

Metodi- ja aineistotutkijan aineistot ja avoimuus

All eyes on data -seminaari
10.5.2019

Lotta Aarikka, lamaar@utu.fi



Kuka olen, mitä teen?

- Lotta Aarikka, tohtorikoulutettava Utuling-tohtoriohjelmassa kieli- ja käännöstieteiden laitoksella, suomen kielen ja suomalais-ugrilaisen kielentutkimuksen oppiaineessa.
- Väitöskirjan työnimi: ”**Murteentutkimuksen metodit ja ideologiat 1871–2018**”
 - Työ aloitettu loppuvuodesta 2015.
 - Rahoittajana oma laitos (1.10.–31.12.2015), Suomalaisen Tiedeakatemian Eino Jutikkalan rahasto (2015–2018) ja Koneen Säätiö (2019–2020), kenttätöitä rahoittanut Suomalaisen Kirjallisuuden Seuran Samuli Paulaharjun rahasto.
 - Pääasiallisena tutkimuskohteena 51 suomen kielen alalla ilmestynyttä väitöskirjaa, joiden otsikossa on sana *murre* tai jotka perustuvat pääasiassa murreaineistolle, sekä 20 tutkijan haastattelut.

Millaisia aineistoja olen kerännyt (ja avannut)?

- Saaren kartanoon liittyvän muistitiedon keruuta keväällä (2009?)
- Satakuntalaisuus puheessa -hankkeeseen liittyviä haastatteluja (2010)
- Pro gradu -haastatteluaineisto (2013)
- Lähdetietokanta (2017)
- Väitöskirja-aineiston haastattelut (2018)



Saaren kartanoon liittyvä muistitiedon keruu

- Folkloristiikan ja kansatieteen kenttäkurssi liittyen yhteistyöhön Koneen Säätiön kanssa.
- Tarkoituksena kerätä Saaren kartanoon ja sen eri vaiheisiin liittyvää muistitietoa paikallisilta.
- Kaksi haastattelua: yksi parin kanssa, toinen yksin.
- Informantit itselle entuudestaan tuntemattomia.
- Käytössä teemarunko.
- **Aineisto arkistoitu HKT-laitoksen arkistoon.**
- Tärkeimmät opit
 - Keskittyminen, jatkokysymyksiä esittäminen, informantin johdattelu.
 - Informantit hyvin erilaisia.
 - Informantit pitivät ”kiinnostavina” eri asioita kuin me.



Satakuntalaisuus puheessa -hankkeeseen liittyviä haastatteluja

- Vuonna 2007 käynnistynyt suomen kielen oppiaineen hanke ([lisätietoja](#)).
- Tutkimuskohteena Satakunnan alueen (nyky)puhekieli.
- Noin kymmenen haastattelua tai nauhoitusta v. 2010.
- Informantit itselle tuttuja sukulaisia ja ystäviä.
- Ei haastattelurunkoa, mutta käytössä ns. tiedostamistesti.
- **Aineisto arkistoitu Lauseopin arkiston SAPU-kokoelmaan.**
- Tärkeimmät opit
 - Haastattelulla ja nauhoituksella on eroa.
 - Osa informanteista on hyvin tietoinen nauhoitustilanteesta ja sensuroi itseään.
 - Tuloksena kandi (2011), jossa tarkastelin nimenomaan tiedostamistestejä.



Pro gradu -aineistonkeruu

- [Pro gradu](#) -aineistonkeruu keväällä 2013.
- Tutkimuskohteena kielenkäyttäjien käsitykset Porin murteesta sekä käsityksien ja kielen tuottamisen erot ja samankaltaisuudet.
- 21 haastattelua, 24 1980-luvulla syntynyttä informanttia.
 - Syntyperäisten porilaisten aineisto (10 informanttia) tuttuja sekä Poriin muuttaneiden aineisto (14 informanttia) itselle tuntemattomia.
- Käytössä itse laadittu haastattelurunko sekä imitaatio- ja tiedostamistesti.
- **Aineisto arkistoitu Lauseopin arkiston SAPU-kokoelmaan.**
- Tärkeimmät opit
 - Kysymyksiä kannattaa testata.
 - Kiinnostavimmat löydökset saattavat tulla haastattelurungon ulkopuolelta.
 - Aineiston hallinta ja analysointi on vaikeaa.



Lähdetietokanta

- 51 suomen kielen alalla ilmestyneen väitöskirjan lähdeluetteloista koostettu tietokanta.
 - Yhdellä tietokannan lähderivillä on tiedot lähteen tekijästä, julkaisuvuodesta, nimikkeestä ja julkaisutiedoista.
 - Lisäksi tietokantarivillä on tieto lähteen tekijän sukupuolesta ja kotimaisuudesta (vs. ei-kotimaisuudesta), ja nämä tiedot olen lisännyt aineistoon manuaalisesti.
- Tutkimustulokulma bibliometrinen: keihin tutkijoihin ja mihin teoksiin on viitattu eniten?
- Lähdetietokannassa on 9783 riviä, ja riveillä mainitaan 2657 uniikkia tekijää ja 5520 uniikkia teosnimikettä. Vanhimman viitatus teosnimikkeen ilmestymisvuosi on 1540 ja uusimman 2010.
- **Aineisto julkaistu avoimena Kielipankissa.**
- Tärkeimmät opit
 - Aineistoa käsitellessä versioiden tallentamisen, tiedostotyypin muuttamisen ja aineiston ”siivoamisen” pitää olla anaalisen järjestelmällistä tai alkaa itkettä.
 - Mitä isompi data, sitä enemmän täytyy sietää virheen mahdollisuutta.



Suomen murteita tutkineiden tutkijoiden haastatteluaineisto

- Liittyy [väitöskirjatyöhön](#), ensimmäinen haastattelu helmikuussa 2018, viimeinen joulukuussa, yhteensä haastatteluja 20 kpl.
- Tutkimuskohteena tutkijoiden suhtautuminen tutkimuskohteeseensa ("murre").
- Käytössä itse laadittu haastattelurunko, piirrostehtävä sekä kysely (termien kuvaileminen), jonka haastateltavat täyttivät haastattelutilanteessa.
- **Aineisto (haastattelut) arkistoidaan Lauseopin arkistoon, piirrosten ja termilomakkeen arkistointia pohdin vielä.**
- Tärkeimmät pohdittavat
 - Oman tutkimuksen tutkiminen on herkkä aihe: miten anonymisoin ilman, että analyysi vesittyy?
 - Miten analysoin piirrostehtävää?



Yhteenvetoa ja pohdintaa

- Omissa tutkimusintresseissä on tutkia aineiston ”kylmäketjua”: miten jonkun yksilön jossain ajassa tuottamasta puheesta tulee tieteellinen väite siitä, miten kieli universaalisti toimii?
- Mikään aineisto ei itsessään ole ”kvalitatiivista” – haastatteluja voi tutkia kvantitatiivisesti ja isoa aineistoa kuten lähdetietokantaani voi tutkia myös kvalitatiivisesti.
- En ole koskaan kerännyt aineistoa ns. omaan jemmaan – kaikki on tallennettu jonnekin. Samaan painostan opiskelijoita ja kaikkia muitakin, koska...
- Erityisesti kvalitatiivisen tutkimuksen kohdalla aineiston avoimuus on ensisijaisen tärkeää ja tieteellisen toistettavuuden ”korvaaja”:
 - Kvalitatiivisin metodein tehtyä tutkimusta ei voi samalla tapaa toistaa uudella aineistolla.
 - Omien analyysien ainoa luotettavuuden kontrolli on se, että joku voi verrata analyysejasi aineistoosi ja arvioida, ovatko johtopäätökset perusteltuja.



Yhteenvetoa ja pohdintaa

- Arkistoon eri tavoin määritellyin luvin arkistoitu haastatteluaineisto ei avoimen datan määritelmän perustein ole avointa, eikä sen mielestäni tarvitsekaan olla.
 - Esim. paikkatiedon erilaisten kaupallisten ja julkisten sovelluksien mahdollisuudet ovat aivan eri maailmasta kuin tutkimustarkoitukseen kerätyn 20 tutkijan haastattelun.
- Haastatteluaineiston jatkokäytön rajoista päättävät haastateltavat, mutta se ei tarkoita, etteikö tutkijan ratkaisulla olisi merkitystä.
- Tutkimusaineistojen keruu, käsittely ja avaaminen vertautuu toivottavasti tulevaisuudessa julkaisemiseen – se nimittäin on usein se työläin osa tutkimuksen teossa.
 - Aineistoihin liittyvä protektionismi roskakoriin heti.
 - Aineiston ja tutkijan työn arvo kasvaa, mitä useampi aineistoa pystyy käyttämään.
- Opetuksen ja tutkimuksen integraatiossa olisi paljon parannettavaa, mitä tulee aineiston keruuseen ja käsittelyyn.





**TURUN
YLIOPISTO**

Kiitos!