

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

aleino@kotus.fi

13. toukokuuta 2003

Tiivistelmä

Tämä pieni artikkeli on varsinaisesti tarkoitettu rungoksi Kielitieteen päivillä 2003 pitämäni esitelmää varten. Aikomukseni on esitellä lyhyesti, millaisia paikannimien esiintymien välisiä maantieteellisiä suhteita Maanmittauslaitoksen paikannimirekisterin avulla voi löytää. Vaikka osa havainnoista on odotuksenmukaisia, aineistosta paljastuu myös muutama mielenkiintoisten uusien tapausten joukko; keskitynkkin tarkastelemaan yhtä tällaista tapausta. Lopuksi hahmotelen aiheita mahdolliselle tarkemmalle tutkimukselle.

Sisältö

1	Johdanto	1
2	Kaksi nimiparia	1
3	Muita havaintoja	3
4	Lopuksi	4
	Lähteet	4

1 Johdanto

Lähtökohtana tämän tutkimusaiheen valinnalle on Maanmittauslaitoksen paikannimirekisteri (Leskinen 2002), joka luovutettiin alkuvuodesta 2001 Kotukseen nimistönhuolto- ja -tutkimuskäyttöön.

- 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

Tällaiseen laajahkoon aineistoon tuntuu luontevalta yrittää soveltaa tietojenkäsittelytieteessä kehitettyjä suurten aineiston analysointimenetelmiä. Helsingin yliopiston tietojenkäsittelytieteen laitoksessa elää voimakas tietämyksen muodostamisen tutkimusperinne, joten yhteistyölle on kohtuullisen hyvä pohja. Aihe on myös sikäli mielenkiintoinen, että paikatiedon käsittelyyn soveltuvia menetelmiä on tutkittu maailmalla suhteellisen vähän.

Paikannimistön tutkimuksessa on jonkin verran tarkasteltu vastakohta- ja variointinimien tematiikkaa (esim. Kiviniemi 1971). Samoin nimien toistumisesta ja läheisyydestä on tehty jokunen opinnäyte (Santakivi 1979, Eskelinen 2002). Systemaattista tutkimusta paikannimien tai nimiesiintymien läheisyydestä ei kuitenkaan ole tähän mennessä tehty.

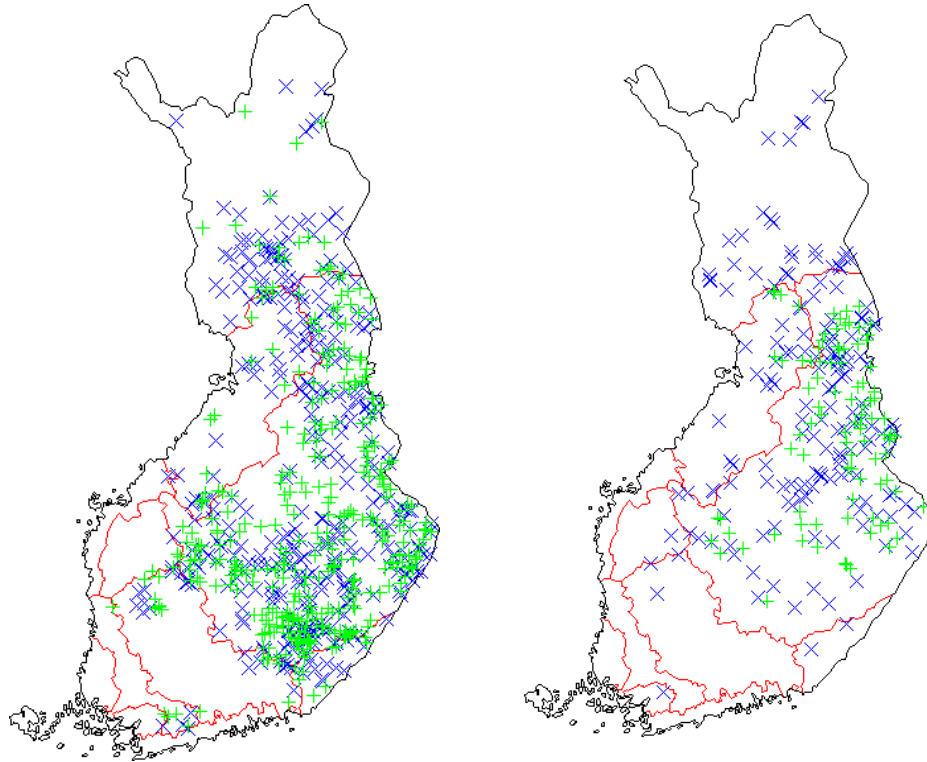
2 Kaksi nimiparia

Kuvassa 1 on kahden nimiparin levikkikartat. Ensivaikutelma on, että nimipareissa *Ahvenlampi—Haukilampi* ja *Hanhilampi—Joutenlampi* nimien välillä on samantapainen merkityssuhde. Myös levikkikarttojen mukaan tapaukset näyttävät suhteellisen samanlaisilta, vaikkakin edellinen pari on selvästi jälkimmäistä yleisempi.

Silmämääräisesti on kuitenkin vaikea arvioida esiintymien maantieteellisiä riippuvuussuhteita. Otetaan avuksi **assosiaatiosäännön** käsite (esim. Mannila — Toivonen 2002): assosiaatiosääntö on sääntö, joka kertoo, että jonkin ilmiön olemassaolosta voidaan päätellä, että jonkin toisen ilmiön esiintymistodennäköisyys on normaalia suurempi. Paikannimien tapauksessa kysymystä voidaan lähestyä kuvan 2 ja kaavan 1 avulla.

$$X \sim \text{Poisson}\left(n(C_B) \frac{N(B)}{N(C_B)}\right) \quad (1)$$

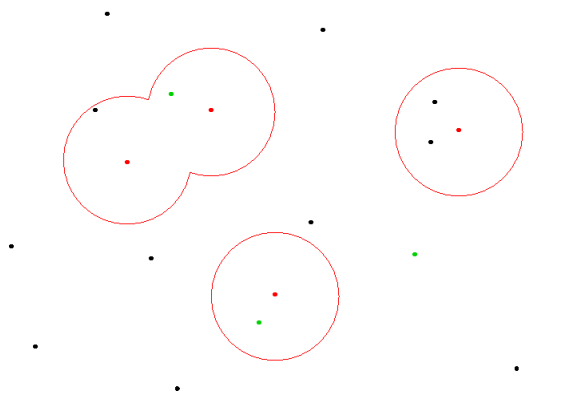
Jos *Ahvenlampi* ja *Haukilampi* esiintyvät toisistaan riippumatta, *Haukilampia* pitäisi löytyä *Ahvenlampien* lähistöllä suunnilleen yhtä tiheässä kuin koko maassa ylipäänsä, ja tässä tapauksessa myös kaavasta 1 lasketun *Haukilampien* lukumäärän pitäisi olla lähellä todellista.



A. Ahvenlampi (x)
ja Haukilampi (+)

B. Hanhilampi (x)
ja Joutenlampi (+)

Kuva 1: Kahden nimiparin levikkikartat



Kuva 2: Assosiaatiosääntöjen etsintä etäisyyksien perusteella

Taulukosta 1 näkyy, että assosiaatiosäännöt *Ahvenlampi* \Rightarrow *Haukilampi* ja *Hanhilampi* \Rightarrow *Joutenlampi* ovat tilastollisesti merkitsevinä voimassa hyvin erilaisilla etäisyyksillä. Edellinen sääntö pätee lyhyillä, jo alle 1 km etäisyyksillä; jälkimmäinen on voimassa vasta selvästi pidemmillä, noin 10–20 km etäisyyksillä.

Ahvenlampi => Haukilampi:	Hanhilampi => Joutenlampi:
+ Säde 1 km: 20; p(n<20) = 1.0000 (korj. 1.00)	Säde 1 km: 0; p(n>0) = 0.2091 (korj. 0.00)
+ Säde 2 km: 40; p(n<40) = 1.0000 (korj. 1.00)	Säde 2 km: 3; p(n<3) = 0.9259 (korj. 0.00)
+ Säde 3 km: 51; p(n<51) = 1.0000 (korj. 0.99)	Säde 3 km: 3; p(n<3) = 0.6418 (korj. 0.00)
+ Säde 4 km: 75; p(n<75) = 1.0000 (korj. 1.00)	Säde 4 km: 5; p(n<5) = 0.6983 (korj. 0.00)
+ Säde 5 km: 92; p(n<92) = 1.0000 (korj. 0.97)	Säde 5 km: 9; p(n<9) = 0.8927 (korj. 0.00)
+ Säde 6 km: 116; p(n<116) = 1.0000 (korj. 0.98)	Säde 6 km: 18; p(n<18) = 0.9990 (korj. 0.00)
+ Säde 7 km: 137; p(n<137) = 1.0000 (korj. 0.95)	Säde 7 km: 21; p(n<21) = 0.9985 (korj. 0.00)
+ Säde 8 km: 170; p(n<170) = 1.0000 (korj. 1.00)	+ Säde 8 km: 31; p(n<31) = 1.0000 (korj. 0.98)
+ Säde 9 km: 181; p(n<181) = 1.0000 (korj. 0.96)	Säde 9 km: 33; p(n<33) = 1.0000 (korj. 0.91)
+ Säde 10 km: 204; p(n<204) = 1.0000 (korj. 0.98)	Säde 10 km: 37; p(n<37) = 1.0000 (korj. 0.91)
Säde 15 km: 263; p(n<263) = 0.9991 (korj. 0.00)	+ Säde 15 km: 60; p(n<60) = 1.0000 (korj. 0.99)
Säde 20 km: 301; p(n<301) = 0.9970 (korj. 0.00)	Säde 20 km: 73; p(n<73) = 1.0000 (korj. 0.86)
Säde 25 km: 326; p(n<326) = 0.9954 (korj. 0.00)	Säde 25 km: 83; p(n<83) = 0.9997 (korj. 0.14)
Säde 30 km: 335; p(n<335) = 0.9831 (korj. 0.00)	Säde 30 km: 90; p(n<90) = 0.9991 (korj. 0.00)
Säde 35 km: 340; p(n<340) = 0.9588 (korj. 0.00)	Säde 35 km: 93; p(n<93) = 0.9965 (korj. 0.00)
Säde 40 km: 344; p(n<344) = 0.9308 (korj. 0.00)	Säde 40 km: 93; p(n<93) = 0.9838 (korj. 0.00)
Säde 45 km: 345; p(n<345) = 0.8857 (korj. 0.00)	Säde 45 km: 93; p(n<93) = 0.9609 (korj. 0.00)
Säde 50 km: 345; p(n<345) = 0.8226 (korj. 0.00)	Säde 50 km: 93; p(n<93) = 0.9348 (korj. 0.00)

Taulukko 1: Nimiparien välinen assosiaatio eri etäisyyksillä

Ilmeisesti parissa *Ahvenlampi*—*Haukilampi* nimenmuodostus on perustunut varsin voimakkaasti saman nimiaiheen variointiin: esiintymiä on niin säännönmukaisesti lähellä toisiaan, että toisenlaiseen tulkintaan on vaikea päätyä. Sen sijaan parissa *Hanhilampi*—*Joutenlampi* esiintymien välinen assosiaatiosääntö selittyy luontevasti jo levikkien päällekkäisyydellä.

3 Muita havaintoja

Edellä esiteltyjen kahden nimiparin lisäksi 54 yleisimmän järven- ja lammennimen aineiston tarkastelu paljasti muutamia muitakin mielenkiintoisia tapauksia. Tässä on esimerkinomainen lajitelma.

- *Mustalampi* \Rightarrow *Valkealampi* ja toisin päin: tämä tapaus oli odotuksenmukainen, mutta sen löytymistä voinee pitää viitteenä menetelmän toimivuudesta.
- *Lehmilampi* \Rightarrow *Likolampi* ja päin vastoin: yhteys oli verraten voimakas, mutta selvästikään tässä ei ole ainakaan kyse varioinnin käytöstä nimenmuodostuskeinona. Ehkäpä taustalla on jonkinlainen kulttuuriin perustuva yhteys?
- *Likolampi* \Rightarrow *Pitkälampi*, *Likolampi* \Rightarrow *Valkealampi* ja päin vastoin: jälleen suhteellisen voimakkaita assosiaatiosääntöjä, mutta ilmi-selvää syytä on vaikea keksiä.
- *Umpilampi* \Rightarrow *Umpilampi* voimakkaasti merkitsevänsä jo 2 km etäisyydellä, ja alle 1 km etäisyyksilläkin jo lähellä merkitsevyyden rajaa. Selvästikään merkittävä osa *Umpilampi*-nimistä ei ole koko kylän käytössä.

4 Lopuksi

Lyhyt katsaus jo yleisimpien järven- ja lammennimien keskinäisiin läheisyysuhteisiin toi esiin muutaman mielenkiintoiselta näyttävän havainnon. Lisätutkimuksiin on selvästi aihetta, ja mahdollisia eri haaroja tulee helposti mieleen ainakin

- 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 sekä
- nimeämistä ja mahdollisesti myös nimitypologiaa koskevien onomastisten teorioiden sovittaminen yhteen tulosten kanssa.

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.