

# Ahvenlammen lähellä on yleensä Haukilampi

näkökulma lähekkäisten vedenkokoumien nimeämiseen

<http://www.cs.helsinki.fi/u/leino/jutut/ktp-03/>

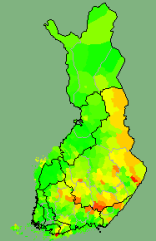
Antti Leino



[leino@cs.helsinki.fi](mailto:leino@cs.helsinki.fi)

[aleino@kotus.fi](mailto:aleino@kotus.fi)

13. toukokuuta 2003



**Ahvenlammen  
lähellä on yleensä  
Haukilampi**

Antti Leino  
Kielitieteen päivät 2003

*Johdanto*

*Kaksi nimiparia*

*Muita havaintoja*

*Lopuksi*

*Lähteet*



Sivu 0 / 6

*Takaisin*

*Koko ruutu*

*Sulje*

*Lopeta*

# Johdanto

## Maanmittauslaitoksen paikannimirekisteri

Leskinen (2002)

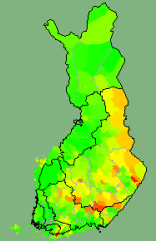
- Koko peruskartan nimistö
- 303 626 suomenkielistä nimeä, 717 747 esiintymää
- 25 178 vedenkokouman nimeä, 58 267 esiintymää
- 54 yleisintä vedenkokouman nimeä:  
kutakin nimeä vähintään 90, yhteensä 9 008 esiintymää

## Tietojenkäsittelytieteen menetelmät

- Tietämyksen muodostaminen, data-analyysi:  
miten etsiä hyödyllistä uutta tietoa suurista aineistoista
- HY:n tietojenkäsittelytieteen laitoksen vankka tutkimusperinne
- Paikkatiedon käsittelyä tutkittu verraten vähän

## Aiempaa nimistöntutkimusta

- Vastakohta- ja variointinimien tematiikkaa: [Kiviniemi \(1971\)](#)
- Nimien toistumista ja läheisyyttä: [Santakivi \(1979\)](#), [Eskelinen \(2002\)](#)
- Ei systemaattista tutkimusta nimi(esiintymi)en läheisyydestä



Ahvenlammien  
lähellä on yleensä  
**Haukilampi**  
Antti Leino  
Kielitieteen päivät 2003

Johdanto

Kaksi nimiparia

Muita havaintoja

Lopuksi

Lähteet



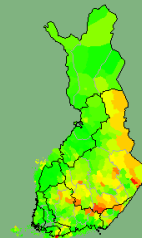
Sivu 1 / 6

Takaisin

Koko ruutu

Sulje

Lopeta



## Ahvenlammen lähellä on yleensä Haukilampi

Antti Leino

Kielitieteen päivät 2003

Johdanto

Kaksi nimiparia

Muita havaintoja

Lopuksi

Lähteet



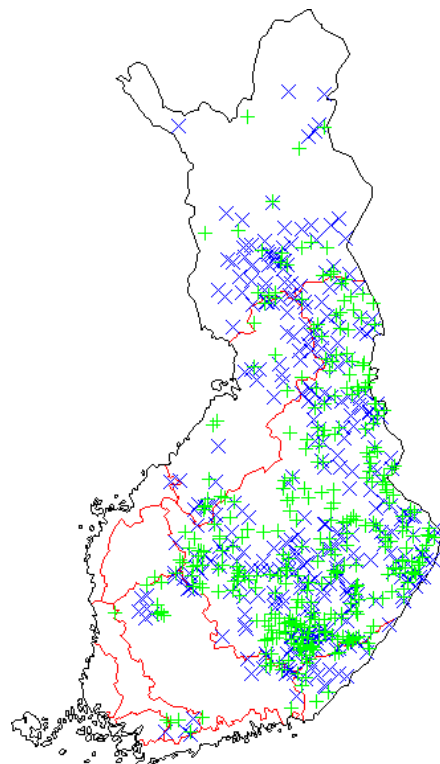
Sivu 2 / 6

Takaisin

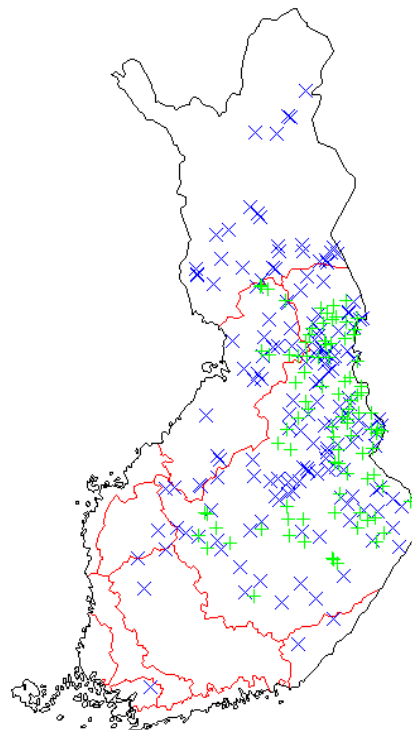
Koko ruutu

Sulje

Lopeta



A. Ahvenlampi (x)  
ja Haukilampi (+)



B. Hanhilampi (x)  
ja Joutenlampi (+)

Kuva 1: Kahden nimiparin levikkikartat

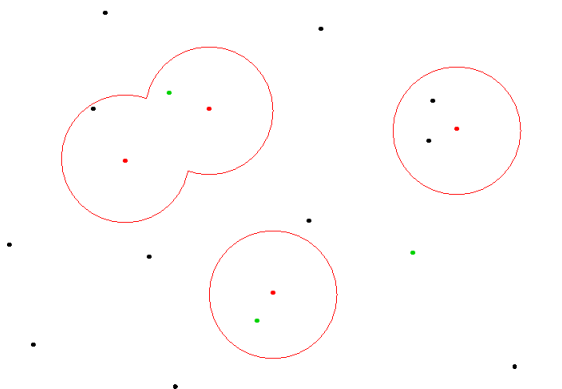
# Kaksi nimiparia

## Ensivaikutelma

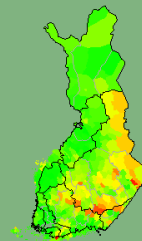
- Nimien välillä samantapainen merkityssuhde
- Levikkien suhde silmämääräisesti osapuilleen samantapainen

## Assosiaatiosääntö

- Tietojenkäsittelytieteen käsite (esim. [Mannila — Toivonen 2002](#))
- Jos  $A$  esiintyy, niin  $B$  normaalia todennäköisempi
- Paikannimien tapaus: verrataan  $B$ :n esiintymätiheyksiä toisaalta  $A$ :n lähistöllä, toisaalta koko aineistossa



Kuva 2: Assosiaatiosääntöjen etsintä etäisyyksien perusteella



Ahvenlampen  
lähellä on yleensä  
Haukilampi  
Antti Leino  
Kielitieteen päivät 2003

Johdanto

Kaksi nimiparia

Muita havaintoja

Lopuksi

Lähteet



Sivu 3 / 6

Takaisin

Koko ruutu

Sulje

Lopeta

## Ahvenlampi—Haukilampi

- Nollahypoteesi: esiintyvät toisistaan riippumatta
- Tällöin *Ahvenlampien* lähistöllä *Haukilampia* suunnilleen yhtä tiheässä kuin koko maassa ylipäänsä
- Näin ei ole, vaan nimien esiintymien välillä on tilastollisesti merkitsevä yhteys
- *Ahvenlampi*  $\Rightarrow$  *Haukilampi* ja *Hanhilampi*  $\Rightarrow$  *Joutenlampi* erilaiset

Ahvenlampi => Haukilampi:

+ Säde 1 km: 20;  $p(n<20) = 1.0000$  (korj. 1.00)  
+ Säde 2 km: 40;  $p(n<40) = 1.0000$  (korj. 1.00)  
+ Säde 3 km: 51;  $p(n<51) = 1.0000$  (korj. 0.99)  
+ Säde 4 km: 75;  $p(n<75) = 1.0000$  (korj. 1.00)  
+ Säde 5 km: 92;  $p(n<92) = 1.0000$  (korj. 0.97)  
+ Säde 6 km: 116;  $p(n<116) = 1.0000$  (korj. 0.98)  
+ Säde 7 km: 137;  $p(n<137) = 1.0000$  (korj. 0.95)  
+ Säde 8 km: 170;  $p(n<170) = 1.0000$  (korj. 1.00)  
+ Säde 9 km: 181;  $p(n<181) = 1.0000$  (korj. 0.96)  
+ Säde 10 km: 204;  $p(n<204) = 1.0000$  (korj. 0.98)  
Säde 15 km: 263;  $p(n<263) = 0.9991$  (korj. 0.00)  
Säde 20 km: 301;  $p(n<301) = 0.9970$  (korj. 0.00)  
Säde 25 km: 326;  $p(n<326) = 0.9954$  (korj. 0.00)  
Säde 30 km: 335;  $p(n<335) = 0.9831$  (korj. 0.00)  
Säde 35 km: 340;  $p(n<340) = 0.9588$  (korj. 0.00)  
Säde 40 km: 344;  $p(n<344) = 0.9308$  (korj. 0.00)  
Säde 45 km: 345;  $p(n<345) = 0.8857$  (korj. 0.00)  
Säde 50 km: 345;  $p(n<345) = 0.8226$  (korj. 0.00)

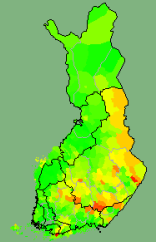
Hanhilampi => Joutenlampi:

Säde 1 km: 0;  $p(n>0) = 0.2091$  (korj. 0.00)  
Säde 2 km: 3;  $p(n<3) = 0.9259$  (korj. 0.00)  
Säde 3 km: 3;  $p(n<3) = 0.6418$  (korj. 0.00)  
Säde 4 km: 5;  $p(n<5) = 0.6983$  (korj. 0.00)  
Säde 5 km: 9;  $p(n<9) = 0.8927$  (korj. 0.00)  
Säde 6 km: 18;  $p(n<18) = 0.9990$  (korj. 0.00)  
Säde 7 km: 21;  $p(n<21) = 0.9985$  (korj. 0.00)  
+ Säde 8 km: 31;  $p(n<31) = 1.0000$  (korj. 0.98)  
Säde 9 km: 33;  $p(n<33) = 1.0000$  (korj. 0.91)  
Säde 10 km: 37;  $p(n<37) = 1.0000$  (korj. 0.91)  
+ Säde 15 km: 60;  $p(n<60) = 1.0000$  (korj. 0.99)  
Säde 20 km: 73;  $p(n<73) = 1.0000$  (korj. 0.86)  
Säde 25 km: 83;  $p(n<83) = 0.9997$  (korj. 0.14)  
Säde 30 km: 90;  $p(n<90) = 0.9991$  (korj. 0.00)  
Säde 35 km: 93;  $p(n<93) = 0.9965$  (korj. 0.00)  
Säde 40 km: 93;  $p(n<93) = 0.9838$  (korj. 0.00)  
Säde 45 km: 93;  $p(n<93) = 0.9609$  (korj. 0.00)  
Säde 50 km: 93;  $p(n<93) = 0.9348$  (korj. 0.00)

Taulukko 1: Nimiparien välinen assosiaatio eri etäisyyksillä

## Johtopäätös

- *Ahvenlampi—Haukilampi* saman nimiaiheen variointia
- *Hanhilampi—Joutenlampi* selittyä levikkien päällekkäisyydellä



Ahvenlammen  
lähellä on yleensä  
Haukilampi

Antti Leino

Kielitieteen päivät 2003

Johdanto

Kaksi nimiparia

Muita havaintoja

Lopuksi

Lähteet



Sivu 4 / 6

Takaisin

Koko ruutu

Sulje

Lopeta

# Muita havaintoja

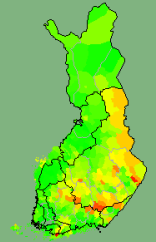
Esimerkkihavaintoja 54 yleisimmän järven- ja lammennimen aineistosta:

- *Mustalampi* ⇒ *Valkealampi* ja toisin päin: odotuksenmukainen, mutta viite menetelmän toimivuudesta
- *Lehmilampi* ⇒ *Likolampi* ja päin vastoin: voimakas yhteys, muttei variointia. Kulttuuriin perustuva yhteys?
- *Likolampi* ⇒ *Pitkälampi*, *Likolampi* ⇒ *Valkealampi* ja päin vastoin: Voimakkaita assosiaatiosääntöjä. Miksi?
- *Umpilampi* ⇒ *Umpilampi* voimakkaasti merkitsevänä lyhyillä etäisyyksillä; selvästi kään ei koko kylän käytössä

## Lopuksi

Lyhytkin tarkastelu paljasti mielenkiintoisia asioita. Entä jatko?

- vastaavanlainen koko vesistönimistöä ja myöhemmässä vaiheessa koko peruskartan nimistöä koskeva tarkastelu
- useamman kuin kahden nimen välisten riippuvuuksien analysointi
- näiden tulosten selittäminen semanttisin, kulttuuri- ja asutushistoriallisin, luonnonmaantieteellisin tai muin perustein
- nimeämistä ja mahdollisesti myös nimitysologiaa koskevien onomastisten teorioiden sovittaminen yhteen tulosten kanssa



Ahvenlammien  
lähellä on yleensä  
**Haukilampi**  
Antti Leino  
Kielitieteen päivät 2003

Johdanto

Kaksi nimiparia

Muita havaintoja

Lopuksi

Lähteet



Sivu 5 / 6

Takaisin

Koko ruutu

Sulje

Lopeta

## Lähteet

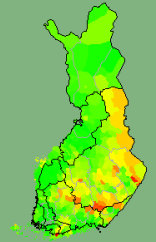
Eskelinen, Riikka 2002: Paikannimien ja määriteosien yleisyys ja toistuminen Tervon kunnassa. Praktikumtyö, Helsingin yliopisto.

Kiviniemi, Eero 1971: Vastakohta- ja variointinimistä. *Virittäjä* s. 123—134.

Leskinen, Teemu 2002: The geographic names register of the National Land Survey of Finland. *Eighth United Nations Conference on the Standardization of Geographical Names*.

Mannila, Heikki — Toivonen, Hannu 2002: Knowledge discovery in databases: The search for frequent patterns. Tietämyksen muodostaminen -kurssin luentomoniste. Helsingin yliopiston tietojenkäsittelytieteen laitos.

Santakivi, Pekka 1979: Paikannimien toistuminen Hauholla, Lammilla ja Tuuloksessa. Pro gradu -tutkielma, Helsingin yliopisto.



**Ahvenlampen  
lähellä on yleensä  
Haukilampi**  
Antti Leino  
Kielitieteen päivät 2003

Johdanto

Kaksi nimiparia

Muita havaintoja

Lopuksi

Lähteet



Sivu 6 / 6

Takaisin

Koko ruutu

Sulje

Lopeta