



HELSINGIN YLIOPISTO  
HELSINGFORS UNIVERSITET  
UNIVERSITY OF HELSINKI

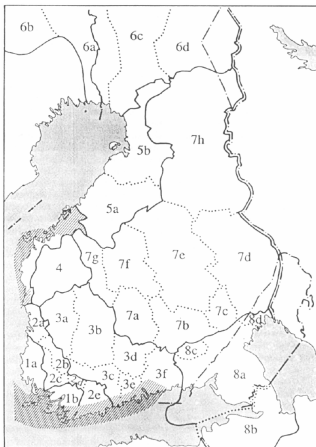
# Mitä murteita Suomessa onkaan?

Antti Leino <[antti.leino@cs.helsinki.fi](mailto:antti.leino@cs.helsinki.fi)>  
9. syyskuuta 2006

Tietojenkäsittelytieteen laitos  
Kotimaisten kielten tutkimuskeskus 



# Perinteinen murrejako



## ■ Länsimurteet

- Lounaismurteet
- Lounaiset välimurteet
- Hämäläismurteet
- Etelä-Pohjanmaan murre
- Keski- ja Pohjois-Pohjanmaan murteet
- Peräpohjolan murteet

## ■ Itämurteet

- Savolaismurteet
- Kaakkoismurteet



# Entäs sitten?

- »Pohja, jonka tähänastinen tieteellinen tutkimus on tarjonnut murteidemme suhteiden selvittelylle on toistaiseksi melko yksipuolisesti äännehistoriallinen. Kuitenkin voidaan ennen pitkää odottaa merkittäviä tuloksia myös murteiden sanavarojen vertailusta, johon toistaiseksi on ollut kovin vähän tilaisuutta.»  
— Martti Rapola, *Johdatus suomen murteisiin*
- Sanojen levikitietoja saatettu tietokonemuotoon Suomen murteiden sanakirjan toimitustyön yhteydessä



# Aineisto ja menetelmät

- Käytössä n. 5500 sanan levikit, yhteensä n. 9000 karttaa
  - Kustakin sanasta (tai sen erillisestä merkityksestä) ne kunnat, joista sana on kerätty
- Tiedon louhinta
  - Tietojenkäsittelytieteen osa-alue
  - Tavoitteena löytää uutta tietoa suuresta aineistomassasta
  - Vähän tai ei lainkaan ennakkokäsityksiä



# Murrevaihtelun jakaminen osiin

- Tarkoitus järjestää aineisto havainnollisemmin
- Jaetaan aineisto komponentteihin
- Käytettävissä vain tieto siitä, mitkä sanat esiintyvät missäkin kunnassa
- Ei kuntien sijaintitietoa
- Piirretään löydetyt komponentit kartalle

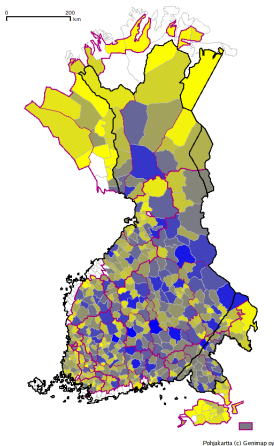


# Pääkomponenttianalyysi

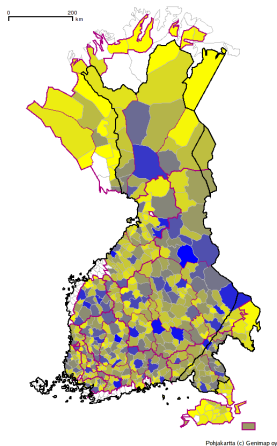
- Tulkitaan aineisto 9000-ulotteiseksi avaruudeksi
  - Kukin sana tämän avaruuden ulottuvuus
  - Kukin kunta piste tässä avaruudessa
- Etsitään voimakkain vaihtelu»suunta»
- Etsitään jäljellä olevasta vaihtelusta voimakkain »suunta»
- Jatketaan niin kauan kuin löytyy järkevänolaisia komponentteja



# 1. komponentti



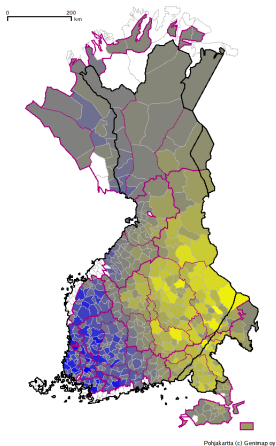
1. komponentti



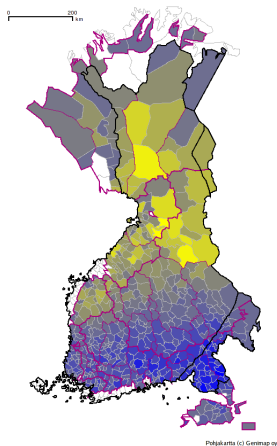
Sanojen lukumäärä



## 2.—3. komponentti



2. komponentti

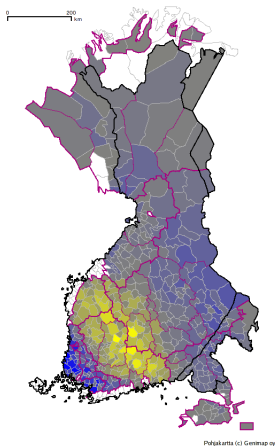


3. komponentti

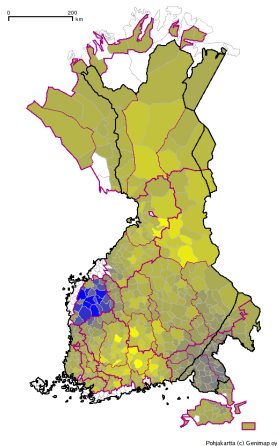




# 4.—5. komponentti



4. komponentti



5. komponentti

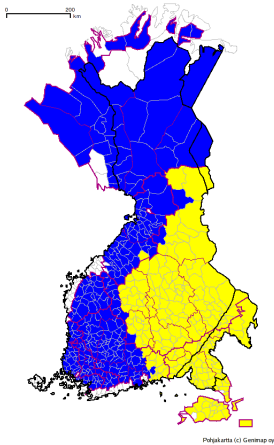


# Ryvästys

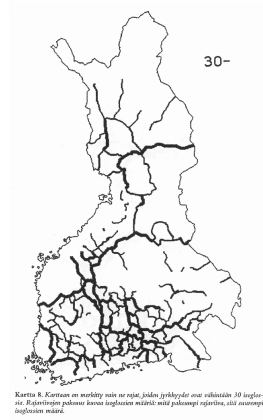
- Tarkoituksena jakaa aineisto rypäisiin:
  - samaan rypäeseen kuuluvat kunnat mahdollisimman samanlaisia
  - eri rypäisiin kuuluvat mahdollisimman erilaisia
- Lähtökohtana alkuperäinen aineisto tai muutama ensimmäinen pääkomponentti
  - Ensimmäisiin pääkomponentteihin keskittyminen »siivoaa» aineistosta satunnaista ja vähämerkityksistä vaihtelua
  - Tällä kertaa myös 1. komponentti mielenkiinnoton, joten jätetty pois



# 2 ryvästä



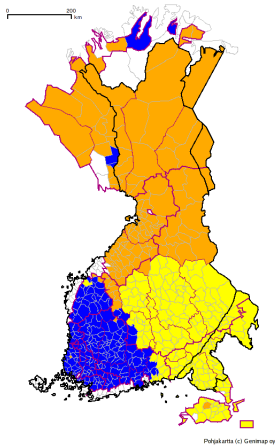
2 ryvästä



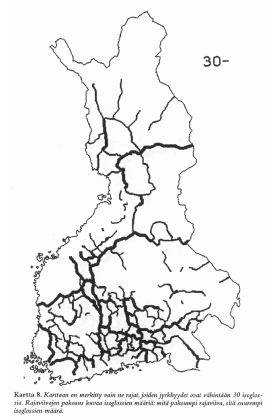
Jyrkimmäät murrerajat  
Wiikin (2004) mukaan



# 3 ryvästä



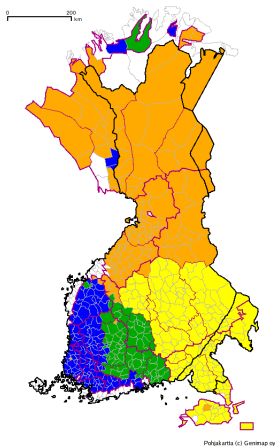
3 ryvästä



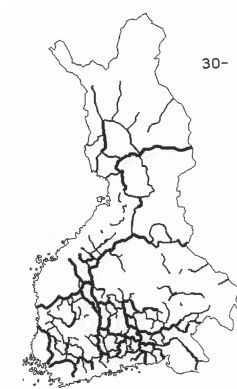
Jyrkimmät murrerajat  
Wiikin (2004) mukaan



# 4 ryvästä



4 ryvästä

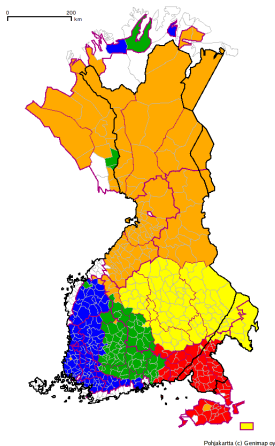


Kartta 8. Karttaan on merkitty ensin ne rajat, joiden jyrkkyydet ovat vähintään 30 asteita. Rajatietojen pohjana on Suomen laajuisen mittätkä: mitä pohjoisempi rajatieto, sitä suurempi laajuisien mittätkä.

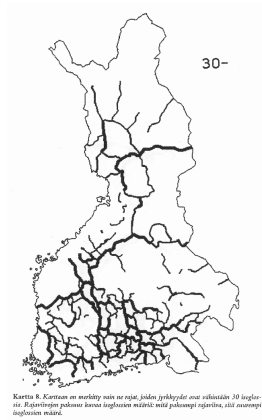
Jyrkimmät murrerajat  
Wiikin (2004) mukaan



# 5 ryvästä



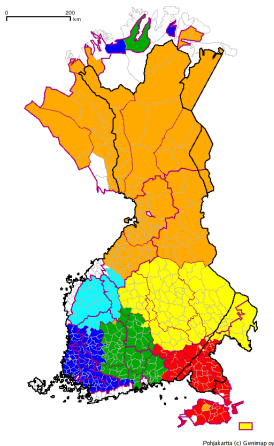
5 ryvästä



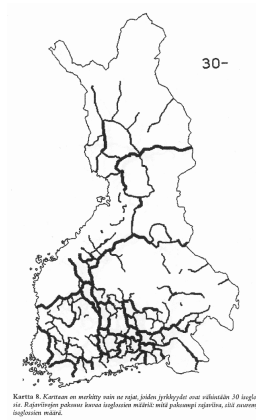
Jyrkimmät murrerajat  
Wiikin (2004) mukaan



# 6 ryvästä



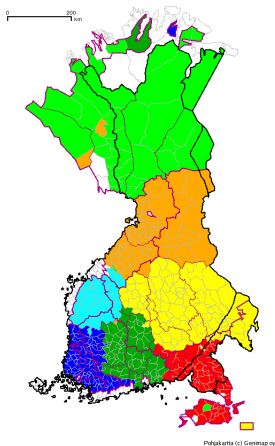
6 ryvästä



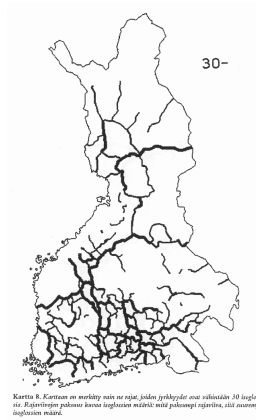
Jyrkimmät murrerajat  
Wiikin (2004) mukaan



# 7 ryvästä



7 ryvästä

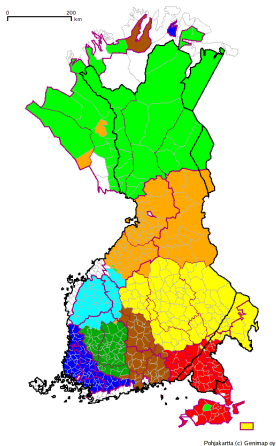


Jyrkimmät murrerajat  
Wiikin (2004) mukaan

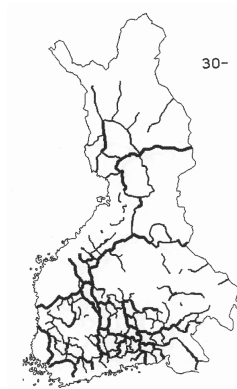




# 8 ryvästä



8 ryvästä

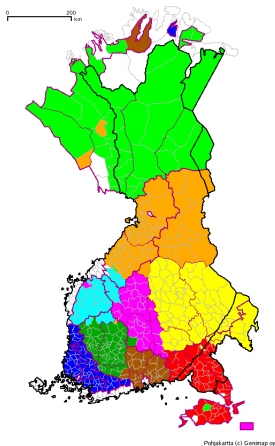


Kartta 8. Karttaan on merkitty vain ne rajat, joiden jyrkkyydet ovat vähintään 30 langla-  
sia. Rajatviivojen paksuus kasvaa langlaasin määrää: mitä paksuampi rajatieto, sitä suurempi  
langlaasin määrä.

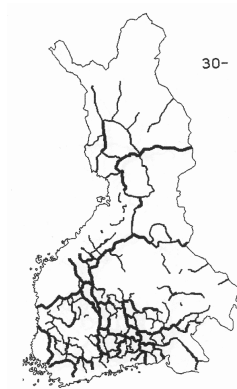
Jyrkimmät murrerajat  
Wiikin (2004) mukaan



# 9 ryvästä



9 ryvästä

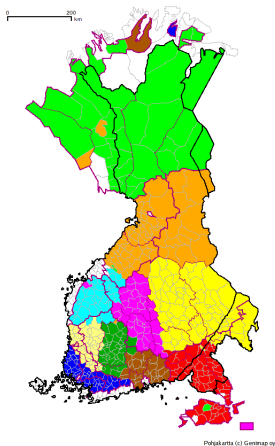


Kartta 8. Karttaan on merkitty vain ne rajat, joiden jyrkkyydet ovat vähintään 30 langla-  
sia. Rajatviivojen paksuus kasvaa langlausten määrää: mitä paksuampi rajatieto, sitä suurempi  
langlausten määrä.

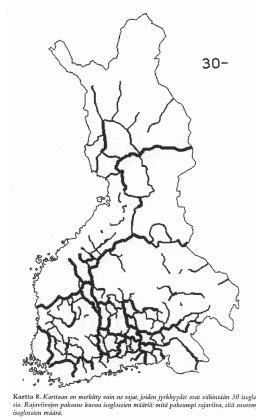
Jyrkimmät murrerajat  
Wiikin (2004) mukaan



# 10 ryvästä



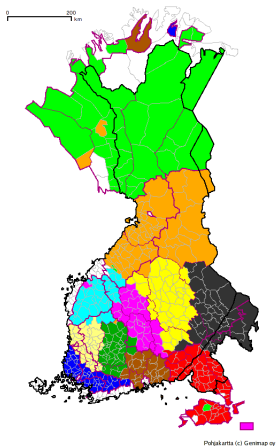
10 ryvästä



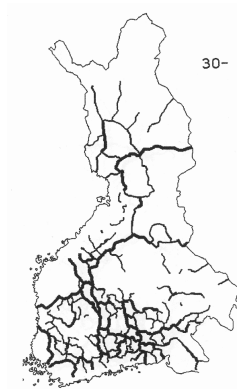
Jyrkimmät murrerajat  
Wiikin (2004) mukaan



# 11 ryvästä



11 ryvästä

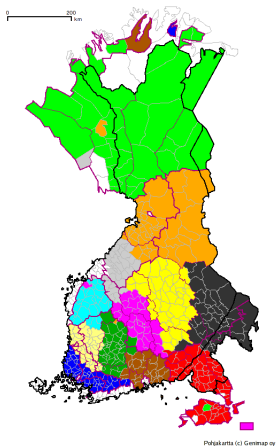


Kartta 8. Karttaan on merkitty vain ne rajat, joiden jyrkkyydet ovat vähintään 30 langla-  
sia. Rajatietojen pohjana on ollut langlaisten mittätkä: mitä pohosampi rajatieto, sitä suurempi  
langlaisten määrä.

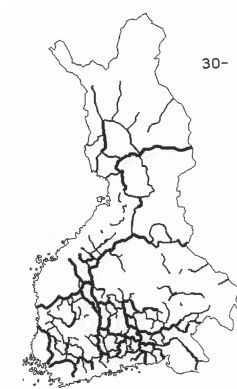
Jyrkimmät murrerajat  
Wiikin (2004) mukaan



# 12 ryvästä



12 ryvästä

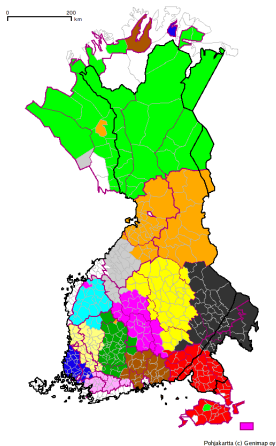


Kartta 8. Karttaan on merkitty vain ne rajat, joiden jyrkkyydet ovat vähintään 30 langla-  
sia. Rajatviivojen paksuus kasvaa langlausten määrää: mitä paksuampi rajatieto, sitä runsaampi  
langlausten määrä.

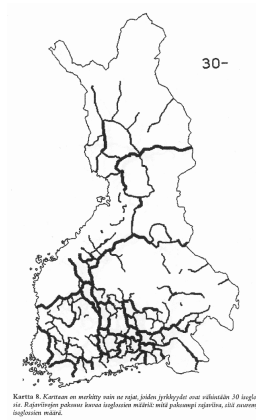
Jyrkimmät murrerajat  
Wiikin (2004) mukaan



# 13 ryvästä



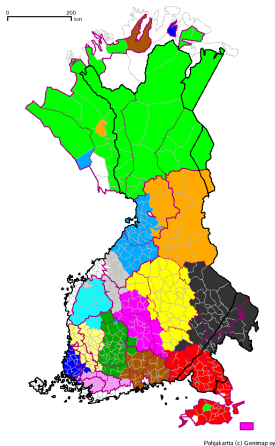
13 ryvästä



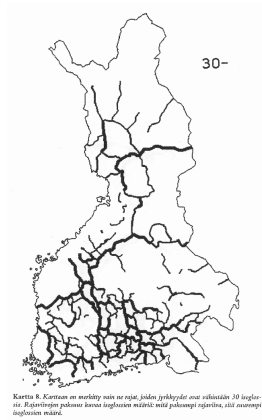
Jyrkimmät murrerajat  
Wiikin (2004) mukaan



# 14 ryvästä



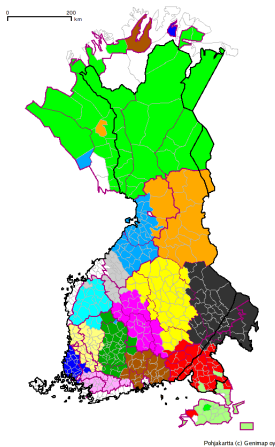
14 ryvästä



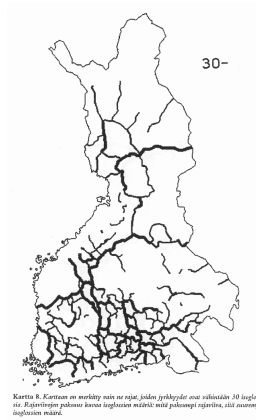
Jyrkimmät murrerajat  
Wiikin (2004) mukaan



# 15 ryvästä



15 ryvästä



Jyrkimmät murrerajat  
Wiikin (2004) mukaan



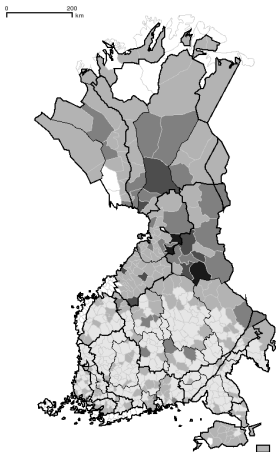


# Riippumattomien komponenttien analyysi

- Etsitään taustalla vaikuttavia tekijöitä, jotka toisistaan riippumattomia
- Komponentit eivät järjestyksessä
- Ei itsestään selvää »oikeaa» lukumäärää
- Mahdollista tulkita eri murteiksi

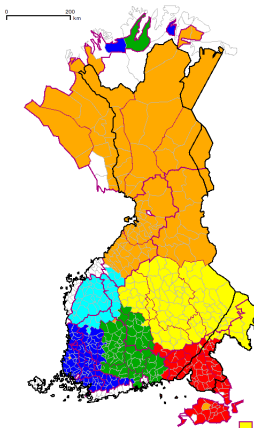


# 1. komponentti



Pohjääkartta © Geomap oy, loppu L6199/05-11

1. komponentti

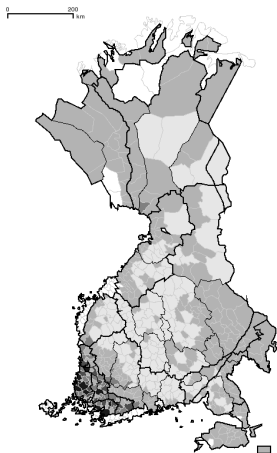


Pohjääkartta (c) Geomap oy

6 ryvästä

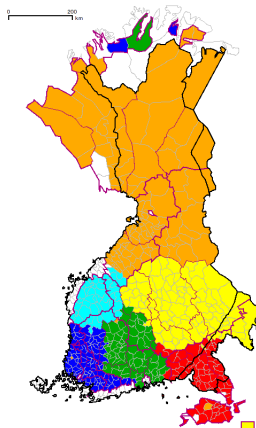


## 2. komponentti



Pohjäkarta © Geomap oy, kupa L6199/05-11

2. komponentti

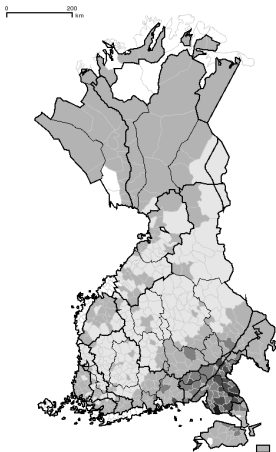


Pohjäkarta © Geomap oy

6 ryvettä

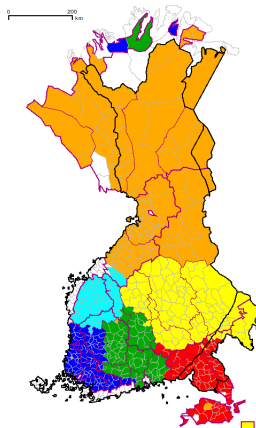


# 3. komponentti



Pohjäkarta © Geomap oy, kupa L6199/05-11

3. komponentti

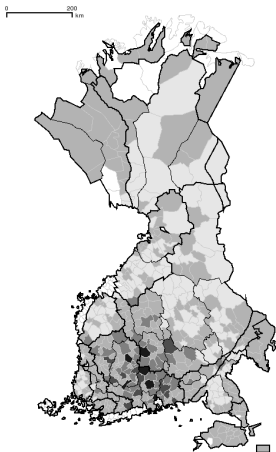


Pohjäkarta © Geomap oy

6 ryvästä

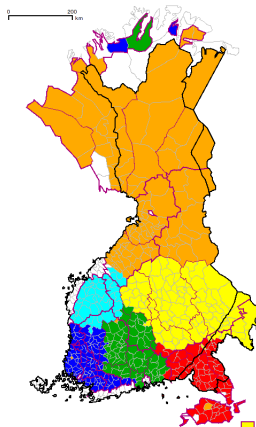


# 4. komponentti



Pohjääkartta © Geomap oy, kupa L6199/05-11

4. komponentti

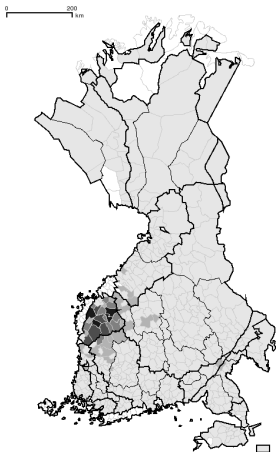


Pohjääkartta © Geomap oy

6 ryvästä

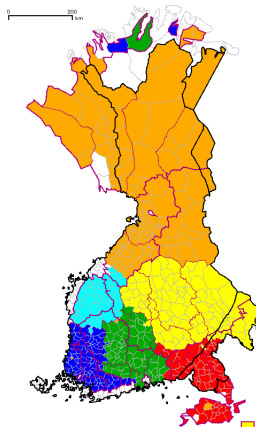


# 5. komponentti



Pohjääkarta © Geomap oy, kupa L6199/05-11

5. komponentti

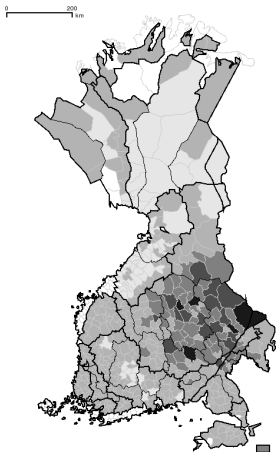


Pohjääkarta © Geomap oy

6 ryvystä

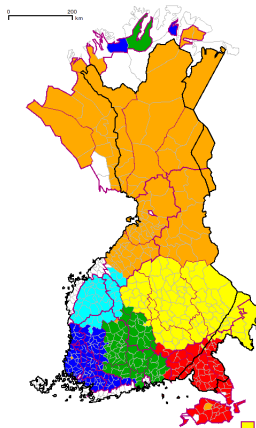


# 6. komponentti



Pohjääkartta © Geomap oy, kupa L6199/05-11

6. komponentti



Pohjääkartta © Geomap oy

6 ryvästä



# Siis mitä?

- Joukko muutoksia perinteiseen murrejakoon
  - Kahden sijasta kolme päämurrealuetta: itä-, länsi- ja pohjoismurteet
    - Kainuun murteet mieluummin pohjois- kuin itämurteita
  - Pohjoispohjalaisten murteiden rajat sekä Keski-Pohjanmaan että Kainuun suuntaan eri kohdassa kuin perinteisesti
  - Ei yhtenäistä lounaisten välimurteiden aluetta
- Rajojen vetäminen ei välttämättä mielekästä
  - Komponentit antavat ehkä todenmukaisemman kuvan vaihtelun luonteesta