaa tässä tutkimuksessa sanomalehtitekstien jakamista sanoiksi, joita käsitellään digitaalisina objekteina.⁶¹ Esimerkiksi *Aistiviälliskoulujen lehdestä* poimittu teksti:

”Sweitsissä kukin kanttoni pitää erikseen huolta kuuromykkäinsä opetuksesta.”

Tokenisoidussa muodossaan tekstistä poistetaan välimerkit ja isot kirjaimet muutetaan pieniksi, jolloin edellinen teksti ilmenee seuraavasti listana sanaobjekteja:

[’sweitsissä’, ’kukin’, ’kanttoni’, ’pitää’, ’erikseen’, ’huolta’, ’kuuromykkäinsä’, ’opetuksesta’]

Näiden muodostettujen sanaobjektien⁶² perusteella lasketaan log-dice-pisteitys kollokaatioanalyysia varten. Objektit ja niistä muodostetut kollokaatit tuottavat tiedonhaun prosessin, jossa yhdistyy tutkimuksen kannalta kiinnostavan ja relevantin tiedon hakeminen tietokone luettavasta tokenisoidusta tekstistä ja pohdinta, miten arkistotieteellisesti tulisi suhtautua tietokone luettavaksi muokattuun aineistoon.⁶³ Tiedonhaku tulee ymmärtää paitsi aineiston käytön ja analyysin välineenä, myös teoreettisena lähtökohtana sille, mitä asioita voidaan ymmärtää digitaalisen aineiston avulla, minkälaisen prosessin läpikäymisen tietynlaisen digitaalisen datan käyttö edellyttää ja miten digitaalinen aineisto eroaa analogisesta. Samalla tiedonhaku liittyy kysymykseen arkistojen hiljaisuudesta, eli nouseeko korpuksista yliedustettuna esille tietynlaista tietoa ja jättääkö käsitelty aineisto katvealueita tai vinoumia⁶⁴ joidenkin tiettyjen aihealueiden osalta.

⁶⁰ Risku, 2020, 17.

⁶¹ Rozenberg 2021, 196–198.

⁶² Token object.

⁶³ Henttonen. 2025, 228.

⁶⁴ Eng: *Bias*.

2 Kuurojen opetuksen historiankuva kollokaatioiden valossa

Tässä luvussa tarkastellaan tutkimusaiheen historiallista kontekstia, aiheen lähdeaineiston eli sanoma- ja aikakauslehtikorpuksien ominaisuuksia ja kollokaatioanalyysin tuloksia. Aluksi tarkastellaan kuurojen opetuksen historiaa aikana, jolloin oralismi rantautui Keski-Euroopasta Suomeen kuurojen koulujen opettajien mukana. Tämän jälkeen tutkitaan sanoma- ja aikakauslehtikorpuksen muodostamista Kansalliskirjaston hakukäyttöliittymällä.

Alaluvussa 2.3 esitellään pro gradu -tutkielmaa varten kehitetyn kollokaatioanalyysin toiminta ja alaluvussa 2.4 analyysistä saadut tutkimustulokset. Keskeisenä tavoitteena on pohtia, miten kollokaatioanalyysin perusteella lähiluetut sanoma- ja aikakauslehtitekstit vastaavat *Maahan lämpimään* ja *Viitottujen muistojen* kuvauksia oralismin saapumisesta Suomeen, sen vaikutuksista kuurojen yhteisöön ja siitä seuranneiden vääryyksien kokemusten kuvauksia.

2.1 Suomen Kuurojen yhteisön opetuksen historiaa

C.O Malm perusti Porvooseen kuurojen koulun 1846.⁶⁵ Malm oli itse opiskellut Tukholman Manila-koulussa, josta hän otti vaikutteita oman koulunsa opetuksen kehittämistä varten. Koulun virallinen opetuskieli oli viittomakieli, mutta oppilaiden tuli tämän lisäksi osata lukea ja kirjoittaa ruotsia sekä osallistua ammattiopetukseen. Tavoitteena oli antaa kuuroille oppilaille valmiudet elättää itsensä.⁶⁶ Malmin oma opetus perustui tähän kaksikieliseen opetustapaan, mutta toisin kuin Manila-koulussa, Malm sisällytti opintoihin myöskin uskonnon, laskennan, maantiedon, piirustuksen ja liikunnan oppitunteja.⁶⁷

Senaatti julkaisi vuonna 1858 säädösasetukset yleisestä kansanopetuksesta ja kuurojen opetuksesta. Malmin koulu oli tätä ennen kohdannut vaikeuksia olemattomasta tai vähäisestä valtionavusta johtuen,⁶⁸ minkä seurauksena joukko vaikutusvaltaisia porvooolaisia perusti koululle johtokunnan edistämään koulun asioita ja auttamaan Malmia koulun hallinnollisten asioiden hoidossa. Tähän johtokuntaan kuuluivat runoilija ja kirjailija J. L. Runeberg,

⁶⁵ Malm ilmoitti Keisarille aloittaneensa opettamisen vasta 1847 Laakso & Salmi 2006, 41.

⁶⁶ Laakso & Salmi 2006, 36

⁶⁷ Laakso & Salmi 2006 43.

⁶⁸ Laakso & Salmi 2006 46.

kapteeni J. R. von Essen, kauppias J. Solitander sekä pastori C. H. Alopaeus.⁶⁹ Johtokunta teki 1857 ehdotuksen valtion varoilla toimivasta kuurojenkoulusta, joka noudatteli sisällöltään keisarin julistusta kansanopetuksen järjestämisestä. Säädösasetus koskien kuurojen opetusta noudatti johtokunnan ehdotusta ja 1860 Turkuun perustettiin Suomen ensimmäinen valtiollinen kuurojen koulu.

Asetus sisälsi maininnan ”sormikielen” opetuksesta, jonka on epäilty tarkoittavan ennemmin sormiaakkosia kuin viittomakieltä. Puheen opetuksesta sen sijaan ei ollut mitään mainintaa vaan koulun johtajia kehoitettiin ”etsimään aktiivisesti sellaisia lahjakkaita kuuroja, jotka koulun käytyään voisivat mahdollisesti opettaa muita kuuroja.”⁷⁰ Anomuksessa kuurojen koulun perustamiseksi määrättiin myös, että koulun johtajalta edellytettiin täydellistä puhetaitoa ja kuuloa. Laakso ja Salmi ovat nostaneet asetukseen liittyen esille, että kyseessä on ainut viittomakielen aseman huomioiva lainsäädäntö sataan vuoteen.⁷¹

Malm menehtyi vuonna 1863. Hänen työnsä kuurojen opetuksen parissa voidaan katsoa vaikuttaneen paitsi kuurojen yhteisön ”kultakauden” katalyyttinä myös ylipäättään muita ”vaivaisia” koskevan koulutuskeskustelun aktivoineena kehityksenä. *Maahan lämpimään* tuo esille myös kuurojen erityisaseman verrattuna muihin vammaisiin. Esimerkiksi vuoden 1852 Vaivaishoitoasetuksessa säädetään muun muassa ”mielettömien ja raajarikkojen” holhouksesta, mutta kuuroja ei mainittu asetuksessa ollenkaan.⁷² Ensimmäinen kuurojen yhdistys perustettiin Malmin koulua käyneen Fritz Hirnin johdolla 1885. Kuurojen oma *Kuuromykkäin lehti* alkoi taas ilmestyä 1896.⁷³ Hirnin tapaan suuri osa kuurojen yhdistyksen perustajajäsenistä oli käynyt Malmin koulua. Koulun opetuksen perustana oleva viittomakielen opetus kytkettynä suomen ja ruotsin kielen kirjoittamiseen ja lukemiseen mahdollisti kuurojen yhteisön oman kulttuurin kukoistamisen sekä viittomakielen aseman vakiintumisen Suomessa.⁷⁴

Oralismi saapui Suomeen vaiheittain. C. H. Alopaeusta, Uno Cygnaeusta ja G. K. Hendelliä on pidetty usein oralismin Suomeen tuoneina henkilöinä. Esikuvana toimi saksalaisen Weissefeldsin malli,⁷⁵ mutta käytännössä opit hankittiin opintomatkoilla muihin

⁶⁹ Laakso & Salmi 2006, 46–47.

⁷⁰ Laakso & Salmi 2006, 48.

⁷¹ Laakso & Salmi 2006, 48.

⁷² Laakso & Salmi 2006, 51–52.

⁷³ Laakso & Salmi 2006, 62–63, 69. Lehden toimittamisesta vastasivat aluksi Mikkelin kuuromykkäinkoulun opettajat Ossian Wichmann ja Konstantin Reponen.

⁷⁴ Katsui et al. 2021, 72.

⁷⁵ Laakso & Salmi 2006, 150.

Pohjoismaihin. Ilmeisesti 1876 Tukholmassa järjestetty aistivalliskoulutajien kokous oli tosiasiallinen tilaisuus, jossa päätettiin kuurojen opettamisesta puhemenetelmällä. Kokouksen päätös kuvastaa hyvin oralismin asteittaista sulautumista osaksi opetusta: ”Koska puhemetodi oli kaikkialla, jossa se oli tullut kokonaan käyttöön, osoittanut omaavansa merkittäviä ja korvaamattomia etuja verrattuna konventionaalisiin viittomiin (---) puhemetodin mukaan tuli opettaa paitsi niin sanottuja ei-varsinaisia kuuromykkiä, myös varsinaisia kuuromykkiä.”⁷⁶ Käytännössä kokouksessa tehtiin kuitenkin vain periaatepäätös puhemetodin käyttöönotosta, vasta 1885 Senaatti lähetti filosofian maisteri Valter Forsiuksen tutustumaan Saksaan ja Sveitsiin ”sokeiden, kuurojen ja tylsämielisten opetukseen”. Kahden vuoden perehtymismatkan jälkeen Forsius teki ehdotuksen Suomen aistivalliskoulutuksen järjestämisestä ja kartoitti koulutuksen tilannetta Suomessa pyytämällä lausuntoja opetuksen järjestäjiltä. Yksi lausuntoihin vastannut oli Kuopion kuurojen koulun johtaja Kustaa Killinen, jonka mukaan koulun 74 oppilaasta oli opetettu ”menestyksekkäästi” puhemetodilla. 1889 aistivalliskoulutuksen suunnittelukomitean kokouksessa ainoastaan Pietarsaaren kuurojen koulun johtajatar Anna Heikel vastusti puhemetodin asettamista pääasialliseksi kuurojen koulujen opetusmetodiksi.⁷⁷

1800-luvun loppupuolen puhemetodia pidettiin käytännönläheisenä ja erityisesti lahjakkaille oppilaille soveltuvana opetusmetodina. Puhemetodista ei määrätty suoranaisesti missään tietyssä laissa tai asetuksessa, mutta 1858 komitean istunnoissa esimerkiksi Alopaeus ja Malm kannattivat puheen opettamista ”siltoin, kun se ei ollut oppilaalle ylivoimaista.”

1892 asetus toteaa tavoitteekseen ”Kasvattaa kuuromykkiä, sokeita ja tylsämielisiä kristillisiksi, siveellisiksi ja höydyllisiksi yhteiskunnan jäseniksi, ---- harjoittaa heitä sellaisissa käytännöllisissä töissä, joilla he vastedes, mikäli mahdollista on, voivat hankkia itselleen elatusta.” Asetus rinnasti kuurot sokeisiin ja tylsämielisiin, mutta käytännön tason merkitys oli vielä suurempi, sillä sen säätämisen myötä aistivalliskoulut siirtyivät kouluylihallituksen alaisuuteen. Käytännössä tämä tarkoitti esimerkiksi Porvoon kuurojen koulun kieliä ja matematiikkaa sisältävien erityisten opetussuunnitelmien lakkauttamista ja viittomakielen syrjäyttämistä puhemetodilla. Kuurojen opettajien kirjoituksia ajalta ei ole säilynyt, mutta käytännössä asetus myös lopetti viittomakielen opetuksen jatkumon, 1892 kouluihin ei enää palkattu kuuroja opettajia.

⁷⁶ Laakso & Salmi 2006, 152, Kuurojen museon suomennos 24.8.1877.

⁷⁷ Laakso & Salmi 2006, 155.

Tultaessa 1920-luvulle 1921 oppivelvollisuus käytännössä tarkoitti kaikkien kuurojen opettamista puhemetodilla. Ulkomailta saadut vaikutteet johtivat siihen, että oralismi oli käytännössä ainoa valtionkouluissa käytetty virallinen kuurojen opetuksen menetelmä.⁷⁸ *Viitottujen muistojen* mukaan vuodet 1920–1970 edustavat aikaa, jolloin kuuroutta pyrittiin aktiivisesti poistamaan rotuhygienian periaatteiden mukaisesti esimerkiksi avioliitto-oikeuden poistamisen muodossa. Tätä aikaa ei tarkemmin käsitellä tässä tutkielmassa, mutta kehityskulku on havaittavissa esimerkiksi puhemetodin vakiintumisena kuurojen koulujen pääasiallisena opetustapana.

2.2 Lähdeaineiston tarkastelu

Digitoitujen sanoma- ja aikakauslehtien käyttö historiantutkimuksen aineistona on yleistynyt 2000-luvulla voimakkaasti. Kansalliskirjaston digitoitujen lehtien käyttöliittymän pitkäjänteinen kehitystyö on tehnyt lehdistä yhden kaikkein käytetyimmän aineistotyypin.⁷⁹ Sanoma- ja aikakauslehtien käyttöön digitaalisessa muodossa liittyy useita tiedonhakuun,⁸⁰ - käyttöön ja laatuun⁸¹ sekä muuntuneeseen materiaalisuuteen liittyviä ongelmia.⁸² Tässä tutkielmassa lähdeaineiston kerääminen perustuu Kansalliskirjaston sanahakuun ja työkaluun korpuksen muodostamiseksi. Korpuksen rakentamisessa ei ole pyritty tarkkaan rajaamiseen ja mallintamiseen,⁸³ koska tavoitteena on kerätä mahdollisimman laaja aineisto sanahaualla. Korpus on tällöin raakadatan kaltainen tutkimusaineisto, jota analysoidaan R-koodilla tarpeellisen tiedon muodostamiseksi.

Sanoma- ja aikakauslehtiaineistoja tarkasteltaessa kuuroihin liittyvä kontekstuaalinen aineisto ei anna lähtökohtaisesti merkittävää näyttöä kärsimyksen tai sorron kokemuksista. Koska kyseessä on erittäin laaja aineisto, 76 943 tekstiä, on selvää, että sen summittainen lähilukeminen harvemmin tuottaa osumia. Sanomalehtikorpus on yleisilmeeltään neutraali. Vaikka käsittelemäni aineisto sisältää kuurojen koulunkäyntiä koskevia nimikkeitä, kuten *Kuuromykkäin lehden* ja *Aistivalliskoulujen lehden* niiden informaatio edustaa vain pientä osaa kaikista osumateksteistä. Kuurojen itsensä tuottama aineisto on toteavan oloista,

⁷⁸ Laakso & Salmi 2006, 172–175.

⁷⁹ Hakkarainen et al. 2022, 105.

⁸⁰ Hakkarainen et al. 2022, 111–113.

⁸¹ Hakkarainen et al. 2022, 111–112, 118–119.

⁸² Hakkarainen et al. 2022, 106–107.

⁸³ Hakkarainen et al. 2022, 114–115.

neutraalia tekstiä. Kuten yleensä korpusaineistoa tutkittaessa, suuri osa sattumanvaraisesti valituista sanahaun tuloksista on itse haettavaan aiheeseen löyhästi liittyviä tyyppitekstejä kuten yleiset ilmoitukset, Kuurojen liiton kokouksien ajankohdat, paikallisissa kuurojen kouluissa järjestetyt rokotukset, vaalit tai vastaavia tutkielman aiheeseen kuulumattomia aihepiirejä.⁸⁴

Kansalliskirjaston korpus tuo myös perspektiiviä lehtien yleisilmeeseen. Sanahaulla ”kuuromyk” koostuvasta 1890–1922 sanomalehtikorpuksista löytyy yli 3000 *Kuuromykkäin lehdessä* olevaa tekstiä, jolloin yleistysten tekeminen on hyvin hankalaa sen suhteen, miten kuurot *yleensä* puhuivat itsestään tai miten voimakkaasti kuulevien ääni todella kuuluu lehdessä. Esimerkiksi käyttämällä tarkempaa hakusanaa ”döfstumsskol” sanahaun tulokseksi tulee uudelleen samoja edellisessä kappaleessa mainittuja yleisluontoisia aiheita esimerkiksi kuurojen koulujen käyttämisestä rokottamispisteinä.⁸⁵ Kuten kappaleessa 2.4.2 selviää, mahdolliset ”väärät positiiviset” osumat eivät tarkkaan luettuna ole välttämättä vääriä.⁸⁶ Myös ilmaisut kuten ”kuuroille korville” käsittävät merkittävän osan yksittäisistä lehtiteksteistä. Kuurojen oma aineisto on lisäksi hankalasti havaittavissa korpusaineistosta. Kursiivisesti lukien tekstien keskeisiä teemoja ovat kuurojen omien aineistojen kohdalla osin samankaltaisia, kuin mitä Muiluvuori on havainnut Pro gradussaan.⁸⁷ Pelkästään *Kuuromykkäin lehden* ilmoitusluontoiset tekstit, kuten *Kuuromykkäinliiton* osoiteilmoitukset⁸⁸ muodostavat huomattavan enemmistön teksteistä omilla hakusanoillani, joten yksittäisten tekstien ulkopuolella on vaikea todeta kovin tarkkoja yleistyksiä sen suhteen, miten ylipäättään kuuroihin suhtauduttiin yksittäisten tekstien perusteella.

Toinen merkittävä seikka sanomalehtien lähiluvun suhteen liittyy lehtien volyyymiin ja kuurojen omaan ääneen. Käyttämilläni hakusanoilla on saatu koottua yli 70 000 lehtitekstiä 32 vuoden ajalta, joista kuuromykkäin lehden tekstejä on noin 4000 kappaletta. Suurin osa teksteistä käsittelee kuuroja jossain yhteyksissä, mutta on selvää, että vääristä positiivisista osumista, kuten ilmoituksista, sanonnoista⁸⁹ tai vain osittain asiaan liittyvistä teksteistä huolimatta kuuroista on puhuttu huomattavasti laajemmin, kuin pelkästään *Kuuromykkäin lehden* sivuilla. Vaikka mukaan otettaisiin *Aistiviälliskoulujen lehti* noin 3000 osumatekstinsä

⁸⁴ Esim. Hufvudstadsbladet 22.5.1908, nro 139, s. 4.

⁸⁵ Esim. Åbo Tidning 6.7.1890, nro 179, s. 4.

⁸⁶ Esim. Hufvudstadsbladet 19.10.1901, nro 283, s. 3.

⁸⁷ Lisää alaluvussa 3.1.

⁸⁸ Helsingin ja ympäristön osoite- ja ammattikalenteri 1.1.1922, s.1078.

⁸⁹ Esim. Kuuroille korville.

kerä, siltikin suurin osa teksteistä on peräisin päivälehdissä, kuten *Helsingin Sanomista* ja *Hufvudstadsbladetista*.

Tässä tutkielmassa ei ole otettu kantaa siihen, miten Kansalliskirjaston digitaalisia sanomalehtikorpuksia tulisi rajata, jotta aineistosta voisi löytää jonkinlaisen punaisen langan sen suhteen, mitkä tekstit liittyvät aidosti oralismin leviämiseen Suomessa. *Maahan lämpimään* sisältää laajan katsauksen oralismin keskeisiin vaiheisiin, joita on kuvattu *Aistivälliskoulujen lehden* kuulevien opettajien kirjoituksissa sekä vastaavasti *Kuuromykkäin lehden* teksteihin viittaamalla. Näissä toisen käden lähteissä käsitellään kuurojen koulunkäyntiä, viittomakielen opettamisen korvaamisesta puhemetodilla⁹⁰ ja kuuromykkien yhdistyksen näkemyksiä puhemetodin vakiintumisesta ja mahdollisista merkityksistä kuurojen yhteiskunnalliseen asemaan.⁹¹ Muiluvuoren pro gradu -tutkielman analyysi opettamiseen liittyvästä diskurssista löytyy myös omasta kollokaatioanalyysistäni. Kielestä riippumatta kuurojen koulut, kuurot koululaiset ja kuurojen oppiminen ovat läsnä useassa kollokaatioanalyysin esille nousseessa teemassa.⁹²

Tutkielmassani korpusaineiston etäluku tapahtuu tietokoneavusteisesti R-ohjelmointikielellä toteutetun kollokaatioanalyysin avulla. Analyysillä on koostettu ja järjestetty eri kollokaatioita sen perusteella, miten todennäköisesti muodostettu sanapari ilmenee ainoastaan kyseisessä muodossa. Sanojen läheisyys toisilleen perustuu tällöin uniikkeihin poikkeuksiin, eli seuraavassa alaluvussa käsiteltävä sanaparien yhdistelmä ei todennäköisesti löydy missään muussa yhteydessä. Sekä suomen- että ruotsinkielisen korpuksen kohdalla on muodostettu lista viidestäkymmenestä korkeimman log-dice-pisteityksen saaneesta kollokaatioparista.

Muiluvuoren pro gradu -tutkielmassa on tarkasteltu *Kuuromykkäin lehden* tekstejä esimerkkeinä auktoriteettien holhoavasta suhtautumisesta kuuroihin. Hänen käyttämässään diskurssianalyyseissä kuurojen opettamiseen liittyvissä teksteissä kuurojen kyvyttömyys ja heidän avun tarpeensa kuulevilta ovat näkyvillä myös *Kuuromykkäin lehden* teksteissä.⁹³ Diskurssianalyysejä varten valitut esimerkkitekstit edustavat ensi sijassa kuulevien näkemystä kuuroista. On kuitenkin huomattava, että *Kuuromykkäin lehden* teksteistä puuttuu usein tiedot niiden kirjoittajista. Esimerkiksi Muiluvuoren tarkemmin analysoima 1.8.1908 julkaistu teksti ”Sananen kuuromykkäin kuuleville ystäville ja auttajille” ei sisällä tietoa kirjoittajasta.

⁹⁰ Laakso & Salmi 2006, 166.

⁹¹ Laakso & Salmi 2006, 194.

⁹² Tarkemmin alaluvuissa 2.4.1 ja 2.4.2.

⁹³ Muiluvuori. 2025, 32.

Kuurojen keskuudessa puhemenetelmään suhtauduttiin toisaalta myönteisesti, koska sen katsottiin parantavan sosiaalisia mahdollisuuksia, mutta toisaalta hyvinkin kielteisesti, koska puheesta alkoi muodostua oralismin aatteiden mukaisesti kuuroja arvottava ominaisuus. Tämä on käynyt esille myös *Maahan lämpimään* esimerkeistä puhemenetelmällä oppineiden kuurojen ja viittomakielellä kommunikoivien kuurojen kohtaamisissa.⁹⁴

2.3 Kollokaatiot digitaalisissa ihmistieteissä ja metodin tulokset

Kollokaatioanalyysi on määrällisen korpustutkimuksen varhaisimpia metodeja, joka perustuu J.R. Firthin teoriaan sanojen hahmottamisesta niiden konseptien mukaan.⁹⁵

Kielentutkimuksessa kollokaatioiden analyysillä on havainnointu muun muassa sanojen taipumuksia esiintyä yhdessä, sanojen merkitysyhteyksiä sekä aineistokokonaisuuksien tutkimista kollokaatiohavaintojen määrällisen analysoinnin avulla.⁹⁶

Käyttämäni kollokaatioiden analyysi perustuu ajatukselle analysoitavien sanojen kanssa esiintyvistä *noodista* eli keskellä olevasta *ytimeistä*,⁹⁷ jonka oikealla ja vasemmalla puolella olevat sanat muodostavat yhdessä kollokaatioparit:

”synnynnäisiä_taipumuksia_kuuromykkyyteen”, eli ”synnynnäisiä_taipumuksia” ja ”taipumuksia_kuuromykkyyteen.” Noodit muodostavat myös useasti ympärillään olevien sanojen kanssa ketjuja, joissa noodi muuttuu vieressä sijaitsevaksi sanaksi:

”perheessänne_synnynnäisiä_taipumuksia”. Uudemmissa kollokaatioanalyyseissa suositaan noodin ympärille asetettavaa ikkunaa, eli itse tarkasteltavan sanan oikealla ja vasemmalle puolelle määritellään tietyn mittainen ikkuna, jolloin kollokaateiksi voidaan määrittää mikä tahansa sanapari ikkunan sisällä. Esimerkiksi sanalle ”synnynnäisiä” hypoteettisesti määritelty kahden mittainen ikkuna olisi ”heillä_oli, oli_synnynnäisiä, synnynnäisiä_taipumuksia, taipumuksia_kuuromykkyyteen”. Tutkielmassani käytetään kuitenkin perinteistä bigrameihin eli sanapareihin perustuvaa tapaa kollokaattien määrittämiseen: ”Tämä_ja”, ”ja_se”.

Hypoteesini on, että aineiston koostuessa teksteistä, jotka rakentuvat tietyille hakusanalle, on mahdollista saada runsaasti lisää tietoa tarkastelemalla hakusanan välittömässä yhteydessä olevia kollokaatioita. Olen myös jakanut analysoidut kollokaatiot vielä trigrameiksi eli

⁹⁴ Laakso & Salmi 2006, 194.

⁹⁵ ”You shall know a word by the company that it keeps.” Firth 1957, 6.

⁹⁶ Kyllönen & Laippala, 2020, 496.

⁹⁷ Kyllönen & Laippala, 2020, 496, Ivaska 2015, 31.

kolmen sanayksikön kokonaisuudeksi ”synnynnäisiä” – ”taipumuksia” – ”kuuromykkyyteen”. Tällöin bigramien keskimmäistä sanaa, esimerkkitapauksessa ”taipumuksia”, on mahdollista analysoida koodin avulla koko aineistosta, mutta trigrameille ei ole tehty erillistä log-dice-pisteytystä.

Tilastollisena tunnuslukuna käytän *log-dice-pisteytystä*⁹⁸, joka kuvaa, kuinka eksklusiivisesti tietyt sanat esiintyvät toistensa läheisyydessä. Tämän suhteen laskeminen tapahtuu antamalla kollokaatioille arvoja 0–14 väliltä. Laskukaavan⁹⁹ mukaan lasketaan kahden sanan (w_1 ja w_2) yhteisen esiintymisen frekvenssi suhteessa kummankin sanan kokonaisfrekvenssiin.

Yhteisesiintymisten määrä kerrotaan kahdella ja jaetaan kummankin sanan yksilöllisten esiintymien summalla. Sen jälkeen osamäärään lisätään logaritimuunnos ja skaalataan tulokset 0–14 välisten lukujen jatkuvaan asteikkoon. Korkean arvon pisteytyksiä ilmaisevat, että kyseiset sanat esiintyvät lähes poikkeuksetta vain toistensa läheisyydessä.

Yhdistelmä ”synnynnäisiä_taipumuksia_kuuromykkyyteen” on aineistoni kollokaatioista ainoa, joka on saanut 14-arvoisen log-dice-pisteytyksen. Tämä tarkoittaa, että tilastollisesti ”taipumuksia_kuuromykkyyteen” sanaparia edeltää lähes poikkeuksetta sana ”synnynnäisiä” eli tulokset voivat olla uniikkeja yhdistelmiä tai poikkeuksellisen vakiintuneita yhdistelmiä. Tällöin kollokaatioanalyysi toimii tiedonhaun menetelmänä, jolla sanahakua rikastetaan lisäämällä haettavalle sanalle sen läheisyydessä esiintyviä poikkeavuuden konteksteja tai sen ympärillä olevia hyvin vakiintuneita sanoja.

Log-dice-pisteytys ei käytännössä reagoi korpuksien sanamääriin, jolloin tyypilliset korkean frekvenssin yhdistelmät, kuten ”sekä_että,” tai ”hän_ja_minä” eivät ole yliedustettuina dataa visualisoidessa. Tarkoituksena on etsiä uniikkeja sanapareja hyödyntämällä hakusanoja, joilla korpuksia on koostettu, eli tässä tutkimuksessa kollokaatioiden muodostamiseksi käytetyt sanat ovat ”kuuromykkä”, ”döfstum” ja ”dövstum.” Esimerkiksi tarkemmassa analyysissä tutkitun ”Johtaja_Killinen” kollokaation ruotsinkielinen versio ”döfstumskoleföreståndaren_Killinen” löytyy sanan ”döfstum” 50 korkeimman log-dice-pisteytyksen saaneista kollokaatioista.¹⁰⁰

Yhdessä log-dice-pisteytysten tulosten ja R-ohjelmointikielen visuaalisointien kanssa kollokaatioilla on mahdollista löytää korpuksen tekstimassassa piileviä yhteyksiä, joita on

⁹⁸ Englanniksi pelkkä log-dice.

⁹⁹ Liite 2.

¹⁰⁰ 11.87259.

mahdotonta havaita esimerkiksi laskemalla datan yleisimmät sanat. Vaikka kyseessä on tilastollinen tunnusluku, en ole käyttänyt log-dice-pisteityksen arvoja tutkimustuloksina vaan tiedonhaun menetelmänä. Korkeat arvot auttavat hahmottamaan sanojen eksklusiivisia yhteyksiä. Yhdistämällä Kansalliskirjaston käyttöliittymän korpuksen rakentamisessa käytetyn hakusanan kollokaation muodostajaksi on myös mahdollista löytää koko korpusta kuvailevia sanapareja.

Kollokaatioiden analyysi tapahtuu R-ohjelmointikielellä, joka soveltuu erityisesti tilastolliseen data-analyysiin.¹⁰¹ Kansalliskirjastosta ladatut Excel-korpuksia on mahdollista ladata suoraan datakehikoksi,¹⁰² joka on vektoririveistä ja -sarakkeista muodostuva tietorakenne. Kehikko on aikajärjestyksessä, eli automaattinen luetteloitu alkaa vanhimmasta lehtitekstistä ja päättyy uusimpaan. Suurin osa tietokoneavusteisista analyyseistä kohdistetaan pystyriviin numero kuusi, ”text”, mutta itse analysoitava kehikko koostuu myös metadataa sisältävistä sarakkeista. Metadatana käytetään Kansalliskirjaston lataustyökalusta saatuja tietoja ja kollokaatioanalyysin tuloksia. Pystyriveittäin tulokset ovat: kollokaatioparit (collocation pair), lehden päänimeke (päänimeke), ilmestymispäivänmäärä (pvm), lehden numero (nro), sivu, jolta kollokaatio on löytynyt (sivunumero), log-dice-pisteitys (log dice) ja tekstintunnistettu teksti (text snippet).¹⁰³

Tulos esitetään taulukkona, jossa kollokaatioiden muodostamiseksi teksti on jaettu bigrameiksi eli kahden sanan pareiksi. Noodin muodostamiseksi bigramit yhdistetään oikealta ja vasemmalta. Seuraavassa esimerkissä on esitetty molempien analyyseiden kymmenen korkeimman log-dice-arvon saanutta kollokaatiota

¹⁰¹ Kopra & Tikka 2022, 17.

¹⁰² ”Dataframe” mm. Kopra & Tikka 2022, 32.

¹⁰³ Liite 3: Kuvat 2 & 3.

collocation	
1	wuosimaksun_joka joka_kuuromykkä-laitoksissa
2	idiotismia_epilepsiaa epilepsiaa_kuuromykkäisyyttä
3	tawasi_samalla samalla_kuuromykkä-aakkosten
4	mielihyvää_yli-ikäisten yli-ikäisten_kuuromykkäraukkojen
5	alkunsa_pieksämäen pieksämäen_kuuromykkäkotikin
6	kuuromykkäinhoitaja_maria maria_rasasen
7	kuuromykkäyhdistystä_olemassa olemassa_kolmiwuotisella
8	kuuromykkäkoulukokouksesta_turussa turussa_innsbruckin
9	waltioawulla_kannatettu kannatettu_kuuromykkäinkoulunsa
10	synnynnäisiä_taipumuksia taipumuksia_kuuromykkyyteen

collocation	
1	döfstumskolan_i i_åbo
2	döfstumskolan_i i_borgå
3	vid_döfstumskolan döfstumskolan_i
4	döfstumskolan_i i_kuopio
5	för_de de_döfstumma
6	döfstumskolan_i i_jakobstad
7	förmån_för för_döfstumma
8	döfstumskolan_i i_uleåborg
9	för_döfstumma döfstumma_i
10	för_döfstumskolan döfstumskolan_i

Ylimmässä esimerkissä i on kollokaation noodi, döfstumskolan vasen kollokaatio ja åbo oikea kollokaatio. Näistä laaditaan kollokaatiovisualisointi, joka on muodostettu valitsemalla yksittäinen kollokaatioissa esiintyvä sana, 50 korkeimman log-dice-pisteystystä saanutta bigram-paria ja käyttämällä visualisoinnin apuna Fruchterman-Reingoldin- menetelmää.¹⁰⁴ Tässä visualisoinnissa x- ja y-akseleilla ei ole kollokaatioiden sijainnin kannalta merkitystä, sen sijaan aseteltu perustuu toistensa kanssa samankaltaisten arvojen ajautumisesta toistensa läheisyyteen. Bigramien oikea ja vasen puoli muodostavat pisteet, jotka yhdistyvät toisiinsa linjalla. Välimatka ja klusterit määräytyvät suhteessa toisiinsa log-dice-pisteityksen

¹⁰⁴ Fruchterman & Reingold 1991, 15.

perusteella. Klustereihin sijoittuvat kollokaatit saavat yleensä samoja log-dice-pisteytyksiä ja sisältävät visualisoidun yhdistelmän lisäksi muita jaettuja sanapareja, tai yhteisen semanttisen kontekstin.¹⁰⁵ Keskimäinen yksikkö klustereissa kuvastaa noodia, koko kuvastaa noodin frekvenssiä. Mitä paksumpi sanat yhdistävä linja on, sitä korkeampi kollokaatioiden log-dice-pisteytyks on, eli periaatteessa sanojen välillä on voimakkaampi yhteys. Keskimäinen sininen noodi kertoo, kuinka monessa kollokaatiossa kaikki siihen liittyvät kollokaatiot esiintyvät.¹⁰⁶ Yksiköiden välillä ei ole semanttista yhteyttä vaan niiden sijainti perustuu Fruchterman-Reingoldin- menetelmän matemaattiseen algoritmiin. Koska kollokaatioiden log-dice-pisteytyks ei perustu niiden esiintyvyyshfrekvenssiin, normaalit visualisoinnit tuottaisivat useita toisistaan irrallaan olevia bigrameja. Fruchterman-Reingoldin menetelmä on vain apukeino visualisoida kollokaatioita, klustereilla ei ole lopulta asiayhteyden perustuvia liitoksia, vaan kielen samankaltaisuuden yhteneväisyys.

Tarkastelin kollokaatioita vertaamalla visualisointeja listattuihin 50 korkeimman log-dice-pisteytyksessä saavaan bigramiin¹⁰⁷. Olisi myös hyvin mahdollista käyttää suurempia arvoja, mutta tällöin nykyiset visualisointimenetelmät muuttuvat liian epäselviksi tulkittaviksi. Visualisoinnin muodostamisessa on myös mahdollista käyttää muita sanoja, kuin *kuuromyk** tai *döfstum** mutta en ole perehtynyt oman aineistoni kanssa tarkemmin tällaisiin tuloksiin.

Kollokaatioiden listaamisen ja visualisoinnin jälkeen tarkastelin kursiivisesti, mitkä kollokaatioista muodostuvat bigramit olivat mielenkiintoisimpia. Aloitin analysoimisen ruotsinkielisten kollokaatioiden korkeimman arvon saaneista ”*döfstumma-blinda + döfstumma-andesvaga*”- yhdistelmistä. Tein tämän listaamalla kaikki bigramit, joissa esiintyvät sanat ”*döfstumma*” Tarkastelemalla pelkästään julkaisuaikaa ja julkaisujen nimikkeitä huomasin, että kollokaatio *döfstum-andesvaga* esiintyi huomattavan usein *Aistivalliskoulujen lehdessä*¹⁰⁸ teksteissä. Luvuissa 2.3.1 ja 2.3.2 on kuvattu tarkemmin, miten kollokaatio yhdessä tekstien kanssa johdatti tekstien lähiluvulla yhdestä sanaparista toiseen. Lähtökohtana oli kuitenkin korkeimman log-dice-pisteytyksen saaneen sanaparin tarkastelu.

¹⁰⁵ Bezina et al. 2024, 17.

¹⁰⁶ https://r.igraph.org/reference/layout_with_fr.html

¹⁰⁷ Liite 3: kuvat 4 ja 5.

¹⁰⁸ Suomen aistivalliskoulujen lehti 1.4.1893, nro 4, s. 1.

Tuloksena saadut visualisoinnit ja log-dice-pisteytykset

Kollokaatioiden visualisoiminen graaffiin pisteytyksien mukaan antaa mahdollisuuden hahmotella erilaisia teemoja koko aineistosta. Osa kollokaatioista on keskittynyt selvästi tiettyihin historian tapahtumiin,¹⁰⁹ kun taas keskellä näkyvä Kustaa Killistä koskeva kollokaatiopari viittaa häneen henkilönä. Näitä tutkielman tuloksena saatuja visualisointeja ei ole tarkasteltu alkuperäisiä sanoma- ja aikakauslehtiaineistoja lähilukemalla lukuunottamatta ”döfstumma-blinda + döfstumma-andesvaga kollokaatioparia ja siitä johdettua ”döfstumma-andesvaga-” kollokaatiota. Kuten alaluvussa 2.3 on havainnoitu, yksittäinen kollokaatiopari liittyy tosiasiaissa satoihin aihepiireihin. Tässä vaiheessa analyysia on hyvin vaikea tehdä, koska jokaisesta rivistä ja parista voidaan muodostaa ainakin kaksi erillistä omaa kollokaatiotaan: Esimerkiksi *Hufvudstadsbladetin* numerossa 139 ”Döfstumma-blinda” saa vasemmalle puolelleen sanan ”för” ja oikealle puolelleen sanan ”och” kun taas ”döfstumma-andesvagan” oikealla puolella ei kyseisessä tekstissä ole sanaa vaan piste ja vasemmalla puolella on vastaavasti sana ”och”.¹¹⁰ Vastaavasti listan ylimpänä olevassa *Borgåbladetin* tekstissä ” döfstumma-blinda” kollokaation vasemmalla puolella on aina sana ”för”, mutta ” döfstumma-andesvagan” oikealla puolella löytyviä sanoja on kaksi, ”beslutit” ja ”äfvansson.”¹¹¹

¹⁰⁹ Kuurojen konferenssit Pariisissa, 1900 *Tidskrift för döfstumma* 1.9.1899, nro 21, s. 15.

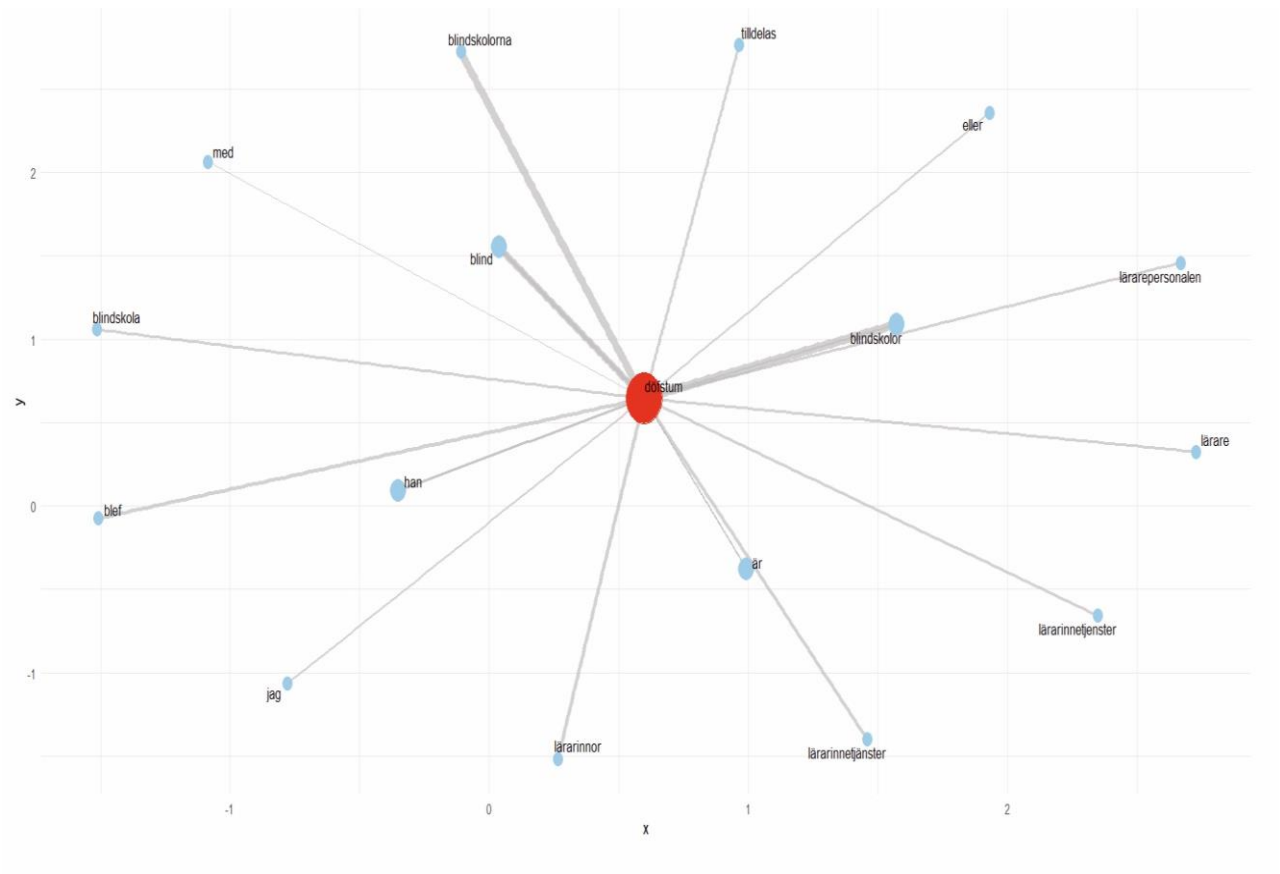
¹¹⁰ *Hufvudstadsbladet* 22.5.1908, nro 139, s. 4.

¹¹¹ *Borgåbladet* 26.5.1908, nro 59, s. 2.

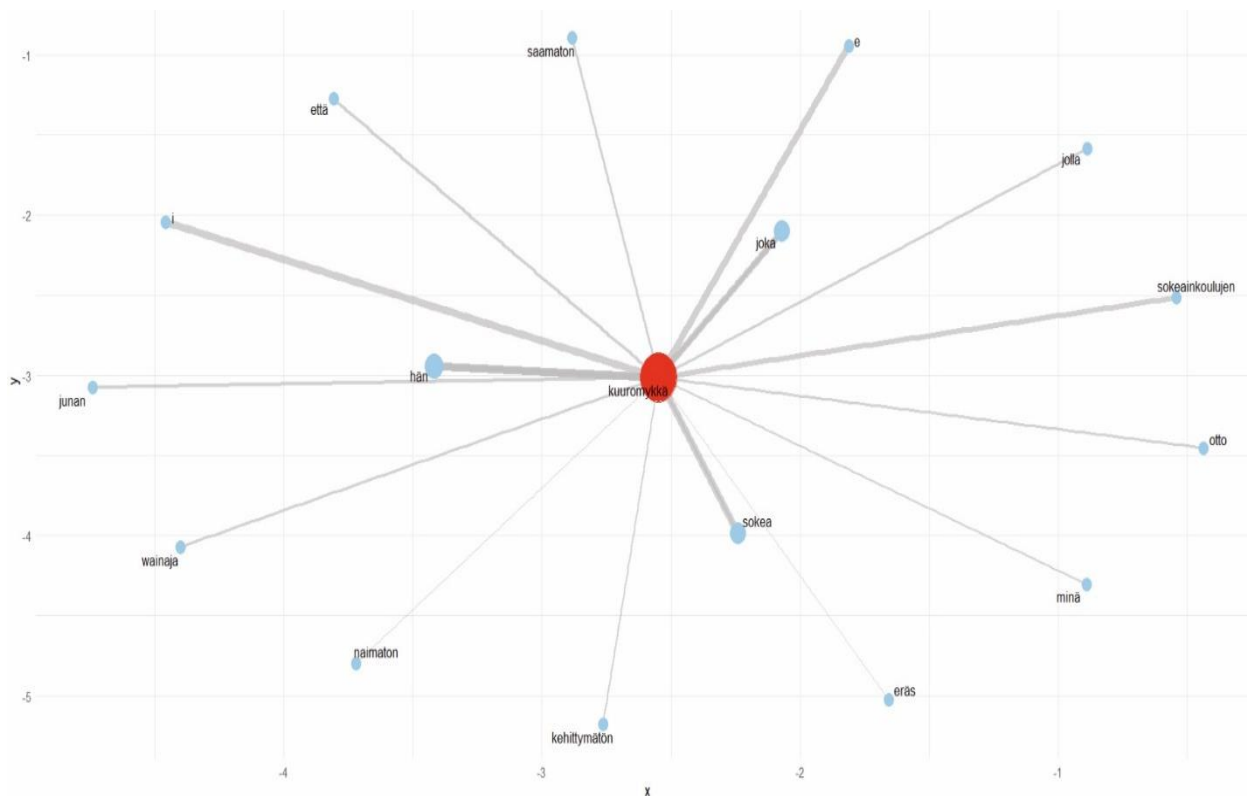
	collocation_pair	Päinimeke	PVM	Nro	Sivu	log_dice
1	döfstumma-blinda + döfstumma-andesvaga	Borgåbladet	1908-05-26	59	2	13.54057
2	döfstumma-blinda + döfstumma-andesvaga	Finlands Allmänna Tidning	1908-05-25	119	1	13.54057
3	döfstumma-blinda + döfstumma-andesvaga	Finlands Allmänna Tidning	1915-10-09	234	1	13.54057
4	döfstumma-blinda + döfstumma-andesvaga	Hufvudstadsbladet	1908-05-22	139	4	13.54057
5	döfstumma-blinda + döfstumma-andesvaga	Hufvudstadsbladet	1910-02-18	47	4	13.54057
6	döfstumma-blinda + döfstumma-andesvaga	Hufvudstadsbladet	1915-10-08	276	4	13.54057
7	döfstumma-blinda + döfstumma-andesvaga	Nya Pressen	1908-05-22	137	3	13.54057
8	döfstumma-blinda + döfstumma-andesvaga	Nya Pressen	1910-02-18	39	3	13.54057
9	kuustö + döfstumslärrarinnan	Finland	1891-11-11	262	2	13.51457
10	kuustö + döfstumslärrarinnan	Hufvudstadsbladet	1891-11-11	307	2	13.51457
11	kuustö + döfstumslärrarinnan	Nya Pressen	1891-11-12	308	2	13.51457
12	kuustö + döfstumslärrarinnan	Åbo Tidning	1891-11-12	308	2	13.51457
13	kuustö + döfstumslärrarinnan	Åbo Underrättelser	1891-11-12	308	2	13.51457
14	utger + döfstumtidskrift	Borgå Nya Tidning	1899-01-31	8	2	13.32193
15	utger + döfstumtidskrift	Hufvudstadsbladet	1899-02-01	31	6	13.32193
16	utger + döfstumtidskrift	Vasabladet	1899-02-04	15	3	13.32193
17	utger + döfstumtidskrift	Västra Nyland	1899-02-03	9	2	13.32193
18	utger + döfstumtidskrift	Östra Finland	1899-02-01	25	3	13.32193
19	döfstumanstalerna + 1891	Aamulehti	1892-03-06	55	3	12.75714
20	döfstumanstalerna + 1891	Finland	1892-03-04	53	3	12.75714
21	döfstumanstalerna + 1891	Hufvudstadsbladet	1892-03-04	62	3	12.75714
22	döfstumanstalerna + 1891	Hufvudstadsbladet	1892-03-05	54	3	12.75714

Tämä tarkoittaa, että pelkästään kollokaatioparit edustavat hyvin pientä osaa kaikista osumista. Ne eivät kerro, mikä kollokaatioparien keskellä oleva noodi on. Jatkoisin kuitenkin seuraavien tulosten tulkitsemista, koska analyysin tarkoituksena ei ole saada tarkkoja tilastollisia tunnuslukuja, joiden avulla aineistosta voidaan tehdä yleistyksiä, vaan apukeinoja uusien tutkimustulosten löytämiseksi. Tarkentamalla hakua muotoon ”döfstumma-andesvaga” kollokaatiopareja löytyi kuitenkin vain 8, jotka kaikki liittyivät pitkälti koulukotiin.

Toinen tapa käyttää visualisointia on keskittää yksi sana keskelle graafia ja tarkastella sen kollokaatioita:



Ruotsinkielisten kollokaatioiden visualisointi



Suomenkielisten kollokaatioiden visualisointi

En lopulta hyödyntänyt keskitetyn kollokaation analyysia enempää tutkielmassa. Keskeinen yhteenveto on, että ruotsinkielisessä aineistossa kuurojen ja koulun yhteys on ilmeistä¹¹² kuten myös yhteys sokeiden opetukseen. Suomenkielisen datasetin visualisointi osoittaa, että 1900-luvun alussa kuuro-sanan yhteyteen on liittynyt negatiivista kuvailua. Tämän tutkimuksen aikana ei ole tutkittu tarkemmin, mistä erot suomen- ja ruotsinkielisten visualisointien välillä johtuvat. Keskitetyissä kollokaatioissa on myös matalammat log-dice-pisteetykset kaikille osumille, koska nykyisellä pisteetyksen laskutavalla tarkastellaan vain kahden sanan välistä suhdetta bigramien sijasta.¹¹³ Esimerkiksi ”döfstumma_blinda” pisteytys on vain noin 8,89, jolloin kyseessä on lopulta melko tavanomainen yhdistelmä. Yksittäisten kollokaatioiden valinnassa ei hyödynnetty tilastollisia tunnuslukuja tai visualisointeja, vaan valinta tapahtui lähiluvun avulla.

Seuraavana askeleena pilkoin kollokaatioparit omiksi yksittäisiksi kollokaatioiksi. Tässä kohtaa en enää käyttänyt log-dice-pisteetystä, vaan turvauduin täysin aineiston lähilukuun.

¹¹² Esim ”lärapersonaler” ja ”lärare.”

¹¹³ Liite 3: kuva 6.

Minua kiinnosti erityisesti ”döfstumma-andesvaga” – sanaparina, ei yhdistelmänä. Pilkkomalla sanaparin sain suuren määrän tuloksia. Tämä tutkimusvaihe on täysin eksploratiivinen ja kokeellinen perustuen lähiluettavaan materiaaliin. Suurin syy valintaan oli, että nyt saaduista osumista osa oli samoja *Maahan lämpimään* oralismia käsittelevän luvun lähteinä käytettyjen sanomalehtitekstien kanssa. Tuloksia on analysoitu tarkemmin alaluvussa 2.3.1.

Päänimeke	PVM	Nro	Sivu	text
Suomen aistivalliskoulujen lehti	1892-04	2	13	andesvaga. En af komiténs medlemmar, direktorn för döfstu...
Suomen aistivalliskoulujen lehti	1892-04	2	24	vid stadens döfstumskola vackert klädda grafven, framträd...
Suomen aistivalliskoulujen lehti	1892-08-15	4	10	f) att ordna och bearbeta de uppgifter rörande alla döfstu...
Suomen aistivalliskoulujen lehti	1892-08-15	4	14	andesvaga. Vid samlandet af primäruppgifter öfver t. ex. döf...
Suomen aistivalliskoulujen lehti	1893-02	1-2	21	abnorme Börns Undervisning" af år 1881. § i- För döfstumm...
Suomen aistivalliskoulujen lehti	1893-03	3	6	just på abnormskolans område. Ty för abnorma barn, döfstu...
Suomen aistivalliskoulujen lehti	1893-09	9	9	bygder barn och unga hoptals strömma till läroverken. Döfs...
Suomen aistivalliskoulujen lehti	1893-12	12	18	12. I hvilken riktning bör en utveckling af vår döfstumskola ...
Suomen aistivalliskoulujen lehti	1894-12	12	13	för Örebro döfstumskola i Sverige, J. Blomkvist utgifven läro...
Suomen aistivalliskoulujen lehti	1895-01	1	15	Skolpligtslagens § 1 lyden: Icke fullsinnade (blinda, döfstu...
Suomen aistivalliskoulujen lehti	1898-01	1	2	Kirj. N. Helsingius 57. Ännu ett ord om de andesvaga döfstu...
Suomen aistivalliskoulujen lehti	1898-02	2	1	frågan om hvad som borde göras för våra andesvaga döfstu...
Suomen aistivalliskoulujen lehti	1898-03	3	6	eller både döfstuinma och blinda, ske i skolor för döfstumm...
Suomen aistivalliskoulujen lehti	1898-03	3	13	aistivallisten luettelossa. 1) Statistiska uppgifter ang. döfstu...
Suomen aistivalliskoulujen lehti	1898-05	5	1	om hvad som borde göras för våra andesvaga döfstumma e...
Suomen aistivalliskoulujen lehti	1898-05	5	2	utomlands finnas speciella anstalter för döfstumma andesva...
Suomen aistivalliskoulujen lehti	1898-05	5	3	tillgäng på bildbara, enbart döfstumma elever, till fullo hunn...
Suomen aistivalliskoulujen lehti	1898-05	5	4	Behöfva de döfstumma skolundervisning och kunna de and...
Suomen aistivalliskoulujen lehti	1898-05	5	5	vore hvarken döfstumskolorna eller de hjälpbehöfvande döf...
Suomen aistivalliskoulujen lehti	1898-05	5	6	större nytta af att förblifva inom den egentliga döfstumskol...
Suomen aistivalliskoulujen lehti	1898-05	5	7	med andra ord just den undervisning, som man på döfstum...
Suomen aistivalliskoulujen lehti	1898-05	5	8	vällottad döfstum inträde i sina skolor. Så länge känddome...
Suomen aistivalliskoulujen lehti	1902-02	2	3	af rätten. Emedan det ganska ofta torde hända, att döfstum...
Suomen aistivalliskoulujen lehti	1902-12	9	3	HEDMAN, döfstumskoleförest. förest, för anstalten för ande...
Suomen aistivalliskoulujen lehti	1904-04	4	1	Om utbredningen af döfstumhet, blindhet och andesvaghet...
Suomen aistivalliskoulujen lehti	1904-04	4	2	skolväsendet af den 4 maj 1897, att statistiken rörande de d...
Suomen aistivalliskoulujen lehti	1904-04	4	3	allmänna befolkningsstatistiken äfven redogör för antalet d...

Kaikki kollokaatioyhdistelmät on mahdollista sijoittaa takaisin teksteihin, joista ne ovat lähtöisin. Tällöin kollokaatioita on mahdollista tarkastella osana lähdetekstejä lähiluvun avulla. Samoin on mahdollista seurata eri sanayhdistelmien esiintymistä koko korpusaineiston kaikissa yksittäisissä teksteissä koko ajanjakson ajalta. Tekstien kiertoa eli saman tekstin toistumista on ollut lähiluetuissa lähteissä vaihtelevasti.

Kollokaatioiden välillä on olemassa yhteyksiä, esimerkiksi tarkemmin tarkasteltavaksi valittu döfstumma-andesvaga- kollokaatioon liittyvissä teksteissä Kustaa Killinen esiintyy nimeltä mainittuna, tekstien kirjoittajana sekä kollokaationa. Korkean log-dice-pisteityksen saaneiden

tekstien teemat vaihtelevat huomattavasti: Toisaalta tekstit ovat ilmoitusluontoisia uutistekstejä, jotka kertovat kuurojen opetuksen tilanteesta 1900-luvun alussa esimerkiksi koulukodin hylättyä apurahahakemusta raportoivassa tekstissä¹¹⁴ sekä kuurojen opetukseen annettavien määrärahojen hylkäämispäätöksissä.¹¹⁵ Toisaalta mukana on laajoja toisessa luvussa käsitellyn *Aistivalliskoulujen lehden* kuurojen opettajien ja kouluvaikuttajien laatimia oppimista käsitteleviä tekstejä.¹¹⁶ Edellä mainitut esimerkit ovat suoria kollokaationalyysin avulla saatuja tuloksia.

Tapa, jolla käytän tutkimusaineistoani tietokoneavusteisesti, vie kokonaisuuden lähemmäksi digisyntyistä tietoa. Käytän tiedonhaun laadun kriteereinä *Archival IR*- artikkelissa mainittuja tiedonhaun menetelmiä: *Sanahakua, tekstin määrällistämistä sekä tietokoneavusteista pisteyttämistä*.¹¹⁷ Aineistoni sitoutuu alkuperäisten hakusanojen ympärille siten, että jokaisesta tekstistä löytyvät jossain muodossa sanat ”döfstumma” ja ”kuuromykkä” taivutuksineen. Itse tiedonhaku perustuu kollokaatiotestin pisteytykseen. Tiedon visualisoiminen ja järjestäminen log-dice-pisteytyksen perusteella tuottaa työkaluja, joiden avulla on mahdollista valita tutkimusaiheen kannalta kaikkein mielenkiintoisimmat kollokaatiot. Haku perustuu painotettuun indeksiin eli aputietorakenteeseen, jonka järjestys perustuu log-dice-pisteytykseen.¹¹⁸ Kollokaatiot pisteytyksineen muodostavat sanahaulle järjestystä kuvaavan numeerisen metatiedon, jonka avulla löysin *Aistivalliskoulujen lehden* kirjoituksista yhteneviä tietoja Kustaa Killisen ja ”döfstumma-andesvaga” kollokaatioiden välillä. Tämä osoittaa, että sanahaun rikastaminen kollokaatioanalyysillä paljastaa piileviä teemoja ja yhteyksiä korpusaineistossa olevien tekstien välillä.

Muiden tietokoneavusteisten analyysien avulla voisi mahdollisesti tutkia paljon tarkemmin, millaisia kokonaisuuksia eri kollokaatiot muodostavat. Suora tekstien analyysi kollokaatioista saatujen tuloksien perusteella on lopulta lähempänä Frédéric Clavertin etä- ja lähiluvun yhdistämistä¹¹⁹ kuin täysin tietokoneavusteista digitaalisen datan tutkimusta. Käytännössä tutkimusaineistoa voidaan tarkastella etäältä visualisoinnin ja tilastollisten tunnuslukujen avulla, mutta tarkempi analyysi vaatii lähilukua. Koska Kansalliskirjaston digitoiduissa aineistoissa on tarkkoja metatietoja esimerkiksi tekstien julkaisuajankohtaan liittyen, koodissa

¹¹⁴ Borgåbladet 26.5.1908, nro 59, s. 2.

¹¹⁵ Tidskrift för döfstumma 1.3.1910, nro 3, s. 12.

¹¹⁶ Esim. Suomen aistivalliskoulujen lehti 1.5.1910, nro 5, s. 15.

¹¹⁷ Furner & Gilliland 2017, 603–605 & 615–617.

¹¹⁸ Kumpulainen 2025, 121.

¹¹⁹ Mähkä et al. 2022, 119.

on myös analysoitu milloin tietyt kollokaatiot ovat alkaneet esiintyä aineistossa ja kuinka pysyviä ne ovat olleet. Tätä analyysia ei ole tehty tarkemmin muiden kuin Kustaa Killiseen liittyvien kollokaatioiden osalta.¹²⁰

Metodini kadottaa kuitenkin samalla tekstien alkuperäisten analogisten ominaisuuksien merkityksen, jos ajatellaan esimerkiksi tekstien sijoittumista lehden sivuilla, niiden palstatilaa ja toisaalta koko lehden kontekstia. Samoin lehtien kontekstia ei ole myöskään pohdittu tarkemmin. Esimerkiksi *Kuuromykkäin lehti* oli kuurojen lehti, mutta sitä ei käytännössä huomioida ollenkaan tässä tutkimuksessa. Sen sijaan arkistojen hiljaisuus eli kysymys siitä, jääkö jonkin tietyn ryhmän ääni aineistossa kuulematta esimerkiksi ryhmän vähemmistöaseman takia, on mielestäni otettava esille aineiston tarkastelun rajautuessa tiettyihin valittuihin kollokaatioihin. Kollokaatioanalyysi antaa huomattavan erilaisia tuloksia esimerkiksi Kustaa Killisen kollokaation osalta verrattuna *Tidskrift för döfstumma*-¹²¹ hakuihin liittyviin osumiin.¹²² Kuten aiemmin mainittu, Killinen on hahmo, joka nousee esille hyvin monen kollokaation lähdetekstien tarkastelussa etsittäessä kuuroihin kohdistuneita vääryyksiä, kun taas kuurojen omaa ääntä on huomattavan paljon vaikeampi hahmottaa korpusaineistosta riippumatta siitä, tarkastellaanko tekstejä lähiluvulla vai etäältä.

Lopullisen kollokaatioanalyysin taustalla on tutkimusprosessin alkuperäinen kysymys pohtia oppivelvollisuuden vaikutusta kuurojen yhteisöön. Alkuasetelma johti aineiston ajalliseen rajaukseen, jonka myötä myös nykyisen tutkimusprosessin asetelma on vakiintunut aineiston aikarajauksessa sekä sanoma- ja aikakauslehtien korpuksessa. Vaikka tutkimusasetelman painotus on muuttunut uuden tiedon löytämistä tiedonhaun tuloksien arviointiin, suunnittelemani analyysi tuottaa silti käytettäviä tuloksia molempien kannalta. Toisin kuin muut kollokaatioanalyysit, log-dice-pisteitys ei ole yhtä herkkä reagoimaan aineiston kokoon ja sanojen absoluuttiseen frekvenssiin,¹²³ mutta metodi ei ole myöskään täysin riippumaton sanojen absoluuttisista määristä.

¹²⁰ Alaluku 2.3.4.

¹²¹ Kollokaatiopari” tidskrift_för_för_döfstumma.”

¹²² Liite 3: kuva, 6.

¹²³ Vaclav & Platt 2021, 61–62.

2.4 Kollokaatioiden tarkastelu

Alun perin ajatukseni oli tutkia listan kaikki kollokaatioparit alkuperäisissä yhteyksissään, mutta lähdettyäni tarkastelemaan ruotsinkielisen korpuksen korkeimmat log-dice-pisteytyksen saanutta osumaa, ”döfstumma-blinda och döfstumma-andesvaga,” kävi ilmi, että kyseisestä yhdistelmästä irrotetut kollokaatiot ”döfstumma-blinda” ja ”döfstumma-andesvaga” muodostavat omanlaisensa kiinnostavat tekstikokonaisuudet.

Kollokaatioyhdistelmään liittyvät tekstit käsittelevät pääasiassa Ossian Wichmanin rahanomusta kuurojen koulun perustamiseksi ja anomuksen hylkäämistä.¹²⁴ Koska yhdistelmän osumia oli vähän, aloin tutkia tarkemmin bigramin oikeaa puolta, ”döfstumma-andesvaga” -yhdistelmää kollokaationa ”döfstumma andesvaga”. Tämä sanapari ei ole käyttämälläni analyysilla noussut tilastollisten tunnuslukujen valossa merkittäväksi tulokseksi, mutta näiden kollokaatioiden lähdetekstien tarkastelu tuotti hedelmällisiä löytöjä. Yksi listan ensimmäisistä osumista oli *Suomen aistivalliskoulujen lehti*, 1.5.1898, nro 5. Lehdessä pohdittiin E. Hedmanin kirjoituksessa kehitysvammaisten kuuromykkien tai kuuromykkien kehitysvammaisten opettamisesta saatavia hyötyjä.¹²⁵ Kirjoituksessa käytetään erityisoppilaitoksista ilmaisua ”epänormaalikoulut”, jossa opiskelevista ”saattoi tulla ihmisiä”. Lehden tekstit vaikuttivat teemoiltaan samankaltaisilta kuin *Viitottujen muistojen* kuvaukset oralismin kannattajien holhoavasta ja määräävästä asemasta kuuroja kohtaan. Tästä syystä lähdin perehtymään tekstin perusteella tarkemmin muihin *Suomen aistivalliskoulujen lehdestä* löydettäviin ”döfstumma-andesvaga”-kollokaatioihin.

2.4.1 Döfstumma-andesvaga – kollokaatio ja *Suomen aistivalliskoulujen lehti*

Suomen aistivalliskoulujen lehti perustettiin vuonna 1892. Lehden päätoimittajana toimi tuolloin Porvoon koulun rehtori ja Suomen aistivalliskoulujen tarkastaja Valter Forsius. Vastaavina toimittajina olivat Edvin Hedman, Hanna Ingman ja Kustaa Killinen. Lehdessä käsiteltiin yleisesti erityisten oppijoiden opettamisen metodiikkaa, opintomatkoja esimerkiksi manner-Eurooppaan ja ulkomailta saatujen pedagogisten vaikutteiden hyödyntämisestä Suomen koulujen erityisopetuksen kehittämisessä.¹²⁶ Kuuromykkien ja tylsämielisten eroja ja yhtäläisyyksiä on korpuksen tekstien perusteella nostettu esille yhteyksissä, joissa pohdittiin

¹²⁴ Esim. Hufvudstadsbladet 22.5.1908, nro 139, s. 4.

¹²⁵ ”Andesvaga döfstumma eller döfstumma andesvaga — huru man vill se saken.”

¹²⁶ Esim. Suomen aistivalliskoulujen lehti 1.4.1892, nro 2, 15.

yleisesti aistiviallisten opetusta.¹²⁷ Kuurouden ja tylsämielisyyden yhteen niputtaminen vaikuttaisi tapahtuneen juuri *Suomen aistivialliskoulujen lehden* teksteissä. On huomattava, että lehden toimittajat olivat aikansa auktoriteetteja erityisoppijoiden koulutuksen parissa. Forsius ja Killinen¹²⁸ olivat kunnostautuneet kuuromykkien koulutuksessa, Hedman taas toimi tylsämielisten Perttulan koulun johtajana ja Ingman sokeain koulun ylijohdajatar. Lehden teksteissä on voimakas kristillissävytteinen holhoamisen ja pelastamisen sävy, jonka Muiluvuori on havainnut tarkastellessaan *Kuuromykkäin lehden* tekstejä.¹²⁹

Aistiviallisten opettamiseen liitettiin ilmaisia kuten ”taakka” ja ”kristillinen velvollisuus”. Koulujen opettajien työ oli samalla oppilaiden suojelemista muun maailman epäoikeudenmukaisuuksilta ja valmistamista kohtaamaan nämä olosuhteet kasvun myötä. Lehdessä vaadittiin ”koko yhteiskunnan” huomiota

”velvollisuuteen tehdä minkä tehdä voi niiden kohtalon lieventämiseksi, jotka kaikkiviisas, mutta myöskin armias taivaallinen Isä on lähettänyt meidän keskuuteemme koetellakseen rakkauttamme Jumalaan ja lähimmäiseen.”¹³⁰

Kristinuskon huolenpidon ja holhoamisen lisäksi *Aistivialliskoulun lehden* voi katsoa alusta lähtien nojautuneen rotuhygieeniin vaikutteisiin. Sama teksti, joka painotti koko yhteiskunnan huolenpitoa, toteaa aistiviallisten olevan mahdollisesti tuleva taloudellinen taakka koko yhteiskunnalle, joten oikeudenmukaisuuden lisäksi opettamisella nähtiin olevan ennalta ehkäiseviä vaikutuksia.¹³¹ Useat lehteen kirjoittaneista henkilöistä olivat käyneet opintomatoilla Keski-Euroopassa. Näitä matkoja myös raportoitiin välillä usean numeron mittaisissa teksteissä. Rotuhygieniä ei jäänyt pelkästään ideaalitasolle, esimerkiksi *Aistivialliskoulujen lehteen* 1800-luvun lopulla kirjoittanut Edvin Hedman ja hänen vaimonsa sovelsivat 1910-luvulla avoimen rotuhygieenisiiä otteita kuten pakkosterilisaatioita Perttulan tylsämielisten koulussa.¹³²

Kuurojen kohdalla oralismi ja puhemetodi jäsenyivät hyvin samantapaisesti osittain laupiaaksi velvollisuudeksi, mutta sitä kuvattiin lehdessä kaikkein tehokkaimpana, mannereurooppalaisena opetustapana, jolla kuurot saattoivat kehittyä lähes tavallisten

¹²⁷ Uleåborgsbladet 16.4.1892, nro 31, s. 2.

¹²⁸ Killisestä lisää alaluvussa 2.4.2.

¹²⁹ Muiluvuori 2025, 40.

¹³⁰ Suomen aistivialliskoulujen lehti 15.8.1892, nro 4, s. 11.

¹³¹ Suomen aistivialliskoulujen lehti 15.8.1892, nro 4, s. 13.

¹³² Heinonen 2021, 29.

ihmisten tasolle. On merkittävää huomata, että kuuroihin liittyi jo 1800-luvun puolella rotuhygienisiä ajatuksia. Esimerkiksi A. Björkmanin tekstissä¹³³ perustellaan puhekyvyn olevan ihmisen tärkein älyllinen ilmaisukeino. Viittomakieltä kritisoitiin etenkin Kustaa Killisen toimesta, mutta vielä pahempaan nähtiin puhekyvyn puuttuminen. Puhumattomuutta verrattiin sairauteen¹³⁴ ja ylipäättään ääneen puhumista pidettiin keskeisimpänä tavoitteena kuurojen koulunkäynnissä. Vuoden 1892 kahdessa viimeisessä numerossa kirjoittanut J. Blomkvist kirjoitti pitkät tekstit puheen luontevuudesta osana ihmismäistä toimintaa liittäen puhemetodin oppimisen osaksi ihmisyyttä. Hänen mukaansa ”kuuroja lapsia ei voi opettaa samoin, kuin täysjärkisiä” mutta opettamisessa puhemetodin ”keinoin ei ole mitään luonnotonta.”¹³⁵ On kuitenkin huomattava, että puhemetodia perusteltiin 1800-luvun lehdissä todisteena kuurojen kyvystä oppia puhumaan, eli ajattelemaan ”täysjärkisten” tavalla. Kuuroja piti toki holhota, mutta heistä saattoi tulla ”kunnan kansalaisia” sekä ”uutteria ja tunnollisia työntekijöitä”, jotka pystyivät omat rajoituksensa tiedostaen¹³⁶ toteuttamaan *Aistivalliskoulujen lehteen* kirjoittavien kuulevien auktoriteettien toiveita.

Toinen puhemetodin keskeinen teema oli, että kuurojen erottuminen edukseen muista aistivallisista mitattiin ensisijaisesti puhetaidossa. Siinä, missä Malm oli painottanut suomen ja ruotsin kielen kirjottamisen sekä lukemisen opettelua viittomakielen rinnalla, *Aistivalliskoulujen lehdessä* kuuroja kannustettiin harjoittelemaan puheen ääntämistä, hampaiden hoitoa¹³⁷ ja hengittämistä. Lehden kirjoittajat ilmaisivat useasti kuurojen soveltuvan parhaiten ruumiillisten töiden pariin ja taidot kuten lukeminen ja kirjoittaminen nähtiin tiettyjen kirjoittajien suunnalta haitallisiksi, koska ne veivät huomiota pois keskeisistä oppimissisällöistä, eli puheen tuottamisesta ja ymmärtämisestä huulilta lukemalla.¹³⁸

Lain tasolla Vaivahoitoasetuksessa kuuroja kohdeltiin useassa yhteydessä paremmin, kuin muita aistivallisia. Kuurojen ei myöskään nähty olevan läheskään yhtä huonossa asemassa verrattuna esimerkiksi sokeisiin, vaikkakin opetusvaikuttajien mielessä kuuroja täytyi auttaa ”aineellisesti ja henkisesti”.¹³⁹ Kuitenkin vuoden 1892 keisarin asetus Aistivalliskouluista enteili jo tulevasta muutoksesta kuurojen asemaan. Asetuksen kohdassa h kehotetaan

¹³³ Suomen aistivalliskoulujen lehti 1.8.1893, nro 7–8, s. 2.

¹³⁴ Suomen aistivalliskoulujen lehti 15.1.1892, nro 4, s. 45.

¹³⁵ Suomen aistivalliskoulujen lehti 15.1.1892, nro 5, s. 5.

¹³⁶ Suomen aistivalliskoulujen lehti 1.4.1893, nro 4, s. 12.

¹³⁷ Suomen aistivalliskoulujen lehti 1.2.1911, nro 2, s. 6–7.

¹³⁸ Suomen aistivalliskoulujen lehti 1.5.1898, nro 5, s. 10.

¹³⁹ Laakso & Salmi 2006, 188.

”huolehtimaan siitä, että kuuromykät, sokeat ja tylsämieliset saavat laissa tai muutoin säädetyn koulutuksen, kasvatuksen tai hoidon.”¹⁴⁰

Tultaessa 1910-luvulle *Aistivalliskoulujen lehd*en lisäksi kuurot, sokeat ja tylsämieliset niputettiin yhä useammin samaksi ryhmäksi. Tämä johtui osittain siitä, että aistivalliset haluttiin keskittää omiin oppilaitoksiinsa,¹⁴¹ ja lehdissä puhuttiin yleensä aistivallisista.¹⁴² Toisaalta vaikuttaa siltä, että raja kuuromykkien, tylsämielisten ja kuuromykkätylsämielisten välillä näkyi lopulta enää erityislaitoksissa, kuten Helsingin kuuromykkäsokeiden ja kuuromykkätylsämielisten koulusta.¹⁴³ ”Döfstumma-andesvaga”-ja ”dövstumma-andesvaga” kollokaatiota löytyy edelleen 1910-luvulla useista *Aistivalliskoulujen lehd*en numeroista, mutta myös päivälehdistä esimerkiksi *Åbo Underrättelseristä* ja kuurojen omista lehdistä.¹⁴⁴ Kuurojen omien lehtien kohdalla useassa kollokaatiossa oli kyse konkreettisesti ”kuuromykistä tylsämielististä” eikä kuurojen niputtamisesta yhteen muiden aistivallisten kanssa.¹⁴⁵ Kuurojen niputtaminen yhteen tai kuuroista puhuminen ”kuuromykkinä-tylsämielisinä” ei sinänsä ole poikkeuksellista. Esimerkiksi Ison-Britannian sanomalehtiaineistossa ”deaf and dumb” oli hyvin tavanomainen termi 1800-luvulla kuuroista puhuttaessa. 1910-luvulla kuurojen fyysisiin ominaisuuksiin, kuten hengittämiseen ja puhetaidottomuuden liittyvä osittain pseudotieteellinen ja rotuhygieenisesti latautunut keskustelu levisi myös *Aistivalliskoulujen lehd*en sivuilta sanomalehtiin¹⁴⁶ samoin kuin säälinsävyiset tekstit.¹⁴⁷

Edeltävät esimerkit valottavat kuurojen sorron olleen pitkälti puhemetodin kehittäjinä ansioituneiden koulun ja opetuksen ammattilaisten toimesta johtuvaa. Killisen ja hänen kollegoidensa aktiivisena pyrkimyksenä oli oikeuttaa puhemetodi ainoana sallittuna opetustapana kuurojen kouluissa. Oralismien mukaan puhumisen oppiminen oli ainoa keino, jolla kuurot sopeutuisivat yhteiskuntaan. Puhemetodin mukana tulleet oralismien asenteet alkoivat vakiintumaan myös kuurojen yhteisön keskuudessa.¹⁴⁸ Yhteisö alkoi hiljaa omaksua kouluissa opetetun asemansa, josta ainoa tie ylöspäin oli puheen oppiminen. *Maahan*

¹⁴⁰ Suomen aistivalliskoulujen lehti 15.8.1892, nro 4, s. 10.

¹⁴¹ Hufvudstadsbladet 28.1.1910, nro 27, s. 3.

¹⁴² Suomen aistivalliskoulujen lehti 1.2.1910, nro 2, s. 3.

¹⁴³ Hufvudstadsbladet 15.9.1916, nro 250, s. 4.

¹⁴⁴ Esim. Tidskrift för döfstumma 1.6.1898, nro 12, s. 12.

¹⁴⁵ Esim. Tidskrift för döfstumma 1.8.1916, nro 8–9, s. 3.

¹⁴⁶ Suomen Urheilulehti 29.10.1914, nro 5, s. 6.

¹⁴⁷ Esim. Kodin kuvasto 25.3.1917, nro 13, s. 4.

¹⁴⁸ Katsui et al. 2021, 148.

*lämpimään*¹⁴⁹ tuo esille erityisesti sen, että yhä useammalle kuuroille ajatus yhteiskunnasta vieraantumisen aiheutti tosiasiaa yhteisön sisäisen konfliktin. Osa yhteisöstä halusi sopeutua puhemetodin ehdoilla yhteiskuntaan, jolloin juuri viittomakielestä tuli leimallisesti välteltävä ”tylsämielisyyden” osoittava rasite. Suomessa puhemetodia hyödyntäneiden laitosten sisäiseksi tavoitteeksi ajettiin aistivalliskoulujen komitean toimesta kuurojen lasten toisistaan vieraannuttaminen tekemällä viittomakielestä tuomittavaa.¹⁵⁰

On kuitenkin huomattava, että *Aistivalliskoulujen lehdessä* kirjoituksista ei voida suoranaisesti todeta kirjoittajien keskittyvän kuurojen ”tylsämielisyyteen” vaan puhemetodin ylivertaisuuteen. Forsiuksen esimerkkiä seuraten kuurojen opettajat argumentoivat lehdessä, että muualla Euroopassa saadut tulokset puhuivat puhemetodin tehokkuuden puolesta. Samalla opetuksen keskeisin tavoite oli sulauttaa kuurot yhteiskuntaan, sillä puhetaito oli kaikki kaikessa myös Suomen puhemetodia käyttävien opettajien keskuudessa. Kuuroja tuli ohjata ja avustaa, jotta he voisivat tulla ”synkstä erikois-asemastaan kohotetuksi edes itsetietoisien ihmisten lukuun”.

Kollokaation tarkentaminen ”döfstumma andesvaga”- muotoon auttoi hahmottamaan samoja teemoja kuin *Maahan lämpimään*¹⁵¹ ja *Viitottujen muistojen*¹⁵² esimerkeissä oralistisen opetuksen soveltamisesta kuuroihin. Kollokaatioiden avulla löytyneet lähdetekstit osoittavat opettajien nähneen puhemetodin ylivoimaisesti parhaana tapana opettaa kuuroja. Kuurous nähtiin heikkoutena ja yksilöiden taakkana, jota puhumaan oppiminen saattoi keventää muttei poistaa kokonaan. Samalla kaikkien muiden aineiden opettamista karsittiin kovalla kädellä. Koko kouluaika palveli lopulta puhumaan oppimista ja puheen ymmärtämistä.

Sekä *Viitotut muistot*, että *Maahan lämpimään* korostavat puhemetodin opettajien kielteistä suhtautumista viittomakieleen.¹⁵³ Tätä ilmiötä ei löytynyt tarkastelemistani ”döfstumma-andesvaga” kollokaatioiden lähdeteksteistä. Sen sijaan puhemetodin toimivuuden kyseenalaistamista kritisoitiin esimerkiksi Björkmanin kirjoituksissa. *Aistivalliskoulujen lehdessä* puhemetodin kritiikkiä pidettiin vastuuttomana kuurojen itsensä kannalta. Perusteluna oli, että koska puhetta pidettiin älykkyyden mittana, puheenopetuksen kieltäminen nähtiin kuurojen heitteillejättönä. *Viitotut muistot* ja *Maahan lämpimään*

¹⁴⁹ Laakso & Salmi 2006, 194.

¹⁵⁰ Katsui et al. 2021, 79.

¹⁵¹ Laakso & Salmi 2006, 164–166.

¹⁵² Katsui et al. 2021, 148.

¹⁵³ Laakso & Salmi 2006, 172–173.

osoittavat opettajien käyttäneen ruumiillista kurittamista, eristämistä ja henkistä väkivaltaa keinoina viittomakielen kitkemiseksi pois luokkatilasta. Myös luokkahuoneiden ulkopuolista oppilaiden keskinäistä kommunikaatiota viittoen pyrittiin vähentämään. Tähän liittyen *Viitottujen muistojen* ainut maininta Kustaa Killisestä koskee hänen vaatimaansa säännönmukaista kuurojen oppilaiden eristämistä toisistaan viittomakielen käytön estämiseksi.¹⁵⁴ Samoin lasten erottaminen vanhemmistaan sijoittamalla heidät asuntoloihin koulun ajaksi oli tavoiteltavaa tapauksissa, joissa kodeissa kommunikoitiin viittomalla. Vastaavasti asuntolassa asuvat oppilaat pyrittiin asuttamaan omiin huoneisiinsa, jotteivät he voisi kommunikoida vapaa-ajalla keskenään viittomalla.¹⁵⁵

”Döfstumma-andesvaga”- kollokaationa löytyi eksploratiivisesti kokeilemalla, minkälaisia lähdetekstejä bigramin ”döfstumma-blinda_och_döfstumma-andesvaga” jälkimmäistä kollokaation perusteella oli löydettävissä. Löydökset eivät siis perustuneet suoraan kollokaatioanalyysin tuloksiin, vaan olivat puhtaasti heuristinen päätelmä jo saaduista tuloksista. Tulkinta johti lopulta samanlaisen kokonaiskuvan äärelle, kuin mitä Muiluvuoren gradussa ja *Maahan lämpimään* havainnoissa kuurojenkoulujen opettajien roolista puhemetodin ja oralististen ajatusten levittäjinä.

Aistivalliskoulujen lehti sisältää useita tekstejä, joista huokuu kirjoittajien omistautuminen aistivallisten ohjaamiseen. Lukiessani kollokaatioiden avulla löytyneitä lähdetekstejä huomioni kiinnittyi ensimmäisenä matkaraportteihin, joissa kuvattiin ihailevaan sävyyn esimerkiksi Saksan kuurojen opetusta. Raportit Saksan oralismista olivat moniosaisia jatkuen usean numeron yli. Näiden kirjoittaja Kustaa Killinen löytyi myös korkeimpien pisteytyksen saaneiden kollokaatioiden joukosta. Heräkkeen mielenkiinnon takia toinen lähiluettavien kollokaatioiden ketju käsittelee siis Killiseen liittyviä tekstejä.

¹⁵⁴ Katsui et al. 2021, 79.

¹⁵⁵ Katsui et al. 2021, 79.

2.4.2 Kustaa Killinen

Maahan lämpimään -teos esittelee Kustaa Killisen (1849–1922) kuurojenopettajana, suomalaisuusmiehenä, valtiopäiväedustajana ja harrastelijarunoilijana. Killinen valmistui kansakouluopettajaksi 1872. Opintomatkat Pohjoismaihin ja Keski-Eurooppaan sekä opetusharjoittelu Turun kuuromykkäinkoulussa johdattivat hänet Kuopion kuuromykkäinkoulun johtajaksi 1880. Tätä virkaa hän hoiti kuolemaansa asti. Killisen kollokaatio onkin tästä syystä ”johtaja-Killinen.”¹⁵⁶ Opetustyönsä ohella Killinen oli Kuopiossa merkittävä kulttuurihenkilö ja myöhemmin poliittinen vaikuttaja nuorsuomalaisten riveissä. Hän kuului myös Kuopion naistenyhdistykseen ja ajoi naisille äänioikeutta aloitteellaan 1897. Killinen oli Valter Forsiuksen ohessa yksi merkittävimmistä pedagogeista, jotka toivat puhemetodin Suomeen. Killisen merkittävydestä kertoo myös se, että hän esiintyy kaikissa muissa käyttämissäni sekundääreissä lähteissä, paitsi Muiluvuoren pro gradu -tutkielmassa.

Kustaa Killinen löytyy kollokaatioanalyysin avulla ruotsinkielisistä korpuksista muodossa ”döfstumskoleföreståndare_K_Killinen” ja suomenkielisestä korpusaineistosta muodossa ”johtaja_K_Killinen.” Hänen saamansa log-dice-pisteetykset eivät ole kummassakaan korpuksessa korkeimman kymmenen joukossa. Hänen nostamisensa osaksi tätä tutkimusta perustuu kollokaatioista muodostuvien tekstien sisällön sidoksisuuteen teknisen yhteenliittävyyden sijasta. Haettuna ”Killinen” esiintyy kollokaatioiden lähdeteksteissä aina vuoteen 1922 asti, mutta suurin osa esiintymistä on vuosina 1896–1905, jolloin hän kirjoitti aktiivisesti *Aistivälliskoulujen lehteen* omasta opettajuudestaan, matkoista ulkomaiden puhemenetelmäkouluihin ja ylipäätään kuurojen opettamisesta. Alun perin aloin tutkimaan Killistä koulun johtajana, mutta luettuani hänen *Aistivälliskoulujen lehden* kirjoituksiaan löysin yhteiskunnallisesti aktiivisen oralismin kannattajan.

Killinen tuotti merkittävän määrän matkapäiväkirjamaista aineistoa oppimatkoistaan. Niissä raportoidaan tarkasti matkoista Keski-Eurooppaan ja muihin Pohjoismaihin, joissa Killinen pyrki vierailemaan puhemetodia soveltavissa kouluissa. Nämä raportit, kuten valta osa muistakin 1890-luvulla Killisen tuottamista teksteistä ovat pääsääntöisesti neutraaleja ja oivaltavia, esimerkiksi liittyen arkipäiväisten tapahtumien käyttämiseen oppilaiden ja opettajan välisenä puheenaiheena kommunikaation opettamiseksi.¹⁵⁷ Killinen myös vertailee

¹⁵⁶ <https://biografiasampo.fi/henkilo/p1428>

¹⁵⁷ Suomen aistivälliskoulujen lehti 15.12.1892, nro 6, s. 14.

melko objektiivisesti Suomen kuurojen kouluja ulkomaisiin ottaen esille esimerkiksi naisopettajien vähäisen määrän Saksan ja Sveitsin kouluissa sekä painottaen samalla suomalaisten naisopettajien ammattitaitoa.¹⁵⁸ Killisen pikkutarkka ja syvällinen tietämys etenkin Saksan puhemetodista antaa kuvan opettajasta, joka pyrki hankkimaan laajasti tietoa ja jakamaan sitä muille asiasta kiinnostuneille. Teksteistä kuuluu myös ihailu puhemetodin tehokkuutta kohtaan erityisesti sen suhteen, miten hyvin kuurot voivat oppia puhumaan erilaisten harjoitusten avulla. Yleisvaikutelma ei kuitenkaan anna samanlaista kuvaa kuin esimerkiksi *Maahan lämpimään*, jossa Killinen on kuvattu merkittäväksi viittomakielen vastustajaksi ja oralismiin kiihkeäksi kannattajaksi. Tämä puoli on kuitenkin löydettävissä tarkasteltaessa Killisen pedagogisia näkemyksiä tarkemmin.

Silmiinpistävää Killisen 1800-luvun puolella kirjoittamissa teksteissä on se, mistä hän ei kirjoita: Matkakertomuksissaan Killinen ei lähes koskaan mainitse viittomakieltä kuurojen opettamiseen liittyvissä teksteissä, vaan painottaa puhemenetelmää ainoana oppimisvälineenä. Tekstissä ”Kuuromyökkäin opettajien valmistumisesta” Killinen suosittaa opettajille laajaa kielitaitoa ja etenkin saksalaiseen metodikirjallisuuteen perehtymistä. Hän myös painottaa, ettei pelkkä kansakoulunopettajan seminaarikoulutus riitä kuurojen opettajalle, vaan suosittaa matkustamaan ulkomaille seminaareihin ja aktiiviseen itseopiskeluun. Opetussisältöjen osalta hän nostaa esille hengitysharjoitukset, mutta muuten kehottaa uusia opettajia opiskelemaan yksinomaan puhemetodin teorioita.¹⁵⁹ Vaikka C.O. Malmin koulu oli tarjonnut oppilaille kirjoittamisen, lukemisen, viittomakielen ja kielen opintoja, Porvoon koulusta tai muista Suomen viittomakieltä käyttävistä laitoksista, Killinen ei näytä mainitsevan matkakertomustensa yhteydessä lainkaan.

Silloin kun Killinen kirjoittaa viittomakielestä ja sen käyttämisestä kuurojen opetuskielenä ja kuurojen keskinäisenä kommunikaation keinona, hän suhtautuu siihen hyvin kielteisesti, jopa vihamielisesti. Hän kuvailee viittomakieltä ”rumana”, ”apinamaisena” ja ”pahana”¹⁶⁰. 1909 Killinen kirjoitti normaalista raportoivasta tyylistään poiketen turhautuneesti ja vihamielisesti saksalaisen M. Schneiderin julkaisusta, jossa käsiteltiin viittomakielen vahvuuksia ja puhemenetelmän heikkouksia.¹⁶¹ Killisen mukaan kirjan näkemyksiä viittomakielestä ilmestyy ”epidemian tapaan” puhemetodia vastustavien keskuudessa. Killisen ajatus

¹⁵⁸ Suomen aistivalliskoulujen lehti 15.10.1892, nro 5, s. 10.

¹⁵⁹ Suomen aistivalliskoulujen lehti 1.8.1893, nro 7–8, s. 17–19.

¹⁶⁰ Suomen aistivalliskoulujen lehti, 1.3.1903, nro 3–4, s. 6.

¹⁶¹ Suomen aistivalliskoulujen lehti, 1.1.1909, nro 1, s. 7–11.

viittomakielestä kiteytyy sitaattiin ”Henki on se, joka eläväksi tekee, eikä viittomiset.” Hän pyrki myös estämään viittomakielen käyttöä jo sen olemassa olevissa muodoissa. Kuurojen yhteisön parissa työskentelevät matkapapit olivat toimineet pitkään viittomakielellä kommunikoiden. Kun 1916 Senaatti julkaisi ohjesäännön Suomen jakamisesta neljään kuurojen matkapapin piiriin, Killinen vastusti ajatusta. Hänen mielestään matkapapeille ei saisi myöntää piirivirkoja, koska Suomesta ei hänen mukaansa löytynyt yhtäkään pätevää henkilöä virkaa varten. *Maahan lämpimään* mukaan Killisen vastustus johtui juuri pappien viittomakielen käytöstä. Hän ei halunnut, että etenkin kouluikäiset altistuisivat enempää viittomiselle.¹⁶² Killinen näki myös, että kaikki kuurot lapset pitäisi saada mahdollisuuksien mukaan kouluun pois kotiseudultaan, koska oletettavasti vanhemmat ja lapset kommunikoivat keskenään viittomalla.¹⁶³

Killinen kehitti ja muokkasi menetelmiä kuurojen opetuksen standardisoimiseksi oralismin periaatteiden mukaisesti. *Aistivalliskoulujen lehden* teksteissä hän ja muut pedagogit viestivät tavalla, joka saa heidät vaikuttamaan pioneereilta erityisopetuksen saralla. Tämä tehtiin tuomalla esille kuurojen opettamiseen liittyviä haasteita ja kuurouteen liittyviä rajoituksia suhteessa ”normaaliin” väestöön samalla, kun Malmin koulun aikana saavutettu osaamisen taso oppiaineissa jätettiin mainitsematta. Killiselle ja muille oralismin aatteen mukaan opettaville puheen oppiminen oli edelleen tärkeintä kuurojen opettamisessa, sillä heidän mukaansa se mahdollisti kuurojen sulautumisen normaaliin väestöön. Silloin harvoin, kun Killinen on maininnut mitään aineopetuksesta, hänen kantansa on ollut yleensä kielteinen. Killinen vaikuttaa pitäneen elämänsä loppuun asti kiinni näkemyksestään, jonka mukaan kuuroilla ei ollut joko valmiuksia tai tarvetta oppia esimerkiksi matemaattisia aineita tai kirjoittamista tapauksissa, joissa oppilas ei oppinut puhumaan ja lukemaan huulilta.¹⁶⁴

Killisen toimet eivät tehneet hänestä elinaikanaan vihattua henkilöä kuurojen keskuudessa. Oli kyse sitten vaali-ilmoituksesta, opetuslautakunnan istunnosta tai *Kuuromykkäin lehden* juhla-artikkelista, hänet on lähes aina tituleerattu koulujen johtajana ja ”kuuromykkien ystävänä.”¹⁶⁵ Hänen ajatuksiaan julkaistiin myös *Kuuromykkäin lehdessä*, joista osa on jopa enemmän oralismin siivittämiä, kuin *Aistivallisten lehden* tekstit.¹⁶⁶ *Maahan lämpimässä* mainitaankin, että Killisen asema ”kuurojen ystävänä” ja luotettavana opettajana edesauttoi

¹⁶² Laakso & Salmi 2006, 218–219.

¹⁶³ Suomen aistivalliskoulujen lehti 1.3.1903, nro 3–4, s. 6.

¹⁶⁴ Suomen aistivalliskoulujen lehti 1.6.1896, nro 6–7, s. 12–22.

¹⁶⁵ Kuuromykkäin lehti 1.9.1919, nro 9–10, s. 1.

¹⁶⁶ Kuuromykkäin lehti 1.9.1912, nro 9, s. 5.

hänen asenteidensa hyväksymistä ja tiettyjen oralististen ajatusmallien normalisoitumista kuurojen yhteisön keskuudessa. *Kuuromykkäin lehdessä* kirjoitettiin 1919 Killisen olevan ”erittäin rakas ja kunnioitettu entisille oppilailleen” ja ”tehneen paljon työtä kuuromykkäin hyväksi”.¹⁶⁷

Oralismo vaikutti kuurojen sisäisesti yhteisöön koulussa opittujen asioiden siirtyessä koteihin ja yhteisön kokouksiin.¹⁶⁸ Yhteisössä alkoi ajan myötä olla viittomiseen kielteisesti suhtautuvia yksilöitä, joiden mukaan kuurojen oli opittava puhumaan tullakseen hyväksytyiksi puhuvien keskuudessa ja puhetaitoa alettiin nähdä viittomista arvokkaampana kommunikoinnin välineenä.¹⁶⁹ *Maahan lämpimään* tuo esille, että viittomakielen puolustajat vaikenivat 1910-luvulla. Osa yhteisöstä halusi sopeutua puhemetodin ehdoilla yhteiskuntaan, mutta viittomakieli miellettiin edelleen pääasialliseksi kommunikoinnin välineeksi kuurojen keskinäisessä kommunikoinnissa.¹⁷⁰ Samalla kuulevien kesken viittomakielen käyttöä alettiin vieroksua, viittomista pidettiin oralismin oppien mukaan merkinä kehittymättömyydestä. Puhetaitoiset kuurot alkoivat vierastaa heidän kanssaan viittomakielellä kommunikoivia.¹⁷¹ *Kuuromykkäin lehdessä* puhemetodilla oli vankka kannattajakuntansa. Kärjekkäimpiin näistä kuului Pekka Poikkiselän kirjoitus ”Kehotan teitä mitä pikemmin lopettamaan viittaamisen.”¹⁷²

Käytetty lähdemateriaali osoittaa Killisen olleen jyrkkä viittomakielen vastustaja.¹⁷³ *Maahan lämpimään* mukaan Killisellä oli merkittävä vaikutus kuurojen yhteisöön. Hänen näkemyksiensä mukaan kuurojen tukeutuminen toisiinsa ja viittomakieleen johti heidän eriytymiseensä yhteiskunnasta. Erityisesti viittominen teki Killisen mukaan kuuroista hylkiöitä, jotka vaihtoehtoisesti ymmärsivät ”synkän asemansa” aivan liian myöhään ollessaan jo yhteiskunnan ulkopuolella¹⁷⁴ tai joutuivat lapsena ”sopimattoman ja raa’an kohtelun alaiseksi.”¹⁷⁵ Killisen teksteissä kuurojen jääminen kristillisten tapojen, kuten

¹⁶⁷ Kuuromykkäin lehti 1.11.1919, nro 11, s. 4.

¹⁶⁸ Katsui et al. 2021, 147–148.

¹⁶⁹ Katsui et al. 2021, 148.

¹⁷⁰ Laakso & Salmi 2006, 194.

¹⁷¹ Laakso & Salmi 2006, 197.

¹⁷² Laakso & Salmi 2006, 196.

¹⁷³ Suomen aistivalliskoulujen lehti 1909, nro 1, s. 7.

¹⁷⁴ Suomen aistivalliskoulujen lehti 1.5.1892 no 2 s. 14.

¹⁷⁵ Suomen aistivalliskoulujen lehti 1.2.1892, nro 1, s. 13–14.

säännöllisen kirkossa käymisen ja esimerkiksi konfirmaation ulkopuolelle on suoraa seurausta puhumattomuudesta, johon vastaavasti liittyy ymmärtämättömyys sekä jumalattomuus.¹⁷⁶

Killisen pedagoginen näkemys selviää melko hyvin *Aistivalliskoulujen lehdessä* sivuilta: Puheen tuottaminen ja huulilta lukeminen oli kaikkein keskeisin oppimistavoite. Lehden julkaisuissa korostuvat opintomatkat ulkomaille, niiden yhteydessä tehdyt ekskursiont paikallisiin kuurojen kouluihin ja saatujen oppien kopioiminen ja mahdollinen soveltaminen Suomessa. Opintomatkojen lisäksi Killinen kirjoitti paljon oppilaiden arvioimisesta ja opetuksen fokuksista. Esimerkiksi puheen tuottaminen oikein oli aihe, johon usea *Aistivalliskoulujen lehteen* kirjoittanut kiinnitti huomiota. Näissä teksteissä korostetaan esimerkiksi ”ääntämisen pehmeyttä” ja ”puheen tasapainoa”.¹⁷⁷ Samalla puheen opetus ja puheen ymmärtäminen vahvistivat asetelman, jossa kuuro oli lähtökohtaisesti holhottava toisaalta yhteiskunnan vaatimusten osalta, toisaalta henkisistä syistä. Opettaminen puhemenetelmällä itsessään asetti kuurot muuta väestöä huonompaan asemaan, vaikka opettaja tähän ei olisikaan pyrkinyt.

Killisen oma kollokaatio ja liittyminen ”döfstumma-andesvaga”- teksteihin osoittaa, että kollokaatioanalyysin perusteella voidaan olla samoilla linjoilla *Maahan lämpimään* kuvauksien kanssa hänen merkityksestään puhemetodin tienraivaajana ja oralismin osalta viittomakielen leppymättömänä vastustajana. Muiluvuoren pro gradu -tutkielman havainto kuulevien auktoriteettien asemasta avuttomien kuurojen pelastajana sopii hyvin Killiseen.¹⁷⁸ Hänen opettajille osoittamissaan teksteissä opettajan oikeutena ja velvollisuutena on holhota kuuroja ja ohjata heidät oppimaan ensisijaisesti puhumaan. Samalla Killisen neuvot hyvästä opettajasta liittyvät pitkälti siihen, miten opettaja hallitsee teorian ja käytännön puheen opettamisen muiden opetettavien aineiden kustannuksella. ”Kaikkien aineiden tulee olla kielenopetuksen palveluksessa in alles ist Spachunterricht.”¹⁷⁹

¹⁷⁶ Suomen aistivalliskoulujen lehti 1.5.1892 no 2 s. 13–16.

¹⁷⁷ Suomen aistivalliskoulujen lehti 1.9.1905, nro 6–7, s. 19–22.

¹⁷⁸ Muiluvuori, 2025, 36.

¹⁷⁹ Suomen aistivalliskoulujen lehti 1.6.1896, nro 6–7, s. 14.

3 Tutkimusprosessin tulokset

Tutkimusprosessin¹⁸⁰ keskeiset tuotokset ovat pro gradu -tutkielma ja kollokaatioanalyysia varten kehitetty R-koodi. Koodia on muunnettu tutkimusprosessin etenemisen myötä vastaamaan eri vaiheissa erilaisia tiedonhaun tarpeita ja lopullinen versio valmistui vasta pro gradu -tutkielman viimeistelyvaiheessa.

Sisällöllisesti tutkimuksen tuotos jäsentyy kahden tarkemmin analysoidun kollokaatioanalyysin tuloksien¹⁸¹ lähilukuun ja niistä saadun tiedon arvioimiseen suhteessa sekundäärisiin lähteisiin. Arkistotieteellisellä näkökulmalla on keskeinen asema pohdittaessa, minkälaista tietoa kollokaatioanalyysilla voidaan muodostaa. Tässä tutkielmassa tarkemmin analysoitujen kollokaatioiden lähdetekstit ovat pääsääntöisesti linjassa muun aihetta koskevan tutkimuksen kanssa.

On kuitenkin huomioitava, että nyt tutkitut kollokaatiot edustavat hyvin kapeaa osaa kokonaisuudesta ja siksi on hyvin vaikea arvioida, miten tarkastelun ulkopuolelle jäävät kollokaatiot lähdeteksteineen olisivat linjassa aiemman tutkimustiedon kanssa, paljastaisivatko ne kenties täysin uusia näkökulmia kuurojen historiasta vai olisiko kyseessä täysin aiheeseen liittymättömästä materiaalista. Käyttämäni tietokoneavusteisesti muodostettu kollokaatioaineisto muodostaa oman osakokonaisuutensa kaikista korpuksen 76 943 tekstistä. Kaikkia kollokaatioita ei ole mitenkään mahdollista käydä läpi, minkä takia ei ole perusteltua väittää tutkielman tarjoavan kaiken kattavaa ratkaisua digitaalisten aineistojen tutkimisen keskeisiin haasteisiin, esimerkiksi aineiston valtavaan määrään rajoittamiseen. Log-dice-pisteytys on yksi tapa arvottaa analyysin muodostamia kollokaatioita, kuten tässä tutkimuksessa on tietyiltä osin tehty. Jos aineistosta haluttaisiin tehdä tarkempia analyyseja tilastollisten tunnuslukujen perusteella, analyysin tuloksena listatuista lähdeteksteistä pitäisi lähiluvun perusteella poistaa vähintään tekstintunnistusvirheiden¹⁸² takia mukaan tulleet väärät osumat. Esimerkiksi tekstintunnistuksesta¹⁸³ johtuvat virheet näkyvät suomenkielisten kollokaatioiden keskitetyssä visualisoinnissa¹⁸⁴ ”i” ja ”0” – sanoina. Kysymyksiin, kuten kuinka paljon relevantteja osumia tutkimuksen kannalta oikeastaan on korkeiden log-dice-pisteytysten kollokaatioissa ja mikä vastaavasti on asiaan kuulumattomien tekstien osuus, ei

¹⁸⁰ S.10, kuva: Tutkimusprosessikaavio.

¹⁸¹ Alaluvut 2.4.1 & 2.4.2.

¹⁸² Tekstintunnistuksesta tarkemmin alaluvussa 3.2

¹⁸³ Tästä enemmän alaluvussa 3.2.

¹⁸⁴ Kuva: Suomenkielisten kollokaatioiden visualisointi.

ole pystytty vastaamaan tässä tutkimuksessa. Sen sijaan löydettyjen tietojen vastaavuutta aikaisempaan tutkimukseen ja tiedonhaun ominaisuuksia on arvioitu tarkemmin seuraavissa alaluvuissa.

3.1 Tapaustutkimuksen analyysi ja kerätyn tiedon luonne

Digitaalisen historiantutkimuksen parissa tehdyissä tapaustutkimuksissa on usein ideana käyttää jotakin tietokoneavusteista ”työkalua,” jolloin tietokoneella analysoidusta aineistosta tai tutkittavasta ilmiöstä tehdään ymmärrettävämpi ihmiselle.¹⁸⁵ Tutkimuksen voidaan haluta tuottavan menetelmän avulla¹⁸⁶ täysin uutta tietoa tai vastaavasti menetelmää testataan tiettyjen haluttujen tulosten saavuttamiseksi.¹⁸⁷ Tässä tutkielmassa keskitytään menetelmän kokeiluun ja tulosten arviointiin vertaamalla niitä aiempiin kuuroihin ja oralismiin liittyviin tutkimuksiin.¹⁸⁸

Itse tutkimusprosessi muotoutui pitkälti aineiston valitsemisella ja koodin suunnittelulla. Humanistisia tieteitä opiskelevan näkökulmasta koodin tuottaminen perustuu ensisijaisesti toivottujen lopputulosten saavuttamiseen ja uuden tiedon löytämiseen. Menetelmän teknisen toteutuksen onnistuminen perustuu uuden tiedon hankintaan, jolloin koodin tuottaminen on perustunut heuristiseen kokeiluun. Koodin laatiminen tapahtui kuitenkin ennen tutkimusnäkökulman tarkentumista. Koodin muokkaamisen lähtökohtana ei ole ollut tekninen laatu vaan mielenkiinto uuden tiedon löytymiselle ja analysoimiselle. Näkökulma siihen, ovatko tulokset linjassa muun tutkimuksen kanssa ja kuinka tarkasti, muodostui vasta tutkimusprosessin myöhempien vaiheiden aikana.¹⁸⁹ Tietokoneavusteisen metodin tulosten analysointi johti tilanteeseen, jossa etä- ja lähiluvun yhdistelevää lukutapaa¹⁹⁰ olisi pitänyt soveltaa useisiin satoihin kollokaatioyhdistelmiin. Se ei ollut käytännössä mahdollista tässä pro gradu -tutkielmassa.

Tutkimuksen alusta lähtien tavoitteena oli kehittää tietokoneavusteinen menetelmä, joka toimisi aputyökaluna tiedon hakemiselle, jäsentelylle ja uuden tiedon löytämiselle. Ajatus

¹⁸⁵ Esim. ”Tekstien uudelleenkäyttö suomalaisessa sanoma- ja aikakauslehdissä 1771–1920. Digitaalisten ihmistieteiden näkökulma” Ranta et. all. 2022.

¹⁸⁶ Hill & Tolonen, 2021.

¹⁸⁷ Elo & Kiiskinen, 2022.

¹⁸⁸ ”Grounded theory.”

¹⁸⁹ S. 10, kuva: ”Tutkimusprosessikaavio.”

¹⁹⁰ Hakkarainen et al. 2022, 118–119.

kollokaatioanalyysistä oli peräisin vuoden 2024 Helsingin digitaalisten ihmistieteiden hackathonista, jossa metodia hyödynnettiin koneoppimisen tuotosten visuaalisessa mallintamisessa. Menetelmä itsessään perustuu digitaalisen historian tutkimuksen paradigmaan tietokoneavusteisuudesta laajojen aineistojen tarkastelun keinona ja uudenlaisten tutkimushypoteesien ja tulkintojen muodostamisen apuvälineenä,¹⁹¹ mutta tuloksia tarkastellaan myös arkistotieteellisesti tiedonhaun kannalta. Tästä näkökulmasta kyseessä on myös tutkimus, miten sanomalehdet toimivat digitoituina eli digitaalisina ilmentyminä. Tällöin arvioidaan, miten ja minkälaista tietoa kollokaatioanalyysi tuottaa, miten informaatio on käytettävissä ja mitä tietoja jää mahdollisesti metodin ulkopuolelle.¹⁹² Koska kyseessä on digitoitu aineisto, alkuperäisestä analogisesta kontekstista ei voi päästää täysin irti. Tarkempi lähilukuun perustuva tarkastelu ja vertailu muuhun tutkimukseen tapahtuu edelleen lähiluvun ja etäluvun vuorottelulla, ”kahteen kertaan”¹⁹³ tarkastellen.

Tiedon todenmukaisuuden tarkistamiseksi muodostettuja kollokaatioita on verrattu edeltävien lukujen tapaan historiategokseen *Maahan lämpimään* ja valtioneuvoston selontekoon *Viitottuihin muistoihin* kuurojen kokemista vääryyksistä. Tämän lisäksi tuloksia on verrattu Milla Muiluvuoren pro gradu -tutkielmaan. Hänen tutkielmassaan on päästy hyvin samankaltaisiin lopputuloksiin paljon pienemmällä aineistolla: Kuurojen kokemien vääryyksien takana voidaan sanoa olleen kuurojen opettajia, joista 1900-luvun puolella suurin osa oli kuulevia.¹⁹⁴ Edellä mainitut kolme teosta toimivat vertailuaineistona, eli tarkoituksena on käydä vielä tarkemmin läpi, mitkä kollokaatioiden lähdeteksteistä nousseet ilmiöt voidaan todeta samankaltaisiksi teemoiltaan suhteessa vertailuaineistoon. Koska kollokaatioista on tässä tutkielmassa käsitelty hyvin pieni osa, ei vertailun perusteella voi tuottaa varmaa tietoa esimerkiksi mahdollisesti analyysin ulkopuolelle systemaattisesti jäävistä teemoista.

Maahan lämpimään on yleisesitys kuurojen menneisyydestä. Teoksen kolmas luku käsittelee 1900-luvun ”ensimmäisiä vuosikymmeniä” aikana, jolloin oralismi levisi Suomeen.¹⁹⁵ Teoksessa viitataan useasti sanomalehtiaineistoon, kuten *Kuuromykkäin lehteen* ja *Aistivalliskoulujen lehteen*. Teoksen yleiskuva ajasta on kontekstuaalisesti hyvin lähellä kollokaatioanalyysin avulla löydettyjä tekstejä ja yhden keskussanan mukaisia

¹⁹¹ Hakkarainen et al. 2022, 119–120.

¹⁹² Furner & Gilliland 2017, 583–584.

¹⁹³ Hakkarainen et al. 2022, 119.

¹⁹⁴ Muiluvuori 2025, 39.

¹⁹⁵ Laakso & Salmi, 2006, 16.

kollokaatioita.¹⁹⁶ Koulu oli keskeisessä asemassa määriteltäessä kuurojen toimijuutta ja yhteiskunnallista asemaa 1900-luvulla.¹⁹⁷ Se, miten hyvin kuuro yksilönä pystyi vastaamaan puhemetodin vaatimuksiin, vaikutti opiskelujen etenemiseen, ”lahjakkuuden” määrittelyyn ja ”sulautumiseen” yhteiskuntaan. *Maahan lämpimään* -teoksessa oralismin Suomeen saapumista käsittelevä luku onkin otsikoltaan ”Pakkosulauttaminen”, mikä sopii yleiskuvaltaan myös sanoma- ja aikakauslehtiaineiston representaatioihin: Kuurojen yhteisön tuli muuttua ja sulautua oralismin vaatimusten mukaisesti osaksi muita yhteiskuntaa. Identiteettitekijät kuten viittomakieli alkoivat kadota yhteisön yhdistävänä tekijänä puhekielen muodostaessa hierarkian puhuvien ja ei-puhuvien kuurojen välillä. Kuurot voidaan visualisoinnin ja historianteoksen avulla hahmottaa toiminnan kohteena, eivätkä he olleet aktiivisia tekijöitä kouluun liittyen. Samoissa lehtiteksteissä on myös arvioitu kuurojen toimintamahdollisuuksia Malmin koulun käyneen sukupolven ja puhemetodilla opiskelleiden kuurojen osalta.

Tarkemmin analysoidut kollokaatioiden lähdetekstit on suurilta osin todettu tietosisällöltään vastaaviksi *Maahan lämpimään* -teoksessa käsiteltyjen teemojen kanssa ristiin lukemalla ja vertailemalla teoksessa lähteinä käytettyjä sanoma- ja aikakauslehtiä. Erityisesti *Aistivalliskoulujen lehden* 1909 numerossa 1 julkaistu Killisen teksti ”Kuuromykkäin ajatus ja puhe” toimii suunnantajana koko tutkimuskysymyksen muotoilulle tiedonhaun mahdollisuuksista. Tekstiä on käsitelty Laakson ja Salmen teoksessa¹⁹⁸ esimerkkinä kuurojenkoulujen opettajien suhtautumisesta viittomakieleen ja Killisen erityisen vihamielisestä näkemyksestä viittomiseen kuurojen kommunikaation välineenä. Se, miten paljon ja kuinka tarkasti kollokaatioiden lähdetekstien informaatio vastaa *Maahan lämpimään* -teoksen luvun tietosisältöä, ei ole helposti arvioitavissa. Kaikkein korkeimman log-dice-pisteytyksen saaneet ruotsinkielisten tekstien kollokaatiot viittaavat useisiin samoihin teemoihin kuin *Maahan lämpimään*: koulun, puhemenetelmän ja oralismin sidoksisuuteen.

Viitotut muistot on ollut keskeinen tässä tutkimuksessa käytetty apuväline vääryyksien hahmottamiseksi tutkimusaineistossa. Se on hallitusohjelmaan kirjattu 2020–2021 toteutettu tutkimus Suomen 1900-luvulta nykypäivään kuuroihin ja viittomakielisiin kohdistuneista vääryyksistä ja niiden käsittelyyn tarkoitettun sovintoprosessin käynnistämistä. Kaikesta

¹⁹⁶ Alaluku 2.3 kuvaa *Suomenkielisten kollokaatioiden visualisointi & Ruotsinkielisten kollokaatioiden visualisointi*.

¹⁹⁷ Laakso & Salmi 2006, 155–156.

¹⁹⁸ Laakso & Salmi 2006, 196.

käytetystä lähdeaineistosta *Viitotuissa muistoissa* on yksityiskohtaisimpia kuvauksia oralismin vääryyksistä kuuroja kohtaan. Tutkimus sisältää runsaasti haastattelumateriaalia, jonka avulla kuvataan 1900-luvun alun puhemenetelmän leviämistä kouluihin viittomakielen kustannuksella.¹⁹⁹

Päädyin lopulta käyttämään *Viitottuja muistoja* enemmän historian kontekstin välineenä, kuin suoranaisena vertailtavana lähdemateriaalina kollokaatioanalyysin tiedonhaun mahdollisuuksien hahmotteluun. Koska 1900-luvun kuurojen kokemista vääryyksistä ei juurikaan ole mainintaa kollokaatioanalyysin poimimissa teksteissä ja koska kuuroja koskevan keskustelun on tosiasiaassa todettu olevan kuurojen opettajien käsissä, ei selonteosta löytynyt yhtä seikkaperäisiä esimerkkejä samojen teemojen havaitsemiseksi digitaalisessa muodossa olevasta aineistosta kuin kahdessa muussa keskeisessä lähteessä. Toisaalta tämä tuo myös esille kuurojen kohdalla arkistoiden hiljaisuuden läsnäolon sanoma- ja aikakauslehtikorpuksessa. Kuurojen kokemia vääryyksiä on ollut olemassa, mutta 1900-luvun alun sanomalehdistä ne eivät välity samanlaisella konkreettisella tavalla, kuin *Viitottujen muistojen* muistitietoaineistossa.

Muiluvuoren pro gradu -tutkielmassa yksi keskeinen tarkastelutapa on diskurssianalyysin muodostaminen *Kuuromykkäin lehden* teksteistä ja niiden avulla hahmoteltujen kuuden arjen kategorian muodostaminen.²⁰⁰ Kategorioista on vastaavasti muodostettu kolme ryhmittelyä, joista ensimmäinen ”koulutus ja kommunikaatio” vastaa aihepiireiltään ja diskursseiltaan pitkälti samankaltaisia toistuvia aiheita sekä teemoja, mitä olen itse löytänyt kollokaatioiden lähdetekstien tarkemmassa tarkastelussa. Näitä ovat esimerkiksi opettajien holhoavana suhtautumisena kuuroihin, kuurojen asema toiminnan kohteena ja kuurojen oppilaiden arvottamisen ensisijaisesti puhetaidon perusteella. Diskurssianalyysissä Muiluvuori on huomannut *Kuuromykkäin lehden* sisältävän kuurojen opettamiseen liittyvien tekstien olevan suuressa määrin kuulevien opettajien tuottamia näkemyksiä²⁰¹. Nämä huomiot kuulostavat hyvin samankaltaisilta, mitä olen itse löytänyt *Aistiviälliskoulujen lehdestä*. Kollokaatioiden ja Muiluvuoren havaintojen välillä on myös eroja. Hänen mukaansa esimerkkitekstit edustavat ensi sijassa kuulevien näkemystä kuuroista. On kuitenkin huomattava, että *Kuuromykkäinlehden* teksteistä puuttuu usein tekstien kirjoittajien tarkemmat tiedot,

¹⁹⁹ Katsui et al. 2021, esim. 74–75.

²⁰⁰ Muiluvuori 2025, 35. Kategoriat: (1) Koulutus ja (2) kommunikaatio (3) Työ ja (4) vapaa-aika (5) Uskonto ja (6) avioliitto.

²⁰¹ Muiluvuori 2025, 32.

esimerkiksi 1.8.1908 julkaistu teksti ”Sananen kuuromyökkäin kuuleville ystäville ja auttajille” ei sisällä tietoa kirjoittajasta. Muiluvuori ei ole myöskään käyttänyt Kustaa Killisen tuottamia tekstejä diskurssiensa muodostamiseksi, vaikka omassa aineiston analyysissä hän nousee muita opettajahahmoja näkyvämmäksi.

Kolmen vertailuaineiston kohdalla oman tutkimukseni ja Muiluvuoren pro gradu -tutkielman jaetuksi teemaksi nousee selvästi koulun ja oralismin yhteys. Koulun ja puhemetodia soveltaneiden opettajien toiminta ja metodilla opettaminen heikensivät selvästi kuurojen asemaa. Puhemenetelmä asetti selvät hierarkiat opettajien, puheella kommunikoivien oppilaiden ja puhumattomien välille. Opettajilla saattoi olla merkittävästi valtaa, joka ei rajoittunut vain koulussa oleviin oppilaisiin, vaan ulottui koskemaan kuurojen yhteiskunnallista asemaa, kuurojen yhteisön sisästä dynamiikkaa ja lopulta kuurojen perusoikeuksia, kuten avioliittoa. Toinen selvä teema on viittomakielen syrjäyttäminen ja siihen liittyvä kuurojen aseman heikkeneminen. Sanomalehdissä viittomakielestä ei puhuta paljoa, mutta esimerkiksi Kustaa Killisen teksteissä se aiheuttaa lähes systemaattisesti äärimmäisen voimakkaan torjumisreaktion kirjoittajassaan. Viittominen nähtiin kuurojen yhtenä ”heikkoutena,” joka oikeutti holhoavaan opettamiseen puhemetodilla. *Maahan lämpimään* ja etenkin *Viitotut muistot* välittävät vielä selkeämmin viittomakielen liittyneen Malmin koulun menestyksekkääseen opetukseen, joka takasi kuuroille hyvät perusedellytykset huomattavasti monipuolisemmalla opetussisällöllä verrattuna puhemetodin yksipuolisuuteen. Kolmas teema on kuuroihin kohdistuva sorto. Sanomalehtiaineistossa on todettu, miten ”döfstumma-andesvaga”- kollokaatio alkaa hiljalleen 1900-luvun edetessä löytyä myös kuurojen omien julkaisujen ja *Aistivälliskoulujen lehden* ulkopuolelta. *Suomen urheilulehden* kirjoitettu teksti kuurojen fysiologisista ongelmista 1914 esimerkiksi sisältää hyvin paljon rotuhygieenisia piirteitä, jotka realisoituivat lainsäädännön tasolla esimerkiksi kuurojen avioliittojen kieltona. *Viitottujen muistojen* muistitietoaineistot osoittavat koulujen soveltaneen huomattavan kovia otteita oppilaisiin, jotka eivät oppineet puhumaan tai kommunikoivat kielloista huolimatta keskenään viittomalla.

3.2 Sanomalehdet digitaalisina ilmentyminä

Kansallisarkisto määrittelee digitoitujen tietoa-aineistot digitaalisiksi ilmentymiksi analogisesta aineistosta.²⁰² Tällä tarkoitetaan, että aineisto säilyttää todistusvoimaisuutensa suhteessa alkuperäiseen. Tätä tutkimusta varten toteutettu sanomalehtikorpus on käytännössä taulukkomuodossa oleva luettelo kaikista lehtiteksteistä, joissa pitäisi olla Kansalliskirjaston käyttöliittymällä tehdyt sanahaun tulokset. Määrällisesti kaikkien saatavilla olevien Kansalliskirjaston lehtiaineistojen tekstidatan kokoaminen korpuksi tekee kokonaisuudesta lähes mahdottoman lähiluvulle. Käyttämäni analyysi muuttaa kaiken sanomalehtiaineiston yksittäiset tekstit valtavaksi datakehikoksi tokeneita: Jokainen kehikon rivi on yksittäinen sana, joka kuitenkin säilyttää tiedot kohdeteksteistä ja sanojen sijainneista niissä. Kollokaatioanalyysi taas liittää jokaiselle tokenille viereisen sanan oikealta ja vasemmalta puolelta, mikäli se on mahdollista, ja laskee, kuinka todennäköisesti sanat esiintyvät toistensa läheisyydessä. Kun sanoma- ja aikakauslehtien tekstit tokenisoidaan, analogisen aineiston olemus katoaa käytännössä kokonaan. Vasta tällöin aineistoa voidaan analysoida tietokoneavusteisin menetelmin. Digitoituilla sanomalehdillä on edelleen niiden alkuperäinen lähdeyhteys ja proveniensi eli järjestys, jonka ne ovat saaneet alkuperäisen käyttötarkoituksensa yhteydessä.²⁰³

Suomen sanoma- ja aikakauslehtien digitointiprojekti juontaa vuoteen 1951.²⁰⁴ Tällöin sanomalehtiä mikrofilmattiin pysyvään säilytykseen tähtävää arkistointia varten. Mikrofilmaus oli kansainvälisesti keskeinen menetelmä analogisten lehtien säilyttämiseksi toisesta maailmansodasta lähtien. Suomessa Kansalliskirjaston digitointiprojekti katsotaan alkaneeksi 1990-luvulla, mutta tällöin ei digitoitu analogisia sanoma- ja aikakauslehtiä vaan mikrofilmejä.²⁰⁵ Kansalliskirjaston primäärinen tallennusala sanoma- ja aikakauslehtiaineistolle oli aina 2000-luvun alkuun saakka mikrofilmi. Digitointityö nähtiinkin jatkumona mikrokuvaukselle.²⁰⁶ Mikrofilmien digitointi oli teknisesti melko yksinkertaista, joten Kansalliskirjasto pystyi toimimaan tehokkaasti: Ensimmäinen digitaalinen kokoelma julkaistiin 2001 ja 2020-luvulle tullessa lehdet on digitoitu vuosilta 1771–1954 ja vuodesta 2017 alkaen nykypäivään kokonaisuudessaan.²⁰⁷ Vuosien 1954–2017

²⁰² KA/20011/30.00.00/2023.

²⁰³ Henttonen 2023, 44–54.

²⁰⁴ Hakkarainen et al. 2022, 106.

²⁰⁵ Paju et al. 2023, 264–266.

²⁰⁶ Hakkarainen et al. 2022, 106–107.

²⁰⁷ Paju et al. 2023, 265.

väliseltä ajalta digitoinnin kattavuuden osalta on vielä puutteita, mutta vuodesta 2017 eteenpäin digitointi on Kansalliskirjaston mukaan kattavaa.²⁰⁸

Lehdet ovat aikansa tuotoksia, joka käy erityisesti ilmi optisen tekstintunnistuksen (OCR, *optical character recognition*) laadun vaihtelevuudesta. Osa teksteistä on edelleen perua 1900–2000 lukujen tunnistusjärjestelmistä, mikä tarkoittaa väärin tunnistettuja kirjaimia ja yleistä ”meteliä” (digital noise) sotkemassa sanahakuja ja tekstiin tehtäviä analyyseja.²⁰⁹ Tässä tutkimuksessa tekstintunnistukseen liittyvät ongelmat ovat suurilta osin ilmenneet vain kollokaatioiden lähdetekstien lähiluvun yhteydessä, mutta on oletettavaa, että tutkimusta varten muodostetut korpuukset sisältävät huomattavan määrän melua.

Amelia Acker on pyrkinyt kuvaamaan teoksen *Research in archival multiverse* artikkelissaan *When is the record* digitaalisten asiakirjojen elinkaaren jatkumon erityispiirteitä: Digisyntyisen tiedon säilyttäminen useilla eri tallennusaloilla hajautetusti haastaa käsityksen siitä, mikä aineiston alkuperä on, onko aineisto olemassa vasta jonkin tietyn teknisen prosessin seurauksena ja onko aineistolla yksittäistä arkistonmuodostajaa vai puhutaanko usean muodostajan järjestelmiin hajaantuneesta tiedosta.²¹⁰ Acker ehdottaa digitaalisen tiedon käsittelylle viitekehystä, joka perustuu infrastruktuuriin, teknologisiin käytäntöihin sekä tiedonhakuun.²¹¹ Tässä tutkielmassa tiedonhakua on käsitelty samasta näkökulmasta kuin Ackerin artikkelin määritelmässä: Kollokaatioiden log-dice-pisteytys tuottaa tokenisoidulle aineistolle tärkeysjärjestyksen, jota verrataan digitaalisiin ilmentymiin lähiluvulla. Tärkeysjärjestys on tuotettu aineiston digitaalisilla ominaisuuksilla, mutta aineiston alkuperäinen yhteys voidaan paikantaa edelleen kollokaatioiden lähdetekstien metatiedoista, jotka ovat käytännössä Kansalliskirjaston metatietoja. Ackerin artikkelissa on nostettu esille moni- ja rinnakkaisprovenienssien teorit, joilla voidaan määrittää hajautettujen digitaalisten ympäristöjen kohdalla tietojen alkuperä ja omistajuus.²¹² Argumentoin, että myös digitoidut aineistot, kuten tässä tutkielmassa käsitellyt korpuukset, voidaan käsitellä laajemman alkuperän perusteella. Esimerkiksi sanomalehtiaineistosta löytyvät osumat perustuvat Kansalliskirjaston metatietoihin kuten lehtien nimikkeisiin, ilmestymispäivämääriin ja sivunumeroihin eli analogisiin ominaisuuksiin. Näihin ominaisuuksiin on yhdistetty kollokaatioanalyysin tuloksena syntynyt konteksti, joka liittyy tekstin sisäisiin nimettyihin

²⁰⁸ Käsitelty esimerkiksi <https://digi.kansalliskirjasto.fi/faq>, haettu 29.4.2026.

²⁰⁹ Paju et al. 2023, 265–266.

²¹⁰ Acker 2017, 297–298.

²¹¹ Acker 2017, 300.

²¹² Acker 2017, 309.

entiteetteihin.²¹³ Tässä tutkielmassa voi huomatta Kustaa Killisen liittyneen kahdella tavalla kollokaatioiden lähdeteksteihin: suoranaisesti hänen mukaansa nimettyjen kollokaatioiden muodossa ja entiteettinä ”döfstumma_andesvaga” - kollokaatioissa. Digitaalisten sidoksien ja analogisiin ilmentymiin perustuvan metadatan yhteiskäytöllä aineistosta paljastuu piilotettuja merkitysyhteyksiä, jotka mahdollistavat tiedonhaun hyödyntämisen digitaalisen ympäristön tietoihin monipuolisesti.²¹⁴

Motivaationa aineiston käsittelemiseksi tietokoneavusteisin menetelmin olen maininnut uuden tiedon hankkimisen sekä suunnittelemani tiedonhaun metodin kokeilemisen käytännössä. Sanna Kumpulainen ja Mika Sihvonen argumentoivat artikkelissaan *Tehtävälähtöinen informaatiovuorovaikutus*, että lähdekeskeisen ”dokumentteihin tallennetun tiedon” tutkimuksen teoreettisen viitekehyksen ja puhtaasti järjestelmälähtöisen teknisen tiedonhaun analysoinnin voi yhdistää tehtävälähtöisen informaatiovuorovaikutuksen teorian keinoin.²¹⁵ Keskeistä on tallennetun teknisen tietosisällön ja tiedonhakijan välinen vuorovaikutus. Kumpulaisen ja Sihvosen mukaan näitä vuorovaikutussuhteita on useita erilaisia. He ovat artikkelissaan nostaneet keskeisiksi tekijöiksi tiedonhakijan, järjestelmälähtöisen teknisen tiedon sekä tiedonhakijan vuorovaikutuksen työkalut eli ”informaatioyksiköiden kanssa työskentelyn”.²¹⁶ Tutkimusprosessissani informaatioyksikkö muodostuu vasta, kun sanoma- ja aikakauslehtikorpus on käsitelty kollokaatioanalyysillä. Tällöin kollokaatiot toimivat keinona kommunikoida tietosisällön kanssa. Sanoma- ja aikakauslehtikorpuksen tekstien tokenisoitu muoto edustaa tässä tutkimuksessa järjestelmälähtöistä teknistä tietoa. Kollokaatioanalyysin tulokset taas edustavat informaatioyksiköitä, joita tarkastellen käyttäjä voi selata ja hahmotella tavoitteensa.²¹⁷

Verratessa omaa tutkimusprosessiani²¹⁸ suhteessa Kumpulaisen ja Sihvosen näkemykseen toimijan ja tiedonhakuprosessin riippuvaisuudesta toisiinsa²¹⁹ argumentoin tämän tutkimusprosessini seuraavan samankaltaista prosessia, kuin tietoyksiköiden hakemisen mallissa.²²⁰ Tutkimuksessani keskeisen tiedonhaun menetelmän käyttäminen vaatii esityönä sanoma- ja aikakauslehtikorpuksen muodostamisen, tutkimuskirjallisuuteen perehtymisen ja

²¹³ Furner & Gilliland 2017, 625.

²¹⁴ Acker 2017, 309–311.

²¹⁵ Kumpulainen & Sihvonen 2025, 88.

²¹⁶ Kumpulainen & Sihvonen 2025, 100.

²¹⁷ Kumpulainen & Sihvonen 2025, 98.

²¹⁸ Kuva: Tutkimusprosessikaavio.

²¹⁹ Kumpulainen & Sihvonen 2025, 99.

²²⁰ Kumpulainen & Sihvonen 2025, 98. (Kuva 2)

sekundääristen lähteiden tuntemuksen. Vuorovaikutus aineiston ja tiedonhakijan välillä tiedonhaulla edellyttää kattavaa ymmärrystä taustatietojen ja lähdeaineiston tietojen osalta eikä pelkästään teknisen järjestelmän kehittämistä ja käyttämistä. Kollokaatioiden muodostaminen tietyn hakusanan ympärille on tiedontarpeen ilmaisua, jonka tuloksena tuotetut kollokaatioiden listaukset ovat selailtavia tietoyksiköitä.²²¹

Tokenisoinnin ja log-dice-pisteytyksen jälkeen käsittelemäni aineisto vertautuu paremmin Yuk Huin *digitaalisiin objekteihin*.²²² Huin määritelmän mukaan tietokone luettavilla ja kvantifioituilla tietosisällöillä on alkuperäisen tiedon tekstuaalisuuteen ja fyysisyyteen liittymättömiä sidoksia ominaisuuksia. Oman aineistoni kohdalla objektinomaisuus on osittain korpuksen koostavan sanahaun ja osittain kollokaatioanalyysin tulosta. Kansalliskirjaston käyttöliittymän sanahauulla muodostettu lehtikorpus rakentuu tekstintunnistuksen mahdollisuuksien mukaan kymmenistä tuhansista teksteistä, joita kaikkia yhdistää sana ”kuuromyk” tai ”döfstum”. Kun nämä sanat järjestetään kollokaatioanalyysin mukaan kaikkein yksilöllisimpien esiintymiensä perusteella, voidaan mielestäni argumentoida, että alkuperäinen yhteys yhden sanan ja sanomalehtitekstikonaisuuden välillä on lakannut olemasta. Kollokaatioilla korpuksen hakusana sitoo²²³ kaikki ympärillä olevat tekstit toisiinsa, jolloin olen pystynyt havaitsemaan itselleni uutta ja hyödyllistä tietoa, kuten edeltävissä luvuissa on demonstroitu.

Historiantutkimuksen kannalta ajatellen digitaalisten ilmentymien tarkastelua on perusteltu perspektiivin hakemisella mutta myös kokonaan uusien tutkimusnäkökulmien ja mahdollisten tutkimusaiheiden hahmottamisella. Arkistotieteessä digitaalisten ilmentymien yhteyttä arkistojen todistusvoimaisuuteen ja alkuperäisyyteen ei ole juurikaan käsitelty näkökulmasta, jossa aineisto irrotetaan tietokoneavusteisesti alkuperäisestä kontekstistaan ja järjestetään uudelleen tiedonhakuvälineille sopivaan muotoon. Tässä tutkimuksessa aineiston irrottaminen eli tokenisointi säilyttää sanoma- ja aikakauslehtikorpuksen metadatan, jolloin aineisto osoittaa edelleen digitoituun ilmentymään. Arkistotieteessä tietokoneavusteisuutta on arvoitettu verrattuna analogiseen arkiston asiakirjahallintaan.²²⁴ Keskeinen kritiikki liittyy vastaavasti tietoaineiston kuvailutietojen mahdollisuuden mallintaa analogista aineistoa alkuperäisessä yhteydessään.

²²¹ Kumpulainen & Sihvonen 2025, 100.

²²² Rozenberg 2021, 198–200.

²²³ Eng. *Bind*.

²²⁴ Furner & Gilliland 2017, 582–584.

Argumenttini aineiston alkuperäiseen ilmentymään liittyy kuurojen näkyvyyteen sanomalehtiaineistoissa ja arkistoiden hiljaisuuteen. 75000 digitoidun sanomalehden kokoaminen hakusanoilla ”kuuromyk*” ja ”döfstumm*” ei nähdäkseni anna ihmisluettuna mitään mahdollisuuksia hahmottaa kuurojen kärsimysten kokemuksia. Oma tapani järjestää aineisto kollokaateiksi mahdollistaa aineiston visualisoinnin ja järjestämisen kollokaatioanalyysin avulla. Järjestys ei ole eksakti eikä myöskään tilastollisesti pätevä keino järjestää eri sanaläheisyyksiä tilastollisten tunnuslukujen mukaan. Järjestys ja visualisoiminen antavat aineistolle yhden esimerkin strukturoidusta tiedosta, jonka avulla olen tehnyt valinnan luvussa 2 mainittujen teemoihin syventyäkseen. Kyseessä ei ole siis aineiston tarkastelu automatisoidulla prosessilla vaan digitaalisen ilmentymän, *objektin* tarkastelusta sähköisessä ympäristössä algoritmien avulla.

3.3 Tiedonhaku ja arkistojen hiljaisuudet

Digitoidut sanomalehdet ovat nousseet erittäin suosituksi lähdeaineistoksi historian tutkimuksen pro gradu -tutkielmissa. Niiden käytölle ominaista on helppous ja saavutettavuus. Samalla kokonaan digitoidut ja tekstitunnistetut aineistot mahdollistavat tietokoneavusteisten menetelmien käytön avulla täysin uusien tutkimusaiheiden- näkökulmien sekä -kysymysten muodostamisen.²²⁵ Helppoutteen liittyy myös riskejä, joita esimerkiksi Taru Hyvönen on nostanut esille pro gradu -tutkielmassaan historian tutkimuksen arkistolähteiden osalta.²²⁶ Itse tiedonhakua sähkösyntyisiin arkistoihin ja hakujen seurauksena syntyneitä ”hiljaisuuksia” on tutkittu hyvin vähän. Hyvösen keskeinen argumentti liittyy aineistojen saatavuuden ja käytön vaivattomuuteen, joka aiheuttaa tutkimuksen köyhtymisen.²²⁷ Analogisen aineiston korvautuminen digitaalisilla ilmentymillä on ollut paljon esillä 2020-luvulla. Esimerkiksi Kansallisarkiston linjaus digitaalisen toimittamisen painottamisesta²²⁸ on aiheuttanut voimakkaita vastareaktioita tutkijoiden keskuudessa. Digitaalinen toimittaminen on realisoitunut toimipaikkojen vähenemisenä ja analogisen aineiston käytön rajoittumisena arkistoiden sulkiessa ovensa. Samalla kysymykset, kuten analogisten ilmentymien

²²⁵ Furner & Gilliland 2017, 623.

²²⁶ Hyvönen 2022, 74–76.

²²⁷ Hyvönen 2022, 80–82.

²²⁸ <https://kansallisarkisto.fi/digitaalinen-toimittaminen>

erityispiirteet ja niiden menettäminen digitoinnin yhteydessä ovat olleet esillä jo Hyvösen tutkielman tutkijahaastatteluissa.

Arkistoiden hiljaisuudella viitataan yleisimmin arkistojen vallankäyttöön. Kaikki arkistoon päätyvä aineisto on päätenyt sinne jonkin valinnan tuloksena, kaikkien ihmisten tarinat eivät koskaan tule kuulluiksi ja näiden ihmiskohtaloiden äänettömyys on mielletty hiljaisuudeksi, joka syntyy vallankäytön seurauksena.²²⁹ Arkistojen hiljaisuuksien voidaan nähdä olevan erityisen haitallista esimerkiksi tietyn ihmisryhmän jäädessä muistamisen ulkopuolelle toisen ryhmän päätöksestä. Sama pohdinta voisi päteä analogisen ja digitaalisen aineiston välillä tilanteessa, jossa analoginen ilmentymä tuhoetaan digitoinnin jälkeen. Toinen hankalampi hiljaisuuteen liittyvistä seikoista liittyy aineiston käytettävyyteen ja löydettävyyteen digitaalisesta säilytysympäristöstä. Esimerkiksi digitaalisten sanomalehtien hakutulosten käyttöliittymä Kansalliskirjastossa antaa mahdollisuuden järjestää tulokset osuvuuden mukaan. Mihin ”osuvuus” perustuu? Miten käyttäjä voi tietää, sisältääkö digitaalinen säilytystila kaikki analogiset aineistot digitoituina, vai vaan osan? Miten digitaalista aineistoa voidaan ylipäätään käyttää niin, että ihminen ymmärtää aineistoa?

Puhuttaessa digitaalisista aineistoista on tehtävä ero digitoitujen ja alkujaan digitaalisten aineistojen välillä. Digitoituilla aineistoilla on tai on ollut olemassa analoginen ilmentymä, johon aineiston metatiedot ja digitaalinen ilmentymä perustuvat. Aineisto on muutettu sähköiseen muotoon jossain vaiheessa elinkaartansa, esimerkiksi skannaamalla tai nauhoittamalla tietokoneelle talletettavaan muotoon.²³⁰ Alkujaan digitaalinen aineisto on taas syntynyt ja kerääntynyt digitaalisessa ympäristössä osana digitaalisia prosesseja.²³¹ Hyvönen on pro gradu- tutkielmansa haastatteluissa havainnoinut digitoitujen lähdeaineistojen olevan nousemassa keskeisimmiksi historianutkimuksen primäärilähteiksi.²³² Oma lähdeaineistoni koostuu digitoituista sanoma- ja aikakauslehdistä, mutta tapa, jolla käytän tutkimusaineistoani tietokoneavusteisesti, vie kokonaisuuden lähemmäksi digisyntyistä tietoa. Tiedonhaun kannalta hyödynnän *Archival IR* menetelmistä *sanahakua, tekstin määrällistämistä* sekä *tietokoneavusteista pisteyttämistä*. Aineistoni sitoutuu alkuperäisten hakusanojen ympärille siten, että jokaisesta tekstistä löytyvät jossain muodossa sanat ”döfstumma” ja ”kuuromykkä” taivutuksineen. Itse tiedonhaku perustuu kollokaatiotestin

²²⁹ Carter 2006, 218–219.

²³⁰ Esim. Lybeck et al. 2006, 159.

²³¹ Eräkaski 2012, 23.

²³² Hyvönen 2022, 61–62.

pisteytykseen: Tiedon visualisoiminen log-dice-pisteytyksen perusteella tuottaa työkalun kaikkein mielenkiintoisimpien tutkimuskysymysten hakemiselle. Haku perustuu painotettuun indeksiin eli aputietorakenteeseen,²³³ jonka järjestys perustuu kollokaatiopisteytykseen. Tiedonhaun aputietorakenteessa muodostuvat osumat on mahdollista taas palauttaa lähiluettaviin teksteihin. Kollokaatiot pisteytyksineen muodostavat sanahaulle lisäkontekstina järjestystä kuvaavan numeerisen metatiedon, jonka avulla löysin tutkimusprosessini aikana yhteneviä tietoja Kustaa Killisen ja ”döfstumma-andesvaga” kollokaatioiden välillä *Aistivalliskoulujen lehden* kirjoitusten muodossa.

Jonathan Furner ja Anne Gilliland käsittelevät teoksessa *Research in the Archival Multiverse* julkaistussa artikkelissaan *Archival IR*, ”arkistollista tiedonhaku” peilattuna analogisen arkiston arkistinhoitajien ammattitaitoon ja kykyyn havaita asiakirjojen sisäistä ja ulkoista provenienssia.²³⁴ Artikkelissa on esitetty useita keinoja kuvailla aineistoja tietokoneuuttavasti siten, että niiden alkuperäinen käyttötarkoitus ja todistusvoimaisuus ovat mahdollisimman tarkasti säilytettävissä esimerkiksi merkitsemällä digitoituihin asiakirjallisiin tietoihin entiteettitietoja kuten aika, paikka, toimija ja organisaatiot.²³⁵ Artikkelin ei kuitenkaan aikanaan huomionnut syntysähköisten aineistojen käytettävyyttä tai löydettävyyttä moderneilla tallennusalustoilta kuten pilvipalveluista tai tietokannoista. Artikkelissa asiakirja on edelleen keskeinen väline, joka tuottaa todistusvoimaisen tiedon²³⁶ tilanteessa, jossa tietoa muodostuu useasta eri tiedonlähteestä eri aikaisesti. Myös Pekka Henttonen toteaa *Informaatiota tiedon vuoksi* -teoksessa tiedonhakuun liittyvän tutkimuksen koskevan aikaa, jolloin aineistojen sähköinen käyttö tarkoitti esimerkiksi ylätasojen kuvaamista verkossa.²³⁷ Acker kysyy ”milloin asiakirja on”, mutta nykyisten digitaalisten alustojen kohdalla kysymys pitäisi koskea myös ei-asiakirjallista tietoa, eli ”milloin tieto on”. Tässä tutkielmassa ei ole onnistuttu irtautumaan asiakirjallisesta tiedosta, vaan lähiluku perustuu edelleen digitoitujen sanoma- ja aikakauslehtien alkuperäiseen analogiseen kontekstiin metatietojen ja digitoitujen ilmentymien käyttöön tiedon lähteinä. Arkistoiden hiljaisuuksia ajatellen, prosessikaavion²³⁸ tietokoneavusteisten välineiden vaiheita²³⁹ ei voi suorittaa ilman historiallisen kontekstin

²³³ Kumpulainen 2025, 121.

²³⁴ Furner & Gilliland 2017, 608.

²³⁵ Furner & Gilliland 2017, Archival IR “Named entities”.

²³⁶ Acker 2017, 316.

²³⁷ Henttonen 2025, 274.

²³⁸ S.10, kuva: Tutkimusprosessikaavio.

²³⁹ S.10, kuva: Tutkimusprosessikaavio, vihreät prosessit.

muodostamista aikalaislähteiden avulla,²⁴⁰ joita puolestaan on taas tarkasteltava lähiluvun keinoin. Pelkästään teknisten prosessien tarkastelun lopputuloksena esimerkiksi puhemetodin käyttöön koulussa liittyvissä teksteissä ei ole lähes lainkaan edustettuna kuurojen omaa ääntä keskustelun ollessa niin voimakkaasti yliedustettuna *Aistivalliskoulujen lehden* lähdeteksteissä.

Suorittamani tiedonhaku korpusaineistoon on mahdollista vain käyttämällä aineiston digitaalisia ominaisuuksia. Kollokaatioanalyysini perustuu aineiston tokenisoimiseen ja erilaisten tokenisoitujen merkkijono-objektien tilastolliseen vertailuun. Käytännössä tapa, jolla muokkaan aineistoa kollokaatioanalyysille ja myöhemmin tiedonhauille sopivaksi datarakenteeksi, hävittää aineiston alkuperäiseen käyttötarkoitukseen ja fyysisyyteen liittyvät ainutlaatuiset erityispiirteet muuttaen kokonaisuuden digitaalisten objektien kokoelmaksi. Tämä tarkoittaa sitä, että alkuperäisen analogisen kontekstin ominaisuudet kuten eri lehtinimikkeiden erityispiirteiden hyödyntäminen menettävät merkityksensä tutkimusprosessin ”primäärin metodin” -vaiheessa, mutta nousevat lopulta omaksi prosessikseen tulosten autenttisuuden tarkastelussa.²⁴¹

Sebastian Rozenberg on kuvannut asiakirjojen olemusta digitaalisina objekteina sähköisissä käyttöympäristöissä niiden suhteiden kautta: Digitaalisessa muodossa olevaa tietoa tulisi ymmärtää niiden teknisten eksistentiaalisten suhteiden avulla. Hänen artikkelissaan käydään läpi Yuk Huin ajatuksia teoksesta *On the Existence of Digital Objects*.²⁴² Teoksessa pohditaan digitaalisten objektien olemusta filosofisesti ja tarkastelee, miten objektit ovat individuaalisia suhteessa toisiinsa. Rozenberg pohtii artikkelissaan tätä yksilöllisyyttä objektien muuttamisena ihmisluettaviksi esimerkiksi metatietojen avulla. Digitaalista tietoa ei tulisi Huin mukaan käsittää pelkästään laskennallisena ja prosessoituna datamassana, vaan tieto on luonnollisessa tilassaan ”verkostoitavissa tiedostojen, tietokantojen ja tietojärjestelmien välillä”.²⁴³ Tätä ajatusta soveltaen pyrin muodostamaan käsittelemistäni digitoiduista sanoma- ja aikakauslehdistä parsittujen tekstien välisiä verkostoja ja samalla löytämään merkityksiä niistä muodostetuista visuaalisista ja sanaläheisyyksiin perustuvista tilastollisista luetteloista. Tällöin tuottamani digitaaliset objektit, tilastollisesti analysoidut kollokaatiot tuottavat uutta tietoa ja jäsentävät kokonaiseksi tekstimassaksi muutetun korpusaineiston sekä

²⁴⁰ S.10, kuva: Tutkimusprosessikaavio, siniset prosessit.

²⁴¹ Sekundääristen lähteiden ilmiöiden vertaaminen, s. 10, kuva: Tutkimusprosessikaavio, valkoiset prosessit.

²⁴² Rozenberg 2021, 199.

²⁴³ Rozenberg 2021, 198.

”etäluettavaksi” tutkimusdataksi sekä metadatan avulla myös mahdollisiksi uusiksi lähiluettaviksi tutkimusaiheiksi. Lehtien digitaalinen ilmentymä on ihmiselle luettava ja olemassa oleva muoto, kun taas tokenisoidussa tekstimassassa digitaaliset objektit ovat jatkuvassa vuorovaikutuksessa sekä toistensa, datarakenteen, että käyttäjän kanssa järjestäytyen prosessoinnin ja algoritmien, eli tässä tapauksessa kollokaatioanalyysin avulla uusiin merkityksien muotoon.²⁴⁴ Soveltaen Huin teoriaa erilaiset tekniset säilytystilat kuten pilvipalvelut, vektorikannat tai tietovarannot voisivat olla hahmotettavissa sisältönsä suhteiden avulla, esimerkiksi tilastollisesti järjestettynä kuten omassa tutkielmassani.

Omassa aineistossani on kyse ihmisen tuottaman aineiston esiprosessoimisesta ja analysoimisesta tietokoneavusteisin menetelmin, mutta argumentoin, että etenkin sähkösyntyisen aineiston kohdalla Huin teoria digitaalisten objektien välisistä suhteista on entistä ajankohtaisempaa. Omassa tutkimuksessani tokenisoitu sanomalehtikorpus on aineistona fundamentaalisti erilaista kuin vastaavat analogiset ilmentymät. Samoin esimerkiksi Kansallisarkiston suositellut siirtomuodot kuten Xml, JSON ja CSV-tiedostoformaattit²⁴⁵ ovat osittain ihmisen luettavia mutta käytännössä tarkoitettuja tietokone-luettaviksi.²⁴⁶ Ackerin mukaan arkistotieteellisesti kyse on aineiston²⁴⁷ kuvailuista metatiedoin, mutta myös sen tunnistamisesta, millaisia relaatioita kaikilla aineiston tiedoilla on suhteessa paikkaan, jossa niitä säilytetään, mistä ne ovat peräisin ja miten tiedon todistusvoimaisuus käy ilmi. Omassa aineistossani olen käyttänyt lehtikorpuksen kuvailuun tekstien alkuperäisiä nimikkeitä ja julkaisupäivänmääriä sekä linkkejä Kansalliskirjaston sanoma- ja aikakauslehtien digitaalisiin ilmentymiin, mutta kokonaisuudessaan tietoja voisi olla myös enemmän kuten esimerkiksi kohdetekstien kirjoittajien tiedot, lehtien sivunumerot ja tekstien sanamäärät. Kaikki edellä mainitut tiedot muuttuvat välttämättömiksi aineiston muodon muuttuessa digitoiduista teksteistä tokenisoinnin seurauksena digitaalisiksi objekteiksi ja myöhemmin kollokaatioanalyysin seurauksena tiedonhakijan haettaviksi informaatioyksiköiksi. Tutkimusprosessin tulokset ovat muotoutuneet asteittain. Aineiston kuvailu visualisoituina ja log-dice-pisteytettyinä kollokaatioina muodostaa metadatan, joka perustuu aineiston digitaalisiin ominaisuuksiin, kun taas Kansalliskirjaston käyttöliittymästä saadut ominaisuudet kuvaavat aineiston analogisen ilmentymän ominaisuuksia kuten

²⁴⁴ Rozenberg 2021, 199–200.

²⁴⁵ <https://arkisto-osaaja.kansallisarkisto.fi/digitaaliset-siirrot/ohjeet/vastaanotettavia-tiedostoja-koskevat-vaatimukset-t>, [Haettu 30.4.2026.]

²⁴⁶ Kumpulainen et al. 120–121, 2021.

²⁴⁷ Acker 2017, 298–300, viittaa asiakirjaan ”record”.

todistusvoimaisuutta ja alkuperäisyyttä. Korpuksen metatiedot toimivat myös haun tarkentavina työkaluina,²⁴⁸ jolloin aineistoa on periaatteessa mahdollista tarkastella pelkästään kollokaatioanalyysin tuloksien avulla. Käytännössä kollokaatioista poimittu tieto vaatii tuekseen vähintään sekundäärisiä lähteitä.

Tutkimusaineistoa olisi voinut lähestyä myös toisin. Jos sanoma- ja aikakauslehtieä tarkastellaan niiden analogisessa muodossa, voimme havaita kuurojen kokemat vääryydet esimerkiksi *Kuuromykkäin lehden* diskurssien muodossa kuten Muiluvuori pro gradu-tutkielmassaan. Jos taas aineistoa tarkastellaan etäältä, voidaan samankaltaiset holhoamisen diskurssit taas havaita esimerkiksi *Aistivalliskoulujen lehden* teksteissä. Tässä tutkielmassa on taas havaittu tietynlaisten hakusanojen muodostavan yhteyden opetusalan voimahenkilöiden puhemetodin edistämiseen vuosina 1898–1920. Samalla tietokonekäsitellystä aineistosta voidaan saada esimerkiksi visualisoinneilla yleiskuva siitä, miten suhtautuminen kuuroihin muuttui 1900-luvun alun Suomessa. Argumentoin, että tutkimuksessa käytetyt visualisoinnit tuovat esille hyvin samankaltaisia teemoja, kuin mitä *Viitottujen muistojen* muistitietoaineistoissa ja niissä ilmenevissä vääryyden kokemuksissa on noussut esille: Kuurojen vääryyksien kokemukset oralistisen opetusmetodin osalta ja ylipäätään koulutuskeskustelun jyrkkenevät asenteet 1900-luvun alussa vähemmistöjä kohtaan ilmenivät oppilaiden kokemuksista esimerkiksi viittomakielen kieltämisen osalta.²⁴⁹ Samalla kovenevat asenteet voidaan lukea esimerkiksi Kustaa Killisen ulostuloissa *Aistivalliskoulujen lehden* kirjoituksissa. Tässä tutkielmassa on myös havaittu, että sanoma- ja aikakauslehtien teksteissä kuuroihin on liittynyt kollokaatioiden tasolla negatiivisia määreitä. Näitä hienovaraisempia vääryyden ilmentymiä on luultavasti huomattavasti paljon enemmän, mutta niiden tutkiminen vaatisi tarkempaa yksittäisten tekstien lähilukua. Vaikka tietokoneavusteinen analyysi valottaa uusia näkökulmia kuurojen kokemista vääryyksistä, ei voida puhua mitenkään kaiken kattavasta yleiskuvasta. Kollokaatioparit tarjoavat uusia tutkimusaihoita lähiluvulle, mutta eivät anna suoraa vastausta siihen, millaisia vääryyden kokemuksia kuurot kohtasivat oralismin rantaudutta Suomeen. Yleiskuvasta ei välity kuurojen yhteisön toimijuus ja mahdollisuudet 1900-luvun alkupuolella.

Aineistojen konversio analogisesta digitaaliseen aiheuttaa lähes aina fyysisten ominaisuuksien katoamisen, oli kyseessä sitten aineiston kunto, haju, sijainti arkistossa tai mahdolliset

²⁴⁸ Acker 2017, 302–303.

²⁴⁹ Liite 3: kuva 3.

käytöstä syntyneet jäljet. Syntysähköisellä aineistolla ei ole näitä ominaisuuksia ollenkaan, ellei niitä erikseen kuvata analogista olemusta jäljittelevillä metatiedoilla.²⁵⁰ Aineistojen ”ihmisluettavuus” saattaa olla etenkin laajoissa sähköisissä dokumenttikokonaisuuksissa hyvin heikko ilman algoritmista lähestymistä esimerkiksi parsinnan avulla, jossa tuotetaan aputietorakenne, indeksi tiedon hakemiselle ja ymmärtämiselle.²⁵¹ Tällaiset laajat dokumenttikokonaisuudet ovat yleensä HTML tai XML-muotoista tietoa. Digitaalisten aineistojen esikäsittelyvaiheiden voidaan kuitenkin yleisesti ottaen katsoa hyödyntävän monipuolisia ja tehokkaita tiedonhakumahdollisuuksia. Tämä on kuitenkin totta ainoastaan siinä tapauksessa, että aineiston käyttäjälle tarjotaan valmissovellusten muodossa työkaluja digitaalisten objektien järjestämiseen, tai vastaavasti joko käyttäjällä tai palvelun tarjoajalla on hallussa osaamista ohjelmointiperustaiseen tiedonhakuun ja tietoaineiston organisointiin.²⁵² Organisointi on tässä yhteydessä myös tiedon muovaamista ihmisluettavaan muotoon, tai omassa tutkimuksessani tiedon tarkastelua analogisten sanomalehtien digitaalisten ilmentymien avulla, mikä on mahdollista ainoastaan riittävän kattavilla aineistokohtaisilla metatiedoilla, jotka Kansalliskirjasto on tuottanut aineistolleen.²⁵³ Edellä mainittuihin tiedon käytettävyyteen liittyvät seikat kuvastavat myös uudenlaisia arkistojen hiljaisuuksia: Digitaalinen aineisto saattaa olla sisällöltään äärimmäisen tiivistä ja relevanttia monen eri tiedonhakijan intressien kannalta, mutta se vaatii runsaasti käsittelyä ennen kuin tiedonhakija saavuttaa vuorovaikutuksen tarkoitustenmukaisten informaatioyksiköiden kanssa.²⁵⁴

Digitaalisten objektien yhteys arkistoiden hiljaisuuksiin liittyy tiedonhaun menetelmien mahdollisuuksiin löytää aineistosta informaatiota, joka ei ole helposti tunnistettavissa lähiluvun keinoin. Käyttämäni kollokaatioanalyysi on tämän tutkielman kehyksessä tarkennettua tiedonhakua perustuen tekstissä havaittavien sanaläheisyyksien visualisoimiseen ja järjestämiseen uusien näkökulmien ja tutkittavan substanssin löytämiseksi. Arkistoiden hiljaisuuksia on käsitetty valtasuhteina ja omassa aineistossani tämä näkyy puhemetodia hyödyntävien opettajien ja kasvatustaikuttajien auktoriteettiasemana suhteessa kuuroihin

²⁵⁰ Esimerkiksi <https://sapa-kayttoliittyma-ohjeet.kansallisarkisto.fi/ohjeet/siirtorajapinnan-kayttoohjeet-ja-rajapintakuvaus>. [Haettu 30.4.2026.]

²⁵¹ Arvola et al. 2025, 121.

²⁵² Henttonen 2023, 163.

²⁵³ Rozenberg 2021, 200.

²⁵⁴ Kumpulainen & Sihvonen 2025, 92.

opetettaviin.²⁵⁵ Toisaalta Muiluvuori on havainnut saman ilmiön tarkastelemalla huomattavasti suppeampaa aineistoa.

Aiemmin on todettu, että kuurojen kokemaa sortoa ei ollut tutkimusprosessin aikana mahdollista havaita tutkimusta varten muodostetuista korpusaineistoista pelkästään lähilukua käyttämällä tai hyödyntämällä pelkästään Kansalliskirjaston käyttöliittymän etälukumahdollisuuksia. On kuitenkin huomattava, että tietyt keskeiset lähteet kuten *Suomen Aistivalliskoulujen* aineisto olisi ollut löydettävissä toisista tulokulmista. *Viitotut muistot* ja *Maahan lämpimään* ovat toimineet ilmiöiden merkityksen selittävinä teksteinä ja niiden alaviitteitä seuraamalla olisi ollut mahdollista paikantaa oralismi kuurojen opettamiseen erikoistuneiden pedagogien pääasiallisille keskustelualustoille. Toisaalta yksittäiset lehtijulkaisut hämärtävät sorron havaitsemista: Kustaa Killisen oma ääni ja myös osittain kuurojen oma aineisto antavat kuvan opetuksen edistäjästä ja kuurojen yhteisön suuresta ystävästä ja ilman *Viitottujen Muistojen* tietoon vertaamista puhemetodin alistava rooli ja Killisen holhoava suhtautuminen kuuroihin ei välity yhtä selkeästi yksittäisistä lehtiteksteistä. Samoin pelkän sanahaun soveltaminen sanomalehtikorpuksen antaa hyvin ympäripyöreitä tuloksia.²⁵⁶

Tutkimusprosessin kokonaisuutta ajatellen lähtökohtani oli ensin koostaa korpus ja suorittaa tietokoneavusteiset analyysit ja vasta sen jälkeen tukeutua toisen käden lähteisiin sekä tutkimuskirjallisuuteen. Tiedonhaku lähdeaineistoon vaikutti suurelta osin valintoihin, miten eri lähteitä on painotettu lopullisessa tutkimuksessa: Kustaa Killisen osoittaminen keskeiseksi oralismin edistäjäksi Suomessa oli ensimmäinen havainto, johon pääsemiseksi vaadittiin ristiinlukua digitoitujen sanomalehtiaineistojen sekä *Viitottujen muistojen* muistitietoaineistojen välillä. Seuraava vaihe oli tutkimusaiheen vaihtuminen oppivelvollisuudesta kuurojen kokemien vääryyksien havaitsemiseen ja lopuksi aineistoa kollokaatioilla tarkastelemalla ja lähilukua yhdistämällä Killisen ja ”döfstumma-andesvaga” käsitteiden välisen yhteyden paikantaminen *Suomen Aistivalliskoulujen lehden* teksteihin. Kuurojen kokeman sorron havaitseminen tapahtui aineiston tietokoneavusteisella analyysillä, joka mahdollisti edellä mainitun tutkimusprosessin kokoamisen.

Digitoiduista sanoma- ja aikakauslehdistä ei löydy samankaltaista kuurojen muistitietoa edustavia vääryyden kokemuksia kuin *Viitottujen muistojen* aineistoista, mutta esimerkiksi

²⁵⁵ Muiluvuori, 2025, 32.

²⁵⁶ Liite 3: kuva 3.

suomenkielisten kollokaatioiden visualisointia²⁵⁷ voi pitää alustavana yleiskuvana suhtautumisesta kuuroja kohtaan. Käyttämäni tiedonhaku vaatii historiallisen kontekstin ymmärtämistä, esimerkiksi oralismin aatteiden havaitseminen *Aistiviälliskoulujen lehden* teksteissä ei olisi ollut mahdollista tuntematta ensin kuurojen opetuksen taustaa ja oralismin laajempaa kontekstia 1800-luvun kuurojen historiassa.

Tämän tutkielman keskeinen näkökulma on ollut tarkastella korpusaineistosta nousseita havaintoja verrattuna selontekoon kuurojen yhteisöön kohdistuneista vääryyksistä. Kollokaatioanalyysin avulla löytyneet havainnot oralismiin liittyen ovat olleet hyvin samankaltaisia, kuin *Maahan lämpimään* kuvaukset, mutta laajempaa kokonaisuutta ei ole voitu hahmottaa nyt tarkasteltavan kahden kollokaatioketjun avulla. Tarkempi analyysi voisi tapahtua esimerkiksi valitsemalla yksi kollokaatio lähdeteksteineen kymmenen korkeimman log-dice-pisteytyksen saaneesta kollokaatiopareista. Tutkimus ei anna ratkaisua digitaalisessa ympäristössä suoritettujen tiedonhaun myötä muodostuneiden arkistojen hiljaisuuksien määrittelyyn tai paikantamiseen nykyisessä aineistossa. Tämä on pikemminkin kysymys tulevaisuuden tutkijoille, koska 75000 tekstistä koostuva sanomalehtikorpus on lopulta varsin pieni aineistokokonaisuus verrattuna pitkäaikaissäilytettäviin tietojärjestelmiin, rekistereihin tai tietovarantoihin.

²⁵⁷ S. 34.

4 Lopuksi

Tutkielmani muodostui pitkälti tietokoneavusteisten menetelmien avulla kerätyn uuden tiedon avulla. Keskeisimmät argumentit kuurojen yhteisön sorron kokemuksen olemuksesta korpusaineistoissa löytyi kollokaatioiden tarkastelun ja sekundääristen lähteiden ristiinvertailun avulla. Tavoittelin tutkimuksessani alusta lähtien digitaalisten ihmistieteiden lähestymistapaa aineiston analyysissä ja mahdollisen uuden tiedon hankkimisessa tietokoneavusteisin menetelmin. Arkistotieteellinen pohdinta tiedonhaun mahdollisuuksista kollokaatioiden lähdetekstien ja sekundäärilähteiden ristiin lukemisella osoitti, että aineiston tietojen todistusvoimaisuuden tarkastaminen vaatii kattavat lähteiden analogisiin ominaisuuksiin perustuvat metatiedot tutkimusprosessin kaikista vaiheista. Näiden tietojen käyttö yhdessä kollokaatioanalyysin kanssa tuotti tiedonhaun sanoma- ja aikakauslehtien digitaalisiin ilmentymiin. Haku paljasti aineistossa piileviä yhteyksiä, joiden ristiin lukeminen aiemman tutkimustiedon kanssa auttoi paikantamaan oralismin aiheuttamia vääryyksiä ja rajaamaan lähiluettavan aineistokokonaisuuden.²⁵⁸

Vertailu lähdeaineiston eli digitoitujen sanoma- ja aikakauslehtien ja vuoden 2022 selonteon sekä kuurojen historiaa käsittelevän yleisteoksen välillä liittää tutkimusaiheeni osaksi ajankohtaista keskustelua historiantutkimuksen lähdeaineiston digitaalisten ilmentymien erityispiirteistä. Digitoitujen aineistojen fyysisten ominaisuuksien häviämistä on käsitelty useaan otteeseen, mutta esimerkiksi sanahakujen yksittäiset tekstit toisiinsa sitova vaikutus ansaitsisi syvemmän analyysin, kuin mitä tässä tutkielmassa on esitetty. Vaikka tutkimuksen tulokset eivät anna vastauksia syntyjään digitaalisen aineiston käyttöön liittyvistä ongelmista uudelleen käytettävyyden osalta, voi prosessia pitää yhtenä mahdollisena lähestymistapana tiedonhaun mahdollistamiseksi. Tutkimusprosessin vaiheet²⁵⁹ kuvaavat, miten tietokoneavusteisen menetelmän ja käytännön tiedonhaun suorittaminen tapahtuu rinnakkain tarkentavien sekundääristen lähteiden kanssa.

Käyttämäni metodi toimi uuden tiedon löytämiseksi laajasta korpusaineistosta ja se todistaa, että laajan aineiston analysoiminen eksploratiivisesti on mahdollista tilastollisten tunnuslukujen avulla uuden historiallisen tiedon muodostamiseksi. Keräämieni tietojen perusteella myös hypoteesini oralismin leviämisestä kouluissa kuurojen opettamisessa sovelletun puhemetodin avulla näyttää pitävän paikkansa. Tietokoneavusteisten menetelmien

²⁵⁸ Alaluku 2.4.

²⁵⁹ S.10, kuva: Tutkimusprosessikaavio.

avulla voidaan tuottaa uudenlaisia näkökulmia ja lähestymistapoja aineistoon. Menetelmät ovat lopulta vain työkaluja uuden tiedon ja näkökulman muodostamiseen, eikä niiden avulla ole tarkoitus poistaa ”perinteisen” historian tutkimuksen vaiheita. Laajempi historiallinen konteksti yhdessä lähiluvun ja lähdekritiikin kanssa ovat edelleen välttämättömiä ilmiöiden, kuten tässä tutkimuksessa oralismin tarkasteluun.

Tiedonhaun kehittyminen tehokkaammaksi hakualgoritmien ja tietokoneavusteisten aineiston järjestämisen sekä suodattamisen avulla ei ole yksiselitteistä ajateltaessa määritelmää arkistoiden hiljaisuudesta. 2025-luvulle tultaessa yhä useampi hakukone sisältää teknologiaa, jonka toimintoja ja hakutuloksien ominaisuuksia ei ole selitetty täysin auki. Avoimuus (transparency) on yksi keskeisimmistä arkistojen toimintaperiaatteista,²⁶⁰ minkä vuoksi myöskin kollokaatioanalyysin tapaisille tilastollisille hakumenetelmille on edelleen paikkansa. Toisessa ääripäässä ovat esimerkiksi laajojen kielimallien (LLM) rikastamat hakumenetelmät, joiden tuottamat tulokset perustuvat todennäköisyyksiin sisältäen huomattavan määrän epävarmuutta riippumatta siitä, tunnetaanko mallien koulutusdata, vai ei.

Menetelmien avulla tuotettu näkökulma antaa myös muille tutkijoille mahdollisuuden arvioida tuloksia. Vastaaviin tulkintoihin olisi voinut päästä myös muiden menetelmien avulla. Esimerkiksi Killisen historiallisen kontekstin hahmottaminen olisi ollut mahdollista myös pelkkään lähilukuun perustuvassa tutkimusprosessissa ottamalla *Maahan lämpimään ja Viitotut muistot* pääasiallisiksi lähteiksi. Oma analyysini toi esille useita muita kiinnostavia teemoja, joita ei ollut tämän tutkielman rajauksesta johtuen mahdollista analysoida tarkemmin. Heuristinen koodin kehittämistyö tuotti lopulta useita mielenkiintoisia kollokaatioihin liittyviä lähdetekstikokonaisuuksia, joista voisi tehdä useamman tarkempaan lähilukuun perustuvan tutkimuksen.

Tiedonhakumenetelmäni ei ole aukoton, sillä aineisto analyysi ei tuota mittareita, joiden avulla siitä löytyisi määrällisyyteen perustuvia yleistäviä piirteitä. Samoin menetelmäni kadottaa eksploraatiivisesti haetun uuden tiedon kustannuksella aineiston analogiseen kontekstiin liittyviä ainutlaatuisia ominaisuuksia. Tällöin niihin kohdistuu teknisestä käsittelystä syntyviä arkistoiden hiljaisuuksia.

Menetelmäni on myös esimerkki etäluvun ja henkilöhistoriallisen tutkimuksen yhdistämisestä. Kustaa Killinen paljastui keskeiseksi hahmoksi tarkastellessa oralismin

²⁶⁰ Furner & Gilliland 2017, 609.

kehittymistä Suomessa, sillä kollokaatioiden visualisointi ja tarkempi tutkiminen osoittivat hänen ilmenemisensä koko tutkimusaineistoni käsittävän aikarajauksen ajan. Visualisointien ja aineiston järjestäminen tilastollisten tunnuslukujen perusteella toimii hyvin juuri tiedonhaun apukeinona, esimerkiksi tarkasteltaessa, olisivatko korkeimman log-dice-pisteytyksen saaneet kollokaatioparien lähdetekstit relevantteja tutkimuksen kannalta.

Tutkielmani sopii myös osaksi *Viitottujen muistojen* sekä *Maahan lämpimään* -teoksen tuottamaa narratiivia 1900-luvun taitteen oralismin noususta, millä on sittemmin katsottu olleen hyvin negatiivisia vaikutuksia kuurojen yhteisön asemaan. Puhemetodin soveltajilla oli vaikutusvaltainen asema esimerkiksi oppivelvollisuuslain taustalla ja heidän näkemyksensä ovat selvästi nähtävillä *Aistivalliskoulujen lehden* teksteissä.

Tutkielmassani käy myös ilmi, että arkistotieteessä ei ole pohdittu tarpeeksi teknisestä näkökulmasta digitaalisten ilmentymien säilyttämisen, ominaispiirteiden ja tiedonhaun suhdetta. Tutkimuksessa käytetyt digitoidut aineistot menettävät tiedonhakuun valmistavassa esikäsitelyssä ihmiselle ymmärrettäviä ominaisuuksia. Esimerkiksi sanomalehden artikkelin tokenisoiminen pilkkoo tekstin listaksi sanoja. Tässä muodossa sanat eivät enää ilmaise niiden alkuperäistä olemusta: siinä ei näy tekstin muotoiluja, sijaintia sivulla ja niin edelleen. Lopulliset tiedonhaun tulokset vaativat alkuperäiseen analogiseen kontekstiin liittyviä teknisesti tuotettuja ominaisuuksia, kuten optisen tekstintunnistuksen, sekä kattavia metatietoja ymmärrettävyyden takaamiseksi.

Tutkimusprosessin aikana kielimalliteknoologiaan ja luonnollisen kielen prosessointiin (NLP) perustuvat tiedonhaun menetelmät ovat kehittyneet huomattavasti. Niiden käytännön soveltaminen ei kuitenkaan ole kohdannut arkistotiedettä teoreettisella tasolla ja joidenkin näkökulmien mukaan ne ovat myös riittämättömiä asianhallinnan näkökulmasta.²⁶¹ Uuteen digisyntyiseen arkistoitavaan aineistoon liittyy useita ongelmia, kuten käyttörajoitteiset ja henkilötietoja sisältävät aineistot.²⁶² Tällaisten aineistojen järjestely ja metatiedottaminen kaupallisilla malleilla, kuten OpenAI:n GPT:n tai Googlen Geminin avulla on syystäkin herättänyt viranomaisorganisaatioiden huolen. Toinen keskeinen ongelma liittyy aineiston suureen määrään. Kielimalliteknologialla järjestettyjen aineistojen läpikäyminen käsin on ollut esimerkiksi Yhdysvaltojen kansallisarkisto NARAn tapauksessa täysin mahdotonta.²⁶³

²⁶¹ Henttonen 2025, 228–230.

²⁶² Henttonen 2025, 224.

²⁶³ Henttonen 2025, 229.

Samalla kielimallien tuottamat tulokset eivät käytännössä ole ikinä täysin virheettömiä. Jos käsitellyn aineiston halutaan täyttävän oikeellisuuden ja alkuperäisyyden ehdot, parhaat tulokset tulisi lähes kaikissa tapauksissa tarkastaa käsin, joka on jo todettu käytännössä mahdottomaksi.²⁶⁴ Digitaalisten ilmentymien verkostojen tutkimus tulee luultavasti ajankohtaiseksi esimerkiksi sähkösyntyisen viranomaisaineiston käytettävyyden kannalta, joten käytännön metodien kehittäminen lienee edelleen tarpeellista myös algoritmisten ratkaisujen kohdalla.

Tämän tutkimuksen tulokset osoittavat, että tietokoneavusteisilla menetelmillä on mahdollista löytää tutkimusaineistosta uudenlaista tietoa ja mahdollisia laajempia historiallisia kokonaisuuksia, jotka ansaitisivat tulla tutkittaviksi. Keskeinen menetelmä aineiston analysoimiseksi ei poista tutkimusprosessista perinteisiä historiantutkimuksen menetelmiä, vaan toimii tiedonhaun välineenä täydentäen Kansalliskirjaston sanahakua tilastollisella menetelmällä. Havaitut tulokset ovat pitkälti linjassa aiemman tutkimuksen kanssa. Aineiston hiljaisuudet vaikuttavat liittyvän pitkälti sanoma- ja aikakauslehtien sisältöön ja analyysin tulosten määrälliseen paljouteen. Siksi tuloksien valossa nyt hahmoteltuja tietokoneavusteisia analyyseja voi ajatella yhtenä lisänä historian tutkimusalan lähdeaineiston lukutapojen joukkoon ja joustavana lisänä osaksi jo käytössä olevia tiedonhaun työkaluja.

²⁶⁴ Henttonen 2025, 229.

Lähteet

Tietokannat, arkistot

Sanoma- ja aikakauslehdet Digitaaliset aineistot. Kansalliskirjasto.

<https://digi.kansalliskirjasto.fi>

Hufvudstadsbladet, 139/1908.

Hufvudstadsbladet, 283/1901.

Hufvudstadsbladet, 283/1901.

Hufvudstadsbladet, 283/1901.

Hufvudstadsbladet, 283/1901.

Hufvudstadsbladet, 27/1910.

Hufvudstadsbladet 250/1916.

Kuuromykkäin lehti 9/1912.

Kuuromykkäin lehti 9–10/1919.

Kuuromykkäin lehti 4/1919.

Suomen aistivalliskoulujen lehti 1/1892.

Suomen aistivalliskoulujen lehti, 2/1892.

Suomen aistivalliskoulujen lehti, 4/1892.

Kustaa Killinen. *Suomen aistivalliskoulujen lehti*, 5/1892.

Kustaa Killinen. *Suomen aistivalliskoulujen lehti*, 6/1892.

Suomen aistivalliskoulujen lehti, 4/1893.

Suomen aistivalliskoulujen lehti, 7–8/1893.

Suomen aistivalliskoulujen lehti 6–7/1896.

Suomen aistivalliskoulujen lehti, 5/1898.

Kustaa Killinen. *Suomen aistivalliskoulujen lehti*, 3–4/1903.

Suomen aistivalliskoulujen lehti 6–7/1905.

Kustaa Killinen. *Suomen aistivalliskoulujen lehti*, 1/1909.

Suomen aistivalliskoulujen lehti, 2/1911.

Kalle Rikala (tohtori). *Suomen urheilulehti*, 5/1914.

Tidskrift för döfstumma, 12/1898.

Tidskrift för döfstumma, 8–9/1916.

Uleåborgsbladet, 31/1892.

Åbo Tidning, 179/1890.

Tutkimuskirjallisuus

Acker Amelia: “When is a record” teoksesta *Research in the Archival Multiverse*. Toim. Gilliland Anne, Sue McKemmish, Lau Andrew. Victoria, Australia: Monash University Publishing, 2016.

Arvola Paavo, Keskustalo Heikki ja Kumpulainen Sanna: ”Hakukoneista kielimalleihin” teoksesta *Informaatiosta tiedon vuoksi, Informaatiotutkimuksen ilmiöitä ja voimia*. Toim. Kumpulainen Sanna ja Sihvonen Mika. Vastapaino, Tampere 2025. DOI: <https://doi.org/10.58181/VP9789523972834> [viitattu 26.4.2026]

Carter, Rodney G. S.: *Of Things Said and Unsaid: Power, Archival Silences, and Power in Silence*. *Archivaria* 61: Special Section on Archives, Space and Power, 2006. 215–233.

Elo Kimmo & Kiiskinen Harri: “Digitaaliset tutkimusmenetelmät kulttuurihistorian tutkimuksessa, semanttinen Web ja verkostanalyysi” teoksesta *Kulttuurihistorian tutkimus: lähteistä menetelmiin ja tulkintaan*. Toim. Mähkä Rami, Ahonen Marika, Niko Heikkilä, Ollitervo Sakari & Räsänen Marika. Turku: Kulttuurihistorian seura, 2022.

Eräkaski, Mikko: Syntysähköinen vai digitoitu? *Akti* 2/2012, 23.

Fruchterman Thomas, Reingold Edward. 1991: Graph Drawing by Force-Directed Placement. *Software, Practice & Experience*, Wiley, 21 1991, pp.1129–1164, doi:10.1002/spe.4380211102. [viitattu 26.4.2026]

Furner Jonathan & Gilliland Anne: "Archival IR" teoksesta *Research in the Archival Multiverse*. Toim. Gilliland Anne, McKemmish Sue, Lau Andrew. Victoria, Australia: Monash University Publishing, 2016.

Gablasova Dana, McEnery Tony and Vaclav Brezina: "Collocations in Corpus-Based Language Learning Research: Identifying, Comparing, and Interpreting the Evidence", *Language learning, A journal of research in language studies* 2017.

Gilliland Anne, Sue McKemmish, and Andrew J Lau: *Research in the Archival Multiverse*. Clayton, Victoria, Australia: Monash University Publishing, 2016.

Hakkarainen Heidi, Kurvinen Heidi, Paju Petri, Salmi Hannu ja Sorvali Satu: "Digitoidut sanoma- ja aikakauslehdet lähdeaineistona" teoksesta *Kulttuurihistorian tutkimus: lähteistä menetelmiin ja tulkintaan*. Toim. Mähkä Rami, Ahonen Marika, Heikkilä Niko, Ollitervo Sakari ja Räsänen Marika. Turku: Kulttuurihistorian seura, 2022.

Heinonen Henriikka: *Valtion ja yksilön valtasuhde vuoden 1935 sterilointilain säätämisprosessissa*. Helsingin yliopisto, pro gradu -tutkielma, 2021.

Hill Mark, Tolonen Mikko: *A Computational Investigation into the Authorship of Sister Peg*. Helsingin yliopiston tutkimusartikkelit, 2021.

Hyvönen Taru: "Digitoitu aineisto oli helppo löytää": *Aiheet, alkuperäisaineistot ja digitalisaation vaikutus Suomen historiaa käsittelevissä pro gradu -tutkielmissa 1995–2017*. Turun yliopisto, pro gradu -tutkielma, 2022.

Henttonen Pekka: *Johdatus asiakirjahallinnan tutkimukseen*. Uudistettu painos, Avain, Helsinki 2023.

Henttonen Pekka: "Sähköinen asiakirjahallinta hakee vielä uomiaan" teoksesta *Informaatiosta tiedon vuoksi, Informaatiotutkimuksen ilmiöitä ja voimia*. Toim. Kumpulainen Sanna ja Sihvonen Mika. Vastapaino, Tampere 2025. DOI: <http://doi.org/10.58181/VP9789523972834> [viitattu 26.4.2026]

Ivaska Ilmari: *Edistyneen oppijansuomen konstruktiopiirteitä korpusvetoisesti: avainrakenneanalyysi*. Turun yliopisto, Utuling 2015.

Katsui Hisayo, Koivisto Maija, Tepora-Niemi Suvi-Maaria, Meriläinen Niina, Rautiainen Pauli, Rainò Päivi, Tarvainen Merja ja Hillamo Heikki: *Viitotut muistot, Selvitys kuuroihin ja viittomakielisiin Suomen historiassa 1900-luvulta nykypäivään kohdistuneista vääryyksistä*

sekä niiden käsitteelyyn tarkoitettun totuus- ja sovintoprosessin käynnistämisen edellytyksistä. Helsinki Valtioneuvoston kanslia, 2021.

Kumpulainen Sanna ja Sihvonen Mika: ”Tehtävälähtöinen informaatiovuorovaikutus” teoksesta *Informaatiosta tiedon vuoksi, Informaatiotutkimuksen ilmiöitä ja voimia*. Toim.

Kumpulainen Sanna ja Sihvonen Mika. Vastapaino, Tampere 2025. DOI:

<http://doi.org/10.58181/VP9789523972834> [viitattu 26.4.2026]

Lybeck, Jari et al.: *Arkistot – yhteiskunnan toimiva muisti. Asiakirjahallinnon ja arkistotoimen oppikirja*. Arkistolaitoksen toimituksia 2. Helsinki 2006.

Tikka Santtu, Kopra Juho, Heinäniemi Merja, López-Pernas Sonsoles ja Saqr Mohammed: *Getting Started with R for Education Research*. 2024 Analytics Methods and Tutorials: A Practical Guide Using R (pp. 67-94). Springer. https://doi.org/10.1007/978-3-031-54464-4_3 [viitattu 26.4.2026]

Laakso Mikko ja Salmi Eeva: *Maahan lämpimään. Suomen viittomakielisten historiaa*. Helsinki Kuurojen liitto, 2005.

Kyröläinen Aki-Juhani ja Laippala Veronika: ”Määrällinen korpuslingvistiikka” teoksesta *Kielentutkimuksen menetelmiä I-IV*. Toim. Hamunen Markus, Nikanne Urpo, Luodonpää-Manni Milla, Miestamo Matti, Sinnemäki Kaius. ja Konstenius Reetta (eds). Helsinki Suomalaisen Kirjallisuuden Seura 487–524, 2020. DOI: <https://doi.org/10.21435/skst.1457> [viitattu 26.4.2026]

McEnery Tony. and Vaclav Brezina: *Fundamental principles of corpus linguistics*. 22/09/2022 Cambridge: Cambridge University Press. ISBN: 9781107624689. Electronic ISBN: 9781107110625.

Nyysönen, Jukka: ”Saamelaisten kouluolot 1900-luvulla” teoksesta *Saamelaisten kansanopetuksen ja koulunkäynnin historia Suomessa*. Toim. Keskitalo Pigga, Lehtola Veli-Pekka ja Paksuniemi Merja. Turku Siirtolaisinstituutti, 63–86, 2014.

Paju Petri, Rantala Heli ja Salmi Hannu. "Towards an Ontology and Epistemology of Text Reuse" teoksesta *Digitised Newspapers – A New Eldorado for Historians?: Reflections on Tools, Methods and Epistemology*. Toim. Bunout Estelle, Ehrmann Maud ja Clavert Frédéric, Berlin, Boston: De Gruyter Oldenbourg, 2023, pp. 253–274.

<https://doi.org/10.1515/9783110729214-012> [viitattu 26.4.2026]

Pivovarova Lidia, Jean-Caurant Axel, Avikainen Jari, Alnajjar Khalid, Granroth-Wilding Mark, Leppänen Leo, Zosa Elaine ja Toivonen Hannu: Personal Research Assistant for Online Exploration of Historical News. in *Advances in Information Retrieval : 42nd European Conference on IR Research, ECIR 2020, Lisbon, Portugal, April 14–17, 2020, Proceedings, Part II. Lecture Notes in Computer Science*, Springer, Cham, 42nd European Conference on IR Research, Lisbon, Portugal, 14/04/2020. https://doi.org/10.1007/978-3-030-45442-5_62 [viitattu 26.4.2026]

Risku Marita: *Johdatus älykkään data-analyysin prosessimalliin, luonnollisen kielen käsittelyyn, koneoppimiseen sekä niihin perustuvien käytötapausten kehittämiseen Valtorin hallinnoiman tekoälyalustan päälle*. Turun yliopisto, pro gradu -tutkielma, 2020.

Rozenberg Sebastian: *Digital records as relational objects—Yuk Hui’s concept of digital objects applied to archival science*. *Archival Science* (2021) 21:193–218. <https://doi.org/10.1007/s10502-021-09357-0> [viitattu 26.4.2026]

Vaclav Brezina, ja Platt William: *Lance box X 5.00 manual*. Lancaster university ac-cs publications. 2024.

Wilkinson Saldaña Zoë: *Sentiment Analysis for Exploratory Data Analysis. Programming historian*, haettu 6.9.2023.

Öksüz Dogus, Vaclav Brezina, Rebuschat Patrick: *Collocational Processing in L1 and L2: The Effects of Word Frequency, Collocational Frequency, and Association*, *Language Learning*, 71, 2021. <https://doi.org/10.1111/lang.12427> [viitattu 26.4.2026]

Liitteet

Liite 1. R-ohjelman koodi

R Notebook

```

if (!require(dplyr)) install.packages("dplyr")
library(dplyr)
if (!require(readxl)) install.packages('readxl')
library(readxl)
if (!require(summarytools)) install.packages('summarytools')
library(summarytools)
if (!require(factoextra)) install.packages('factoextra')
library(factoextra)
if (!require(ggpubr)) install.packages('ggpubr')
library(ggpubr)
if (!require(ggeffects)) install.packages('ggeffects')
library(ggeffects)
if (!require(tidyverse)) install.packages('tidyverse')
if (!require(quanteda)) install.packages("quanteda")
if (!require(quanteda.textstats)) install.packages("quanteda.textstats")
library(quanteda)
library(quanteda.textstats)
if (!require(writexl)) install.packages("writexl")
library(writexl)
if (!require(tidytext1)) install.packages('tidytext')library(tidytext)
if (!require(wordspace)) install.packages("wordspace")
library(wordspace)
if (!require(textTinyR)) install.packages("textTinyR")
library(textTinyR)
library(stringr)
if (!require(igraph)) install.packages("igraph")
library("igraph")
if (!require(ggraph)) install.packages("ggraph")
library("ggraph")
if (!require(tidygraph)) install.packages("tidygraph")
library("tidygraph")

```

```

datatest <- read_excel("../Data/dvdf_combined.xlsx")
View(datatest)
colnames(datatest)[6] <- "text"
subdata <- datatest %>%
  select(1,3,4,5,6)
View(subdata)
corpus <- quanteda::corpus(subdata$text)
tokenit <- tokens(corpus, remove_punct = TRUE)
bigrams <- tokens_ngrams(tokenit, n = 2)
collocations_all <- textstat_collocations(bigrams, min_count = 5, size = 2)
total_tokens <- sum(n_token(tokenit))
word_freqs <- textstat_frequency(dfm(tokenit)) %>%
  select(feature, frequency) %>%
  mutate(feature = trimws(tolower(feature)))
collocations_all <- collocations_all %>%
  mutate(
    word1 = str_extract(collocation, "^[^_]+"), # Everything before the first "_"
    word2 = str_extract(collocation, "[^_]+$") # Everything after the last "_"
  )

```

```

)
word1_freqs <- word_freqs %>%
  rename(word1 = feature, freq1 = frequency)

word2_freqs <- word_freqs %>%
  rename(word2 = feature, freq2 = frequency)

processedcollocations <- collocations_all %>%
  left_join(word1_freqs, by = "word1") %>%
  left_join(word2_freqs, by = "word2")

median_freq <- median(word_freqs$frequency, na.rm = TRUE)
processedcollocations <- processedcollocations %>%
  mutate(
    freq1 = ifelse(is.na(freq1), median_freq, freq1),
    freq2 = ifelse(is.na(freq2), median_freq, freq2)
  )

processedcollocations <- processedcollocations %>%
  mutate(
    p_xy = count / total_tokens,      # Joint probability P(x, y)
    p_x = freq1 / total_tokens,       # Probability of word1
    p_y = freq2 / total_tokens,       # Probability of word2
    mi_score = log2(p_xy / (p_x * p_y)), # Mutual Information Score
    log_dice = 14 + log2((2 * count) / (freq1 + freq2)) # Log Dice Score
  )
View(processedcollocations)

#Find Top 50 collocations overall. The results are often quite ambiguous

top_collocations <- processedcollocations %>%
  slice_max(log_dice, n = 50)

results_list <- list()

# Loop through top collocations using collocation as corresponding item
for (i in 1:nrow(top_collocations)) {
  colloc <- top_collocations$collocation[i]
  word1 <- top_collocations$word1[i]
  word2 <- top_collocations$word2[i]

  pattern <- phrase(c(word1, word2))

  # Find documents containing this pattern
  docs_with_pattern <- tokens_select(tokenit, pattern, selection = "keep")

  # Get document IDs that have tokens left (meaning they contain the pattern)
  doc_ids <- which(n_token(docs_with_pattern) > 0)

  if (length(doc_ids) > 0) {
    # Create temporary df for hits
    temp_results <- data.frame(
      collocation = colloc,
      word1 = word1,
      word2 = word2,
      log_dice = top_collocations$log_dice[i],
      doc_id = doc_ids,
      stringsAsFactors = FALSE
    )
  }
}

```

```

    # Add to results list
    results_list[[i]] <- temp_results
  }
}

# Combine all results
all_matches <- do.call(rbind, results_list)

# Now join with metadata
final_results <- all_matches %>%
  left_join(
    subdata %>%
      select(1, 2, 3, 4, 5) %>%
      mutate(doc_id = row_number()),
    by = "doc_id"
  )

View(final_results)

# Define your node word. The node should be in the item "word1_node + node_word2"
node_word <- "i"

# Find all collocations
search_collocations <- processedcollocations %>%
  filter(word1 == node_word | word2 == node_word)

top_node_collocations <- search_collocations %>%
  filter(!is.na(log_dice)) %>%
  slice_max(log_dice, n = 50)

# Error-handling in case of no hits.
if (nrow(top_node_collocations) == 0) {
  stop("No collocations found for the node word.")
}

print(paste("Found", nrow(search_collocations), "collocations containing 'node_word'"))
## [1] "Found 1442 collocations containing 'node_word'"

print(paste("Top log-dice score:", max(search_collocations$log_dice, na.rm = TRUE)))
## [1] "Top log-dice score: 11.8646563436002"

# Organize collocations by whether node appears as word1 or word2
node_as_word1 <- search_collocations %>%
  filter(word1 == node_word) %>%
  arrange(desc(log_dice)) %>%
  select(collocation, word2, log_dice)

node_as_word2 <- search_collocations %>%
  filter(word2 == node_word) %>%
  arrange(desc(log_dice)) %>%
  select(collocation, word1, log_dice)

# Rename for clarity and add positional information
colnames(node_as_word1)[2] <- "collocate"
colnames(node_as_word2)[2] <- "collocate"
node_as_word1$position <- "after" # word2 comes after
node_as_word2$position <- "before" # word1 comes before

# Combine the results and find collocations
all_node_collocations <- rbind(node_as_word1, node_as_word2) %>%

```

```

arrange(desc(log_dice))

# Empty list for results
results_list <- list()

for (i in 1:nrow(top_node_collocations)) {
  colloc <- top_node_collocations$collocation[i]
  word1 <- top_node_collocations$word1[i]
  word2 <- top_node_collocations$word2[i]

  pattern <- phrase(c(word1, word2))

  # Find documents containing the pattern
  docs_with_pattern <- tokens_select(tokenit, pattern, selection = "keep")

  # Get document IDs that have tokens left (meaning they contain the pattern)
  doc_ids <- which(ntoken(docs_with_pattern) > 0)

  if (length(doc_ids) > 0) {
    # Create temporary results
    temp_results <- data.frame(
      collocation = colloc,
      word1 = word1,
      word2 = word2,
      log_dice = top_node_collocations$log_dice[i],
      doc_id = doc_ids,
      stringsAsFactors = FALSE
    )

    # Add to results list
    results_list[[i]] <- temp_results
  }
}

# Combine all results
all_matches <- do.call(rbind, results_list)

# Now join with metadata
final_results <- all_matches %>%
  left_join(
    subdata %>%
      select(1,2, 3, 4, 5) %>%
      mutate(doc_id = row_number()),
    by = "doc_id"
  )

View(final_results)

#Visualization
#Dimensional visualization akin to LancsBox 2024
central_word <- "döfstum"

# Filter collocations containing word1 or word2
# Add here & Log_dice > 4 if results have too much noise
filtered_edges_all <- processedcollocations %>%
  filter(grepl(central_word, word1) | grepl(central_word, word2)) %>%
  arrange(desc(log_dice)) %>% # Sorting by log_dice instead of count
  slice_max(log_dice, n = 50) # Using log_dice as the measure

# Ensure central_word stays in the dataset (if missing)
if (!any(central_word == filtered_edges_all$word1 | central_word == filtered_edges_all$word

```



```

central_word <- "döfstum"

# Get edges where one node is exactly the central word
filtered_edges <- processedcollocations %>%
  filter(word1 == central_word | word2 == central_word) %>%
  # Create a new column with the collocate (non-döfstumma word)
  mutate(
    collocate = ifelse(word1 == central_word, word2, word1)
  ) %>%
  arrange(desc(log_dice)) %>%
  slice_max(log_dice, n = 20)

# Create a new edge list that has the central word and collocates
edge_list <- data.frame(
  from = rep(central_word, nrow(filtered_edges)),
  to = filtered_edges$collocate,
  log_dice = filtered_edges$log_dice
)

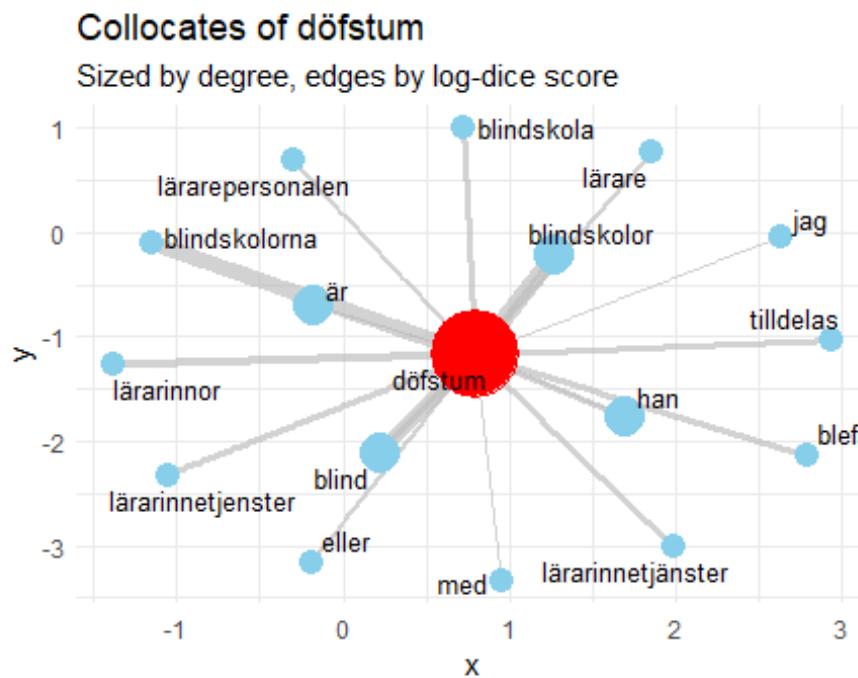
# Create the graph from our custom edge list
graph_tbl <- as_tbl_graph(edge_list, directed = FALSE)

# Now we'll add node attributes
graph_tbl <- graph_tbl %>%
  activate(nodes) %>%
  mutate(
    # Get the name from the graph
    node_name = .N()$name,
    # Set size and color
    size = centrality_degree() * 3,
    color = ifelse(node_name == central_word, "red", "skyblue")
  )

# Check if we have data
if(nrow(edge_list) > 0) {
  # The plot with a more suitable layout
  p <- ggraph(graph_tbl, layout = "nicely") + # 'nicely' tries to find the best layout
    geom_edge_link(aes(edge_width = log_dice), alpha = 0.7, color = "gray") +
    geom_node_point(aes(size = size, color = color)) +
    geom_node_text(aes(label = node_name), repel = TRUE,
                  size = 3.5, max.overlaps = 20) +
    scale_color_identity() +
    scale_edge_width(range = c(0.5, 3)) +
    scale_size(range = c(4, 15)) + # Ensure nodes are visible
    labs(title = paste("Collocates of", central_word),
         subtitle = "Sized by degree, edges by log-dice score") +
    theme_minimal() +
    theme(legend.position = "none",
          plot.margin = margin(20, 20, 20, 20))

  # Print the plot
  print(p)
} else {
  cat("No connections found for the central word:", central_word, "\n")
}

```



```

if(nrow(filtered_edges) == 0) {
  cat("No matches found for", central_word, "in the dataset.\n")

  # Look for similar words that might be in the dataset
  possible_matches <- collocations_all %>%
    filter(grepl(central_word, word1) | grepl(central_word, word2)) %>%
    select(word1, word2) %>%
    unique()

  if(nrow(possible_matches) > 0) {
    cat("However, found similar words containing", central_word, ":\n")
    print(possible_matches)
  }
}

# Get all unique collocation pairs
collocation_pairs <- filtered_edges_all %>%
  mutate(pair_id = row_number()) %>%
  select(pair_id, word1, word2, log_dice)

# New frame for storing the results
results_df <- data.frame()

# Loop through each collocation pair
for (i in 1:nrow(collocation_pairs)) {
  word1 <- collocation_pairs$word1[i]
  word2 <- collocation_pairs$word2[i]
  pair_id <- collocation_pairs$pair_id[i]
  log_dice_score <- collocation_pairs$log_dice[i]

  # Find all documents that contain both words
  matching_docs <- subdata %>%
    filter(grepl(word1, `text`, ignore.case = TRUE) &
           grepl(word2, `text`, ignore.case = TRUE)) %>%
  # Add the collocation information
  mutate(

```

```

    collocation_pair = paste(word1, word2, sep = " + "),
    log_dice = log_dice_score,
    pair_id = pair_id
  )

  # Add to our results
  results_df <- bind_rows(results_df, matching_docs)
}

# Now results_df contains all the original metadata plus the collocation information
# You can view a summary of which documents contain which collocations
collocation_summary <- results_df %>%
  group_by(collocation_pair, `Päänimeke`, `PVM`, `Nro`, `Sivu`) %>%
  summarize(
    log_dice = first(log_dice),
    .groups = 'drop'
  ) %>%
  arrange(desc(log_dice))

View(collocation_summary)

# Calculate timeseries for collocations

# Extract year from date column
Year_df <- results_df %>%
  mutate(Year = substr(PVM, 1, 4)) # Takes first 4 characters (YYYY)

# Group by and count occurrences into new dataframe
collocation_yearly_counts <- Year_df %>%
  group_by(Year, collocation_pair) %>%
  summarize(
    count = n(),
    log_dice = first(log_dice), # Keep log_dice for reference
    .groups = 'drop'
  ) %>%
  arrange(Year, desc(count))
View(collocation_yearly_counts)

top_10_collocations <- collocation_yearly_counts %>%
  group_by(collocation_pair) %>%
  summarize(total_count = sum(count), max_log_dice = max(log_dice)) %>%
  slice_max(max_log_dice, n = 10)

top_10_with_metadata <- top_10_collocations %>%
  left_join(results_df %>% select(collocation_pair, Päänimeke, PVM, Nro, SivU), by = "collocation_pair") %>%
  distinct() # Ensure unique rows

view(top_10_with_metadata)

ggplot(collocation_yearly_counts %>% filter(collocation_pair %in% top_10_collocations$collocation_pair),
  aes(x = as.numeric(Year), y = count, color = collocation_pair, group = collocation_pair)) +
  geom_line(size = 1) +
  geom_point(size = 2) +
  theme_minimal() +
  labs(title = "Top Collocations Over Time",
    x = "Year",
    y = "Frequency per Year") +
  scale_x_continuous(breaks = seq(min(as.numeric(collocation_yearly_counts$Year)),
    max(as.numeric(collocation_yearly_counts$Year)), by = 5))
+
  theme(legend.position = "right")

```

```

# Tool to find and visualize individual collocations

# Define the target collocation
target_collocation <- "döfstumskoleföreståndaren + killinen"

# Filter for only this collocation
single_collocation_df <- collocation_yearly_counts %>%
  filter(collocation_pair == target_collocation)

ggplot(single_collocation_df, aes(x = as.numeric(Year), y = count)) +
  geom_line(size = 1, color = "blue") + # Line graph
  geom_point(size = 3, color = "red") + # Points for clarity
  theme_minimal() +
  labs(title = paste("Individual collocation Over Time :", target_collocation),
       x = "Year",
       y = "Frequency per Year") +
  scale_x_continuous(breaks = seq(min(as.numeric(single_collocation_df$Year)),
                                max(as.numeric(single_collocation_df$Year)), by = 1))

#Some general visualizations

# Frequency
processedcollocations %>%
  arrange(desc(count)) %>%
  top_n(10, count) %>%
  ggplot(aes(x = reorder(collocation, count), y = count)) +
  geom_col(fill = "steelblue") +
  coord_flip() +
  labs(title = "Most Frequent Collocations", x = "Collocation", y = "Count")

#Relative frequency
ggplot(collocations_df %>% top_n(10, freq), aes(x = reorder(collocation, freq), y = freq))
+
  geom_col(fill = "steelblue") +
  coord_flip() +
  labs(title = "Top 10 Collocations with relative frequency", x = "Collocation", y = "Relative Frequency") +
  theme_minimal()

```

#The visualization for parts that require for-loops is omitted from word document, since running the code while knitting causes crashes due to heavy computation.

Liite 2. Matemaattiset kaavat

$$14 + \log_2 \text{Dice} \left(\frac{\|w_1, R, w_2\|}{\|w_1, R, *\|}, \frac{\|w_1, R, w_2\|}{\|*, *, w_2\|} \right) = 14 + \log_2 \frac{2 \cdot \|w_1, R, w_2\|}{\|w_1, R, *\| + \|*, *, w_2\|}$$

Log-uskottavuuden laskukaava.

Liite 3. Kollokaatioiden visualisoinnit ja tiedonhaun indeksit

collocation_pair	Päänimeke	PVM	Nro	Sivu	log_dice	text_snippets
1 döfstumskoleföreståndaren + killinen	Aftonposten	1898-06-09	130	2	11.872592	mark. Döfstumskoleföreståndaren i Kuopio K. Killinens anhåll...
2 döfstumskoleföreståndaren + killinen	Aftonposten	1898-10-12	236	2	11.872592	i går såsom kandidater för landtdagsvalet döfstumskoleföre...
3 döfstumskoleföreståndaren + killinen	Aftonposten	1898-10-17	240	1	11.872592	kommunalrådet F. W. Lagerstedt samt för Kuopio stad döfst...
4 döfstumskoleföreståndaren + killinen	Aftonposten	1899-09-18	190 Landsortsupplagan	2	11.872592	landtdagsmän för Kuopio stad valdes i lördags döfstumskol...
5 döfstumskoleföreståndaren + killinen	Aftonposten	1899-09-18	190 Stadsupplagan	2	11.872592	landtdagsmän för Kuopio stad valdes i lördags döfstumskol...
6 döfstumskoleföreståndaren + killinen	Björneborgs Tidning	1891-06-19	47	3	11.872592	festplan, hwarest wählilet hälsningstal hölls af döfstumskol...
7 döfstumskoleföreståndaren + killinen	Björneborgs Tidning	1898-11-02	85	2	11.872592	och hofrättsassom F. V. Eriksson. Kuopio: döfstumskoleföres...
8 döfstumskoleföreståndaren + killinen	Björneborgs Tidning	1900-10-25	124	2	11.872592	8, kollegan mag. S. Westerlund 8, direktorn för döfstumskol...
9 döfstumskoleföreståndaren + killinen	Björneborgs Tidning	1900-11-20	135	2	11.872592	blifvit återvald med 245 röster, hvarjämte döfstumskoleföres...
10 döfstumskoleföreståndaren + killinen	Björneborgs Tidning	1912-10-08	75	4	11.872592	Till doktor Nordbergs minne talade vackert döfstumskolefö...
11 döfstumskoleföreståndaren + killinen	Björneborgs Tidning	1916-01-21	6	2	11.872592	utredning. Stadsfullmäktiges i Kuopio ordförande, döfstums...
12 döfstumskoleföreståndaren + killinen	Borgå Nya Tidning	1899-11-21	91	2	11.872592	Bankkontorsföreståndaren Herman Roschier (ny). Kuopio: D...
13 döfstumskoleföreståndaren + killinen	Borgåbladet	1899-11-18	90	2	11.872592	Bankkontorsföreståndaren Herman Roschier (ny). Kuopio: D...
14 döfstumskoleföreståndaren + killinen	Borgåbladet	1904-10-19	82	3	11.872592	landtdagsmän för Kuopio stad valdes i lördags döfstumskol...
15 döfstumskoleföreståndaren + killinen	Dagens Press	1914-10-25	213	3	11.872592	E. Bonsdorff, Helsingfors, 12 röster och döfstumskoleförestå...
16 döfstumskoleföreståndaren + killinen	Dagens Press	1916-01-18	13	4	11.872592	utredning. Stadsfullmäktiges i Kuopio ordförande, döfstums...
17 döfstumskoleföreståndaren + killinen	Finland	1891-09-10	209	3	11.872592	Trfors, handl. J. V. Söderlund från Raumo och döfstumskolef...
18 döfstumskoleföreståndaren + killinen	Finlands Allmänna Tidning	1905-11-30	273	1	11.872592	Brofeldt med 1,924 röster af 256 röstande, döfstumskolefö...
19 döfstumskoleföreståndaren + killinen	Flyttfågeln : ströskrift	1899-04-07	1	2	11.872592	åter fanfarer och talarestolen bestogs af döfstumskoleförest...
20 döfstumskoleföreståndaren + killinen	Folket	1904-10-22	34	3	11.872592	Aliq; och apotekar M. Björklund; för Kuopio: döfstumskolefö...
21 döfstumskoleföreståndaren + killinen	Fredrikshamns Tidning	1904-10-19	80	2	11.872592	valdes i lördags till sta dens representanter döfstumskolefö...
22 döfstumskoleföreståndaren + killinen	Fredrikshamns Tidning	1905-12-02	92	4	11.872592	Bro teldt med 1,924 röster af 256 röstande, döfstumskolefö...
23 döfstumskoleföreståndaren + killinen	Hangö	1898-10-16	124	3	11.872592	Till landtdagsman för Kuopin walde i går döfstumskoleföre...
24 döfstumskoleföreståndaren + killinen	Hangö	1904-10-18	123	2	11.872592	landtdagsmän för Kuopio stad walde i lördags döfstumskol...
25 döfstumskoleföreståndaren + killinen	Hangö	1904-11-03	130	2	11.872592	Hammarström. Nystad: konsul Erik Sawón. Kuopio: döfstum...
26 döfstumskoleföreståndaren + killinen	Helsingfors Aftonblad	1893-08-31	61	2	11.872592	samma tid lärarinnan vid döfstumsskolan i Kuopio L. Aschan...
27 döfstumskoleföreståndaren + killinen	Helsingfors-Posten	1903-02-03	32	2	11.872592	viceordf. dr Ernst Wendelin. Kuopio: ordf. döfstumskoleföres...
28 döfstumskoleföreståndaren + killinen	Helsingfors-Posten	1904-01-05	4	3	11.872592	stadsfullmäktiges i Kuspio ordförande walde tlån 2 dennes ...
29 döfstumskoleföreståndaren + killinen	Helsingfors-Posten	1904-10-16	281 A	2	11.872592	landtdagsmän för Kuopio atad walde i går döfstumsskolefö...
30 döfstumskoleföreståndaren + killinen	Helsingfors-Posten	1905-11-30	318	2	11.872592	Brofeldt med 1,924 röster af 256 röstande, döfstumskolefö...
31 döfstumskoleföreståndaren + killinen	Helsingfors-Posten	1905-12-23	341	4	11.872592	deras bekantskap och emottaga råd af dem. Döfstumskolef...
32 döfstumskoleföreståndaren + killinen	Hufvudstadsbladet	1890-09-11	246	2	11.872592	röster af 71 röstande. Närmast i röstetal voro döfstumskolef...
33 döfstumskoleföreståndaren + killinen	Hufvudstadsbladet	1891-05-01	116	5	11.872592	hofrättsrådet D. V. Åkerman, handl. F. Neppenström, döfstu...
34 döfstumskoleföreståndaren + killinen	Hufvudstadsbladet	1891-06-19	163	2	11.872592	Anslående, wählilet hälsningstal hölls af döfstumskoleförest...
35 döfstumskoleföreståndaren + killinen	Hufvudstadsbladet	1891-09-10	245	2	11.872592	Trfors, handl. I. V. Söderlund från Raumo och döfstumskolef...
36 döfstumskoleföreståndaren + killinen	Hufvudstadsbladet	1898-10-14	278	4	11.872592	Kuopio har uppståit som landtdagskandidater döfstumskole...
37 döfstumskoleföreståndaren + killinen	Hufvudstadsbladet	1898-10-16	280	5	11.872592	en lärarinnetjänst vid döfstumskolan i Borgå samt förestånd...

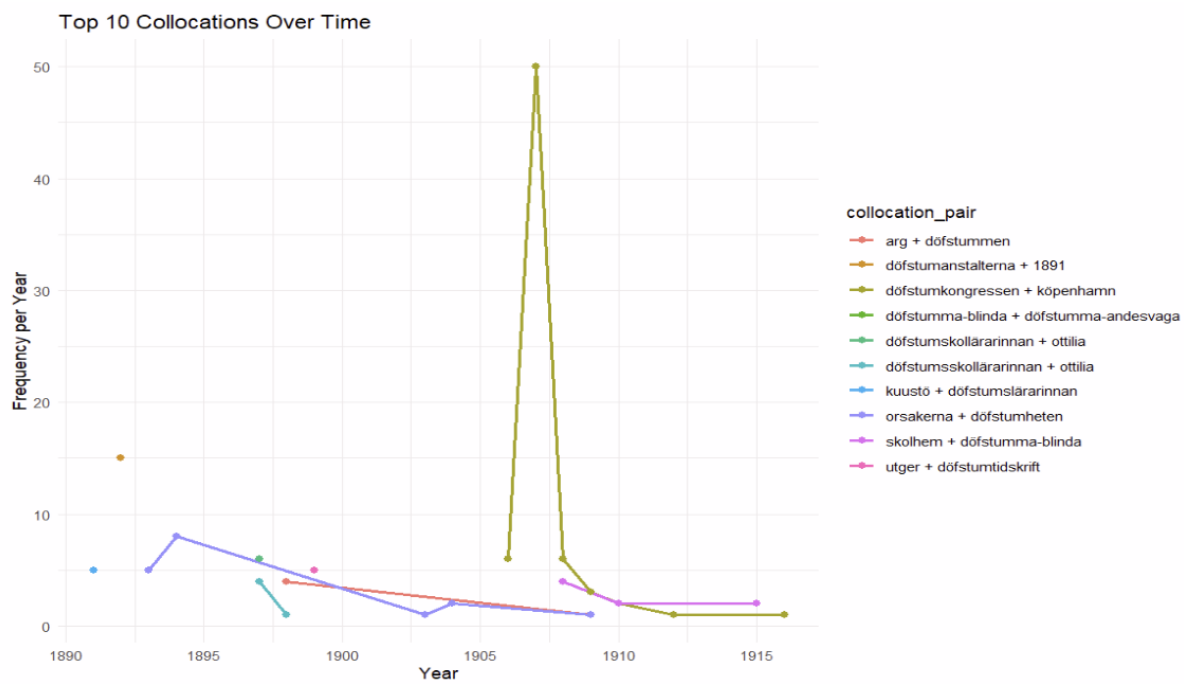
Kuva 1 Kustaa Killinen

	collocation_pair	Pääniemeke	PVM	Nro	Sivu	log_dice	text_snippets
1071	döfstumma + andesvaga	Suomen aistivalliskoulujen lehti	1893-02	1-2	21	4.814899	abnorme Börens Undervisning" af år 1881. § i- För döfstumm...
1072	döfstumma + andesvaga	Suomen aistivalliskoulujen lehti	1893-03	3	6	4.814899	just på abnormskolans område. Ty för abnorma barn, döfstu...
1073	döfstumma + andesvaga	Suomen aistivalliskoulujen lehti	1893-09	9	9	4.814899	bygder barn och unga hoptals strömma till läroverken. Döfs...
1074	döfstumma + andesvaga	Suomen aistivalliskoulujen lehti	1893-12	12	18	4.814899	12. I hvilken riktning bör en utveckling af vår döfstumskola ...
1075	döfstumma + andesvaga	Suomen aistivalliskoulujen lehti	1894-12	12	13	4.814899	för Örebro döfstumskola i Sverige, J. Blomkvist utgifven läro...
1076	döfstumma + andesvaga	Suomen aistivalliskoulujen lehti	1895-01	1	15	4.814899	Skolpligtslagens § 1 lyder: Icke fullsinnade (blinda, döfstum...
1077	döfstumma + andesvaga	Suomen aistivalliskoulujen lehti	1898-01	1	2	4.814899	Kirj. N. Helsingius 57. Ännu ett ord om de andesvaga döfstu...
1078	döfstumma + andesvaga	Suomen aistivalliskoulujen lehti	1898-02	2	1	4.814899	frågan om hvad som borde göras för våra andesvaga döfstu...
1079	döfstumma + andesvaga	Suomen aistivalliskoulujen lehti	1898-03	3	6	4.814899	eller både döfstuinma och blinda, ske i skolor för döfstumm...
1080	döfstumma + andesvaga	Suomen aistivalliskoulujen lehti	1898-03	3	13	4.814899	aistivallisten luettelossa. 1) Statistiska uppgifter ang. döfstu...
1081	döfstumma + andesvaga	Suomen aistivalliskoulujen lehti	1898-05	5	1	4.814899	om hvad som borde göras för våra andesvaga döfstumma e...
1082	döfstumma + andesvaga	Suomen aistivalliskoulujen lehti	1898-05	5	2	4.814899	utomlands finnas speciella anstalter för döfstumma andesva...
1083	döfstumma + andesvaga	Suomen aistivalliskoulujen lehti	1898-05	5	3	4.814899	tillgång på bildbara, enbart döfstumma elever, till fullo hunn...
1084	döfstumma + andesvaga	Suomen aistivalliskoulujen lehti	1898-05	5	4	4.814899	Behöfva de döfstumma skolundervisning och kunna de and...
1085	döfstumma + andesvaga	Suomen aistivalliskoulujen lehti	1898-05	5	5	4.814899	vore hvarken döfstumskolorna eller de hjälpbehöfvande döf...
1086	döfstumma + andesvaga	Suomen aistivalliskoulujen lehti	1898-05	5	6	4.814899	större nytta af att förblifva inom den egentliga döfstumskol...
1087	döfstumma + andesvaga	Suomen aistivalliskoulujen lehti	1898-05	5	7	4.814899	med andra ord just den undervisning, som man på döfstum...
1088	döfstumma + andesvaga	Suomen aistivalliskoulujen lehti	1898-05	5	8	4.814899	välottad döfstum inträde i sina skolor. Så länge kännedome...
1089	döfstumma + andesvaga	Suomen aistivalliskoulujen lehti	1902-02	2	3	4.814899	af rätten. Emedan det ganska ofta torde hända, att döfstum...
1090	döfstumma + andesvaga	Suomen aistivalliskoulujen lehti	1904-04	4	1	4.814899	Om utbredningen af döfstumhet, blindhet och andesvaghete...
1091	döfstumma + andesvaga	Suomen aistivalliskoulujen lehti	1904-04	4	2	4.814899	skolväsendet af den 4 maj 1897, att statistiken rörande de d...
1092	döfstumma + andesvaga	Suomen aistivalliskoulujen lehti	1904-04	4	3	4.814899	allmänna befolkningsstatistiken äfven redogör för antalet d...
1093	döfstumma + andesvaga	Suomen aistivalliskoulujen lehti	1904-04	4	11	4.814899	antecknade såsom döfstumma och 26 såsom andesvaga. Af ...
1094	döfstumma + andesvaga	Suomen aistivalliskoulujen lehti	1905-01	1	14	4.814899	tillika döfstumma och 29 såsom andesvaga. Af hela antalet ...
1095	döfstumma + andesvaga	Suomen aistivalliskoulujen lehti	1905-01	1	15	4.814899	rundligt stipendium för studium af blinda och andesvaga dö...
1096	döfstumma + andesvaga	Suomen aistivalliskoulujen lehti	1905-09	6-7	31	4.814899	för döfstumma afsedd snickareyrkeskola i Jakobstad. Lyseik...
1097	döfstumma + andesvaga	Suomen aistivalliskoulujen lehti	1907-01	1	17	4.814899	lärarepersonal. Förordnad på prof. Utstående läraren vid dö...
1098	döfstumma + andesvaga	Suomen aistivalliskoulujen lehti	1911-02	2	14	4.814899	anstalter för döfstumma. För upprätthållande av ett skolhe...

Kuva 2 döfstumma-andesvaga

	Pääniemeke	PVM	Nro	Sivu	teksti
1574	Finlands Allmänna Tidning	1892-11-01	Register	6	Kustaa Killinens utn. t. föreståndare vid döfstumssko lan i Ku...
1711	Finlands Allmänna Tidning	1893-02-01	26	1	1893 utnämnt och förordnat t. f. föreståndaren för döfstums...
1729	Östra Finland	1893-02-02	31	2	1993 utnämnt och förordnat t. f. föreståndaren för döfstumsk...
6984	Tidskrift för döfstumma	1898-03	10	4	glädje mer. Jag tänkte: det blir väl döfstumt det också. Och l...
7604	Finlands Allmänna Tidning	1898-10-17	240	1	den 15 dennes föreståndaren för läroanstalten för döfstum...
7621	Finlands Allmänna Tidning	1898-10-21	244	3	tillförordnade läraren wid döfstumskolan i Kuopio Juho Kust...
7633	Finlands Allmänna Tidning	1898-10-24	246	3	tillförordnade läraren wid döfstumskolan i Kuopio Juho Kust...
7636	Finlands Allmänna Tidning	1898-10-25	247	3	tillförordnade läraren wid döfstumskolan i Kuopio Juho Kust...
7717	Finlands Allmänna Tidning	1898-11-16	266	1	filosofiedoktorn Albin Lönnbeck; för Kuopio: föreståndaren f...
7733	Wasa Nyheter	1898-11-19	269	4	filosofiedoktorn Albin Lönnbeck; för Kuopio: föreståndaren f...
7747	Österbottningen	1898-11-29	92	2	filosofiedoktorn Albin Lönnbeck; för Ruopio: föreståndaren f...
7821	Finlands Allmänna Tidning	1898-12-21	296	1	skolväsendet förordnat på prof tillförordnade läraren vid dö...
8946	Aftonposten	1900-02-28	49	2	Wessberg, döfstumskolelärarinorna Aino Pulkkinen, Emilia ...
8968	Wiborgsbladet	1900-03-01	49	2	Strunck och Sigrid Liljefors, döfstumsskoleföreståndaren Kus...
9330	Lördagen	1900-06-23	25	8	en insiktsfull och sympatisk ståndsmemlem, döfstumskolefö...
11141	Tidskrift för döfstumma	1902-10	54	8	Kiikka. Döfstumme Kustaa Ritola afled i slutet af september i...
11330	Hufvudstadsbladet	1903-01-03	3	5	stadsfullmäktiges i Kuopio ordförande har valts döfstumskol...

Kuva 3 Sanahaku Killisellä



Kuva 4 Esimerkki kollokaatioiden ajallisesta jatkumosta

collocation	count	count_nested	length	lambda	z	word1	word2	freq1	freq2	p_xy	p_x	p_y	mi_score	log_dice
1 wuosmaksun_joka_kuuroymyksi-laikissa	7	0	2	17.259784	8.489538	wuosmaksun	kuuroymyksi-laikissa	7	7	6.108576e-06	6.108576e-06	6.108576e-06	17.320733	14.000000
2 idiootmia_epilepsiaa_epilepsiaa_kuuroymykytta	6	0	2	17.116684	6.398358	idiootmia	kuuroymykytta	6	7	5.235922e-06	5.235922e-06	6.108576e-06	17.320733	13.884523
3 tavasi_samalla_samalla_kuuroymyksi-aakkosten	7	0	2	15.650344	9.632810	tavasi	kuuroymyksi-aakkosten	7	9	6.108576e-06	6.108576e-06	7.853883e-06	16.958162	13.807355
4 mielihyvaa_yli-ikisten_yli-ikisten_kuuroymyksi-aakkosten	10	0	2	16.497640	9.926886	mielihyvaa	kuuroymyksi-aakkosten	11	12	8.726537e-06	9.599190e-06	1.047184e-05	16.405621	13.798366
5 aikunsa_pleksimaan_pleksimaan_kuuroymyksi-aakkosten	12	0	2	16.671992	10.059686	aikunsa	kuuroymyksi-aakkosten	16	12	1.047184e-05	1.396246e-05	1.047184e-05	16.128087	13.777608
6 kuuroymyksi-aakkosten_maria_maria_rassan	6	0	2	15.507244	9.703696	kuuroymyksi-aakkosten	rassan	6	8	5.235922e-06	5.235922e-06	6.981229e-06	17.128087	13.777608
7 kuuroymyksi-aakkosten_olemassa_olemassa_kolmiwuotisia	5	0	2	16.949631	8.288532	kuuroymyksi-aakkosten	kolmiwuotisia	7	5	4.363268e-06	6.108576e-06	4.363268e-06	17.320733	13.736966
8 kuuroymyksi-aakkosten_turussa_turussa_innsbruckin	14	0	2	14.700143	15.769578	kuuroymyksi-aakkosten	innsbruckin	17	18	1.221715e-05	1.483511e-05	1.570777e-05	15.678035	13.678072
9 waltiovalia_kannatettu_kannatettu_kuuroymyksi-aakkosten	11	0	2	16.077765	10.195113	waltiovalia	kuuroymyksi-aakkosten	15	14	9.599190e-06	1.308980e-05	1.221715e-05	15.873274	13.601451
10 synnynnaisia_taipumuksia_taipumuksia_kuuroymykyteen	8	0	2	17.384946	8.567934	synnynnaisia	kuuroymykyteen	10	12	6.981229e-06	8.726537e-06	1.047184e-05	16.221197	13.540568
11 kuuroymyksi-aakkosten_kotini_kotini_wietäviksi	9	0	2	13.604345	16.537958	kuuroymyksi-aakkosten	wietäviksi	13	12	7.853883e-06	1.134450e-05	1.047184e-05	16.012610	13.526069
12 rikosasiaseen_syytteen_syytteen_kuuroymyksi-aakkosten	6	0	2	14.718784	9.630893	rikosasiaseen	kuuroymyksi-aakkosten	11	6	5.235922e-06	9.599190e-06	5.235922e-06	16.668636	13.497500
13 supistunut_turun_turun_kuuroymyksi-aakkosten	7	0	2	15.062558	12.437494	supistunut	kuuroymyksi-aakkosten	8	13	6.108576e-06	6.981229e-06	1.134450e-05	16.235003	13.415037
14 räpäässä_avioliittoon_avioliittoon_kuuroymyksi-aakkosten	10	0	2	16.497640	9.926886	räpäässä	kuuroymyksi-aakkosten	13	17	8.726537e-06	1.134450e-05	1.483511e-05	15.662113	13.415037
15 kuuroymyksi-aakkosten_opettajan_ja_painosiamiehen	5	0	2	15.851018	9.391828	kuuroymyksi-aakkosten	opettajan_ja_painosiamiehen	6	9	4.363268e-06	5.235922e-06	7.853883e-06	16.691128	13.415037
16 siirtymään_opettajaksi_opettajaksi_kuuroymyksi-aakkosten	5	0	2	16.949631	8.288532	siirtymään	kuuroymyksi-aakkosten	7	8	4.363268e-06	6.108576e-06	6.981229e-06	16.642661	13.415037
17 kuuroymyksi-aakkosten_e_e_ernberg	1881	0	2	9.704280	120.022303	kuuroymyksi-aakkosten	ernberg	3373	2510	1.641462e-03	2.943461e-03	2.190361e-03	7.992083	13.354946
18 kuuroymyksi-aakkosten_ajaja_kuuroymyksi-aakkosten	5	0	2	13.653790	13.195252	kuuroymyksi-aakkosten	kuuroymyksi-aakkosten	10	6	4.363268e-06	6.108576e-06	5.235922e-06	16.543125	13.321928
19 kuuroymyksi-aakkosten_syyt_syyt_paiseista	19	0	2	16.018036	10.623358	kuuroymyksi-aakkosten	paiseista	42	20	1.658042e-05	3.665145e-05	1.745307e-05	14.661769	13.293731
20 erinomaisia_todistuksia_todistuksia_kuuroymyksi-aakkosten	9	0	2	14.931216	9.934053	erinomaisia	kuuroymyksi-aakkosten	16	14	7.853883e-06	1.396246e-05	1.221715e-05	15.490638	13.263034
21 oikeusministeriä_olemaan_olemaan_kuuroymyksi-aakkosten	12	0	2	12.860244	18.768243	oikeusministeriä	kuuroymyksi-aakkosten	22	19	1.047184e-05	1.919838e-05	1.658042e-05	15.005691	13.227410
22 kuuroymyksi-aakkosten_waan_waan_ullottui	7	0	2	16.161171	9.659145	kuuroymyksi-aakkosten	ullottui	13	13	6.108576e-06	1.134450e-05	1.134450e-05	15.534563	13.106915
23 kuuroymyksi-aakkosten_kavattamasta_kavattamasta_kr...	6	0	2	15.170771	9.712964	kuuroymyksi-aakkosten	kr...	6	17	5.235922e-06	5.235922e-06	1.483511e-05	16.040625	13.061401
24 siitonen_10-wuotiaan_10-wuotiaan_kuuroymyksi-aakkosten	7	0	2	15.650344	9.632810	siitonen	kuuroymyksi-aakkosten	17	10	6.108576e-06	1.483511e-05	8.726537e-06	15.526051	13.052467
25 morsmannin_nimisen_nimisen_kuuroymyksi-aakkosten	7	0	2	14.861884	9.767515	morsmannin	kuuroymyksi-aakkosten	7	20	6.108576e-06	6.108576e-06	1.745307e-05	15.806159	13.052467
26 pohdittu_heisingin_heisingin_kuuroymyksi-aakkosten	5	0	2	15.851018	9.391828	pohdittu	kuuroymyksi-aakkosten	5	15	4.363268e-06	4.363268e-06	3.089800e-05	16.221197	13.000000
27 kuuroymyksi-aakkosten_ulla_ulla_asiainen	26	0	2	13.453078	15.616328	kuuroymyksi-aakkosten	asiainen	34	71	2.266899e-05	2.967022e-05	6.195841e-05	13.591317	12.986194
28 kuuroymyksi-aakkosten_rahastonhoitaja_rahastonhoi...	15	0	2	15.152493	10.257473	kuuroymyksi-aakkosten	julkaisi	19	43	1.308980e-05	1.658042e-05	3.752411e-05	14.368786	12.952694
29 kuuroymyksi-aakkosten_ajaja_ajaja_ugust_ugust_helälle	6	0	2	15.170771	9.712964	kuuroymyksi-aakkosten	helälle	14	11	5.235922e-06	1.221715e-05	9.599190e-06	15.446263	12.941106
30 kuuroymyksi-aakkosten_wällyksellä_wällyksellä_todistajat	6	0	2	16.018036	9.537754	kuuroymyksi-aakkosten	todistajat	6	19	5.235922e-06	5.235922e-06	1.658042e-05	15.880160	12.941106
31 kuuroymyksi-aakkosten_lujin_lujin_piirteitä	15	0	2	14.941187	14.816527	kuuroymyksi-aakkosten	piirteitä	35	30	1.308980e-05	3.054288e-05	2.617961e-05	13.998804	12.884523
32 kuuroymyksi-aakkosten_ovat_ovat_nousseet	19	0	2	15.817384	10.584746	kuuroymyksi-aakkosten	nousseet	54	29	1.658042e-05	4.712330e-05	2.530896e-05	13.763147	12.872888
33 kuuroymyksi-aakkosten_tule_tule_wähentämään	8	0	2	13.721378	14.166386	kuuroymyksi-aakkosten	wähentämään	22	14	6.981229e-06	1.919838e-05	1.221715e-05	14.861301	12.830075
34 kuuroymyksi-aakkosten_signe_signe_hoimistin	8	0	2	16.286333	9.760322	kuuroymyksi-aakkosten	hoimistin	28	8	6.981229e-06	2.443430e-05	6.981229e-06	15.320733	12.830075
35 waikuttamaan_suhteisiin_suhteisiin_kuuroymyksi-aakkosten	9	0	2	15.098271	9.983580	waikuttamaan	kuuroymyksi-aakkosten	20	22	7.853883e-06	1.745307e-05	1.919838e-05	14.516653	12.777608
36 kuuroymyksi-aakkosten_kristina_kristina_hakmanin	6	0	2	13.620161	9.152695	kuuroymyksi-aakkosten	hakmanin	6	24	5.235922e-06	5.235922e-06	2.094369e-05	15.543125	12.678072
37 limassa_krokotillimartta_krokotillimartta_kuuroymyksi-aakkosten	12	0	2	16.671992	10.059686	limassa	kuuroymyksi-aakkosten	26	35	1.047184e-05	2.268899e-05	3.054288e-05	13.883327	12.654225
223 223.550_mik_mik_kuuroymyksi-aakkosten	14	0	2	15.210966	10.249875	223.550	kuuroymyksi-aakkosten	21	51	1.221715e-05	1.832579e-05	4.450534e-05	13.870700	12.637430
39 kuuroymyksi-aakkosten_ääkkäillä_ääkkäillä_ajaksi_ajaksi	6	0	2	13.109344	15.264134	kuuroymyksi-aakkosten	ajaksi_ajaksi	14	17	5.235922e-06	1.221715e-05	1.483511e-05	14.818232	12.630766
40 käytin_äännetömästi_äännetömästi_kuuroymyksi-aakkosten	5	0	2	15.851018	9.391828	käytin	kuuroymyksi-aakkosten	20	6	4.363268e-06	1.745307e-05	5.235922e-06	15.543125	12.621488
41 kuuroymyksi-aakkosten_käytin_äännetömästi_äännetömästi_kuuroymyksi-aakkosten	5	0	2	16.018036	9.537754	kuuroymyksi-aakkosten	morsmannilla	20	12	5.235922e-06	1.745307e-05	1.047184e-05	14.806159	12.584963
42 kuuroymyksi-aakkosten_rinnakkaisosastoja_rinnakkaisosastoja	2	0	2	15.851018	9.391828	kuuroymyksi-aakkosten	rinnakkaisosastoja	20	8	4.363268e-06	1.745307e-05	6.981229e-06	15.128087	12.514573
43 kuuroymyksi-aakkosten_halutaan_halutaan_vakavaa	8	0	2	16.286333	9.760322	kuuroymyksi-aakkosten	wakavaa	30	15	6.981229e-06	2.617961e-05	1.308980e-05	14.314306	12.508147
44 syyksi_paietti_paietti_paietti_kuuroymyksi-aakkosten	7	0	2	15.313871	9.846065	syyksi_paietti	kuuroymyksi-aakkosten	10	30	6.108576e-06	8.726537e-06	2.617961e-05	14.706624	12.485427
45 määrärahat_kustannukset_kustannukset_kuuroymyksi-aakkosten	17	0	2	15.709172	10.498665	määrärahat	kuuroymyksi-aakkosten	44	54	1.483511e-05	3.839676e-05	4.712330e-05	13.001231	12.472753
46 viht_peltonen_peltonen_kuuroymyksi-aakkosten	8	0	2	14.340414	9.640114	viht	kuuroymyksi-aakkosten	8	42	6.981229e-06	6.981229e-06	3.665145e-05	14.735770	12.336144
47 kuuroymyksi-aakkosten_jonka_jonka_perustamiseen	11	0	2	15.489998	12.997587	kuuroymyksi-aakkosten	perustamiseen	12	58	9.599190e-06	1.047184e-05	5.061391e-05	14.144576	12.330149
48 kuuroymyksi-aakkosten_virginassa_virginassa_morsmannilla	5	0	2	15.851018	9.391828	kuuroymyksi-aakkosten	morsmannilla	20	12	4.363268e-06	1.745307e-05	1.047184e-05	14.543125	12.321928
49 _savolainen_savolainen_kuuroymyksi-aakkosten	1690	0	2	9.533932	121.037425	_savolainen	kuuroymyksi-aakkosten	7445	3373	1.474785e-03	6.496906e-03	2.943461e-03	6.269022	12.321661
50 viht_peltonen_peltonen_kuuroymyksi-aakkosten	8	0	2	14.340414	9.640114	viht	kuuroymyksi-aakkosten	11	42	6.981229e-06	9.599190e-06	3.665145e-05	14.276338	12.272080

Kuva 5 50 korkeimman log-dice-pisteytyksen saaneet suomenkieliset kollokaatioparit

collocation	count	count_nested	length	lambda	z	word1	word2	freq1	freq2	p_xy	p_x	p_y	mi_score	log_dice
döfstumma-blinda_och_och_döfstumma-andesvaga	12	0	2	14.483946	14.254981	döfstumma-blinda	döfstumma-andesvaga	17	16	1.376943e-05	1.950669e-05	1.835924e-05	15.230634	13.54057
kuusto_gård_gårds_döfstumskölarinnan	5	0	2	16.707501	8.170128	kuusto	döfstumskölarinnan	5	9	5.737261e-06	5.737261e-06	1.032707e-05	16.563210	13.51457
utger_en_en_döfstumtidsskrift	5	0	2	13.999447	12.529076	utger	döfstumtidsskrift	9	9	5.737261e-06	1.032707e-05	1.032707e-05	15.715213	13.15200
döfstumstanstämman_höstterminen_höstterminen_1891	15	0	2	16.134141	10.277320	döfstumstanstämman	1891	38	36	1.721178e-05	4.360318e-05	4.130828e-05	13.222173	12.69744
döfstumskölarinnan_selma_selma_ottilla	5	0	2	11.945315	15.992788	döfstumskölarinnan	ottilla	13	12	5.737261e-06	1.491688e-05	1.376943e-05	14.769660	12.67807
arg_andre_andre_döfstummen	5	0	2	16.707501	8.170128	arg	döfstummen	17	8	5.737261e-06	1.950669e-05	9.179618e-06	14.967600	12.67807
orsakerna_till_till_döfstumheten	13	0	2	12.574961	19.545025	orsakerna	döfstumheten	22	44	1.491688e-05	2.524395e-05	5.048790e-05	13.514711	12.63605
döfstumskölarinnan_selma_selma_ottilla	6	0	2	12.279423	16.440104	döfstumskölarinnan	ottilla	23	12	6.884713e-06	2.639140e-05	1.376943e-05	14.209573	12.45568
medtaga_någon_någon_döfstumskölarare	7	0	2	17.017654	8.370462	medtaga	döfstumskölarare	9	37	8.032165e-06	1.032707e-05	4.245573e-05	14.161111	12.28379
skoihen_för_döfstumma-blinda	12	0	2	11.476330	16.172982	skoihen	döfstumma-blinda	68	17	1.376943e-05	7.802675e-05	1.950669e-05	13.143171	12.17557
döfstumkongressen_i_köpenhamn	62	0	2	8.563544	42.155531	döfstumkongressen	köpenhamn	143	320	7.114204e-05	1.640857e-04	3.671847e-04	10.205531	12.09933
vårkstalles_i_döfstumskolans	72	0	2	9.326828	45.833027	vårkstalles	döfstumskolans	180	360	8.261656e-05	2.065414e-04	4.130828e-04	9.919353	12.09311
döfstumskolans_lokal_lokal_kaskisgatan	71	0	2	9.150238	46.106368	döfstumskolans	kaskisgatan	360	179	8.146911e-05	4.130828e-04	2.053939e-04	9.907213	12.07561
döfstumkongress_eller_eller_enskilt	7	0	2	17.017654	8.370462	döfstumkongress	enskilt	33	21	8.032165e-06	3.786592e-05	2.409650e-05	13.103778	12.05247
reseombud_för_för_döfstumföreningarna	9	0	2	13.362212	16.243610	reseombud	döfstumföreningarna	26	52	1.032707e-05	2.983376e-05	5.866751e-05	12.502180	11.88452
döfstumskoleföreståndaren_k_k_killinen	38	0	2	10.608803	35.686667	döfstumskoleföreståndaren	killinen	399	405	1.009758e-04	4.578334e-04	4.647181e-04	8.890543	11.80838
döfstumkongressen_i_paris	60	0	2	8.220555	34.78100	döfstumkongressen	paris	143	139	3.442357e-05	1.640857e-04	1.594959e-04	10.361213	11.76734
döfstumundervisning_besittit_besittit_föreläs	5	0	2	14.142545	9.253045	döfstumundervisning	föreläs	12	35	5.737261e-06	1.376943e-05	4.016033e-05	13.340817	11.76734
döfstumskolan_i_åbo	1528	0	2	4.756025	127.031449	döfstumskolan	åbo	9552	4856	1.753307e-03	1.096046e-02	5.572028e-03	4.843424	11.76285
döfstumstanstämman_ha_ha_stigit	5	0	2	12.112370	15.824584	döfstumstanstämman	stigit	38	10	5.737261e-06	4.360318e-05	1.147452e-05	13.485207	11.73697
döfstumskolan_i_borgå	1384	0	2	4.826548	121.037015	döfstumskolan	borgå	9552	3821	1.588074e-03	1.096046e-02	4.384415e-03	5.046441	11.72759
döfstumhemmet_i_pleksämäki	11	0	2	12.083757	13.666458	döfstumhemmet	pleksämäki	49	61	1.262197e-05	5.622516e-05	6.999458e-05	11.647119	11.67807
drag_ur_ur_döfstumundervisningens	11	0	2	13.513256	14.477156	drag	döfstumundervisningens	42	73	1.262197e-05	4.619299e-05	8.376401e-05	11.610424	11.61394
döfstumhem_och_och_economien	7	0	2	14.184431	9.454225	döfstumhem	economien	71	7	8.032165e-06	8.146911e-05	8.032165e-06	13.583387	11.52195
tredje_älmännä_älmännä_döfstummötet	27	0	2	11.142861	21.350545	tredje	döfstummötet	118	183	3.098121e-05	1.353994e-04	2.099838e-04	10.089679	11.52127
utgifterna_för_för_döfstumstanstämman	5	0	2	13.674278	9.149489	utgifterna	döfstumstanstämman	16	38	5.737261e-06	2.065414e-05	4.360318e-05	12.637210	11.51457
döfstumkongressen_i_paris	16	0	2	8.366866	26.237501	döfstumkongressen	paris	55	139	1.835924e-05	6.310987e-05	1.594959e-04	10.832834	11.40009
döfstumföreningens_lokal_lokal_ännegatan	27	0	2	8.674695	32.635443	döfstumföreningens	ännegatan	238	90	3.098121e-05	2.730936e-04	1.032707e-04	10.101351	11.39734
döfstumföreningens_lokal_lokal_ännegatan	14	0	2	9.099259	25.312950	döfstumföreningens	ännegatan	81	90	1.606433e-05	9.294363e-05	1.032707e-04	10.708786	11.38950
döfstumskolan_i_kuopio	944	0	2	5.026654	99.264250	döfstumskolan	kuopio	9552	2287	1.083195e-03	1.096046e-02	2.624223e-03	5.234950	11.35138
döfstumskolehuset_å_å_ryssbacken	5	0	2	13.411658	12.961250	döfstumskolehuset	ryssbacken	17	46	5.737261e-06	1.950669e-05	5.278280e-05	12.444038	11.34465
berättelser_om_om_döfstumskolornas	8	0	2	12.363679	17.128717	berättelser	döfstumskolornas	48	53	9.179618e-06	5.507771e-05	6.081497e-05	11.420252	11.34179
döfstumskolan_i_uleåborg	870	0	2	5.391263	91.938981	döfstumskolan	uleåborg	9552	1468	9.882834e-04	1.096046e-02	1.684460e-03	5.756783	11.33703
döfstumföreningens_lokal_lokal_richardsgatan	27	0	2	8.525474	32.822765	döfstumföreningens	richardsgatan	238	106	3.098121e-05	2.730936e-04	1.121629e-04	9.865284	11.32862
döfstumskolan_i_jämsböstad	275	0	2	5.054276	69.377165	döfstumskolan	jämsböstad	1719	1903	3.155494e-04	1.972470e-03	2.183602e-03	6.195009	11.28072
döfstumskolan_i_jämsböstad	668	0	2	4.937140	95.677248	döfstumskolan	jämsböstad	9552	1903	9.959885e-04	1.096046e-02	2.183602e-03	5.379039	11.27786
döfstumföreningar_jämsböstad_jämsböstad	10	0	2	14.309595	13.905587	döfstumföreningar	jämsböstad	86	50	1.147452e-05	9.866089e-05	5.737261e-05	10.964942	11.23447
ställa_det_det_döfstumhem	8	0	2	17.142816	8.448071	ställa	döfstumhem	42	71	9.179618e-06	4.619299e-05	8.146911e-05	11.191070	11.17982
döfstumheten_beträffar_beträffar_anses	6	0	2	15.775941	9.393580	döfstumheten	anses	44	44	6.884713e-06	5.048790e-05	5.048790e-05	11.399234	11.12553
döfstumundervisningens_historia_historia_marie	6	0	2	11.831108	15.043392	döfstumundervisningens	marie	73	16	6.884713e-06	8.376401e-05	1.835924e-05	12.128272	11.10923
döfstuminstitutet_i_fredericia	10	0	2	9.232268	22.308275	döfstuminstitutet	fredericia	82	70	1.147452e-05	9.409108e-05	8.032165e-05	10.568228	11.07400
döfstumhem_och_och_fastigheternas	5	0	2	13.662967	9.054965	döfstumhem	fastigheternas	71	5	5.737261e-06	8.146911e-05	5.737261e-06	13.583387	11.07400
bevista_internationella_internationella_döfstumkongressen	5	0	2	13.210974	8.822178	bevista	döfstumkongressen	21	55	5.737261e-06	2.409650e-05	6.310987e-05	11.881385	11.07400
döfstumhem_detta_detta_övertogs	5	0	2	15.608888	9.248364	döfstumhem	övertogs	71	6	5.737261e-06	8.146911e-05	6.884713e-06	13.320333	11.05514
döfstumbildningen_i_amerika	8	0	2	11.265007	12.585840	döfstumbildningen	amerika	16	108	9.179618e-06	2.065414e-05	1.239246e-04	11.808322	11.02272
banditer_ett_ett_döfstumt	5	0	2	12.629920	8.484840	banditer	döfstumt	5	77	5.737261e-06	5.737261e-06	8.835382e-05	13.466348	10.96438
döfstumskolan_i_jyväskylä	657	0	2	5.316165	80.359210	döfstumskolan	jyväskylä	9552	1282	7.538761e-04	1.096046e-02	1.4711034e-03	5.547116	10.95647
döfstumhem_i_pleksämäki	8	0	2	9.965741	19.496890	döfstumhem	pleksämäki	71	61	9.179618e-06	8.146911e-05	6.999458e-05	10.652650	10.95561
döfstumsaken_i_ryssland	5	0	2	10.618422	15.920229	döfstumsaken	ryssland	32	51	5.737261e-06	3.671847e-05	5.852006e-05	11.382637	10.94689
döfstumskolans_lokal_lokal_kaskisgatan	74	0	2	9.037208	46.616571	döfstumskolans	kaskisgatan	1084	179	8.491146e-05	1.243838e-03	2.053939e-04	8.376623	10.90681

Kuva 6 50 korkeimman log-dice-pisteytyksen saaneet ruotsinkieliset kollokaatioparit

	collocation_pair	Päänimiike	PVM	Nro	Sivu	log_dice	text_snippet
1	kuuromyökkä + kehittymätön	Betania : aikakauskirja uskonelämän ja -toiminnan edistämis...	1904-07	8	10	6.65677	2 sokenin, 1 tyysämielinen, 1 muuten sruutnton, 1 kuuromyk...
2	kuuromyökkä + kehittymätön	Forssan Sanomat	1905-09-05	67	3	6.65677	2 sokeata, 1 tyysämielinen, 1 muuten saamaton, 1 kuuromyk...
3	kuuromyökkä + kehittymätön	Inkeri	1905-09-02	35	2	6.65677	2 sokeata, 1 tyysämielinen, 1 muuten saamaton, 1 kuuromyk...
4	kuuromyökkä + kehittymätön	Itä-Karjala	1905-09-05	69	3	6.65677	2 sokeata, 1 tyysämielinen, 1 muuten saamaton, 1 kuuromyk...
5	kuuromyökkä + kehittymätön	Kajaanin Lehti	1905-06-16	65	3	6.65677	sokeata, 1 tyysämielinen, 1 saamaton, 1 kehittymätön kuuro...
6	kuuromyökkä + kehittymätön	Keski-Savo	1905-01-19	7	1	6.65677	saamaton, 1 kuuromyökkä ja kehittymätön, 1 halvattu mielen...
7	kuuromyökkä + kehittymätön	Koitar	1904-04-09	39	2	6.65677	2 sokeaa, 1 tyysämielinen, 1 muutoin saamaton, 1 kuuromyk...
8	kuuromyökkä + kehittymätön	Kokkola	1905-09-09	72	3	6.65677	2 sokeata, 1 tyysämielinen, 1 muuten saamaton, 1 kuuromyk...
9	kuuromyökkä + kehittymätön	Kyiväjä	1905-09-14	37	1	6.65677	2 sokeata, 1 tyysämielinen, 1 muuten saamaton, 1 kuuromyk...
10	kuuromyökkä + kehittymätön	Laatokka	1904-06-17	63	3	6.65677	lasta 2 sokeata, 1 tyysämielinen, 1 muuten saamaton, 1 kuur...
11	kuuromyökkä + kehittymätön	Laatokka	1905-09-05	100	2	6.65677	2 sokeata, 1 tyysämielinen, 1 muuten saamaton, 1 kuuromyk...
12	kuuromyökkä + kehittymätön	Louhi	1905-09-07	107	2	6.65677	lasta, 2 sokeaa, 1 tyysämielinen, 1 muuten saamaton, 1 kuur...
13	kuuromyökkä + kehittymätön	Mikkelin Sanomat	1905-09-05	100	1	6.65677	tyttöä, 106 poikaa, ja alemmassa 99; yhteensä 285. Kuuromy...
14	kuuromyökkä + kehittymätön	Otava	1904-06-13	77	3	6.65677	2 sokeata, 1 tyysämielinen, 1 muuten saamaton, 1 kuuromyk...
15	kuuromyökkä + kehittymätön	Pohjois-Karjala	1905-08-19	94	2	6.65677	sokeata, 1 tyysämielinen, 1 saamaton, 1 kehittymätön kuuro...
16	kuuromyökkä + kehittymätön	Pohjois-Karjala	1905-09-05	101	3	6.65677	2 sokeata, 1 tyysämielinen, 1 muuten saamaton, 1 kuuromyk...
17	kuuromyökkä + kehittymätön	Rauman Lehti	1904-03-24	35	2	6.65677	2 sokeata, 1 tyysämielinen, 1 muuten saamaton, 1 kuuromyk...
18	kuuromyökkä + kehittymätön	Rauman Lehti	1905-09-07	101	3	6.65677	2 sokeato, 1 tyysämielinen, 1 muuten saamaton, 1 kuuromyk...
19	kuuromyökkä + kehittymätön	Suomalainen	1905-09-06	103	2	6.65677	2 sokeata, 1 tyysämielinen, 1 muuten saamaton, 1 kuuromyk...
20	kuuromyökkä + kehittymätön	Suomen aistiviälliskoulujen lehti	1916-05	5	3	6.65677	että hänen rakkautella tekemänsä työn kautta monen kuuro...
21	kuuromyökkä + kehittymätön	Suomen aistiviälliskoulujen lehti	1921-05	5-6	11	6.65677	tarpeen tullessa kääntyä tällaiset kuuromyökät itse tahi heidä...
22	kuuromyökkä + kehittymätön	Suupohjan Kaiu	1905-09-09	102	3	6.65677	2 sokeata, 1 tyysämielinen, 1 muuten saamaton, 1 kuuromyk...
23	kuuromyökkä + kehittymätön	Toivo	1904	1905	107	6.65677	2 sokeata, 1 tyysämielinen, 1 muuten saamaton, 1 kuuromyk...
24	kuuromyökkä + kehittymätön	Uusi Aura	1904-03-18	65	3	6.65677	2 sokeata, 1 tyysämielinen, 1 muuten saamaton, 1 kuuromyk...
25	kuuromyökkä + kehittymätön	Uusi Suometar	1904-08-10	183	3	6.65677	lasta, 2 sokeata, tyysämielinen, 1 muuten saamaton, 1 kuuro...
26	kuuromyökkä + kehittymätön	Uusi Suometar	1905-06-10	183	2	6.65677	sokeata, 1 tyysämielistä, 1 muuten saamakonta, 1 kuuromyk...
27	kuuromyökkä + kehittymätön	Uusimaa	1905-09-01	100	2	6.65677	2 sokeata, 1 tyysämielinen, 1 muuten saamaton, 1 kuuromyk...
28	kuuromyökkä + kehittymätön	Wiipuri	1904-08-13	186	2	6.65677	lasta, 2 sokeata 1 tyysämielinen, 1 muuteu saamaton, 1 kuur...

Kuva 7, Keskitetyn kollokaatio-analyysin tulosten listaus