



# Miten tietokone näkee suomen murteet?

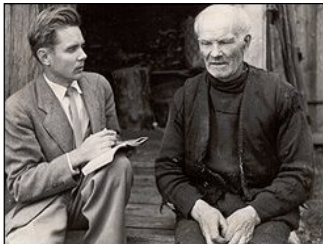
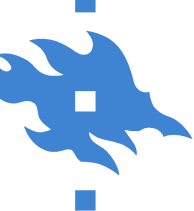
Antti Leino <[antti.leino@cs.helsinki.fi](mailto:antti.leino@cs.helsinki.fi)>

suomen kielen ja kotimaisen kirjallisuuden laitos  
tietojenkäsittelytieteen laitos

Kotimaisten kielten tutkimuskeskus



2.11.2009



Kuva: Kotus



Kuva: popgadget.net

- Johdanto: aineistot tietokoneelle
- Menetelmiä: mitä sitten?
- Murteita: löytyykö niitä?



# Johdanto

## Suomalaiset murreaineistot

- Pitkät perinteet aineistojen keruussa
- Jotkut aineistot tietokonemuodossa
  - Suomen murteiden sanakirjan konseptikarttoja
  - Muoto-opin arkisto
  - Lauseopin arkisto
  - Suomen murteet III a osittain
- Osa aineistoista sopii alueellisen vaihtelun etsimiseen
  - kunhan käytössä oikeat menetelmät...



# Johdanto

## Dialektometria

- Laskennallista murteentutkimusta
  - murrepiirteistä laskettuja eroja
  - usein ääntämyseroja
  - joskus (ja tässä) myös yleisemmin
- Tavoite: lisää objektiivisuutta
  - murteentutkijan intuitio ja asenteet taustalle
  - tietokone laskee
- Löytääkö tietokone murrealueet?



# Menetelmiä

## Kuinka erilaisia murteet ovat?

- Levenštein-etäisyys
  - sama teksti eri murteenpuhujilta
  - tekstit litteroidaan
  - kahden murteen ero = yhden kirjaimen kokoisten erojen (poisto / lisäys / vaihto) lukumäärä
  - keskieurooppalaisessa tutkimuksessa tavallinen
- Päijäthämäläisen murrealueen kolme ryhmää
  - 7a: *Äete taht**oo** männä ott**amaan** paalla maet**oo** kellarista.*
  - 7b: *Äete taht**uu** männä ottaan pajalla maet**uu** kellarista.*
  - 7c: *Äete taht**oo** männä ott**amaan** paalla maet**ova** kellarista.*

(Wiik 2006)

- 7a–7b: 7 eroa
- 7a–7c: 2 eroa
- 7b–7c: 8 eroa



# Menetelmiä

## Kuinka erilaisia murteet ovat?

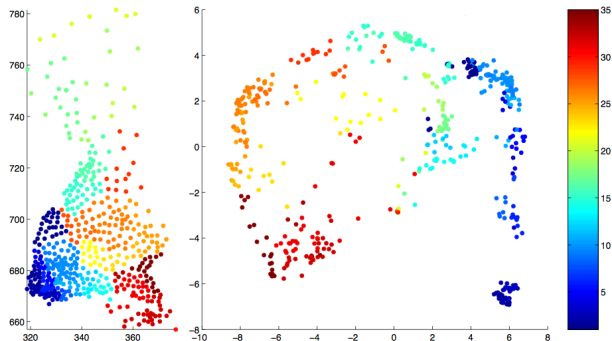
- Yhteiset murrepiirteet
  - paljonko kahden pitäjän välillä on eroja?
  - murrekartta-aineistoahan on
- Mitä etäisyysmitoilla voi tehdä?
  - $600 \times 600$  kahden pitäjän välistä etäisyyttä on aika hankala lukea
  - erilaisia tapoja havainnollistaa murteiden samanlaisuutta kuvin



# Menetelmiä

## Murre-erot kartaksi

- Sovitetaan kuntaparien väliset etäisyydet mahdollisimman hyvin tasolle





# Menetelmiä

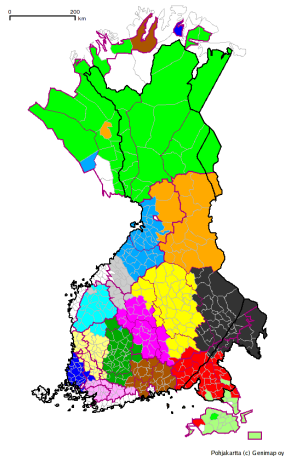
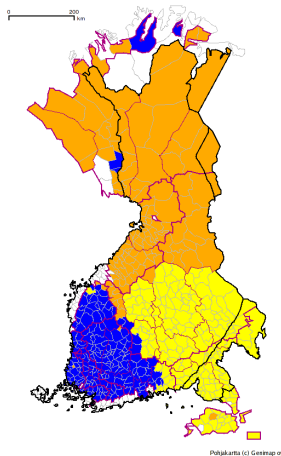
## Murrerajat

- Ryvästys: jaetaan pitäjät luokkiin
  - samaan rypäeseen kuuluvat mahdollisimman samanlaisia
  - eri rypäisiin kuuluvat mahdollisimman erilaisia
- Tätähän murteentutkijat ovat tehneet!
- Ei kumminkaan aina kauhean hyvä idea
  - murre-erot harvoin jyrkkiä
  - kummalle puolelle rajaa siirtymämurteistot?





# Menetelmiä Murrerajat





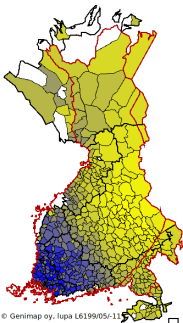
# Menetelmiä

## Murteiden diffuusio

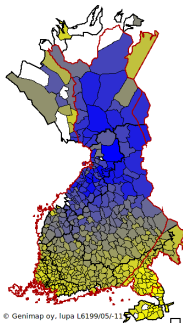
- Entäpä jos rajoja ei olisi?
  - ydinalueen ympärillä vähitellen laimenevaa vaikutusta
  - kuvaa paremmin siirtymämurteistoja
- Tilastotieteen monimuuttujamenetelmiä
  - yhdistellään piirteitä eri painokertoimilla
  - painotukset niin, että vaihtelu mahdollisimman selvää
- Tätäkin murteentutkijat tehneet!



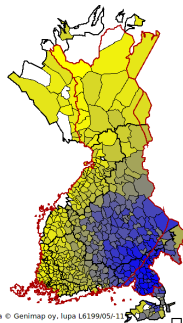
# Menetelmiä Murteiden diffuusio



Pohjakartta © Genimap oy, lupa L6199/05-11



Pohjakartta © Genimap oy, lupa L6199/05-11



Pohjakartta © Genimap oy, lupa L6199/05-11



# Murteita

## No mitä se tietokone tietää?

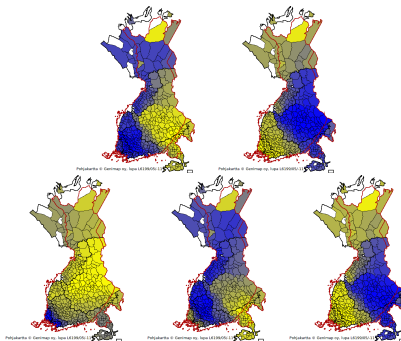
- Eri aineistoista saa vähän erilaisia tuloksia
  - sanastosta kolmijako itä-länsi-pohjoinen
  - äänne- ja muotopiirteistä perinteisemmin itä-länsi
- Eri menetelmillä saa vähän erilaisia tuloksia
  - ryvästyksellä saa rajoja näkyviin, jos niitä haluaa
  - komponenttianalyseillä näkee paremmin vähittäisen muutoksen
- Murteita on niin monta kuin haluaa nähdä
  - tarkastelutarkkuuden voi valita tarpeen mukaan



# Murteita

## Erilaiset päämurrejaot

- Suomen murteet III a:n aineistosta itä-länsi-jako

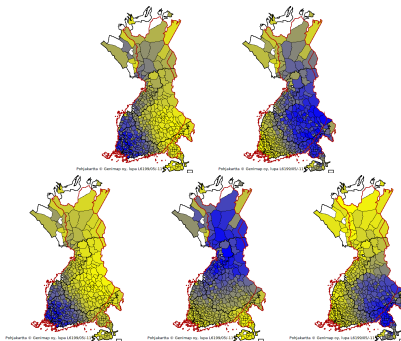




# Murteita

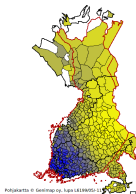
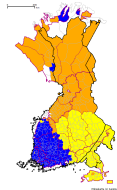
## Erilaiset päämurrejaot

- Murresanakirja-aineistosta pikemminkin itä-länsi-pohjoinen

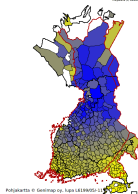




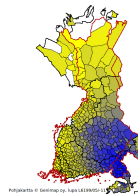
# Murteita Rajoja vai ei?



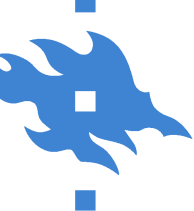
Pohjakkanta © Geonmap oy, Lupa LÄ198050-11



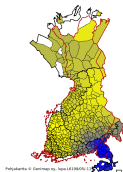
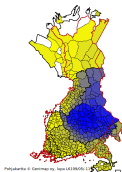
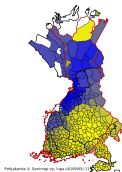
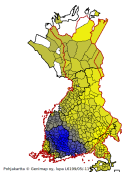
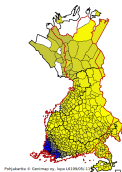
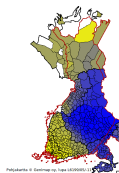
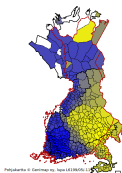
Pohjakkanta © Geonmap oy, Lupa LÄ198050-11



Pohjakkanta © Geonmap oy, Lupa LÄ198050-11



# Murteita Montako saisi olla?







# Murteita Lopuksi

- Murrevaihtelua saa esiin tietokoneella
  - kohtuullisen lähellä perinteistä näkemystä
  - etenkin, kun lähtöaineisto on sama
- Ei yhtä ainoaa totuutta
  - eroja eri aineistojen kesken
  - eroja eri menetelmien kesken
  - yllättyikö joku?

