

Katsaus

Suullisen runon sähköisestä lukemisesta

Kati Kallio ja Eetu Mäkelä

Aineiston tuottamisen, käsittelyn ja lukemisen käytännöt vaikuttavat monin tavoin siihen, minkälainen käsitys ja minkälaisia tutkimustuloksia tietystä ilmiöstä on saatavissa. Kalevalamittaisen runouden kohdalla se, miten aineistoa on tallennettu, tekstualisoitu, valikoitu, jaoteltu, arkistoitu, kortistoitettu, luetteloidu, julkaistu ja analysoitu luo pohjan sille, mitä nykytutkija aineistosta ja itse ilmiöstä voi ymmärtää ja tavoittaa. Vaikka kalevalamittaisen runon keskeisimmät aineistot ovat olleet viisitoista vuotta käytössä sähköisinä, ei tutkimuksessa ole juurikaan pohdittu, mitä tämä käyttömuodon muutos tarkoittaa tutkimuksen käytäntöjen ja mahdollisuuksien kannalta (ks. kuitenkin Harvilahti 2013; Ilyefalvi 2018; Tangherlini 2016).

Kalevalamittaisen runon laajojen ja varioivien aineistojen kohdalla sähköiset haut nopeuttavat aineiston läpikäymistä huomattavasti, jos päämääränä on vaikkapa tarkastella, minkälaisissa runoissa tietyn tyyppinen säe tai ilmaisu esiintyy tai mitä ovat runotyyppien suhteet toisiinsa (ks. Kallio et al. 2017, 40–55). On mahdollista vaikkapa katsoa huomattavasti manuaalista selaamista nopeammin, miten Ritvalan kylän helkajuhlien Annikaisen virsi asettuu suhteessa Annikki- ja Anni-neitoihin muilla runoalueilla, tai miten runoaineistoa on 1900-luvun alkupuoliskon tulkinnoissa jaoteltu hypoteettisten historiaan projisoitujen ”alkumuotojen” ja joskus pelkkien yksittäisten säevastaavuuksien perusteella (Kallio et al. 2017, 413–447). Toisaalta käyttöliittymän ominaisuudet, aineiston ennakkotuntemus ja jopa tutkijan mielikuvitus sekä puhdas sattuma vaikuttavat siihen, minkälaisen aineiston sähköisellä käyttöliittymällä saa esiin luettavaksi (Kallio 2019).

Tämä katsausartikkeli tarkastelee historiallisen suullisen runouden lukemista sähköisillä välineillä. Yli 89 000 karjalaista, inkeriläistä ja suomalaista tekstiä sisältävä *Suomen Kansan Vanhat Runot* -kirjasarja (SKVR) on ollut käytettävissä sähköisenä *SKVR-tietokantana* vuodesta 2004. Aineiston sähköisen käytön tekee haastavaksi sen kielellinen, ortografinen ja temaattinen monimuotoisuus, variaation monitasoisuus sekä käytön edellyttämien taustatietojen runsaus. Tekstit edustavat neljän eri sukukielen murteita ja ne on pantu muistiin moninaisin eri menetelmin ja kirjoitusasuin lähes neljänsadan vuoden aikana. Nähdäksemme SKVR-korpuksen kaltaista aineistoa käytettäessä keskeistä on riittävä tieto aineiston luonteesta sekä mahdollisuus liikkua mahdollisimman helposti määrällisten tulosten, metatietojen ja varsinaisten tekstien yksityiskohtien välillä.



Analysoimme SKVR-tietokannan ja kehitteillä olevan Octavo-käyttöliittymän mahdollisuuksia suhteessa suullisen runon rakenteita koskevaan teoreettiseen keskusteluun. Esittelemme ensin kalevalamittaisen runon aineiston, SKVR-tietokannan ja Octavon, ja käsittelemme sitten tarkemmin sanamuotoihin, formuloihin eli kiteytyneisiin ilmauksiin sekä runoteemoihin liittyviä variaation kysymyksiä suhteessa käyttöliittymien hakumahdollisuuksiin. Mitä SKVR-korpuksen sähköinen lukeminen tämänhetkisillä välineillä edellyttää ja mitä metodologisia mahdollisuuksia se antaa? Tarkoituksemme on myös kirjoittaa auki aineiston sähköisen käytön perusedellytyksiä.

Kalevalamittaisen runouden aineisto

Suurin osa itämerensuomalaisista kieliryhmistä – inkeroiset, karjalaiset, vatjalaiset, virolaiset ja suomalaiset – on käyttänyt runomitaltaan, poetikaltaan ja monilta teemoiltaan samankaltaista suullista runoutta. Sitä on kutsuttu monilla nimillä: vanha runo, runolaulu, *regilaul*, kalevalamittainen tai kalevalainen runous, tetrametrinen trokee (ks. tarkemmin Kallio, Frog ja Sarv 2017). Runokieli oli joustava väline, jonka pohjalta rakennettiin myös uusia runoja (Hautala 1945). Samaa runomuotoa käytettiin laajassa kirjossa runoja ja lajeja, jotka risteytyivät ja viittasivat toisiinsa monin tavoin (Tarkka 2004; 2011; Timonen 2004).

Kalevalamittaisen runouden aineisto on laaja ja poikkeuksellisen perusteellisesti editoitu ja järjestetty. Tämä johtuu siitä, että tallennusaikanaan aineistoa pidettiin kansallisesti keskeisenä: sen varaan rakennettiin suomalaista identiteettiä ja kansakuntaa (Anttonen 2005; Sihvo 2003; Wilson 1976). Sekä tallennus- ja kuratointihistoria että runokulttuurien alueelliset erot selittävät monia aineiston painotuksia ja vääristymiä. Aineisto ei ole kattava kuvaus siitä, mitä 1800-luvun rahvas lauloi ja uskoi, vaan tietyin intressein tallentunut valikoima (ks. esim. Gröndahl 1997). Tallennusperiaatteet vaihtelivat aikakausittain ja tallentajittain. Joidenkin alueiden, lajien ja aikakausien suhteen kokoelman voi olettaa olevan hyvinkin kattava, toisten kohdalla on huomattavia aukkoja.

Kalevalamittaisten runojen tallennus- ja toimitusperiaatteet olivat sidoksissa runojen kaukaisia alkumuotoja ja leviämisreittejä tavoittelevaan, suuria aineistoja yksityiskohtaisesti vertailevaan maantieteellis-historialliseen menetelmään. Sen reunaehdot alkoivat 1900-luvun puolivälin jälkeen romahtaa, kun ymmärrys suullisen perinteen variaation monitasoisuudesta ja ennakoimattomuudesta lisääntyi (esim. Apo 1986; Honko 1986; 1998; Bendix ja Hasan-Rokem 2012). 1900-luvun loppupuoliskon teoreettiset ja metodologiset käänneet mullistivat myös suullisen runouden tutkimuksen. Kalevalamittaisen runouden tutkimuksessa huomio alkoi kiinnittyä tallennusaikaisiin paikalliskulttuureihin, tallennusprosesseihin, marginaaliin jääneisiin lajeihin ja ilmiöihin sekä yksittäisiin laulajiin ja esityksiin, mikä on tuottanut syvempiä käsityksiä myös paikallisista ilmaisukulttuureista ja runokielen käytöstä.¹ Etenkin Anna-Leena Siikala (esim. 1992; 2012), Mari Sarv (2008; 2011), Senni Timonen (2004), Frog (esim. 2013; 2016) ja Joonas Ahola (2014) ovat jatkaneet kalevalamittaisten runojen laajojen vertailevien tai syntetisoivien tutkimusasetelmien parissa uusista teoreettisista ja metodologisista lähtökohdista käsin.

1 Esim. Arukask 2003; DuBois 2006; Huttu-Hiltunen 2008; Oras 2008; Saarinen 2018; Siikala et al. 2004; Siikala ja Vakimo 1994; Särg 2008; Seppä 2015; Timonen 2004; Tarkka 2005; Västrik 2008; ks. myös Stepanova 2014; Sykäri 2011.



SKVR-tietokanta ja Octavo-järjestelmä

Suurin ja vanhin osa tallennetusta karjalaisesta, inkeriläisestä ja suomalaisesta kalevalamittaisesta runoudesta eli 89 247 tekstiä sisältävä *Suomen Kansan Vanhat Runot* -kirjasarja digitoitiin 1990-luvulla ja muokattiin vapaasti käytössä olevaksi SKVR-tietokannaksi, jonka tekstit vastaavat mahdollisimman tarkkaan alkuperäisen kirjasarjan tekstejä. Suuria aineistokokonaisuuksia on muitakin. Virolaisten runojen yhä täydentyvä tietokanta *Eesti regilaulude andmebaas* (ERA) sisältää tällä hetkellä 92 134 runoa vanhimmista 1800-luvun kokoelmista alkaen, mikä on noin kaksi kolmannesta virolaisten arkistojen aineistosta. Petroskoin Kielen ja Kirjallisuuden Instituutin aineistoja on julkaistu antologioissa (esim. Evseev 1950; 1976; Mironova 2006). Tämän lisäksi eri arkistoissa on julkaisemattomia aineistoja – Suomalaisen Kirjallisuuden Seuran (SKS) arkistossa on noin 60 000 digitoitua runotekstiä – sekä ääniteaineistoja.

Sähköisesti käytössä olevalle suomalaiselle ja virolaiselle aineistolle on ominaista toisaalta samanlainen yksikkökohtainen perusrakenne eli toisiaan vastaavat metatiedot ja itse runoteksti, toisaalta suuri kielellinen, murteellinen, keruuhistoriallinen ja ortografinen variaatio. Kaikki runot on jaoteltu tallennuspaikkakunnan (tai joskus informantin koti- tai synnyinpaikkakunnan), tallentajan ja tallennusajankohdan mukaan, milloin nämä vain ovat olleet tiedossa tai pääteltävissä, ja alkuperäisen käsikirjoituksen arkistoviite kulkee viitetiedoissa mukana. Mahdolliset konteksti- tai lisätiedot on typografisesti eroteltu itse runotekstistä. Paikoin käsikirjoituksessa voi olla enemmän taustatietoa kuin julkaisuihin on aikanaan otettu.

Tietokantamuoto nopeuttaa erilaisten osa-aineistojen kokoamista ja selaamista. Sanahauilla voi etsiä osumia laajasta aineistosta, vaikka toisaalta kielellinen, poeettinen ja ortografinen kirjo on aineistossa niin suurta, että hakeminen edellyttää moninaisia taustatietoja ja usein sanakirjojenkin käyttöä. SKVR-tietokannan tämänhetkiset sanahakumahdollisuudet kattavat sanan alkuosalla sekä yhdellä tai useammalla kokonaisella sanalla hakemisen, tiettyjen sanojen pois sulkemisen sekä fraasihaun (ks. SKVR-tietokanta, Ohjeita, <https://skvr.fi/ohje>). Osumat annetaan metatietoriveinä, joista saa yksitellen auki itse runotekstit. Tietokanta ei pysty näyttämään sanahaun osumakohtia, vaan ne täytyy erikseen itse etsiä osumista selaimen haulla tai lukemalla koko teksti.

Keväällä 2018 Mäkelä sijoitti SKVR-datan kehitteillä olevaan Octavo-järjestelmäänsä. Octavo on yhdistettyihin teksti- ja metadata-aineistoihin keskittyvän humanistisen tutkimuksen tueksi kehitettävä järjestelmä. Järjestelmän ideana on tarjota käyttäjälleen laajat mahdollisuudet rajata haluamansa osa aineistoa sekä tekstin sisäisten että siihen liittyvien metadata-piirteiden avulla. Rajattua aineistoa on tämän jälkeen mahdollista tarkastella sekä lähiluvun että erilaisien tilastollisten menetelmien kautta. Lisäksi järjestelmä on suunniteltu tukemaan iteratiivisia työkulkuja, joissa tutkija voi saamiensa tulosten perusteella helposti muokata rajaus- ja vertailuehtojaan. Osa näytöistä on myös erityisesti suunniteltu tukemaan variaation ja vaihtoehtoisien rajausten löytämistä. Näin tutkija voi aloittaa selkeimmistä tapauksista ja ehdoista, mutta vähitellen varmistaa löytävänsä ja huomioivansa myös ne tutkimansa ilmiön vaihtelut, jotka eivät välittömästi tulleet alussa mieleen. Yleisimpiin käyttötapauksiin Octavo tarjoaa graafisen käyttöliittymän, mutta ohjelmointia osaaville myös kattavat rajapinnat. Näin järjestelmä pystyy tarjoamaan tärkeimmät työkulkunsa käyttöön kaikille, muttei myöskään rajoita osaavaa käyttäjää luomasta omia työkulkuja.



Jo pelkästään SKVR-aineiston hakukoneena käytettynä järjestelmässä on nähdäksemme useita etuja aikaisempaan verrattuna. Hakumahdollisuudet ovat laajemmat ja ne antavat mahdollisuuksia etsiä myös vaihtelevia sanamuotoja sekä useammasta varioivasta sanasta koostuvia formuloita. Rajauksia on mahdollista tehdä monipuolisemmin – esimerkiksi poistaa hakutulosista kaikki tietyn runotyypin tai merkkijonon sisältävät tai kaikki tietyn kerääjän tallentamat tekstit. Osumat annetaan itse valittavien metatietojen rivinä sekä haluttuna määränä itse tekstiä. Käyttäjä voi itse määrittellä, minkä määrän tekstiä hän halua nähdä: ainoastaan tietyn määrän merkkejä, sanoja, rivejä tai koko tekstin. Sanahakuja tehdessä osumakohdat näytetään lihavoituna, eli on mahdollista hahmottaa suurikin joukko osumia nopealla vilkaisulla. Verrattuna SKVR-tietokantaan nämä mahdollistavat huomattavasti nopeampia ja usein myös tarkempia hakuprosesseja.

Sanamuotojen variaatio

Merkittävin SKVR-aineiston tekstihakuja vaikeuttava tekijä on aineiston rikas ja monitasoinen kielellinen ja ortografinen variaatio. Tämä on sidoksissa aineiston luonteeseen ja tallennushistoriaan. SKVR-aineistossa kirjoitusasuihin liittyvällä variaatiolla onkin monia toisiinsa vaikuttavia syitä:

1) Esittäjät käyttivät erilaisia sanamuotoja, joskus yhdenkin runon sisällä. Tämä liittyi sekä sanamuotojen variaatioon ja erilaisiin rinnakkaisiin sanamuotoihin että esitystapoihin. Sanellessa tietty sana saattoi esiintyä joko täyspitkänä tai lyhyempänä ja puhekielisempänä aiheutuen tavallista lyhyempiä ja mitaltaan poikkeavia säkeitä. Lauletaessa sanat esiintyivät usein pidemmissä, arkaaisemmissa, runokielelle ominaisissa tai tiettyyn sävelmään tai esitystapaan liittyvissä muodoissa. Yleensä ei ole tiedossa, onko runoesitys kirjoitettu muistiin sanellusta vai laulettua esityksestä: ilmeisesti aika monet kerääjät kokivat helpommaksi kirjoittaa muistiin saneltua esitystä. (Lauerma 2004; Palola 2009; Timonen 2004; Kallio 2013.)

2) Esittäjät puhuivat eri kieliä ja niiden murteita. Runokielten suhdetta paikallisiin puhekieliin ei ole perinpohjaisesti tutkittu, mutta on selvää, että runojen mukana saattoi kulkea myös sellaisia kielen piirteitä, joita esittäjän puhekieleen ei kuulunut – joko arkaaisia, runokieleen liittyviä tai muille alueille ominaisia muotoja. Toisaalta esittäjän puhekieli vaikutti myös runokielen saamiin muotoihin. SKVR-korpuksessa on runoja ainakin karjalan, inkeröisen, vatjan ja suomen kielillä ja murteilla. Esittäjien joukossa oli myös joitain romaneja ja venäjää äidinkielenään puhuvia (Blomster ja Mikkola 2017; Kallio 2013, 45–46) sekä mitä luultavimmin myös suomenruotsalaisia ja saamelaisia. Sanojen variaation kohdalla aineiston kirjon ja kielimuotojen – vaikkapa tyypillisten päätteiden, diftongien ja konsonanttityhtymien – ennakkotuntemus auttaa paljon.

3) Tallennuksen menetelmät ja tavoitteet ovat olleet moninaisia. Suurin osa aineistosta on tallennettu arkistointia, tutkimusta tai julkaisuja varten, mutta joukossa on omaa käyttöä varten ylös pantuja runoja, tuomiopöytäkirjoihin tallentuneita loitsuja ja tuntemattomia tarkoituksia varten luotuja käsikirjoituksia. Kunkin tallennuksen tavoitteista, käytettävistä olevista menetelmistä ja tallentajan taidoista riippuen runoja on tallennettu muistiin sanamuotoja lyhentäen tai kirjakielistäen, pikakirjoituksella, erilaisilla tarkekirjoituksilla ja ortografioilla, sanelusta, laulusta, vahaliერიolle. Siinä missä Elias Lönnrot saattoi merkitä itselleen jo ennalta tutut runosäkeet muistiin murremuodot hävittävinä lyhenteinä (vuv = Vaka vanha Väinämöinen), kirjoitti tutkimukseen suuntautunut A. A. Borenius runot foneettisella tarkekirjoituksella, joskus



vielä analysoiden esityskertojen eroja marginaalimerkinnöissä. Toisaalta tarkkakaan kirjoitusasu ei aina kerro alkuperäisestä esityksestä: esimerkiksi Volmari Porkka käytti muistiinpanoihinsa epätarkkaa pikakirjoitusta, mutta kirjoitti runot puhtaaksi tuntemiensa murremuotojen mukaisesti (Bono 2003). Paikoin tallenteissa on myös väärinymmärryksiä.

4) SKVR-kirjasarjan toimittajat tulkitsivat paikoin epäselvätkin käsikirjoitukset ja kirjoittivat esimerkiksi lyhennetyt sanat ja säkeet auki hakasulkeisiin, yleensä melko yleiskielisiin muotoihin, mutta kunkin kerääjän kirjoituskäytäntöjä mukaillen. Lisämerkinnät rivien väleissä tai marginaaleissa sijoitettiin alaviitteiksi ja metatietoja täydennettiin. Lisäksi SKVR-aineiston digitointiprosessi on tarkistuksista huolimatta tuottanut aineistoon joitain merkkivirheitä.

SKVR-tietokannassa osa kirjoitusasuun liittyvästä variaatiosta on ohitettu tekstien normaalistetuilla versioilla (SKVR-tietokanta, Ohjeita, <https://skvr.fi/ohje>). Tekstihaut kohdistuvat tekstiversioon, josta diakriittiset erikoismerkit on normaalistettu tavallisille aakkosille ja sulkeiden, hakasulkeiden ja välimerkkien kaltaiset merkit poistettu. Murremuotoihin liittyvää variaatiota normaalistus ei ohita. Hakutuloksissa runo näytetään alkuperäismuodossaan, mutta normalisoidun tekstin saa halutessaan näkymään alkuperäisen tekstin jälkeen. Normaalistuksesta johdun selaimen haku ei välttämättä löydä kaikkia hakutermejä alkuperäisestä tekstistä.

Octavossa SKVR-tietokannan normaalistussäännöt eivät tämän katsausartikkelin esimerkkihakuja tehdessä olleet vielä käytössä, mutta liittymän hakuominaisuudet antavat muitakin keinoja löytää erilaisia varioivia sanamuotoja ja kirjoitusasuja. Nämä pätevät myös nyt jo käytössä olevien normaalistettujen tekstien kohdalla. Käyttäjän kannalta on hyödyllistä, että haut saa kohdistettua tarpeen mukaan joko alkuperäiseen tai normaalistettuun tekstiin.

Itämerensuomalaisten kielten sanat varioivat eri tavoin eri kohdista. Esimerkiksi sananlopuille tyypillinen morfologinen vaihtelu on helppoa ohittaa hakemalla ainoastaan sanojen alkutavuuilla. Octavo tarjoaa sanamuotojen variaation kartoittamiseen muutaman muunkin apuvälineen. Yksinkertaisin näistä on hakea merkkijonoa, joissa yksi tai kaksi merkkiä saa olla jotain muuta kuin annettu tai puuttua kokonaan. Tekstihaku väinämöinen~2 etsii kaikki Väinämöistä muistuttavat sanat, joissa voi kaksi merkkiä olla toisin tai puuttua. Variaatiolle voi rakentaa myös tarkempia hakuja, joissa merkkien pysyvyys tai vaihtuvuus on määritelty tarkemmin. Esimerkiksi haku /väi?n.* / etsii kaikki vä-alkuiset sanat, joissa kolmas kirjain voi vaihdella, neljäs on n ja loppu voi jatkua miten vain (hakusyntaksista ks. Mäkelä 2019). Näillä kahdella haulla löytyvät jo esimerkiksi sanamuodot *Väinö*, *Väinämö*, *Väilämöinen*, *Viänämöinen*, *Vainämöinen*, *Wäinämöisen*, *Väinämöisen*, *Väinämyösen*, *Väinämöinji*, mutta myös naispuolinen *Väinätär*, paikkaan viittaava *Väinälä* eri muotoineen sekä aivan eri hahmo *Väntäriikki*.

Sanamuotoihin liittyvä variaatio ja rinnakkaisuus ovat kuitenkin niin moninaisia, että kaikkia variaation piirteitä on mahdotonta ennakoita ja etsiä tietoisesti. Mikael Agricolan jumal-luettelossaan käyttämää muotoa *Äinemöinen* ei esiinny varsinaisissa runoteksteissä, mutta Christfried Ganander ja Eric Lencqvist mainitsevat sen suorasanaaisessa tekstissään *Väinämöinen*-sanana kanssa samassa lauseessa – tätä muotoa eivät yllä mainitut haut tunnista. Ilmeisesti Joukahaisen sijalla inkeriläisessä runossa esiintyvä, *Äinemöistä* muistuttava *Einemoin* tai runon minän epiteettinä savolaisessa runossa ja Christfried Gananderilla toimiva *Vangamöinen* eivät ole suoraa sukua *Väinämöiselle*, sillä ne eivät esiinny samoissa rooleissa samankaltaisissa runoteemoissa. Tilannetta monimutkaistaa se, että henkilöhaamot ja keskeiskäsitteet voivat



samaakin tarinaa laulettaessa vaihtua, tai joku voi esittää pätkän yleisesti tunnetusta juonesta mainitsematta tiettyä siihen usein kuuluvaa keskeistäkään jaksoa tai henkilöahmoa.

Mitä joustavampi tai laajempi haku on, sitä enemmän tulee hakutuloksiin myös osumia, jotka eivät liity haettavaan aiheeseen mitenkään. Näistä toistuvimpia ja selvimpiä on mahdollista rajata tulosjoukon ulkopuolelle tiettyjen sanojen tai runotyyppien perusteella. Väinämöisen kohdalla pois on tarpeen sulkea esimerkiksi paikannimiä rajauksella -Väinölä* ja verbejä rajauksella -vänkyt* -vänäht*. Hakutuloksia voi myös rajata tietyillä runoteemoilla, jotka tuottavat turhia osumia eivätkä sisällä Väinämöistä, kuten -theme:"Marketan runo".

Kiinteät sanayhdistelmät eli formulat

Suullisen runon tutkimuksessa poeettisesti ja metrisesti motivoituja toistuvia kiteymiä kutsutaan *formuloiksi*. Näille laulun perusrakenteille on ominaista luonteva, helppo asettuminen runomittaan. Usein niiden kantamat laajemmat, assosiativiset merkitykset aukeavat vasta niiden käyttöyhteyksien kirjoa tarkasteltaessa (Foley 1991). Nämä suulliselle runolle tyypilliset muutaman sanan kiinteähköt toistuvuudet toimivat hyvin tekstihakujen lähtökohtana.

Lauri Harvilahti (erit. 1992) näytti, että – toisin kuin aiemmin oli oletettu – monet Albert Lordin (2002 [1960]) ja hänen työtään jatkaneiden tutkijoiden (ks. Foley 1988) näkemykset formuloista istuvat myös itämerensuomalaiseen suulliseen runouteen. Samalla hän (Harvilahti 1992, 141–143) totesi, että formulaa käsitettä on kansainvälisessä tutkimuksessa käytetty ja sovellettu hyvin monin, usein melko jäykin tavoin. Variaatio ja toistuvuus on niin monitasoista ja hienovireistä, että mikä hyvänsä yksittäinen määritelmä jättää siitä suuren osan ulkopuolelle. Kalevalamittaiseen runoon istuvia, eri tutkijoiden kehittämiä formulaa määritelmiä ovat Harvilahden (mts., 142) mukaan esimerkiksi 1) sanastolliseen toistoon perustuvat, 2) yhden sanan korvautumiselle rakentuvat (formulasysteemiksikin kutsuttu), 3) samankaltaiselle lauseopille perustuvat (formulatyyppi tai formulamainen ilmaus), 4) toistuville lauseopillisille, runomitallisille ja/tai säeparallelismille rakentuvat (rakenneformula) sekä 5) väljän mitallisen tai säeparallelismiin pohjautuvan muotin, äänteellis-rytmisen samankaltaisuuden tai morfologisen piirteen pohjalta tuotetut formulat. Osuvin formulaa määritelmä riippuukin paljon tutkimusasetelmasta eli siitä, minkälaista samankaltaisuutta ja millä tarkkuudella halutaan löytää. Sähköisille menetelmille variaation ja toistuvuuden monitasoisuus aiheuttaa haasteita.

Esimerkiksi puolikkaan säkeen pituinen formula *oro juoksi* esiintyy kalevalamittaisessa runoudessa yleisesti osana pidempää matkantekoa kuvaavaa formulaista säettä "oro juoksi, matka joutui", ja voi jatkua vaikkapa toisella säkeellä "reki vierä, tie lyheni". Formula kuuluu niiden joukkoon, joita on käytetty matkantekoa kuvaavina hyvin monissa erilaisissa temaattisissa yhteyksissä.

Octavon kollokaatiohaku "or* juok*"~1 etsii kaikki tekstit, joissa or- ja juok-alkuiset sanat esiintyvät peräkkäin tai enintään yhden sanan erottamina. Tuloksena on laaja kirjo formulaperheeseen *ori juoksi, matka joutui* kuuluvia osumia – ja samalla muita osumia, kuten "kuukauden orava juoksi", jotka voi olla tarpeen seuloa pois. Esimerkkiformulassa suhteellisen kiinteänkin osan muoto vaihtelee. *Ori* voi esiintyä muodoissa *oroi, ori, orit, orih, Oršo, oroba* tai *Orsho* ja verbi *juosta* ainakin muodoissa *juoks, juoksi ja juoksoo*. Yleensä puolen säkeen pituinen formula esiintyy myös pidempänä koko säkeen versiona "lähti oro juoksemahan". Toisaalta esiintyy myös orin korvaavia muotoja, kuten "virkku juoksi, matka joutui", "laivoi juoksi, matka joutui",



”reki juoksi, matka joutui” tai ”hebo juoksi, matka joutui”. Näitä puolestaan löytyy esimerkiksi *ori juoksi, matka joutui* -formulan jälkiosalla hakemalla (esim. ”mat* jou*~1”).

Yksi ja sama asia voidaan usein sanoa vaihtelevin tavoin suurenkin formulaperheen eri osia käyttäen. Lauri Harvilahti (2004) puhuu pidempien toistuvien formulayhdistelmien kohdalla *vakiojaksosta*. Frog (2016, 62–63) on puolestaan ehdottanut *multiformin* käsitettä muutaman sanan vakiintunutta yhdistelmää kuvaavan formulan ja *Oral Formulaic* -teoriassa yleensä sisällöllisesti motivoituneeseen pienehköön kokonaisuuteen viittaavan teeman käsitteen välille. Esimerkiksi livana Shemeikka lauloi A. A. Boreniukselle 172-säkeisen Lemminkäisen virren vuonna 1897. Laulussa toistuu samankaltainen hevosen juoksemista kuvaava jakso viisi kertaa. Vaikka jakso on sangen kiinteä, monet yksittäiset sanamuodot (Helkkäin(ba)–Helkkäi(m) ba; oro–oroba) vaihtelevat, säkeet voivat esiintyä eri järjestyksessä ja kaikki formularyppään säkeet eivät tule joka kerta lauletuiksi:

Iškinbä virkkuista vitšalla,
Helkkäin(ba) helmi ruoškazella,
Lähtiba virkku vieremähä,
Oro helhetyttämähe.

Virkkuba vieri, tie lyheni,
Oroba juoksi, matka joudu.

Jo hebo hörösteleksi,
Luppabo korva luondeleksi.
([SKVR VII1 803](#), säkeet 77–84.)

Iški(m)bä virkkuist on vitšalla,
Helkkäi(m)ba helmiruoškazella,
Lähtibä virkku vierömäh[e],
Oroba helhetyttämäh[e],
Oroba juoksi, matka joudu,
Virkku vieri, tie lyheni.

Ajoimba matkuvo vähäsen,
Sembom tiedä pikkaraizen,
Jo hebo hörösteleksi,
Luppaba korva luondelekse.
([SKVR VII1 803](#), säkeet 94–103.)

Variaatio voi olla esityskerta-, runo-, laulaja-, kylä- tai aluekohtaista. Esimerkiksi anonyymin, Shemeikan tavoin Laatokan Karjalasta kotoisin olevan laulajan Kojosen pojan runossa samantapainen jakso esiintyy lyhyempänä ja järjestykseltään varioivammin:

Iski ruoskalla hevoista,
Ori juoksi, ohjat seiso,
Reki juoksi, tie lyheni.
Ajo matkoa vähäsen
Sekä tietä pikkaraizen:
([SKVR VII1 373](#))

Ajon matkoa vähäsen,
Iskin ruoskalla hevosta,
Ori juoksi, ohjat seiso,
Reki vieri, tie lyheni:

([SKVR VII1 373](#))

Joku toinen esittäjä voi laulaa vaikkapa ainoastaan ”oro juoksi, matka joutu” tai ”löi hään ruosall orroi, kamahutti kannoillaan.” Vaihtoehtoisesti matkan teon voi kuvata vaikkapa säeparilla ”ajoit matkaa vähhäisen, tekkiit tietä pikkaraizen”. (SKVR I1 108; III 1258; IV2 2055; VII1 790.) Samaa asiaa kuvaavilla, samankaltaisissa runokokonaisuuksissa lomittain ja vaihtoehtoisina esiintyvillä, toisiinsa vaihtelevin perheyhtäläisyyden sitein liittyvillä kiteytyneillä ilmauksilla ei siis aina ole minkäänlaisia sanatason yhteyksiä.

Tietyt formulat ja formularyhmät voivat olla *oro juoksi* -formulan tavoin laajalti käytettyjä tai ainoastaan tietylle yksittäiselle runolle tyypillisiä, tunnusomaisia. Esimerkiksi formula *savu soarella palauve* esiintyy Karjalassa ja Kainuussa hyvin tunnuksenomaisesti Lemminkäisen virren aloituksena tai sen alkupuolella (Frog 2016). Vähemmän säännöllisesti säe lähiformuloineen esiintyy myös häälaulujen aihevana (SKVR-tietokannan runohakemiston otsikko Savu



saarella palaa), ketjurunossa (SKVR-hakemiston otsikko Kenpä saarella savua polttaa) sekä muutaman kerran erilaisissa muissa kertovissa tai syntyrunoissa (SKVR I1 163; I2 1119; I2 erill.), myös Arhippa Perttusen runossa Viisaampansa vieressä maannut (SKVR 13 2008). Muuttuvimmiskin käyttöyhteyksissään formularypäs näyttää viittaavan Lemminkäisen virren myyttisten Päivölän pitojen kaltaisiin juhla- ja muisteluun.

Runotyyppi

Kalevalamittaisen runon sisällöillä ja kieliasuilla on vaihtelevia keskinäisiä suhteita. Runotyyppi tai juonenkulku, kuten Kultaneito, voi esiintyä eri alueilla pintatasoltaan hyvin erilaisena, ja toisaalta tietty teema tai formula voi esiintyä pintatasoltaankin hyvin samanlaisena kautta koko itämerensuomalaisen alueen (Frog 2016; Harvilahti 1992; Kouvola 2019; Siikala 2012, 120–126, 181–185). Formula- ja teemarakenteiden variaation ja vaihtoehtoisuuden takia sisällöllisen samankaltaisuuden kattava etsiminen pelkillä sanahauilla on vaikeaa, vaikka prosessi etenisi-kin laajempien formularyppäiden ja monien erilaisten niille tyyppillisten sanojen ja ilmausten vuorottelevilla hauilla. Ääritapauksessa saattaa olla, että runot ovat sisällöllisesti samankaltaisia vailla yhtään suoraa formulatason yhteyttä.

Sisällöllisten samankaltaisuuksien nopeassa kartoittamisessa SKVR-tietokannan ja virolaisen ERA-tietokannan runo- eli runotyyppihakemistot ovat korvaamattomia – myös SKS:n julkaisemattomien runojen aineistosta on olemassa runotyypeittäin järjestetty manuaalinen kortisto. ERA-tietokannan hakemisto pohjautuu täysin vanhemmille kortistoluokituksille, SKVR-tietokannan hakemistoa puolestaan on työstetty lajikohtaisesti uusiksi kirjasarjan hakemistoja yhdenmukaistaen ja paikoin aineistoa uudelleen analysoiden. Näissä kahdessa hakemistossa on jonkin verran yhteisiä otsikoita, kuten Kultaneito – Kuldnaine, mutta monin kohdin niiden jaottelut ja periaatteet eroavat toisistaan. Esimerkiksi ERA-tietokannan hakemiston runotyyppi Laulu voim vastaa lukuisia SKVR-korpuksen lyyrisiä laulun voimaan liittyviä otsikoita.

SKVR-tietokannan runohakemiston periaatteet vaihtelevat jonkin verran lajeittain: kaikkein tarkimmin, yksittäisten aiheilmien tasolla, on analysoitu lyyriset runot (Senni Timonen, suullinen tiedonanto). Tapauksesta riippuen otsikot kuvaavat

1. runokokonaisuutta, teemaa tai kertomusrakennetta (esim. kertovissa runoissa);
2. aihetta tai aiheilmää (esim. lyriikassa);
3. funktiota tai tilannetta (esim. loitsuissa); tai
4. se voi olla näiden yhdistelmä (esim. häärunoissa) (<https://skvr.fi/skvr-runohakemisto>).

Kertovien runojen kohdalla otsikon alla on siten melko laajojakin juonellisia kokonaisuuksia, joiden voisi katsoa myös koostuvan useista pienemmistä teemoista (esim. Sammon ryöstö). Lyriikassa taas aiheilmät voivat olla sangen lyhyitä (esim. Alahalla allin mieli) – lyriikalle on myös tyyppillistä kertovia runoja varioivampi ja ketjumaisempi rakentuminen (Timonen 2004). Loitsut on jaoteltu ennemminkin sen mukaan, minkä vaivan parantamiseen tai minkälaiseen parantamistilanteeseen teksti liittyy kuin tarkan sisällöllisen tai säerakenteen vastaavuuden perusteella (esim. Riisi; häärunoista esim. Kylvetyruno).

Kuten usein on todettu, runoaineiston monitasoisesta variaatiosta on usein mahdotonta erottaa yksiselitteisiä, tarkkarajaisia, poikkeuksettomia laji- tai runokohtaisia kategorioita (Tarkka 2005; Timonen 2004; Kallio 2013, 93–95). Käytännössä hakemisto kuvaakin hyvin monen



tasoisia kokonaisuuksia, ja siihen on syytä suhtautua jatkuvasti kyseenalaistettavana ja pohdittavana työvälineenä. Sen avulla on mahdollista löytää yhdelle tekstile nopeasti juonen tai käytön kannalta lähimpiä vastineita tai tarkastella erilaisten teemojen esiintymistä vaikkapa eri alueilta tallennetuissa aineistoissa, mutta joidenkin tutkimuskysymysten yhteydessä se voi johtaa myös harhaan.

Runotyypinhakemisto on tällä hetkellä käytettävissä lähinnä SKVR-tietokannan ja Octavon hakutoimintoina ja niiden näyttämänä runokohtaisina otsikoina, joita voi yhdellä runotekstillä olla useita. SKVR-tietokannassa on esillä hakemiston yleiskuvas sekä listaus ylätasoinen jaotteluista, mutta ei kuvausta lähes 7 500 yksittäisestä otsikosta – yhä täydentyvä hakemisto on tarkoitus julkaista yksityiskohtineen myöhemmin. (SKVR-tietokanta, SKVR:n runohakemisto, <https://skvr.fi/skvr-runohakemisto>.) Tekstiedoston sisältämät selitykset sekä jaotteluvalintojen ja runotyyppien välisten yhteyksien ja päällekkäisyyksien kuvaukset ovat olennaisia runohakemiston tulkinnan ja käytön kannalta.

Ketjuina etenevät formula- ja teemahaut

Runokielen kiinteidenkin ilmausten variaatio, rypäsmäisyys, vaihtoehtoisuus ja ketjuuntuminen tarkoittaa sitä, että jos tietyn teeman, aiheelman tai ilmauksen haluaa kartoittaa jotakuinkin kattavasti, hakuja täytyy jatkaa ketju- tai verkostomaisesti ja lähestymistapoja vaihdellen. Näin on mahdollista tunnistaa ne tiettyyn aiheeseen tai formularyppäeseen liittyvät keskeiset, tunnusomaiset, pysyväluonteiset termit, formulat ja sanamuodot, joiden kautta on mahdollista löytää suurin osa haluttuja tekstitoisintoja.

Havainnollistamme asiaa esimerkinomaisesti yhden teeman hakuprosessin kuvauksella, jossa lähtökohtana on keminsaamelaisen Olaus Sirman vuonna 1672 kirjoittaman Kulnasatz-runon linnun siivillä lentämisen teeman vertailu SKVR-korpukseen. Keminsaamelaisesta suullisesta perinteestä ei ole tallentunut montakaan esimerkkiä, joten runojen vertailu lähialueiden kulttuureihin on keskeinen väline toisaalta niiden luonteen, toisaalta keminsaamelaisen suullisen kulttuurin piirteiden pohtimisessa. Sirman kahdella runolla onkin selkeitä sisällöllisiä ja muodollisia yhteyksiä etenkin myöhempisiin saamelaisperinteisiin, mutta myös joitain yhtymäkohtia kalevalamittaiseen runostoon ja vienankarjalaisiin joikuihin. (Kallio, Valtonen ja Jouste 2019; Kallio 2019.)

Tämänkaltaisessa sisällöllisessä vertailussa on pakko aloittaa melko satunnaisilla sanahauilla. Sirman Kulnasatz-runossa keskeistä on haave linnun siivillä lentämisestä. Erilaiset formulahaut tarkasti Sirman runoa vastaavilla sanapareilla *variksen siivet* tai *sotkan siivet* variaatioineen eivät tuota tuloksia: vastine löytyy ainoastaan lastenrunosta Mikä tuolla näkyy, jossa kätkyt on usein ”sotkan siiven alla”. *Linnun siivet* on Sirman käyttämien ilmausten pohjalta johdettu yleisempi kokeiluterminä, joka tässä tapauksessa onnekaasti tuottaa osuvia. Sen kautta taas on mahdollista päästä kalevalamittaisessa aineistossa yleisempiin ja keskeisempiin ilmauksiin.

Sanojen alkuosalla hakeminen antaa itämerensuomalaisissa kielissä laajemman tuloksen kuin koko sanoilla. Tulosjoukon määrä ja laatu riippuvat usein tapauskohtaisesti siitä, mistä kohtaa sana katkaistaan, kuinka pitkä se on ja paljonko sanastossa on merkeiltään sitä läheisesti muistuttavia sanoja. Vaikkapa haku enintään kolmen sana erottamista sananalusta linnu* ja siiv* antaa tulokseksi 63 osumaa, joiden joukossa viisi *linnun siivet*, *sääsken sääret* -tyyppistä



säettä. Tämä vaikuttaa lupaavalta, sillä Sirman runossa puhutaan ensin linnun siivistä, sen jälkeen jaloista ja kintereistä.

Ensimmäistä hakua kannattaa testata ja jatkaa osumissa kiinteästi haettuun termiin liittyvillä ilmauksilla. Haku tulokseksi tulleen formulan *linnun siivet, sääsken sääret* jälkiosalla ("sääs*säär*"~2) antaa 55 osumaa. Näihin kuuluu 38 inkeriläistä runoa, jossa säkeen alkuosa *linnun siivet* esiintyykin yleisemmässä muodossaan *sirkun siivet*, joka puolestaan esiintyy yleisenä myös muilla alueilla. Yleisin ilmaus löytyy siis mutkan kautta. Formulana tiiviys eli sille keskeisten sanojen etäisyys toisistaan voi myös vaihdella. Joskus *sirkun siivet* ja *sääsken sääret* eivät esiinny samassa säkeessä, vaan ainoastaan paralleelisten sanojen parina peräkkäisissä säkeissä: "Ku miul olliis linnun siivet / Ku miul olliis sääsen sääret."

Tuloksia kannattaa selata myös laajempia runokokonaisuuksia lukien. Hakujen epätarkkuuden, laajuuden ja satunnaisuuden hyvä puoli on siinä, että ne pakottavat lukemaan tietyn ilmauksen tai teeman käyttöä etsittyä laajemmalla temaattisella ja maantieteellisellä alalla. Tässä tapauksessa ilmauksen esiintymisessä on selviä alue- ja lajikohtaisia jakautumia, joiden ymmärtäminen auttaa hauissa ja analysissa eteenpäin. *Sääsken sääret* liittyvät myös esimerkiksi humoristiseen aiheeseen, jossa kerskutaan aitassa tai tarjolla olevien ruokavarojen määrällä ja joka esiintyy varioiden useiden eri runoteemojen kuten esimerkiksi Kyll' on kystä aitassani puitteissa: "Kissan kintut, sääsken sääret / Peiposen peräpakarat". Humoristisen aiheen yhteydessä *sääsken sääret* esiintyvät muillakin maantieteellisillä alueilla, *sirkun siipien* yhteydessä taas lähinnä vain Inkerissä. Toisaalta *sirkun siivet* esiintyvät moninaisissa yhteyksissä lähes koko suomalais-karjalais-inkeriläisessä aineistossa, toisin kuin *linnun siivet* tai *sääsken sääret*. *Sirkun siivet* on siis yleisin ja laajakäyttöisin karjalais-inkeriläis-suomalaisen alueen ilmaus.

Suuri osa inkeriläisistä *sirkun siivet* -tyyppisistä formuloista liittyy runotyyppiin Haudasta nostaisin emoni, muutama myös runotyyppiin Jos olis sirkun siivet, lentäisin armaan luo. Näillä ja muilla tulokseksi tulleilla runotyyppi-otsikoilla hakeminen on yksi keino etsiä edellisten hakujen ulkopuolelle jääneitä formularyyppeään variaatioita. Onko joku vaihtoehtoinen ilmaus yleinen tietyissä runotyyppissä? Haut näin löytyneiden formula- ja säevaihtoehtojen yksittäisillä sanoilla ja ilmauksilla antavat taas lisää esimerkkejä variaation kirjosta – ja samalla suuren joukon osumia aivan toisiin runotyyppisiin.

Jos tavoittelee mahdollisimman kattavaa tulosjoukkoa, täytyy tarkastella erikseen myös formularyyppeään keskeisimpien sanojen vaihtelevia kirjoitusasuja, tässä esimerkiksi haulle *siivet*~2. Etenkin lyhyissä sanoissa näin löytyvä sanamuotojen kirjo voi olla pääosaltaan aika kaukana etsitystä. Esimerkkihaku löytää varioivia sanamuotoja kuten *siiven, šiived, siibet, siive*, mutta myös aivan muita sanoja kuten *liivat, viijet, kuivet, siit, sarvet* ja *hirvet*. Ylimääräisiä osumia on helpointa rajata pois kustakin hausta ainoastaan niissä esiintyviä mahdollisimman kiinteämuotoisia sanoja tai sanan alkuja rajaamalla (-kuiv*, -liiva* jne.). Octavolla on helppo rajata hausta ja tuloksista pois myös kokonaisia runoteemoja (-theme:"Kyll on kystä aitassani"). Väli-lehdellä *Term discovery* on myös käytössä listaus sanamuotojen variaatioista, joista voi valita hakuun haluamansa muodot.

Kielten, kirjoitusmuotojen ja aineiston tuntemus vaikuttaa hakumahdollisuuksiin suuresti. Viron kieltä ja murteita kunnolla taitamattomana Kallion oli työlästä löytää inkeriläis-karjalaisen *sirkun/linnun siivet* -formulan yleisin virolainen vastine *tedre tiivad* eli "teeren siivet". Tämä edellytti ERA-tietokannan aineiston haarukointia inkeriläisten ja karjalaisten toisintojen



pohjalta lähtevin yksittäisin sanahauin ja lisähakuja niiden tuloksina tullein uusin säemuodoin ja runotyypein, samalla viron kielen sanakirjaa ja murre-sanakirjaa selaten. Nopeinta olisi ollut kysyä asiasta suoraan virolaisen aineiston tuntevalta kollegalta ja tehdä sitten itse täydennys-hakuja. Toisaalta aineistojen takeltelevakin käyttäminen johtaa hitaasti myös niiden parempaan tuntemukseen.

Riippuu suuresti yksittäisestä runosta ja ilmaisusta, miten sanahaut toimivat. Tietty sanat ja ilmaukset esiintyvät vain tietyn runoaihelman puitteissa, toiset taas kymmenien tai jopa satojen erilaisten aiheiden ja ilmausten puitteissa. Tietty ilmaus voi toistua tietyssä runotyypissä alueelta toiselle, tai samankaltainen ajatus voidaan esittää eri alueilla eri sanoin. On sanoja, jotka ovat hyvin samankaltaisia tai jopa samoja lähisukukielissä, ja niitä, jotka esiintyvät vain osassa. Joidenkin sanojen kirjoitusasu vaihtelee suuresti murteista ja käytetystä ortografiasta riippuen, toiset taas on kirjoitettu hämmästyttävän samanlaisina läpi aineiston.

Hakustrategioiden toimivuus vaihtelee sen mukaan, minkälainen, miten varioiva, kuinka yleinen tai miten monessa eri merkityksessä tai yhteydessä esiintyvät sana, formula tai aihe on kyseessä, ja sen mukaan, miten tarkka tai kattava hakutulos on tarpeen. Aineiston kielellisestä ja poeettisesta variaatiosta johtuen ei voi olettaa, että kaikki erilaiset variaatiot löytyisivät välttämättä minkäänlaisilla hauilla, ja aineistojen käyttö sanahakujen tai runotyypihakemistojenkin kautta edellyttää melko laajaa runojen ja käytettyjen kielten tuntemusta. Mitä tutumpi aineisto, sitä helpommin ja perinpohjaisemmin se on selattavissa. Osansa on myös tutkijan tilannekohtaisella mielikuvituksella ja sattumalla, joskin riittävän pitkäaikainen ja monipuolinen aineiston läpikäyminen luultavasti johtaa samankaltaisiin tuloksiin. (Kallio 2019.) Tämä tarkoittaa myös sitä, että tietyn hakuprosessin tulokset eivät välttämättä ole suoraan toistettavissa.

Aineiston ennakkotuntemus ei vaikuta vain siihen, minkälaisia hakuja käyttäjä keksii tehdä, vaan myös siihen, miten hän saatuja hakutuloksia tulkitsee. Runotyypihakemiston otsikoiden ohella muut SKVR-aineiston metatiedot vaikuttavat keskeisesti hakutulosten tulkintaan. Satunnaisen käyttäjän ongelmana kuitenkin on, että keskeisiä taustatietoja ei ole kattavasti kirjattu mihinkään yksittäiseen lähteeseen. Osa näistä löytyy esimerkiksi kirjasarjan osien esipuheista, vanhasta tutkimuskirjallisuudesta tai arkistossa olevista oheistiedoista. Kokenut lukija tarkastelee jokaista yksittäistä tekstiä suhteessa siihen, mitä tietää tallennushistoriasta, alueellisista runokulttuureista ja niiden ajallisista muutoksista, tietyn tallentajan menetelmistä ja tavoitteista, julkaistujen runojen vaikutuksista suulliseen kulttuuriin ja runoteemojen intertekstuaalisesta verkostosta.

Hakujen toimivuus ja työläys riippuvat suuresti tutkimus- ja hakuasetelmasta. Esimerkiksi yhden Lönnrotin *Kantelettareen* (1840) koostaman runon tarkastelu suhteessa SKVR-korpukseen paljasti myös lähteitä ja vaikutteita, joita aiempi manuaalinen luenta ei ollut tavoittanut (Kallio 2019; samankaltaisesta asetelmasta ks. myös Hämäläinen ja Sykäri 2019). Matthias Salamniuksen vuonna 1690 sepittämän kristillisen kalevalamittaisen eepoksen *Ilo-Laulu Jesuxelle* ja SKVR-aineiston vertailu Octavolla tuottaa samoin uusia näkökulmia ja tulkintamahdollisuuksia sekä siihen, miten Salamnius on käyttänyt tuntemaansa pohjalaista, karjalaista ja inkeriläistä perinnettä, että siihen, miten hänen runonsa säkeet ovat vaikuttaneet myöhemmän suulliseen perinteeseen (Bastman ja Kallio [tulossa]). Tilanne on hankalampi ja osuvat haut vaikeammin löytyviä, jos etsitään toisiinsa liittyvien, osin vaihtoehtoisten formuloiden ennalta tuntematonta rypästä, tai jos joku keskeinen ajatus tai säe voi esiintyä useammassa



hyvin erilaisessa rinnakkaisessa muodossa, kuten *linnun siivet – tedre tiivad* (Kallio, Valtonen ja Mäkelä [tulossa]). Laajemman aiheelman tai juonenkulun käsittely puolestaan edellyttää perinpohjaisempaa analyysia ja aikaisempien oletusten kriittistä tarkastelua: näillä premisseillä myös runotyyppihakemistot ovat toimiva osa hakuvälinepalettia (Kallio et al. 2017, erit. 428–436).

Lopuksi

Kalevalamittaisen runouden aineistoille on ominaista hyvin monitasoinen muuntelu ja toisaalta monien tasojen vahva toistoisuus. Aineisto varioi niin monimuotoisesti ja runsaasti, että on vaikea keksiä tutkimusasetelmaa, jossa sen tarkastelu puhtaasti laskennallisilla menetelmillä olisi mielekästä. Sähköiset välineet voivat kuitenkin olla monin tavoin avuksi aineiston käsittelyssä, analyysissä sekä osa-aineistojen luonnissa. Hedelmällisimmiltä vaikuttavat lähestymistavat, joissa suuren aineiston ja metatietotason käsittely yhdistyy mahdollisuuksiin tarkastella myös yksittäisiä tekstejä ja säätason osumia.

Suomalaisissa ja virolaisissa arkistoissa oleva kalevalamittaisen aineisto on suuri, yli 250 000 runotekstiä. Kalevalamittainen runous on määrittynyt kansallisesti siinä määrin keskeiseksi sekä Virossa että Suomessa, että sen tallentamiseen, järjestämiseen, kortistamiseen ja tematisointiin on käytetty poikkeuksellisen suuria resursseja. Tämä tekee aineistosta harvinaisen monikäyttöisen. Toisaalta pitkä tallennus- ja kuratointihistoria aiheuttaa aineistoon moninaisia vääristymiä. Aineisto ei ole kattava kuvaus siitä, mitä rahvas ennen toista maailmansotaa itämerensuomalaisilla alueilla kalevalamitalla esitti, vaikka kertoo siitä paljon.

Digitoitu, hakutoiminnot mahdollistaviin käyttöliittymiin sijoitettu aineisto antaa aiempaa nopeampia mahdollisuuksia analysoida aineiston kokonaisuutta ja jakautumia sekä runojen ja ilmaisujen välisiä yhteyksiä ja toistuvuuksia. Sähköistä aineistoa käyttävä nykytutkija kykenee suhteellisen nopeasti tarkistamaan, minkälaisissa käyttöyhteyksissä tietyn tyyppinen yksittäinen ilmaus aineistossa löytyy, tai saamaan kokonaiskuvan siitä, miten tietty aihe tai runotyyppi eri alueilla, aikoina ja käyttöyhteyksissä varioi. Laajempien sisällöllisten samankaltaisuuksien kohdalla olemassa olevien runotyyppihakemistojen käyttäminen on usein tarpeellinen apuväline. Samalla niiden käyttö edellyttää myös kriittistä tarkastelua: temaattisen jaottelun tarpeet riippuvat kulloisestakin tutkimusasetelmasta, ja hakemistot ovat rakentuneet suhteessa pitkiin tutkimusperinteisiin.

Hakutoimet ovat eräänlaisia yrityksiä hahmottaa ja jäljittää aineiston monitasoisia, varioivia toistuvuuksia. Jokaisen haun taustalla on jonkinlainen tutkimuskysymys: miten tämä ilmaus, merkkijono, runotyyppi, tallentaja, alue tai esittäjä esiintyy korpuksessa tai miten eri piirteet suhteutuvat toisiinsa. Haut voi käsittää myös tapauskohtaisten tutkimusaineistojen muodostamisena (tutkimusaineistojen muodostamisesta ks. Knuutila 2010).

Aineiston ennakkotuntemus, tutkimuskysymyksen ja haettavan osa-aineiston luonne, mielikuvitus ja puhdas sattuma vaikuttavat siihen, kuinka kattavasti ja nopeasti tietty osa-aineisto on sähköisestä korpuksista mahdollista löytää lähemmin tarkasteltavaksi. Manuaalisen kalevalamittaisen runon maantieteellis-historiallisen tutkimuksen aikana vallitsi kaiken mahdollisen käytettävissä olevan aineiston huomioimisen ideaali. Tätä omana aikanaankin hypoteettista tavoitetta ei sähköisissä hauissa aina ole mahdollista täyttää, tai ainakaan sen täyttymisestä ei ole mahdollista varmistua. Jotain saattaa aina jäädä hakujen ulkopuolelle. Toisaalta sähköisellä



luennalla voi saada myös esiin vaikkapa sellaisia sana- ja formulatason intertekstuaalisia suhteita, joita manuaalinen näin ison aineiston luenta ei onnistuisi tavoittamaan. Sähköinen käyttöliittymä luo myös mahdollisuuksia nopeisiin aineiston läpileikkauksiin ja jaotteluihin, joiden pohjalta yleiskuva suurenkin aineiston luonteesta on mahdollista muodostaa paljon aiempaa pikaisemmin.

Toisaalta sähköisessä luennassa hakuprosessi, ennako-oletukset ja aineiston ennakkotuntemus voivat vaikuttaa tuloksiin vahvemmin kuin manuaalisessa, etenkin jos edetään pelkien säetason osumakohtien varassa ilman laajempien runokokonaisuuksien tai -aineistojen lukemista. Manuaalinen luenta auttaa havaitsemaan piirteitä, joita ei olisi itse tullut etsineeksi. Toisaalta yhtä lailla sähköinen haku saattaa heittää silmien eteen samankaltaisuuksia, joita ei laajempaakaan aineistokokonaisuutta läpi lukiessaan tullut huomanneeksi. Parhaimmillaan sähköisten aineistojen käyttö onkin liikettä läheltä ja kaukaa lukemisen eli tilastollisten, määrällisten tai visuaalisten esitysten ja yksityiskohtaisten alkuperäistekstien välillä (ks. Jänicke et al. 2015; Moretti 2013).

Lähteet

Tutkimusaineistot

- ERA. *Eesti regilaulude andmebaas*. 2003–. Toimittaneet Janika Oras, Mari Sarv ja Liina Saarlo. Tarto: Eesti Kirjandusmuuseumi Eesti Rahvaluule Arhiiv. Käytetty 2018–2019. <http://www.folklore.ee/regilaul/andmebaas>.
- SKVR I–XV = *Suomen kansan vanhat runot*. 1908–1997. Eri toimittajia. Helsinki: Suomalaisen Kirjallisuuden Seura. Julkaistu sähköisenä: *SKVR-tietokanta*. 2004–. Käytetty myös Octavo-liittymän kautta 2018–2019. <http://www.skvr.fi>.

Kirjallisuus

- Ahola, Joonas. 2014. "The Other Island: Kalevalaic Epic and History." Teoksessa *The Viking Age in Åland: Insights into Identity and Remnants of Culture*, toimittaneet Joonas Ahola, Frog ja Jenni Lucenius, 55–82. Helsinki: Suomalainen Tiedeakatemia.
- Anttonen, Pertti J. 2005. *Tradition through Modernity: Postmodernism and the Nation-State in Folklore Scholarship*. Helsinki: Suomalaisen Kirjallisuuden Seura. <https://doi.org/10.21435/sff.15>.
- Apo, Satu. 1986. "Questions Arising in the Comparative Study of Magic Tales." *Journal of folklore research* 23(2/3): 177–186.
- Arukask, Madis. 2003. *Jutustava regilaulu aspektid: 19. sajandi lõpu setu lüroepiliste regilaulude žanr ja struktuur*. Tartu: Tartu Ülikooli kirjastus.
- Bastman, Eeva-Liisa ja Kati Kallio. [Tulossa.] "'Turning simple speech into beautiful song': Imitative poetics and the combination of registers in *Ilo-Laulu Jesuxesta* (1690)." Teoksessa *Networks, Poetics and Multilingual Society in the Early Modern Baltic Sea Region*, toimittaneet Ilkka Leskelä, Kati Kallio, Tuomas M. S. Lehtonen ja Anu Lahtinen. Leiden: Brill.



- Blomster, Risto ja Kati Mikkola. 2017. "Kenen perinnettä? Romani-informantit Suomen Kansan Vanhat Runot -aineistoissa." *Etnomusikologian vuosikirja* 29: 1–40. <https://doi.org/10.23985/evk.63130>.
- Bono, Marie. 2003. *Volmari Porkan vuonna 1883 Inkeristä keräämien runojen pikakirjoitusmuistiinpanot*. Julkaisematon pro gradu -tutkielma. S 545 I–III. Helsinki: Helsingin yliopisto, Kulttuurien tutkimuksen laitos.
- DuBois, Thomas A. 2006. *Lyric, Meaning, and Audience in the Oral Tradition of Northern Europe*. Notre Dame, IN: University of Notre Dame Press.
- Evseev, V. J. 1950. *Karelskije epitseskije pesni*. Leningrad: Akademija nauk SSSR, Karelo-finski filiaal AN SSSR, Institut istorii, jazyka i literatury.
- Evseev, V. 1976. *Karjalan kansan runot: 1, Kalevalanaiheiset kertovat runot*. Tallinn: Eesti raamat.
- Foley, John Miles. 1988. *The theory of oral composition: History and methodology*. Bloomington and Indianapolis: Indiana University Press.
- Foley, John Miles. 1991. *Immanent Art: From structure to meaning in traditional oral epic*. Bloomington and Indianapolis: Indiana University Press.
- Frog. 2013. "The Parallax Approach: Situating Traditions in Long-Term Perspective." Teoksessa *Approaching Methodology*, toinen uudistettu painos, toimittaneet Frog ja Pauliina Latvala sekä Helen F. Leslie, 99–129. Helsinki: Academia Scientiarum Fennica.
- Frog. 2016. "Linguistic Multiforms in Kalevalaic Epic: Toward a Typology." *The Ecology of Metre. RMN Newsletter* 11: 61–98.
- Gröndahl, Satu M. 1997. *Den ofullkomliga traditionen: Bilden av Ingermanlands kvinnliga runotraditionen*. Uppsala: Acta Universitatis Upsaliensis.
- Harvilahti, Lauri. 1992. *Kertovan runon keinot: Inkeriläisen runoepiikan tuottamisesta*. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Harvilahti, Lauri. 2004. "Vakiojaksot ja muuntelu kalevalaisessa epiikassa." Teoksessa *Kalevala ja laulettu runo*, toimittaneet Anna-Leena Siikala, Lauri Harvilahti ja Senni Timonen, 194–214. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Harvilahti, Lauri. 2013. "The SKVR Database of Ancient Poems of the Finnish People in Kalevala Meter and the Semantic Kalevala." *Oral Tradition* 28(2): 223–232.
- Hautala, Jouko. 1945. *Lauri Lappalaisen runo: Vertaileva kansanrunoudentutkimus*. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Honko, Lauri. 1986. "Types of Comparison and Forms of Variation." *Journal of Folklore Research* 23(2/3), 105–124. www.jstor.org/stable/3814443.
- Honko, Lauri. 1998. *Textualising the Siri Epic*. Helsinki: Suomalainen Tiedeakatemia.
- Huttu-Hiltunen, Pekka. 2008. *Länsivienalainen runolaulu 1900-luvulla: Kuuden runolaulajan laulutyylin kulttuurisensitiivinen musiikkianalyysi*. Kuhmo: Juminkeko.
- Hämäläinen, Niina ja Venla Sykäri. 2019. "Kalevala ja kansanrunousaineisto." Teoksessa *Avoin Kalevala: Kansalliseepoksen digitaalinen, kriittinen editio*, toimittaneet Niina Hämäläinen, Reeta Holopainen, Marika Luhtala, Juhana Saarelainen ja Venla Sykäri. Helsinki: Suomalaisen Kirjallisuuden Seura. <http://kalevala.finlit.fi>.
- Ilyefalvi, Emese. 2018. "The theoretical, methodological and technical issues of digital folklore databases and computational folkloristics." *Acta Ethnographica Hungarica* 63(1): 209–258. <https://doi.org/10.1556/022.2018.63.1.11>.



- Kallio, Kati. 2013. *Laulamisen tapoja: Esitysareena, rekisteri ja paikallinen laji länsi-inkeriläisessä kalevalamittaisessa runossa*. Helsinki: Helsingin yliopisto. <http://urn.fi/URN:ISBN:978-952-10-9566-5>.
- Kallio, Kati. 2019. "Oisko linnun lentoneuot: Lönnrotin taustavaikutteista ja runokorpusten sähköisestä lukemisesta." Teoksessa *Eurooppa, Suomi, Kalevala*. Kalevalaseuran vuosikirja 98, toimittaneet Ulla Piela, Pekka Hakamies ja Pekka Hako, 175–199. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Kallio, Kati, Frog ja Mari Sarv. 2017. "What to Call the Poetic Form: Kalevala-Meter or Kalevalaic Verse, regivärss, Runosong, the Finnic Tetrameter, Finnic Alliterative Verse, or Something Else?" *RMN Newsletter* 12–13: 94–117.
- Kallio, Kati, Tuomas Lehtonen, Senni Timonen, Irma-Riitta Järvinen ja Ilkka Leskelä. 2017. *Laulut ja kirjoitukset: Suullinen ja kirjallinen kulttuuri uuden ajan alun Suomessa*. Helsinki: Suomalaisen Kirjallisuuden Seura. <https://doi.org/10.21435/skst.1427>.
- Kallio, Kati, Taarna Valtonen ja Marko Jouste. 2019. "Olaus Sirman runojen vertailevaa luentaa: runojen poetiikka ja rakenne ympäröivien kulttuurien tekstien näkökulmasta." *Suomalais-Ugrilaisen Seuran Aikakauskirja* 97, 111–155. <https://doi.org/10.33340/susa.75266>.
- Kallio, Kati, Taarna Valtonen ja Eetu Mäkelä. [Tulossa.] "Olaus Sirman runot ja itämerensuomalainen suullinen perinne." Artikkelikäsikirjoitus.
- Knuuttila, Seppo. 2010. "Tutkimusaineistojen muodostaminen." Teoksessa *Vaeltavat metodit*, toimittaneet Jyrki Pöysä, Helmi Järviluoma ja Sinikka Vakimo, 19–42. Joensuu: Suomen Kansantietouden Tutkijain Seura.
- Kouvola, Karolina. 2019. The artificial bride on both sides of the Gulf of Finland: The Golden Maiden in Finno-Karelian and Estonian folk poetry." Teoksessa *Contacts and Networks in the Baltic Sea Region: Austmarr as a Northern Mare Nostrum, ca. 500-1500 CE*, toimittaneet Maths Bertell, Frog ja Kendra Willson, 211–234. Amsterdam: Amsterdam University Press.
- Lauerma, Petri. 2004. *Larin Parasken epiikan kielellisestä variaatiosta*. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Lord, Albert B. 2002 [1960]. *The Singer of Tales*. Cambridge: Harvard University Press.
- Mironova, V. P. M. 2006. *Èpičeskie pesni Ūžnoj Karelii: Anuksen karjalazien eppizet pajot*. Petrozavodsk: Periodika.
- Mäkelä, Eetu. 2019. "Octavo API." *SwaggerHub*. Boston: Smartbear. <https://app.swaggerhub.com/apis/comhis/octavo/1.0.0>.
- Oras, Janika. 2008. *Viie 20. sajandi naise regilaulumaailm: Arhiivitekstid, kogemused ja mälestused*. Tartu: Eesti Kirjandusmuuseumi Teaduskirjastus.
- Palola, Elina. 2009. *Arkaismi, konteksti ja kirjuri: Sananloppuisen k:n merkintä Christfrid Gananderin kansanrunokoelmassa*. Oulu: Oulun yliopisto.
- Saarinen, Jukka. 2018. *Runolaulun poetiikka: Säe, syntaksi ja parallelismi Arhippa Perttusen runoissa*. Helsinki: Helsingin yliopisto. [http://urn.fi/URN:ISBN:ISBN978-951-51-3919-1 \(PDF\)](http://urn.fi/URN:ISBN:ISBN978-951-51-3919-1(PDF)).
- Sarv, Mari. 2008. *Loomiseks loodud: Regivärsimööd traditsiooniprotsessis*. Tartu: Eesti Kirjandusmuuseum.
- Sarv, Mari. 2011. "Metrical Universals in Oral Poetry." Teoksessa *Current Trends in Metrical Analysis*, toimittanut Christoph Küper, 329–337. New York: Peter Lang Verlag.



- Seppä, Tiina. 2015. *Kohtaamisia menneen kanssa: Tutkimus kansanrunousaineistojen synnystä ja myöhemmistä tulkinnoista*. Joensuu: Kansantietouden Tutkijain Seura.
- Sihvo, Hannes. 2003 (1973). *Karjalan kuva: Karelianismin taustaa ja vaiheita autonomian aikana*. 2. tarkistettu ja täydennetty painos. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Siikala, Anna-Leena. 1992. *Suomalainen šamanismi – mielikuvien historiaa*. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Siikala, Anna-Leena. 2012. *Itämerensuomalaisten mytologia*. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Siikala, Anna-Leena, Lauri Harvilahti ja Senni Timonen (toim.). 2004. *Kalevala ja laulettu runo*. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Siikala, Anna-Leena ja Sinikka Vakimo (eds.). 1994. *Songs beyond the Kalevala: Transformations of Oral Poetry*. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Stepanova, Eila. 2014. *Seesjärveläisten itkijöiden rekisterit: Tutkimus äänellä itkemisen käytänteistä, teemoista ja käsitteistä*. Joensuu: Suomen Kansantietouden Tutkijain Seura.
- Sykäri, Venla. 2011. *Words as events: Cretan mantinades in performance and composition*. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Särg, Taive. 2005. *Eesti keele prosoodia ning teksti ja viisi seosed regilaulus*. Tartu: Tartu Ülikooli Kirjastus.
- Tangherlini, Timothy R. 2016. "Big Folklore: A Special issue on computational Folkloristics." *Journal of American Folklore* 129(511): 5–13.
- Tarkka, Lotte. 2005. *Rajarahvaan laulu: Tutkimus Vuokkiniemen kalevalamittaisesta runokulttuurista 1821–1921*. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Tarkka, Lotte. 2010. "Lajisiirtymiä kalevalamittaisessa runossa: Kuivajärven Moarien tapaus." Teoksessa *Kalevalamittaisen runon tulkintoja*, toimittaneet Seppo Knuuttila, Ulla Piela ja Lotte Tarkka, 13–34. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Timonen, Senni. 2004. *Minä, tila, tunne: Näkökulmia kalevalamittaiseen kansanlyriikkaan*. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Vana Kannel*. 1875–2019. Eri toimittajia ja kustantajia. Tartu: Mattiesen/Teaduslik kirjandus/ Kirjandusmuuseum/Eesti keele instituut; Tallinn: Kulttuurikoondis/Eesti raamat.
- Västrik, Ergo-Hart. 2007. *Vadjalaste ja isurite usundi kirjeldamine keskajast 20. sajandi esimese pooleni: Alliktekstid, representatsioonid ja tõlgendused*. Tartu: Tartu Ülikooli Kirjastus.
- Wilson, William Albert. 1976. *Folklore and nationalism in modern Finland*. Bloomington: Indiana University Press.

FT Kati Kallio toimii tutkijatohtorina Suomalaisen Kirjallisuuden Seurassa ja Helsingin yliopistossa (Suomen Akatemian projektit 308381, 322071 ja 288119).

TkT Eetu Mäkelä toimii ihmis- ja tietojenkäsittelytieteiden välisen vuorovaikutuksen apulaisprofessorina Helsingin yliopistossa.