

Suomen suomenkielisten paikannimien alueellinen variaatio

Emilia Virtanen

Pro gradu -tutkielma

Kieliasiantuntijuuden tutkinto-ohjelma, suomen kieli

Kieli- ja käännöstieteiden laitos

Humanistinen tiedekunta

Turun yliopisto

Huhtikuu 2025

Turun yliopiston laatu järjestelmän mukaisesti tämän julkaisun alkuperäisyys on tarkastettu

Turnitin OriginalityCheck –järjestelmällä

Pro gradu -tutkielma

Kieliasiantuntijuuden tutkinto-ohjelma, suomen kieli

Emilia Virtanen

Suomen suomenkielisten paikannimien alueellinen variaatio

Sivumäärät: 66, 9

Tämän tutkielman aiheena on paikannimien alueellinen variaatio. Tavoitteena on selvittää, millaisia eri murrealueiden paikannimet ovat ja millaisen ennustuksen paikannimen sijaintialueesta voi tehdä ohjatun koneoppimisen menetelmällä. Menetelminä ovat perinteinen nimien syntaktis-semanttinen analyysi määrälliseen aineistoon soveltaen sekä tilastolliset menetelmät ja *satunnaismetsä*-niminen ohjatun koneoppimisen menetelmä.

Tutkielman aineisto on paikannimirekisteristä ja kattaa koko Suomen. Koko aineistossa on yli 700 000 paikannimeä. Nimistä on erotettu puoliautomaattisesti viimeinen leksikaalinen osa ja nimenalkuja, ja leksikaalisesti moniosaisien nimien yleisimpien piirteiden jakautumista on tarkasteltu murrealueittain ja paikanlajeittain. Leksikaalisesti moniosaisista maasto- ja vesistönimistä koostuvaan murrealueittain tasattuun 108 000 nimen aineistoon pohjautuen on lisäksi luotu ohjatun koneoppimisen menetelmällä luokittelumalleja.

Paikannimet eroavat tulosten perusteella toisistaan murrealueittain, mutta alueilla on paljon myös yhteisiä piirteitä. Viiden yleisimmän viimeisen leksikaalisen osan tarkastelu paljasti vaihtelua paitsi esiintymistaajuudessa myös viimeisen osan funktiossa. Vaihtelu liittyy alueiden luonnonmaantieteeseen ja kielenpiirteisiin. Paikanlajeittaisessa tarkastelussa selvisi, että joillain yleisimmillä osilla näyttäisi toisilla alueilla korostuvan paikan lajiin viittaamisen funktio. Murrealueittain yleisimmistä kolmen merkin nimenaluista löytyi aikaisemmasta tutkimuksesta tuttua alueellista vaihtelua. Vaihtelu liittyy toisaalta kieleen ja toisaalta erilaisiin nimeämisaiheisiin.

Luokittelumalleista saatujen tulosten perusteella nimet muodostavat murrealueittaisia ryhmittymiä. Etenkin pohjoisin Suomi erottuu johdonmukaisesti kaikissa malleissa muista alueista. Sekä paikanlajeittain ryhmitellyt viimeiset leksikaaliset osat että nimenalut ovat merkityksellisiä sijaintialueen ennustamisessa. Kaikkein tärkeimpiä mallien tekemien ennustusten kannalta ovat kohoumien nimissä esiintyvät viimeiset leksikaaliset osat. Erilaisilla otoksilla on vaikutusta tuloksiin etenkin silloin, kun malli ennustaa sijaintialuetta perustuen sekä nimen viimeiseen leksikaaliseen osaan että nimenalkuun.

Luokittelumallien tärkeimpien muuttujien tarkempi tarkastelu olisi vielä jatkossa tarpeen. Lisäksi tarkasteltujen piirteiden jakaumaa kannattaisi selvittää esimerkiksi ohjaamattoman koneoppimisen menetelmin, kuten klusterianalyysillä. Paikannimistä voisi lisäksi erottaa johdinaineiksia, morfofonologisia piirteitä tai sanastokerrostumia ja luoda vielä useampien piirteiden pohjalta luokittelumalleja.

Avainsanat: nimistöntutkimus, paikannimitutkimus, paikannimet, paikannimirekisteri, alueellinen variaatio, murrealueet, ohjattu koneoppiminen, satunnaismetsä

Sisällys

1	Johdanto	2
1.1	Tutkimuksen aihe	2
1.2	Tutkimuksen tavoitteet	3
1.3	Aiempi nimistöntutkimus	4
1.4	Tutkimuksen kulku	5
2	Paikannimitutkimus	7
2.1	Paikannimien tehtävät	7
2.2	Paikannimien osat	8
2.3	Paikannimien muodostaminen ja muuttuminen	10
2.4	Paikannimet ja suomen kielen alueellinen variaatio	13
2.5	Paikannimistön alueellisen variaation tutkimus	15
3	Tutkimuksen aineisto ja menetelmät	18
3.1	Maanmittauslaitoksen paikannimirekisteri ja digitaalinen nimiarkisto	18
3.2	Murrealueet, maakunnat ja kunnat	21
3.3	Piirteiden erottaminen ja luokittelu	23
3.4	Sijainnin ennustaminen ja muuttujat	26
4	Paikannimien alueellinen variaatio	31
4.1	Paikannimet paikanlajikategorioittain ja murrealueittain	31
4.1.1	Koko paikannimi	31
4.1.2	Yleisimmät viimeiset leksikaaliset osat	35
4.1.3	Yleisimmät kolmen merkin nimenalut	42
4.2	Paikannimen viittauskohteen sijaintialueen ennustaminen	46
5	Päätelmät	55
6	Lopuksi	59
	Lähteet	61
	Liitteet	

1 Johdanto

1.1 Tutkimuksen aihe

Tämä tutkimus keskittyy suomenkielisten paikannimien alueelliseen variaatioon. Valitsin tutkimukseni aiheeksi paikannimet ja alueellisen variaation oman kiinnostukseni vuoksi. Tein kandidaatintutkielman lapsuuden kotipaikkani lähellä sijaitsevien peruslohkojen eli viljelysmaiden nimistä ja halusin jatkaa paikannimien tutkimista myös maisterintutkielmassani. Maskunjoen yläjuoksun ja laajemminkin Varsinais-Suomen viljelysnimien taustojen kartoittaminen sai minut kiinnostumaan siitä, millaisia erityispiirteitä muiden alueiden paikannimistössä on ja miten alueet eroavat toisistaan. Kandidaatintutkielmani aineisto oli maantieteellisesti hyvin pieneltä alueelta, mutta tämän tutkimuksen aineisto on paikannimirekisteristä ja kattaa koko Suomen.

Aikaisemman tutkimuksen perusteella tiedetään, että paikannimissä on alueellista vaihtelua ainakin idän ja lännen sekä idän, lännen ja pohjoisen välillä. Esimerkiksi jotkut ominaisuuksia kuvaavat sanat noudattelevat itä–länsi-jakoa: *iso* ja *pieni* ovat verrattain yleisiä koko maassa, mutta *suur(i)* on yleisempi idässä ja *vähä* 'pieni' lännessä. Muitakin sanoja käytetään alueellisesti eri tavoin: *Neva*-sanaa käytetään eniten läntisessä Suomessa, kun taas *rahka*-sanan tiheimmät esiintymät ovat lounaisessa Suomessa. *Suo* sen sijaan on yhtä yleinen pohjoista Suomea lukuun ottamatta kaikkialla, mutta *aapa*-sanan esiintymiä on lähinnä pohjoisessa. (Salmenkivi–Pulkinen–Tuominen 2006, 219–226.)

Nimistöntutkimuksessa paikannimet voidaan jakaa toisaalta ihmisten luonnollisesta nimeämistarpeesta syntyneisiin perinnäisiin paikannimiin ja toisaalta suunniteltuihin nimiin (Ainiala 2018, 96). Varta vasten suunniteltuja nimiä ovat esimerkiksi monet kaupunkien katujen ja puistojen nimet. Edellä kuvattu sanastossa esiintyvä alueellinen variaatio liittyy ennen kaikkea perinnäisiin paikannimiin: nimet ovat syntyneet niiden ihmisten parissa, jotka ovat käyttäneet nimettyjä alueita, ja sillä kielellä, jota ihmiset ovat käyttäneet.

Paikannimirekisterissä on kuitenkin myös jonkin verran suunniteltuja paikannimiä, kuten hallintopaikkojen (esim. erilaiset virastot ja keskuskeskukset), liikennepaikkojen (esim. jotkin rautatie- ja metroasemat) ja kulttuuripaikkojen (esim. museot ja muistomerkit) nimiä, sillä rekisteri on karttojen tekemisen sivutuote. Suunnitellut nimetkin voivat toki pohjautua perinnäiseen paikannimistöön (ks. esim. Hyvän nimistön opas).

1.2 Tutkimuksen tavoitteet

Tavoitteenani on tutkia paikannimien murrealueittaista variaatiota kartoittamalla paikannimissä esiintyvien viimeisten leksikaalisten osien eli paikannimissä viimeisenä esiintyvien sanojen (terminologiasta ks. luku 2.2; piirteistä ks. luku 3.3) alueellista ja paikanlajeittaista jakautumista. Toinen tavoite on selvittää, millaisen ennustuksen paikannimen sijaintialueesta voi eri piirteiden perusteella tehdä ohjatun koneoppimisen menetelmän avulla. Tutkimuskysymykseni ovat

1. Millaisia eri murrealueiden paikannimet ovat, entä millaisia maasto- ja vesistönimien yleisimmät viimeiset leksikaaliset osat ja yleisimmät kolmen merkin nimenalut ovat?
2. Millaisen ennustuksen paikannimen viittaushohteen sijaintialueesta voi tehdä ohjatun koneoppimisen menetelmällä eri paikanlajeissa esiintyvien leksikaalisten viimeisten osien sekä nimenalkujen ja nimien kulumisen perusteella?

Aikaisemmassa paikannimien alueelliseen variaatioon liittyvässä tutkimuksessa on usein keskitytty paikannimien leksikaalisiin piirteisiin. Tarkastelussa ovat olleet esimerkiksi erilaiset paikan lajiin viittaavat tai ominaisuuksia kuvaavat lekseemit tai sanat. Pääpaino tässäkin tutkimuksessa on nimissä esiintyvillä sanoilla, mutta tarkastelen lisäksi nimenalkuja ja nimien kulumista. Kerron tarkemmin luvussa 3.3 siitä, miten olen piirteistänyt aineistoa. Paikannimissä esiintyvistä piirteistä on tehty määrällisiä vertailevia havaintoja ainakin maakunnittain (ks. esim. Salmenkivi ym. 2006). Koska tämä on kielitieteellinen tutkimus ja koska tutkimusaineisto koostuu lähinnä perinnäisistä paikannimistä, halusin tarkastella aineistoa murrealueittain. Kuvaan luvun 2.4 alussa vielä lyhyesti murrealuejaottelun taustoja.

Tutkimukseni uusi näkökulma on selvimmin menetelmissä. Käytän tutkimuksessani ohjatun koneoppimisen menetelmää, jolla selvitän, millaisen ennustuksen paikannimen sijaintialueesta voi erottamieni piirteiden perusteella tehdä. Tutkimus on kuitenkin monimenetelmäistä, sillä aineiston saattaminen automaation ymmärtämään muotoon vaatii myös aineiston laadullista analyysia eli piirteiden erottamista ja luokittelua. Menetelmällisesti samankaltaista kielitieteellistä tutkimusta ovat hiljattain tehneet ainakin Ilmari Ivaska, Mirva Johnson ja Tommi Kurki (2023) tutkiessaan amerikansuomalaisten murretaustaa.

Tutkimuksessa selvitettiin Pohjois-Amerikassa puhutun suomen ja Suomessa puhuttujen suomen murteiden vastaavuutta ja sitä, miten puhujien perheiden murretausta heijastuu heidän puheeseensa. Käytännössä puhujien kielinäytteistä tunnistettiin erilaisia murrepiirteitä, joiden

perusteella heidät sijoitettiin Suomen kartalle. Ohjattua koneoppimismenetelmää ovat kielitieteellisessä tutkimuksessa käyttäneet myös muun muassa Ilmari Ivaska ja Kirsti Siitonen (2017) tutkiessaan, miten erilaisten morfologisten muotojen (esimerkiksi monikon genetiivin tai peruslukujen yksikön nominatiivin) esiintymistajuus suomenoppijoiden teksteissä on yhteydessä kirjoittajan ensikieleen. Kuvaan tämän tutkimuksen metodologiaa tarkemmin luvussa 3.

Ensimmäinen, nimien sijainnin ennustettavuuteen liittyvä hypoteesini on, että jotkin paikannimien leksikaaliset viimeiset osat poikkeavat toisistaan alueellisesti niin merkittävästi, että niiden esiintyessä voidaan ennustaa nimen viittauskohteen sijaintialue. Hypoteesi perustuu etenkin Salmenkiven ja kumppaneiden (2006, 198) maakunnittaisiin havaintoihin (yksittäisten osien levikeistä ks. myös SPNK = Suomalainen paikannimikirja). Mikäli samassa nimessä yhdistyy monta alueellisesti toisistaan erottuvaa piirrettä, viittauskohteen sijainnin ennustaminen on tarkempaa. Suuri osa suomenkielisestä nimistöä koostuu hyvin yleisistä leksikaalisista ja morfologisista rakenneosista, joita esiintyy koko maassa, ja pelkästään tällaisista piirteistä koostuvien nimien viittauskohteiden sijainnin ennustaminen on hyvin epätarkkaa.

Toinen hypoteesini liittyy alueiden eroihin ja yhtäläisyyksiin. Aiemman tutkimuksen perusteella (esim. Salmenkivi ym. 2006) oletan, että suuri osa leksikaalisten viimeisten osien alueellisesta variaatiosta liittyy alueiden luonnonmaantieteellisiin eroihin. Samaan käsitteeseen viittaavien eri sanojen (esimerkiksi suoalueen nimitykset *aapa* ja *rahka* tai maankohouman nimitykset *vuori* ja *vaara*) tarkastelu paljastanee kuitenkin myös kielellistä vaihtelua eri alueiden välillä siten, että ainakin itä, länsi ja pohjoinen erottuvat toisistaan (ks. esim. Leino–Hyvönen–Salmenkivi 2006). Nimenalut taas muodostavat tiheään nimetyillä alueilla maantieteellisesti pienempiä alueellisia ryhmittymiä ja harvaan nimetyillä alueilla maantieteellisesti laajempia alueellisia ryhmittymiä (ks. esim. Kiviniemi 1990, 85–86). Aiemman tutkimuksen havaintoja on koottu tarkemmin lukuun 2.5.

1.3 Aiempi nimistö tutkimus

Nimistö tutkimus on alkanut Suomessa 1800-luvun lopulla. Ensimmäinen tutkimus käsitteli henkilönnimiä eli antroponyymejä, mutta tutkimuskenttä laajeni nopeasti paikannimiin eli toponyymeihin. (Itkonen 1997, 15.) Vaikka muitakin nimiä on tutkittu, paikannimet on ainakin Virittäjän kirjoitusten valossa ollut kaikkein tutkituin kohde koko nimistö tutkimuksen historian ajan. Metodologisesti tutkimus painottui alkuaikoina

etymologiseen analyysiin eli nimien alkuperän tutkimiseen. Typologinen analyysi, jossa tutkitaan nimisysteemejä, ja kontaktionomastiikka, jossa tutkitaan kielten ja kulttuurien välisten kohtaamisten vaikutuksia, ovat lisänneet pikkuhiljaa suosiotaan. 2000-luvulle tultaessa etymologinen analyysi on edelleen pitänyt pintansa, mutta nimistöä on 1900-luvun lopulta alkaen tutkittu aiempaa enemmän myös sosio-onomastisin keinoin, eli nimistöön liittyviin ilmiöihin on etsitty selitystä sosiaalisesta todellisuudesta. Kvantitatiivinen korpusanalyysi, joka on suurten aineistojen tilastollista tarkastelua, on myös saanut suosiota, kun tietotekniikan kehitys on mahdollistanut yhä suurempien aineistojen käsittelyn. (Ainiala–Sjöblom 2021, 525–527; 538–539.) Tässä tutkimuksessa hyvin tärkeä näkökulma on typologinen analyysi, sillä erotan nimistä rakenneosia.

Yksi syy paikannimitutkimuksen suosioon lienee poikkeuksellisen korkealaatuinen ja Suomen hyvin kattava paikannimiaineisto (ks. Itkonen 1997, 20). Nimestys eli nimien kerääminen sai alkusysäyksen jo 1800-luvun lopulla, mutta nimestäjien eli nimien kerääjien työtä vaikeutti 1960-luvulle asti se, ettei käytössä ollut riittävän hyviä karttoja (Itkonen 1997, 14–18). Suullista perimätietoa nimistä kerättiin etenkin 1960–1980-luvulla erilaisten stipendien turvin ja tutkijoiden talkootyönä valtava määrä (Itkonen 1997, 17; 20). Aineisto on tallessa nimiarkistossa, johon pääsee käsiksi nykyään myös verkossa (ks. Nimiarkisto). Nimien keruulla oli vaikutusta myös karttanimestykseen, jota ajantasaistettiin ja korjattiin käyttöä vastaavaksi nimien keruun yhteydessä (Kotus = Kotimaisten kielten keskus, Karttanimien tarkistus).

Paikannimien rakenteen ja semantiikan eli merkityssisällön tutkimuksen selkärankana on ollut pitkään Kurt Zilliacuksen (1966) esittelemä ja Eero Kiviniemen edelleen kehittänyt syntaktis-semanttinen malli (ks. Kiviniemi–Pitkänen–Zilliacus 1974, 8; 44). Mallia on käytetty jo yli 50 vuotta esimerkiksi lukuisissa opinnäytetöissä, joissa keskitytään jonkun rajatun alueen tietyn paikanlajin nimiin (esim. Skinnarila 1994) tai tietyn lekseemin sisältäviin nimiin (esim. Törmälä 2020). Tässä tutkimuksessa mallia ei sellaisenaan pysty käyttämään, koska aineisto on niin suuri. Nimiä on kuitenkin mahdollista pelkän kirjoitusasun perusteella analysoida semanttisiin osiin, ja näitä nimenosia voi tietyin varauksin (esim. homonyymiset ilmaukset huomioiden) analysoida syntaktisesti, kun tietää nimetyn paikan lajin.

1.4 Tutkimuksen kulku

Tutkimus etenee siten, että ensin kerron luvussa 2 tutkimukseni kannalta tärkeästä paikannimitutkimuksen teoriasta: paikannimien tehtävistä, paikannimien jaottelusta

rakenneosiin sekä paikannimien muodostamisesta ja muuttumisesta. Alaluvussa 2.4 käsittelen paikannimien ja suomen kielen alueellisen vaihtelun suhdetta, ja alalukuun 2.5 olen koonnut paikannimistön alueellisesta variaatiosta tehtyjen tutkimusten tuloksia.

Luvussa 3.1 kerron paikannimirekisteristä ja siitä, miten olen rajannut tämän tutkimuksen aineistoa. Käyn luvussa 3.2 läpi tässä tutkimuksessa käytetyn murrealuejaon ja vertailen sitä paikannimirekisterin valmiiseen maakuntajakoon ja kuntajakoon. Luvussa 3.3 kerron siitä, miten olen erottanut aineistosta tarkasteltavia piirteitä, ja luvussa 3.4 siitä, miten hyödynnän tässä tutkimuksessa ohjatun koneoppimisen menetelmää, satunnaismetsää.

Luvussa 4 analysoin paikannimiä. Aloitan siitä, millaisia paikannimet ovat paikanlajeittain ja murrealueittain koko aineistossani ja miten viisi aineistoni yleisintä viimeistä leksikaalista osaa jakautuu murrealueittain ja paikanlajeittain. Analysoin myös nimenalkuja murrealueittain kolmen merkin pituisiin nimenalkuihin pohjautuen. Sen jälkeen kuvaan ohjatulla koneoppimismenetelmällä luotuja nimen viittauskohteen sijaintialueen ennustamisen malleja. Luvussa 5 vastaan tutkimuskysymyksiini ja esitän tuloksista tekemiäni päätelmiä, ja lopuksi luvussa 6 arvioin tutkimustani ja kerron jatkotutkimusajatuksiani.

2 Paikannimitutkimus

2.1 Paikannimien tehtävät

Paikannimi on ilmaus, jolla paikka yksilöidään ja jonka avulla paikka erotetaan muista samanlajisista paikoista (Kiviniemi 1978, 73). *Nimi*-sanalla viitataan nimistöntutkimuksessa eli onomastiikassa aina propreihin eli erisnimiin (Sjöblom–Ainiala 2020, 803). *Nimellä* on kuitenkin myös appellatiivin eli yleisnimen merkitys lajiaan edustavana henkilönä, oliona, esineenä tai asiana (KS = Kielitoimiston sanakirja). Propreja ja appellatiiveja erottaa se, että propreilla viitataan vain yhteen tarkoitteeseen eli referenttiin (VISK = Iso suomen kielioppi § 553). Käytännössä propri ja appellatiivi voivat olla täysin samanmuotoiset. Pääasia on, että propri tunnustetaan propriksi ja ohitetaan samanmuotoisen sanan eli homonyymin appellatiivinen merkitys. (Kiviniemi 1975, 9–10; Kiviniemi 1978, 73.) Kirjoitetussa kielessä proprin tunnistaa kirjoitusasusta: esimerkiksi *kotka* on appellatiivi ja *Kotka* propri. Puhutussa kielessä proprin tunnistaa usein kontekstista (*matkustetaan Kotkaan* mutta *kotka lentää*), mutta aina eron tekeminen proprin ja appellatiivin välille ei ole yhtä yksinkertaista (*olin valokuvaamassa kotkaa viikonloppuna* ja *olin valokuvaamassa Kotkaa viikonloppuna* ovat molemmat mahdollisia).

Paikannimiä on olemassa kaikissa kielissä ja kulttuureissa, eli ne ovat universaaleja. Suomessa yleisin paikannimi on rakenteeltaan kaksiosainen nimi, jossa toinen osa ilmaisee paikan lajin ja toinen paikan erityispiirteen. Perinnäiset paikannimet ovat syntyneet, kun ihmisillä on ollut tarve yksilöidä ympärillään ja kulkureiteillään olevia paikkoja. Usein paikkaa on kuvailtu jollain tavalla yhdistämällä paikan lajiin viittaava ilmaus ja jokin paikan erityispiirrettä kuvaava ilmaus. (Kiviniemi 1990, 86; 90–91.) Esimerkiksi viljelysmaan nimi *Latopelto* sisältää paikan lajiin viittaavan sanan *pelto* ja paikan erityispiirrettä kuvaavan sanan *lato*. Tarpeeksi pienessä yhteisössä jo se, että pellon läheisyydessä on lato, riittää yksilöimään pellon muista viljelysmaista.

Varsinkin paikannimissä on säilynyt sellaisia appellatiiveja, jotka ovat hyvin harvinaisia, paikallisia tai jopa muutoin kadonneita (Itkonen 1997, 50). Nimisampoon (2019) tehty haku esimerkiksi osoittaa, että vain Varsinais-Suomen alueella esiintyy paikannimissä sanaa *tahdon*, kun taas *kytö* esiintyy laajemmalti, mutta erityisesti Ylä-Satakunnasta Etelä-Pohjanmaalle ja läntiseen Keski-Suomeen ulottuvalla alueella. *Tahdo(i)n* on lounaismurteissa esiintynyt sana, joka on tarkoittanut 'aidattua peltoaluetta' (SMS = Suomen murteiden

sanakirja s.v. *kauratahdon*). *Kytö* taas tarkoittaa ’viljelysmaata, joka on tehty suolle kydöttämällä’ (KS). Sanojen appellatiivisuus on hämärtynyt ja käyttö harvinaistunut, sillä *tahdo(i)n*-sanaa ei löydy nykyistä yleiskieltä kuvaavasta Kielitoimiston sanakirjasta ja kytöjäkään tuskin luodaan lisää.

Paikannimet ovat syntyhetkellään aina olleet paikkaa kuvailevia ja nimenantajille ymmärrettäviä ilmauksia. Nimenantajat ovat muodostaneet paikannimet olemassa olevista kielenaineksista eli sanoista, päätteistä ja johtimista. Paikannimiin sisältyy myös toisia propreja, niin paikannimiä kuin henkilönnimiäkin. Nimenantajat ovat käyttäneet nimeä antaessaan nimenmuodostusmalleja, jotka ovat muotoutuneet nimenantajien tuntemien proprien pohjalta. Suuri osa paikannimistöstä on sellaista, että ymmärrämme yhä niissä olevien kielenainesten merkityksen. Esimerkiksi nimet *Haapajärvi*, *Rantapelto*, *Niemelä* ja *Niemistö* sisältävät tunnistettavia kielenaineksia. (Kiviniemi 1975, 10–13.)

Ajan myötä joko nimet tai niitä ympäröivä kieli ja kulttuuri muuttuvat, jolloin osa nimissä olevista kielenaineksista ei enää ole ymmärrettäviä. Paikannimien rakenteen ja merkityssisällön tutkiminen edellyttää nimen etymologian eli alkuperän selvittämistä varsinkin silloin, kun nimessä on nykykielen kannalta jotakin hämärää. Tällöin yritetään selvittää, kuinka vanha nimi on ja mitä kieltä ja murretta nimi on. (Kiviniemi 1990, 36–39.) Esimerkiksi saamen kielestä on jälkiä paikannimistössä eniten pohjoisessa mutta myös muualla Suomessa (ks. Aikio 2003; Räisänen 2005). Yksi eteläisimmistä saamelaisetyologioista on esitetty Nuuksion kansallispuiston nimessä esiintyvälle *nuuksio*-sanalle, jossa ajatellaan näkyvän nykyään pohjoissaamessa *njukča*-asuinen ’joutsenta’ merkitsevä sana (SPNK 296). Jotkut suurten vesistöjen ikivanhat nimet ovat siinä määrin hämärtyneitä, etteivät edes kielentutkijat osaa sanoa, mitä nimien taustalla on. Tällaisia nimiä ovat esimerkiksi *Saimaa* ja *Päijänne*. (Kiviniemi 1990, 13; ks. myös SPNK 2007, 360–361, 396.) Piirteiden erottaminen nimistä edellyttää sitä, että nimissä olevat rakenneosat tunnistetaan.

2.2 Paikannimien osat

Paikannimien rakenteen ja semantiikan eli merkityssisällön selvittämisessä on suomalaisessa nimistöntutkimuksessa käytetty jo pitkään alun perin Kurt Zilliacuksen (1966) esittelemää syntaktis-semanttista analyysia. Analyysia varten paikannimistä tunnistetaan erilleen syntaktiset nimenosat, joita on kahdenlaisia: paikkaan viittaava perusosa ja ilmausta tarkentava määriteosa. Kaksiosaisissa nimissä on sekä perusosa että määriteosa, kun taas

yksiossaisten nimien ainoa osa luokitellaan perusosaksi, sillä se viittaa paikkaan. (Kiviniemi 1975, 17–18.) Mallissa nimenosan funktio eli tehtävä voi olla joko paikan lajia ilmaiseva, nimittävä tai erityispiirrettä kuvaava (Kiviniemi ym. 1974, 8). Alla olevaan kuvaan (ks. kuva 1) on havainnollistettu kehittelemäni esimerkin avulla syntaktiset luokat: *Haavistonmäki* viljelysmaan nimenä sisältää ainoastaan erityispiirrettä kuvaavan perusosan, mutta kohouman nimenä se sisältää paikan lajiin viittaavan perusosan *mäki* ja erityispiirrettä kuvaavan määriteosan *Haaviston*. Lähellä sijaitseva matalampi kohouma *Vähä Haavistonmäki* taas sisältää erityispiirrettä kuvaavan määriteosan *Vähä* ja nimittävän perusosan *Haavistonmäki*.



Kuva 1 Paikannimen luokittelu syntaktisiin osiin

Edellä käsitellyn esimerkin *mäki* on merkitykseltään hyvin rajatusti vain yhdenlaiseen paikkaan viittaava, joten se on kohouman nimen perusosana tulkittavissa paikan lajiin viittaavaksi. Paikannimissä esiintyy kuitenkin paljon myös sellaisia perusosia, joiden tulkinta ei ole yhtä yksinkertaista. Vaikkapa *saari* tai *niemi* voi esiintyä viimeisenä leksikaalisena osana hyvin monenlajisten paikkojen nimissä. Silloin täytyy selvittää, onko kyseessä alun perin jonkun muun paikan nimi vai paikalle uutena annettu nimi. Juuri *saaren* ja *niemen* kaltaisia vesistötermejä käytetään usein metaforisesti myös paikoista, jotka muistuttavat vesistöjen saaria tai niemiä (Kiviniemi 1990, 58–59; 84). Tällaisia nimiä ei ole yhtä suoraviivaista analysoida syntaktis-semanttisesti.

Mikäli halutaan erottaa nimistä leksikaalisia ja morfologisia osia, voidaan käyttää termiä *nimielementti*, jolla viitataan nimiin sisältyviin sanoihin tai sanavartaloihin sekä suffikseihin (Kiviniemi 1975, 17). Esimerkiksi johtimet *-nen* ja *-la* sekä sanat *Sanna* ja *Koivu* nimissä *Sannanen* ja *Koivula* ovat nimielementtejä. Käytän tässä tutkimuksessa myös Salmenkiven ja kumppaneiden (2006, 191) esittelemää käsitettä *nimenalku*. Nimenalku voi olla myös nimielementti (ks. Salmenkivi ym. 2006, 191–192). Esimerkiksi *Sannanen*-nimessä nimenalkuja ovat *s*, *sa*, *san*, *sann* ja *sanna*, mutta näistä ainoastaan *sanna* eli 'hiekkä' on myös nimielementti. Viimeisenä esiintyvistä leksikaalisesta osasta eli nimen viimeisestä sanasta käytän tässä tutkimuksessa nimitystä *viimeinen leksikaalinen osa* erotuksena perusosasta, joka on määritelty syntaktis-semanttisten suhteiden perusteella. Niitä nimiä, joista on erotettavissa viimeinen leksikaalinen osa kutsun leksikaalisesti moniosaisiksi nimiksi.

Paikannimet voivat olla joko transparentteja eli läpinäkyviä tai opaakkeja eli läpinäkymättömiä. Läpinäkyvien nimien nimeämisperuste on näkyvissä, kun taas läpinäkymättömien nimien nimeämisperuste ei ole. Läpinäkyvyys voi olla leksikaalista eli sanasemanttista, jolloin kyseessä on nimeen sisältyvien yksittäisten sanojen läpinäkyvyys. Lausesemanttinen eli syntaktis-semanttinen läpinäkyvyys viittaa nimen asiataustan läpinäkyvyyteen. Läpinäkyvyys on kummassakin tapauksessa jatkumo läpinäkyvästä monitulkintaisen kautta läpinäkymättömään. Esimerkiksi järvennimi *Päijänne* edustaa opaakkia ääripäätä, kun taas *Pitkäjärvi* on malliltaan pitkän järven nimenä transparentti. *Kettukivi* taas on sanasemanttisesti läpinäkyvä, mutta syntaktis-semanttisesti monitulkintainen, sillä paikalla on voinut esimerkiksi olla ketun pesä tai kivi voi muistuttaa muodoltaan kettua. Nimen asiatausta selviää usein, kun paikkaa ja sen lähiympäristöä tarkastelee esimerkiksi kartalla. Usein on hyödyksi myös tieto siitä, mikä paikannimissä on yleistä ja mikä harvinaista. (Kiviniemi 1990, 12–13.) Tässä tutkimuksessa ei ole tarkoitus kartoittaa kaikkien nimien asiataustaa, koska aineisto on siihen liian suuri. Etenkin nimissä olevien sananmuotojen ja viimeisten leksikaalisten osien paikanlajeittaisen tarkastelun kannalta läpinäkyvyys on kuitenkin olennaista.

2.3 Paikannimien muodostaminen ja muuttuminen

Paikannimien muodostaminen nojaa vahvasti nimeämismalleihin (Kiviniemi 1971, 131–132; Kiviniemi 1977, 207–208). Paikannimet voidaan ajatella omana systeeminään, jonka sisällä vallitsee niin kutsuttu nimikielioppi ja nimeämismalleihin perustuva säännöstö, jota

noudatetaan, kun uusia paikannimiä muodostetaan (Kiviniemi ym. 1974, 32–33; Kiviniemi 1978, 73–74). Yleisintä nimityyppiä edustavat nimet eli erikoispiirrettä kuvaavasta määriteosasta ja paikan lajiin viittaavasta perusosasta koostuvat syntaktisesti kaksiosaiset nimet ovat nimikieliopin ytimessä (Kiviniemi 1975, 11–15).

Mallin mukaan on muodostettu esimerkiksi sellaiset ääntä kuvaavat eli onomatopoeettiset vA-partisiippiset vesistönimet kuin *Kohiseva* ja *Kuohuva* (Kiviniemi 1971, 23; 36). Muita morfologiaan nojaavia malleja ovat esimerkiksi verbikantaisen mA-partisiipin ja teonnimen sisältävät nimet (esim. *Elämä-* ja *Vieremä-*alkuiset nimet) sekä lukuisilla muillakin johtimilla muodostetut nimet (esim. *lA*-paikannimijohtimella muodostetut *Niemelä* ja *Koivula* tai *stO*-johtimella muodostetut *Haavisto* ja *Koivisto*). Paikannimissä voi esiintyä tällaisenkin nimenmuodostuksen tuloksena myös sellaisia muotoja, joilla ei tunnu nykykielen kannalta olevan tunnistettavaa merkitystä. Kiviniemi (1971, 10) esittää esimerkiksi *Onsava*-nimen muodostuneen vA-aineksen ja sanan *onsi* 'ontto (paikka)' yhdistelmästä. Muotohan ei ole yleiskielisten sananmuodostussääntöjen mukainen, sillä vA-partisiippi on verbikantainen (ks. VISK § 62). Paikannimisysteemissä on myös sellaisia johtimia, joilla on alun alkaenkin ollut paikkaan viittaamisen merkitys, kuten *nkO* tai *lA* (Kiviniemi 1975, 55; Aikio 2007, 169). Kiviniemi (1975, 37; 41) esittää, että eräät paikannimissä esiintyvät suffiksit (ainakin vedenkokoumien nimissä *nen* ja talonnimissä *lA*) ovat ainakin osassa tapauksista muodostuneet korvaamaan perusosaa, jolloin nimet ovat muuttuneet lyhyemmiksi ja siten taloudellisemmiksi.

Nimisysteemin keventämisen kannalta edullinen tapa muodostaa nimiä on sisällyttää niihin jokin jo olemassa oleva nimi. Uuden paikannimen määriteosaksi voidaan ottaa toisen paikan koko nimi, jolloin nimi on kokoliitynnäinen (*Ryti|jänkä* > *Rytijängän|harju*). Jos taas uuden paikannimen määriteosana on vanhan paikannimen määriteosa, kyseessä on määriteosaliitynnäinen (*Ryti|jänkä* > *Ryti|harju*). Perusosaliitynnäisessä taas uuden paikannimen määriteosana on vanhan paikannimen perusosa (*Ryti|jänkä* > *Jängän|harju*). Tällä tavoin toisiinsa liittyviä nimiä kutsutaan nimirypääksi ja rypään alkulähdettä kantanimeksi. (Kiviniemi 1990, 36; 124–125.) Kiviniemen (1990, 125) mukaan keskimäärin 30 prosenttia paikannimistä on jonkinlaisia liitynnäisiä, mutta osuus on alhaisempi viljelysvaltaisilla ja asutukseltaan tiheillä alueilla. Harvaan asutuilla ja harvaan nimetyillä paikoilla on yleensä maantieteellisesti laajalle levittäytyvää nimenosien ketjuuntumista (Kiviniemi 1990, 35–36). Tämän tutkimuksen aineistossa nimitiheys on kokonaisuudessaan alhaisempi kuin Kiviniemen tutkimuksen aineistossa, joka koostuu mikrotoponyymeistä eli

pienen yhteisön käyttämistä paikannimistä. Asutustiheyden vaikutus liitynnäisten levittymiseen lienee kuitenkin samansuuntainen, koska tässäkin tutkimuksessa valtaosa nimistä on perinnäisiä paikannimiä. Tätä tutkimusta ajatellen etenkin kokoliitynnäiset ja määriteosaliitynnäiset voivat vaikuttaa erilaisten nimenalkujen esiintymistaajuuteen eri alueilla.

Kokonainen paikannimi voidaan myös ottaa jonkun toisen paikan nimeksi, jolloin puhutaan siirrynnäisistä. Metonyymisiksi siirrynnäisiksi kutsutaan sellaisia nimiä, jotka ovat syntyneet kosketusassosiaatioon perustuen jonkin läheisen paikan nimestä, ja vertaileviksi siirrynnäisiksi sellaisia nimiä, jotka perustuvat paikkojen yhtäläisyyteen (Kiviniemi ym. 1974, 48–49). *Koivumäki*-nimisen maankohouman kupeeseen rakennettu asumus on voinut esimerkiksi saada nimekseen metonyymisesti *Koivumäki*. Vertailunimiä ovat esimerkiksi Suomen parikymmentä *Kapernaumia*, joiden esikuvana on ollut Uudesta testamentista tuttu Jeesuksen kotikaupunki (ks. SPNK, 135). Metonyymiset siirrynnäiset kasvattavat samanasuisten nimien määrää, jos kantanimi säilyy.

Vaikka paikannimi on syntyhetkellään nimenantajan näkökulmasta ollut paikkaa kuvaileva, nimen yksilöivän tehtävän voi ajatella olevan nimen käyttäjille olennaisempi kuin kielellisten elementtien merkitys on (Kiviniemi 1975, 13; erisnimien semantiikasta ks. esim. Sjöblom 2004). Joskus nimen käyttäjät kuitenkin selittävät nimiä läpinäkyvämmiksi: esimerkiksi *Elimäki*-nimen alkuperäiseen asuun *Elimä* (joka tulee pohjoissaamen ’ylimmäistä’ merkitsevästä *alimus*-sanasta) on jossain kohtaa sen historiaa liitetty kansanetymologisesti *mäki*-sana, kun nimi on hahmotettu kohouman eikä enää järven nimeksi (SPNK, 46). Silloin kun paikanlajiin viittaavaa ilmaus lisätään puhtaasti tarkoitetta selventämään, ilmiötä kutsutaan epeksegeesiksi (Kiviniemi ym. 1974, 13; Kiviniemi 1975, 62). Esimerkiksi kun ’ojaa’ merkitsevää *puru*-sanaa ei ole enää *Raitinpuru*-nimessä hahmotettu paikan lajiin viittaavaksi, siihen on lisätty *oja*-sana, jolloin nimestä on tullut *Raitinpurunoja*. Vieraalle nimien kerääjälle eli nimestäjälle on myös voitu antaa nimestä läpinäkyvämpi muoto, vaikka muutoin on käytetty lyhyempää nimeä. (Ainiala 1997, 174; 177; 192.) Näin jotkut uudet tai yleiskielisemmät paikan lajiin viittaavat ilmaukset ovat voineet yleistyä varhaisempien tai murteellisiksi ajateltujen ilmausten kustannuksella.

Nimien muuttuminen lyhyemmäksi ja siten taloudellisemmaksi on kuitenkin tavallisempaa. Nimi voi muuttua esimerkiksi äänteellisesti: nimissä *Mustio* ja *Teijo* on alun perin ollut lopussa paikan lajia ilmaiseva *joki*, mutta *-ki* on hävinnyt molemmista (SPNK, 277; 451).

Nimi voi kulua myös keskeltä siten, että paikan lajia ilmaiseva perusosa pysyy ennallaan ja määriteosa katkeaa lopusta. Usein kulumisen seurauksena nimi on helpompi ääntää, kuten vaikkapa muutoksessa *Kaislalahti* > *Kaislahti*. (Kiviniemi ym. 1974, 22.) Näin syntyy joskus muotoja, joiden alkuperäistä määriteosaa voi olla hankala selvittää pelkän kieliasun perusteella. Esimerkiksi *hawk*-määriteosan sisältävät nimet voisivat yhtä hyvin olla peräisin sanasta *hauki* tai *haukka*. Vesistöön viittaava perusosa puoltanee *hauki*-tulkintaa ja maankohoumaan viittaava *haukka*-tulkintaa, mutta esimerkiksi liitynnäisyyttä tai siirrynnäisyyttä ei voi sulkea pois tarkastelematta lähialueen nimiä tai nimen aikaisempia muotoja vaikkapa asiakirjoista.

Suomenkielisiä kirjallisia lähteitä ei ole kovin pitkältä ajalta, mutta paikannimiä on esiintynyt kirjoitetussa muodossa jo ihka ensimmäisissä suomen kieltä sisältävissä pääosin ruotsin- ja latinankielisissä teksteissä (Häkkinen 1994, 77). Esimerkiksi nimiarkistossa on lueteltu kootusti etenkin Kansallisarkistossa säilytettävistä 1500-luvun voudintileistä kerättyjä henkilön- ja paikannimiä (ks. Nimiarkisto, Asiakirjanimet). Asiakirjat ja mahdollisesti myös kirjurit olivat vielä ainakin 1800-luvulla pitkälti ruotsinkielisiä (Häkkinen 1994, 12), mikä on vaikuttanut nimien kirjoitusasuun (nimien vanhoista kirjoitusmuodoista ruotsinkielisissä lähteissä ks. Pitkänen 1986, 67–70). Varhaisempien asiakirjamuotojen tutkiminen ei kuulu tämän tutkimuksen piiriin, mutta nimien vakiintumiseen ja siten murrepiirteiden säilymiseen on voinut vaikuttaa niiden esiintyminen jo varhain kirjoitetussa muodossa (murrepiirteiden säilyttämisestä ks. Itkonen 1997, 56–57).

2.4 Paikannimet ja suomen kielen alueellinen variaatio

Fennistiseen murteentutkimukseen vakiintuneessa murrealuejaossa itämurteet ja länsimurteet erotetaan omiksi ryhmikseen. Länsimurteita katsotaan usein olevan kuusi: lounaismurteet, lounaiset välimurteet, hämäläismurteet, Etelä-Pohjanmaan murteet, Keski- ja Pohjois-Pohjanmaan murteet ja Peräpohjolan murteet (huom. kuitenkin esim. Kettusen vuoden 1940 jaottelussa lounaiset välimurteet ovat osa hämäläismurteita). Itämurteita taas ajatellaan olevan kaksi: savolaismurteet ja kaakkoismurteet. Lisäksi Etelä-Pohjanmaan murteita lukuun ottamatta kullakin murrealueella on vielä alaryhmiä. (Kotus, Suomen murteet; murrealuejaon vakiintumisesta ks. Aarikka 2023, 281–282.) Perinteinen murrealuejako perustuu useiden äänne- ja muotorakenteeseen liittyvien piirteiden levinneisyyteen ja asutushistorialliseen tietoon (murrepiirteistä ks. Rapola 1969; asutushistoriasta ks. Lehtinen 2007). Paikannimistöä on myös käytetty historian tutkimuksen aputieteenä ja etenkin juuri asutushistorian

tutkimisessa (ks. Virrankoski 1978; Vahtola 1980). Liitteenä on kartta ja lista murrealuejaosta (ks. liite 1). Liitteessä on kartan jälkeen lueteltu sekä murrealueiden pääluokat että alaluokat. Tämän tutkimuksen aluejaosta kerron vielä tarkemmin luvussa 3.2.

Murresanojen alueellisen variaation tilastollinen tarkastelu toisaalta vahvistaa ja toisaalta haastaa joitakin perinteisiä jakolinjoja. Kahteen ryhmään jaettaessa itä–länsi-kontrasti on selvä, vaikka varsinaiset rajalinjat poikkeavatkin totutusta, mutta kolmeen jako erottaa idän ja lännen rinnalle pohjoisen yhtä selkeäksi omaksi ryhmäkseen. Tarkempien rajalinjojen vetäminen haastaa myös esimerkiksi lounaisten välimurteiden olemassaolon ja erottaa sen sijaan satakuntalaiset murteet omaksi ryhmäkseen. (Leino ym. 2006, 34–35; 42.) Rajojen vetäminen eri murteiden välille on kuitenkin aina keinotekoista, sillä murteet ovat jatkumo. Yhden piirteen raja toki kulkee verrattain selkeästi yhdestä kohdasta, mutta useampiin piirteisiin perustuvat aluerajat ovat aina sovinnaisia. (Rapola 1969, 14–16.)

Koska paikannimet ovat ihmisten puhetta paikoista, paikannimistöissä esiintyvää alueellista vaihtelua koskevat periaatteissa samat lainalaisuudet kuin muussakin kielessä esiintyvää alueellista vaihtelua. Vaikka paikannimirekisteri on paikallismurteeseen perustuvaa kirjoitettua kieltä, perinteisessä murteentutkimuksessa tarkastellut äänne- ja muotorakenteen piirteet eivät välttämättä ollenkaan näy paikannimissä. Esimerkiksi voidaan ottaa vaikkapa ensitavun pitkien avointen vokaalien toteutumisen itämurteissa (esim. *pää* > *piä/peä*, *maa* > *mua/moa*; piirteen levikistä ks. Rapola 1969, 57). Nimiarkistossa on esimerkiksi Kiuruvedeltä tallennettu metsäalueen nimi *Ahompää*, jolle on paikalliseksi ääntöasuksi esitetty *ahompiä* (Nimiarkisto, Nuutinen 1960). Lemiltä tallennetun *Uudismaa*-nimisen pellon ääntöasu taas on *uuvismoa* (Nimiarkisto, Kiiski 1965). Paikannimirekisterissä nimet kuitenkin ovat muodossa *Ahompää* ja *Uudismaa*.

Vaikka paikannimirekisterin nimet siis ovat paikallisesti syntyneitä, nimistöntarkistaja on normalistanut eli yleiskielistänyt nimet, jotta ne olisivat valtakunnallisissa kartoissa selkeitä muillekin kuin paikalliselle väestölle (Kotus, Karttanimistön tarkistus). Normalistaminen on ollut osin sattumanvaraista, mutta etenkin nimistöneruiden yhteydessä tehdyssä karttanimien tarkistuksessa on noudatettu samaa yhteneväistä tapaa kuin arkistoon kerättyjen nimien normalistamisessa. Esimerkiksi paikalliseen asuun perustuvat loppuheittoiset muodot (esim. *Kuparsaari*) on säilytetty silloin, kun ne ovat vakiintuneet kirjalliseen käyttöön. Normalistetussa asussa säilytettäväksi paikallismuodoiksi mainitaan muun muassa sanat *lammi* ja *niittu*, länsisuomalaisten nimien alkujen konsonanttiyhtymät (esim. *Krepaali*) ja *f-*

konsonantti (esim. *Faarinpelto*) sekä *eA*-nominien murreasut, kuten *kapia* ja *valkia*. (Itkonen 1997, 56–57.) Paikannimirekisterin nimissä pyritään siis säilyttämään mahdollisuuksien mukaan paikallisia murrepiirteitä, jotta nimet tuntuisivat paikallisille asukkaille omilta yleiskielistämisestä huolimatta. Vaikka *soejinsuo* siis kirjataan rekisteriin muodossa *Soidinsuo*, *Ahvenlammensuo* voi olla myös *Ahvenlamminsuo* paikallisen taivutusparadigman mukaisesti. Kartoissa esiintyvien nimien on tarkoitus sekä auttaa paikantamaan että säilyttää paikallista kulttuuria. (Kotus, Karttanimistön tarkistus.)

Paikannimien alueellisen variaation kannalta laajimmin tutkitut piirteet liittyvät sanastoon. Lisäksi sanaston alueellisiin eroihin liittyvää tutkimusta voi soveltaa myös paikannimien sanoista tai lekseemeistä tehtyihin havaintoihin. Etenkin eräillä paikan lajiin viittaavilla sanoilla on paikannimissä alueellisia variantteja. Esimerkiksi edellä kuvatussa *lammen–lammin*-parissa näkyy alueellinen vaihtelu (*lampi*-sanan asusta eri murteissa ks. SMS s.v. *lampi*). Myös muun muassa *uusi*-sanaa käytetään eri alueilla eri tavoin: ainakin Pohjois-Savossa se on nimen alussa muodossa *uudis*, kun muualla muoto on *uusi* (Salmenkivi ym. 2006, 204). Joitain merkityksiä ilmaistaan eri alueilla perinteisesti eri lekseemeillä: etenkin kaakossa käytetään sanaa *suuri*, muualla sanaa *iso* (Klemettinen 2010).

2.5 Paikannimistön alueellisen variaation tutkimus

Eräiden leksikaalis-morfologisten piirteiden alueellista levinneisyyttä on tarkastellut jo varhain Kiviniemi ensin väitöskirjassaan (1971), jossa käsitellään ensimmäisen partisiipin eli *vA*-partisiipin sisältäviä henkilö- ja paikannimiä. Kiviniemen (1977) nimimalleihin syventyvässä tutkimuksessa taas tarkastellaan sellaisia vedenkokoumien nimiä, joissa on ’väärrä’ merkitsevä ilmaus (esim. *väärä*, *koukku* tai *haara*). Tutkimusten mukaan paikannimet muodostavat erilaisiin malleihin perustuvia alueellisia ja paikanlajeittaisiakin ryhmittymiä (Kiviniemi 1971, 134–135; 1977, 207–209). Alueellisesti *vA*-partisiippisissa nimissä esiintyy muun muassa erilaisia kantasanoja ainakin asutushistoriallisista syistä: esimerkiksi *Kohiseva*- ja *Tohiseva*-nimet ovat keskittyneet Pohjois-Savon huomattavimpien erä- ja asutusreittien läheisyyteen ja *Koliseva*- ja *Holiseva*-nimet taas keskiaikaiseen Hämeeseen ja sen erämaa-alueille (Kiviniemi 1971, 78; 85–92). Paikanlajeittain taas talonnimi esimerkiksi on usein yksiosainen *la*-johtiminen nimi ja luonnonpaikkojen nimistä valtaosa on kaksiosaisia nimiä, joissa toinen osa yksilöi paikan ja toinen viittaa paikan lajiin (Kiviniemi 1990, 111; ks. myös luku 2.2).

Kiviniemen (1990) paikkoja luokittelevia sanoja ja yleisimpiä määriteosia koskeva tutkimus on ensimmäinen Suomen paikannimistöä tilastollisesti tarkastellut esitys. Tutkimus käsittelee Kotuksen nimiarkistosta poimitun aineiston nimien rakennetta ja nimien rakenneosien yleisyyttä paikanlajeittain ja maantieteellisesti. Nimiarkiston paikannimistö on rakenteellisesti hieman erilainen aineisto kuin paikannimirekisteri: arkistossa on valtavasti sellaisia mikrotoponyymejä, joita ei ole päätynt karttoihin. Yleisyshavainnot eivät siis ole suoraan sovellettavissa tämän tutkimuksen aineistoon, mutta havainnot antavat osviittaa siitä, millaisia alueellisesti rajautuneita rakenneosia paikannimistössä voi olla ja miten niitä voi ryhmitellä. Kiviniemi (1990, 123) on esimerkiksi määriteosien alueellisuutta tarkastellessaan niputtanut yhteen alueellisesti toisistaan eroavat saman käsitteen variantit siten, että esimerkiksi lekseemit *peltö*, *vainio* ja *kytö* kuuluvat kaikki samaan *viljelysmaan* käsitteeseen. Tällä tavalla on Kiviniemen (1990, 123) mukaan mielekkäämpää kuvata nimisysteemejä. Tässä tutkimuksessa olen ainoastaan palauttanut monikolliset sanat yksikkömuotoon, sillä olen kiinnostunut nimenomaan alueellisista varianteista ja tarkastelen viimeisiä leksikaalisia osia.

Salmenkivi, Pulkkinen ja Tuominen (2006) tarkastelevat tilastotieteellisessä tutkimuksessaan vuoden 2002 Maanmittauslaitoksen paikannimirekisterin suomenkielisten paikannimien leksikaalisten ja syntaktisten nimenosien yleisyyttä ja levinneisyyttä maakunnittain. Tutkimuksen aineistona on 717 746 suomenkielistä nimeä (Salmenkivi ym. 2006, 192). Salmenkiven ja kumppaneiden tuloksissa on selvitetty esimerkiksi maakunnittain koko Suomen keskiarvosta merkittävästi eroavia nimenalkuja eli nimen aloittavia merkkijonoja. Tulosten perusteella etenkin pohjoisimpaan Suomeen ja lounaaseen on keskittynyt paljon muusta Suomesta eroavia nimenalkuja (Salmenkivi ym. 2006, 219; 226). Vähiten keskittyneitä nimenalkuja taas on Päijät-Hämeessä, Uudellamaalla, Kanta-Hämeessä, Keski-Pohjanmaalla ja Kymenlaaksossa (Salmenkivi ym. 2006, 220; 221). Pohjoisen Suomen nimistö eroaa muusta Suomesta myös leksikaalisten viimeisten osien osalta (Salmenkivi ym. 2006, 216–217). Osa merkittävästi keskittyneistä nimenaluista on samoja usealla alueella: esimerkiksi *Koto*-alkuisia nimiä on keskittynyt merkittävästi Satakuntaan, Kymenlaaksoon, Etelä-Pohjanmaalle ja Keski-Pohjanmaalle. Toiset nimenalut taas ovat keskittyneet vain yhdelle alueelle, kuten *Koli*-alku Satakuntaan. (Salmenkivi ym. 2006, 219). Alueiden luonnonmaantieteeseen liittyvät erot näkyvät Salmenkiven ja kumppaneiden (2006, 217) mukaan heidän esittelemissään tuloksissa.

Nimirypäät eli kantanimet liitynnäisineen ja mallinimeäminen kaikkine mahdollisuuksineen muodostavat alueellisia ryhmittymiä (ks. luku 2.3). Saman lekseemin sisältäviä nimiä on kyllä

nimirypäissä, mutta saman lekseemin sisältäviä erilaisia rypäitä saattaa tietysti esiintyä monessakin paikassa. Paikallinen kielimuoto tarjoaa raamat mahdollisille lekseemeille ja sananmuodoille. Salmenkivi ja kumppanit (2006, 212) ovat nimenalkujen tarkastelussa ottaneet yksittäisinä esiintyminä huomioon vain sellaiset samanalkuiset nimet, jotka esiintyvät vähintään 1500 metrin etäisyydellä toisistaan, jotta yhdestä rypästä tulisi mukaan vain yksi esiintymä. Tässä tutkimuksessa ei ole poistettu samoja lekseemejä sisältäviä lähekkäin sijaitsevia esiintymiä. Nimirypäät voivat siis nostaa joidenkin viimeisten leksikaalisten osien ja ehkä etenkin nimenalkujen esiintymistäajuuksia ja vaikuttaa näin mallintamisen tuloksiin.

Antti Leino hyödyntää monissa tutkimuksissaan tietokoneavusteisia menetelmiä, joiden avulla hän pyrkii selvittämään, millaisia ryhmittymiä nimet muodostavat. Etenkin järvien nimet näyttäisivät muodostavan alueellisia ryhmittymiä, mutta jokien nimissä ei ole yhtä vahvaa systemaattisuutta (Leino 2004, 244). Leino (2006) on myös tarkastellut järvien ja maankohoumien nimien alueellista variaatiota ja havainnut, että suurin ero nimissä näyttäisi liittyvän maantieteellisiin seikkoihin, mutta myös itse nimiin liittyviä eroja tulee esiin.

3 Tutkimuksen aineisto ja menetelmät

3.1 Maanmittauslaitoksen paikannimirekisteri ja digitaalinen nimiarkisto

Tutkimusaineistoni on Maanmittauslaitoksen paikannimirekisteristä, joka on vapaasti käytettävissä Nimeä 4.0 Kansainvälinen (CC BY 4.0) -lisenssillä (ks. Maanmittauslaitos, Nimistö). Aineistoa saa jakaa, kopioida ja muokata vapaasti mitä tahansa tarkoitusta varten, kunhan mainitsee lisenssinantajan nimen (Maanmittauslaitos), aineiston nimen (Paikannimet) ja ajankohdan, jolloin Maanmittauslaitos on luovuttanut aineiston (tässä tapauksessa lokakuu 2024). Lisäksi jos aineistoa muokkaa jotenkin, tulee raportoida, miten aineistoa on muokannut. Kerron muokkaamisesta jäljempänä tässä luvussa.

Aikaisemmassa tutkimuksessa ei ole käytetty tismalleen samaa aineistoa kuin tässä tutkimuksessa. Paikannimirekisterissä on kuitenkin pääosin samat tiedot kuin parikymmentä vuotta sitten, kun Salmenkivi ja kumppanit (2006) käyttivät sitä tutkiessaan leksikaalisten ja syntaktisten nimenosien yleisyyttä maakunnittain. Vuonna 2002 paikannimirekisterissä oli 717 746 suomenkielistä paikannimeä (Salmenkivi ym. 2006, 192). Suomenkielisiksi merkittyjä paikannimiä oli lokakuussa 2024 rekisterissä 729 207 eli 11 461 nimeä (1,6 %) enemmän kuin parikymmentä vuotta aikaisemmin.

Koko Suomen kaikki nimetyt paikat sisältävä Paikannimet-tiedosto on ladattavissa Maanmittauslaitoksen Karttapaikka-palvelusta (ks. Maanmittauslaitos, Karttapaikka). Rekisterin tiedot päivitetään joka yö, mutta tiedostotuote ajantasaistetaan aina tammikuussa, toukokuussa ja syyskuussa (Maanmittauslaitos, Nimistö). Tiedoston lataamisen jälkeen se on muokattu R-ohjelmointiympäristössä taulukkomuotoon siten, että jokainen nimetty paikka on omalla rivillään ja jokainen tieto omassa sarakkeessaan. Paikkojen tietoja voi tarkastella myös Nimistön tietokortit -työkalulla tai selainpohjaisessa rajapintapalvelussa, joten olen pystynyt halutessani tarkastelemaan myös esimerkiksi korkeus- ja koordinaattitietoja, jotka eivät kaikki jaottuneet oikeisiin sarakkeisiin taulukointivaiheessa (ks. Maanmittauslaitos, Nimistön kyselypalvelu; Nimistön tietokortit).

Maanmittauslaitoksen paikannimirekisterissä on kustakin nimetystä paikasta 28–30 erilaista tietoa, joista tämän tutkimuksen kannalta olennaisimpia ovat seuraavat:

- paikannimen kirjoitusasu
- paikannimen kieli (suomi, pohjoissaame, inarinsaame, koltansaame tai ruotsi)

- paikanlaji (tarkin luokitus, 120 luokkaa)
- paikkakategoria (karkein luokitus, 9 luokkaa)
- kunta (yhteensä 309)
- maakunta (yhteensä 19).

Keskeisin tarkasteltava kohde on paikannimen kirjoitusasu eli varsinainen paikannimi, josta erotettuja piirteitä kuvaan luvussa 3.3. Tarkastelen tässä tutkimuksessa vain suomenkielisiksi rekisteriin merkityjä nimiä. Kaikkiaan rekisterissä on 817 886 nimeä. Suomenkielisiä nimiä on siis 729 207 (89,2 %), ruotsinkielisiä 76 730 (9 %), pohjoissaamenkielisiä 6 400 (0,8 %), inarinsaamenkielisiä 4 847 (0,6 %) ja koltansaamenkielisiä 702 (0,09 %). Rajanveto kielten välillä on rekisterissä siinä mielessä häilyvä, että nimiä ei nykyään enää käännetä kielestä toiseen tai mukauteta toisiin kieliin (ks. SPNK, 13). Suomenkielisiksi merkityissä nimissä onkin pieni määrä ruotsinkielisiksi ja saamenkieliksi tunnistettavia nimiä. Tällaisia ovat esimerkiksi inarinsaamen *Ákkunämmir* (kohouma Inarissa), pohjoissaamen *Livvadievát* (kohouma Enontekiöllä), koltantsaamen *Ku'vžžluubbâl* (vedenkokouma Inarissa) ja ruotsin *Mjölkvägen* (muu kulttuurikohde Pietarsaassa).

Paikanlajien tarkimmassa luokittelussa on yhteensä 120 erilaista luokkaa (ks. Maanmittauslaitos 2022), joista olen rajannut tarkasteltavaksi karkeimman luokituksen eli paikkakategorioiden mukaisia maasto-, vesistö- ja asutuspaikkojen nimiä. Kokonaan tarkastelun ulkopuolelle jää siis kuusi paikkakategoriaa:

- hallinto ja julkiset palvelut,
- suojellut kohteet,
- liikenne, tuotanto, teollisuus ja kauppa,
- kulttuuri-, liikunta ja vapaa-aika sekä
- muut kulttuurikohteet.

Olen edelleen jättänyt etenkin mallinnusta ajatellen maastokohteista pois pensaikot tai varvikot ja puut, vesistökohteista pois Itämeren ja sen pääosa-alueet, tekojärvet tai -lammet ja tekojärven osat sekä asutuskohteista pois puistot ja torit tai aukiot ja muut asutuskohteet. Tein tällaisen rajauksen, sillä otaksun suuren osan pois rajattujen paikanlajien nimistä olevan

suunniteltuja nimiä (ainakin puistojen ja torien tai aukkioiden nimet) tai nimettyjä paikkoja on vain muutamia (esim. Itämeren ja sen pääosa-alueiden nimiä). Olen lisäksi jättänyt aineistosta pois vielä sellaiset paikanlajit, joita ei esiinny kaikilla murrealueilla (murrealuejaosta ks. luku 3.2). Aineistoni paikanlajiluokat on lueteltu liitteessä (ks. liite 3).

Liitteeseen (ks. liite 3) on myös listattu tässä tutkimuksessa käytetty paikkakategorialuokitus maasto-, vesistö- ja asutusnimiin, joka eroaa hieman Maanmittauslaitoksen kategorialuokituksesta. Nimistöntutkimuksessa on perinteisesti käytetty jakoa luonnonpaikkojen ja kulttuuripaikkojen nimiin. Luonnonpaikkoja ovat kaikki luonnontilaiset paikat, ja kulttuuripaikoiksi luetaan sellaiset paikat, jotka ovat syntyneet ihmisen toiminnan tuloksena. Luonnonpaikkojen nimet on vielä yleensä tapana jakaa maastopaikkojen ja vesistöpaikkojen nimiin ja kulttuuripaikat asutuspaikkojen ja viljelysten nimiin. Jako perustuu siis nimetyn paikan lajiin. (Kiviniemi ym. 1974, 22; 27; 56.) Maanmittauslaitoksen luokista esimerkiksi talot ja pellot ovat selkeimmin kulttuuripaikkoja. Olen kuitenkin jättänyt viljelysnimet eli pellot ja niityt tai laitumet maastopaikkojen luokkaan. Tämä johtuu siitä, että viljelysnimet eroavat maastopaikkojen nimistä rakenteellisesti vähemmän kuin asutuspaikkojen nimistä. Asutuspaikkojen nimet olen jättänyt mallinnuksen ulkopuolelle siksi, että leksikaalisesti moniosaisia nimiä on niissä aika vähän (ks. Kiviniemi 1990, 166–167; ks. myös luku 4). Sen sijaan esimerkiksi alkuperäisessä luokituksessa maastopaikkoihin kuuluvat saaret, syvänteet, rannat ja niemet kuuluvat uudessa kategorialuokituksessa vesistönimiin. Kiviniemi (1990, 58) on paikkoja luokittelevien sanojen luokituksessa käyttänyt tällaisiin paikkoihin viittaavista sanoista nimitystä *veteen rajautuvat paikat*. Varsinaiset vesipaikkojen, kuten järvien tai lampien ja koskien, nimet ja veteen rajautuvien paikkojen nimet olen siis yhdistänyt vesistönimiin.

Paikannimirekisteriin sisältyvä kuntajako 309 kuntaan ja maakuntajako 19 maakuntaan perustuu hallinnollisiin aluejakoihin, ja Maanmittauslaitoksen lisäksi esimerkiksi Tilastokeskus käyttää tilastoinnissa samoja aluejakoja (ks. Tilastokeskus 2025). Kerron seuraavassa alaluvussa 3.2 tarkemmin aineistoon valmiiksi koodattujen aluejakojen ja tässä tutkimuksessa käytetyn aluejaon suhteista. Olen joka tapauksessa rajannut ruotsin- ja kaksikielisten alueiden nimet vielä pois aineistostani (pois rajatut kunnat on lueteltu liitteessä 3). Paikanlaji- ja aluerajausten jälkeen aineistossani on 705 712 nimeä, mikä kattaa edelleen 96,8 prosenttia paikannimirekisterin suomenkielisistä nimistä. Kerron mallinnusta varten tehdystä aineistorajauksesta vielä luvussa 3.4.

Käytän tutkimuksessani myös digitaalista nimiarkistoa, johon on digitoitu keskeisimpiä Suomesta ja Suomen lähialueilta kerättyjä suomen-, saamen- ja ruotsinkielisiä nimiaineistoja (ks. Nimiarkisto). Nimiarkistoon on tallennettu suullista perimätietoa lähinnä maaseudun paikannimistä. Nimestäjät ovat paikannimikokoelmaa varten haastatelleet niin sanottuja nimioppaita, jotka ovat kertoneet kotiseutunsa paikannimistä. Nimestyksen pääasiallinen tarkoitus on ollut tallettaa mahdollisimman vanhaa ja perinnäistä paikannimistöä. Kustakin nimestä on kokoelmaan laadittu nimilippu, jolle on kirjattu muun muassa paikannimen murremuoto, nimetyn paikan sijainti ja nimetyn paikan laji. (Närhi 1978, 55–57; Itkonen 1997, 53–68.) Nimilippuja pääsee digitaalisessa arkistossa tarkastelemaan rekisteröitymällä käyttäjäksi ja kirjautumalla sisään. Arkistosta poimittua aineistoa käytetään Kotuksen kieliaineistojen sääntöjen mukaan (ks. Kotus, Kieliaineistojen käyttö). Sähköiset aineistot ovat sääntöjen puitteissa vapaasti käytettävissä samalla Nimeä 4.0 (CC BY 4.0) -lisenssillä kuin paikannimirekisterikin. Käytän tässä tutkimuksessa nimiarkistoa paikannimien taustan selvittämiseen ja paikannimen osien levinneisyyksien tarkasteluun.

3.2 Murrealueet, maakunnat ja kunnat

Paikannimirekisterin pohjalta yksinkertaisimmat mahdollisuudet aluejakoon ovat kunnat, seutukunnat ja maakunnat, koska kussakin nimessä on jo valmiiksi tieto näistä. Kuntia on aineistossa 309, seutukuntia 69 ja maakuntia 19. Aikaisemmassa tutkimuksessa on tarkasteltu paikannimirekisterin nimiä kunnittain ja maakunnittain (ks. Leino 2004; Salmenkivi ym. 2006). Koska kyseessä on kielitieteellinen tutkimus ja koska tutkimusaineisto koostuu lähinnä perinnäisistä paikannimistä, halusin tarkastella aineistoa murrealueittain. Käytännössä tekemäni jako perustuu rekisterissä valmiina olevaan kuntajakoon. Olen poiminut murrealueisiin yhdistettävät kunnat (ent. pitäjät) Suomen murteiden sanakirjan listauksesta (ks. SMS, Alueet, pitäjät ja murteistot). Ne kunnat, jotka jäävät esimerkiksi nimenmuutoksen takia listauksen ulkopuolelle, olen käynyt läpi yksitellen ja yhdistänyt johonkin oikealla murrealueella olevaan kuntaan (esim. entinen Kymi on yhdistetty Kotkaan, pitäjistä puuttuva Uusikaupunki Kalantiin jne.). Luomani luokitusavain ja ohjelma R-ohjelmointiympäristöön ovat saatavilla verkossa (ks. Virtanen 2025).

Nykyiset kunnat eivät siis täydellisesti vastaa murrealuejaon pitäjiä. Muutoksia on tapahtunut sekä nimissä että alueissa. Suurimmat erot murrealueiden ja kuntien välillä liittyvät siihen, että alueita jää jonkin verran Suomen rajojen ulkopuolelle etenkin Venäjän vastaisella rajalla kaakossa sekä Ruotsin vastaisella rajalla luoteessa. Toinen suurehko eroavaisuus on se, että

Ranua, Vaala, Kokkola, Kauhava ja Jämsä eivät sijoitu kokonaisuudessaan millekään yhdelle kahdeksasta murrealueesta. Ranuaa, Vaalaa ja Jämsää halkoo jopa itä- ja länsimurteiden raja. Kauhava on lähinnä Etelä-Pohjanmaan murteiden mutta hieman myös savolaismurteiden alueella (niin ikään itä-länsi-raja), ja Kokkola on osittain ruotsinkielisellä alueella. Kuten luvusta 2.4 kuitenkin käy ilmi, murrealuejako on kokoelma eri tavoin painotettuja piirteitä ja rajojen määrittämisessä on tukeuduttu kielenpiirteiden lisäksi myös asutushistoriaan. Havaitsemiani entisten pitäjien ja nykyisten kuntien välisiä eroja on koottu liitteeseen (ks. liite 1). Lisäksi liitteenä on kuntia ja murrealueita havainnollistava kartta (ks. liite 2).

Salmenkivi ja kumppanit (2006) tarkastelevat tutkimuksessaan aineistoa maakunnittain. Liitteenä olevasta kartasta (ks. liite 2) voi rajoja vertailemalla havaita, että itä-länsi-jaottelun kannalta etenkin länsirannikolle ulottuva niin sanottu savolaiskiila jää maakuntarajojen pohjalta osaksi Etelä-Pohjanmaan maakuntaa ja pohjoissavolaisten murteiden pohjoisosat jäävät osittain etenkin Pohjois-Pohjanmaan puolelle. Eroja on myös hämäläismurteiden ja savolaismurteiden rajalla sekä lounaismurteiden ja lounaisten välimurteiden rajoilla. Länsirannikon ruotsinkieliset alueet ja kaakkoismurteet sen sijaan erottuvat maakuntajaottelullakin pääpiirteittäin samoin kuin murrealuejaottelussa (pl. ainakin Laihia ja Vaasa länsirannikolla sekä Miehikkälä ja Virolahti kaakossa).

Ruotsinkieliset ja kaksikieliset alueet eivät sisälly perinteiseen murrealuejakoon. Kielisyistä murrealuejaon ulkopuolelle jäävät kunnat on lueteltu liitteessä (ks. liite 1). Luokittelun ulkopuolelle jäävien kuntien alueella on niin vähän suomenkielisiä nimiä (kaikki kategoriat mukaan lukien 6 622 nimeä), että aluetta ei oikein voi verrata muihin, joten olen rajannut sen tutkimuksen ulkopuolelle. Osittain nimien vähyys johtuu siitä, että ruotsinkielisten alueiden paikannimistö on pääosin ruotsinkielistä eikä kaikista nimetyistä paikoista ole olemassa nimeä kaikilla alueen virallisilla kielillä (ks. SPNK, 12). Esimerkiksi Ahvenanmaalla on vain 72 suomeksi nimettyä paikkaa, joista ainoastaan 12 (16,7 %) on luonnonpaikkojen tai asutuspaikkojen nimiä. Ahvenanmaan maakunta tosin poikkeaa muista Suomen ruotsinkielisistä alueista, sillä se on edelleen vahvasti yksikielinen ruotsinkielinen alue, kun muilla alueilla etenkin kaupunkiympäristöt ovat suomenkielistyneet (Pitkänen 2009, 9–10). Kaupunkiympäristöissä on kuitenkin paljon suunniteltuja nimiä. Keskimäärin muualla Suomessa maastokohteiden, vesistöjen ja asutuskohteiden nimiä on 98,8 prosenttia kaikista nimistä. Esimerkiksi juuri Ahvenanmaalla on nimimäärään suhteutettuna poikkeuksellisen paljon erilaisten hallintopaikkojen, liikennepaikkojen ja kulttuurikohteiden nimiä. Ruotsinkielisiä nimiä Ahvenanmaalla on 11 867, ja niistä 11 735 (98,9 %) on

luonnonpaikkojen ja asutuspaikkojen nimiä. Myös Uudellamaalla, Varsinais-Suomessa ja Pohjanmaalla on kuntia, jotka kuuluvat aluejaottelussani ruotsin- ja kaksikielisten ryhmään. Nimien määrä vaihtelee myös murrealueittain. Kerron murrealueiden nimimääristä tarkemmin luvussa 4.1.1.

3.3 Piirteiden erottaminen ja luokittelu

Ennen kuin paikannimien sijaintialuetta voi alkaa ennustaa ohjatun koneoppimisen menetelmillä ja ennen kuin kielenpiirteiden määriä voi tarkastella, aineistosta pitää erottaa piirteitä eli se pitää piirteistää. Piirteistyksen tarkoitus on muuttaa tutkimusaihe eli tässä tapauksessa paikannimien kielenpiirteet mitattavaan ja ohjelman käsiteltävissä olevaan muotoon. (Kyröläinen–Laippala 2020, 498–499.) Käytännössä kustakin nimestä siis tehdään havaintoja, jotka tallennetaan muuttujiin. Havainnot voivat olla joko kategorisia (esimerkiksi nimen *Isosuo* viimeinen leksikaalinen osa on *suo* ja nimen *Ritakankare* viimeinen leksikaalinen osa on *kankare*) tai numeerisia (esimerkiksi *Isosuo*-nimessä on 33,3 % konsonantteja ja *Ritakankare*-nimessä on 54,5 % konsonantteja). Tässä tutkimuksessa on ainoastaan kategorisia havaintoja.

Valitsin erotettaviksi piirteiksi Salmenkiven ja kumppaneiden (2006) mallin mukaisesti paikannimien leksikaaliset viimeiset osat ja nimenalut (ks. luku 2.5), koska niissä on maakunnittaista vaihtelua. Erottelin myös leksikaalista viimeistä osaa edeltävät merkkijonot (yksi ja kaksi merkkiä) ja ryhmittelin leksikaaliset viimeiset osat valittuihin paikanlajiluokkiin (ks. luku 3.4). Nimenalut on hyvin helppo erottaa aineistosta, etenkin jos ei ota huomioon sitä, muodostaako nimenalku leksikaalisen kokonaisuuden vai ei (eli onko nimenalku myös nimielementti). Olen silti aloittanut piirteistyksen leksikaalisista viimeisistä osista, jotta nimenalkuja voi tarkastella niistä erillään. Leksikaalisten viimeisten osien erottelu on kuitenkin työläämpää.

Omat haasteet (ks. luku 6) ja Salmenkiven ja kumppaneiden (2006, 194) kuvaamat haasteet automaattisen lauseenjäsentäjän erottelutarkkuuden kanssa saivat minut valitsemaan leksikaalisten viimeisten osien erottamiseksi vähän enemmän käsityötä vaativan menetelmän. Erottelin siis nimien lopusta listoihin keräämiäni viimeisiä leksikaalisia osia säännöllisillä lausekkeilla. Aloitin kokoelmalla Salmenkiven ja kumppaneiden (2006, 196–197) mainitsema paikan lajiin viittaavia viimeisiä leksikaalisia osia. Erottelin aluksi sellaisia sananmuotoja, joiden tarkistin Nimisampoon tehtävillä hauilla erottuvan oikein tietyt ehdot täyttävinä merkkijonoina. Esimerkiksi *alus* on erotettava vain silloin, kun sitä edeltää *n*, sillä

Taipalus-nimeä ei haluta erotella osiin *Taip* ja *alus*, mutta *Mäenalus* halutaan erotella osiin *Mäen* ja *alus*. Erottelu oli myös tehtävä osissa, sillä sama merkkijono voi olla osa toista merkkijonoa (esim. *aro* sisältyy sanaan *saparo*), jolloin käyttämäni funktio (*stringr*-nimisen R-kirjaston *str_match*; ks. Wickham 2023) erotti aina lyhimmän mahdollisen jonon vertailukohteestaan.

Salmenkivi ja kumppanit (2006, 196–197) esittelevät tutkimuksessaan 398 nimenloppuista lekseemiä, joista jokainen sijoittuu jossain kohdetyypissä (esim. vedenkokouma, virtavesi tai kohouma) 15 yleisimmän sananloppuisen lekseemin joukkoon. Näiden vajaan 400 nimenloppuisen lekseemin avulla olen ensin jaotellut 3 000 ensimmäistä ja viimeistä nimeä koko aineistosta. Tämän osa-aineiston kukin erilainen nimi on käyty läpi, ja alkuperäisiä säännöllisiä lausekkeita on muokattu siten, että niiden avulla leksikaalinen viimeinen osa erottuu oikein 6 000 ensimmäisestä ja viimeisestä nimestä (0,9 % koko aineistosta). Alkuperäisiin listoihin on siis lisätty sananmuotoja tai ehtoja, joiden avulla osa-aineisto on saatu eroteltua viimeiseen leksikaaliseen osaan ja nimen alkuosaan.

Erottelu on tämän jälkeen tehty koko aineistoon, minkä jälkeen on käyty tarkemmin läpi niitä nimiä, jotka eivät erottuneet (eli joista ei löytynyt määriteltyä merkkijonoa lopusta).

Erottomattomia nimiä oli tässä vaiheessa noin puolet koko aineistosta. Jäljelle jääneistä erilaisista nimiasuista koostuvaa listaa on käyty läpi ensin aakkosjärjestyksessä ja sitten käänteisessä aakkosjärjestyksessä. Näin paljastuneita erottomattomia viimeisiä leksikaalisia osia on erotettu aineistosta ja käyty taas listaa läpi, kunnes jäljellä olevassa listassa on ainakin näennäisesti vain yksiosaisia nimiä. Näennäisesti siksi, että olen jättänyt viimeisen leksikaalisen osan erottamatta sanasemanttisesti läpinäkymättömistä tai hyvin hämäristä nimistä sekä selkeästi muunkielisistä nimistä (esim. inarinsaamen *Ákkunämmir*, pohjoissaamen *Livvadievát*, koltantsaamen *Ku'vžžluubbâl* ja ruotsin *Mjölkvägen*, jotka siis rekisteriin on merkitty suomenkielisiksi).

Läpinäkymättömissä ja hyvin hämärissä nimissä vaikuttaa esimerkiksi vokaaliharmonian tai muun fonotaktisen seikan perusteella olevan useampia leksikaalia osia: esimerkiksi taka- ja etuvokaaleita sisältävä *Vaambääru*-niminen saari Porissa tai neljän eri konsonantin klusterin sisältävä *Humpsteeti*-niminen pelto Pyhäjoella. Esimerkiksi lainasanoissa, affektiivisissä ilmauksissa ja murteissa poikkeukset fonotaksiin ovat kyllä melko tavallisia (ks. VISK § 1), ja paikannimissä voi olla kulumisen seurauksena esimerkiksi erikoisia konsonanttiyhdistelmiä. Toisinaan nimet selittyvät kielikontakteilla, kuten *Fofängi* (saari Porissa), joka tulee ruotsin

ilmauksesta *fåfånga* 'turha'. Joissain nimissä on myös ristiriitaisuuksia nimiarkiston ja paikannimirekisterin välillä: esimerkiksi *Arästi* (saari Paltamossa) on nimiarkistossa sekä hakusanana että nimilipussa muodossa *Ärästi* (Nimiarkisto, Aarni 1956). Tällaisia hämäriä nimiä on kuitenkin alle 200, mikä on noin 0,02 prosenttia koko aineistosta.

Aineistossa on myös sellaisia leksikaalisesti moniosaisia nimiä, joissa viimeinen leksikaalinen osa on erisnimi, kuten vaikkapa Juuassa sijaitseva saari *Suuri-Nauha* ja Toholammilla sijaitseva vedenkokouma *Keski-Raippo*. Tällaisia nimiä on yhteensä 14 705. Olen tunnistanut aineistosta tällaiset nimet vasta viimeisen leksikaalisen osan erottamisen jälkeen: nimet ovat siis sellaisia, joista ei löydy määrittelemiäni viimeisiä leksikaalisia osia. Kirjoitusasun perusteella erisnimisen viimeisen osan sisältäviä nimiä on 29 570. Luvussa on siis mukana myös sellaiset nimet kuin vaikkapa Kannonkoskella sijaitseva suo *Pieni Kylkineva*, jossa *neva* on viimeinen leksikaalinen osa. Pelkän erisnimisen viimeisen leksikaalisen osan nimissä viimeisessä leksikaalisessa osassa on lähes yhtä paljon vaihtelua kuin koko aineistossa: uniikkien asujen osuus on jopa 34,3 prosenttia kaikista asuista, kun yleisnimisissä viimeisissä leksikaalisissa osissa uniikkien asujen osuus on vain 0,5 prosenttia. Olen siksi erottanut aineistosta nämä erisnimisen viimeisen leksikaalisen osan sisältävät nimet. Kun siis puhun leksikaalisesti moniosaisista nimistä, tarkoitan niitä nimiä, joissa leksikaalinen viimeinen osa on yleisnimi.

Viimeisten leksikaalisten osien erottelun jälkeen olen palauttanut monikolliset viimeiset osat yksikkömuotoon (esim. *saaret* > *saari*). Kaikki viimeiset osat ovat nominatiivissa, joten käytännössä tämä tarkoitti kaikkien *t*-loppuisten sananmuotojen tunnistamista. Tuloksista piti lisäksi poistaa joitain poikkeuksia (*herännyt*, *katkennut*, *palanut*). Yhteensä monikollisia viimeisiä osia on aineistossa 13 091, ja erilaisia monikollisia viimeisiä osia on aineistossa 370. Ennen monikkomuotoisten viimeisten osien muuttamista yksikkömuotoisiksi erilaisia viimeisiä osia oli 2 982. Erilaisia yksikkömuotoisia viimeisiä osia on koko aineistossa 2 656 (analyysiluvussa ilmoittamani luvut eroavat tässä kerrotuista, koska aineisto on rajattu luvussa 3.1 kuvatulla tavalla vain maasto- ja vesistönimien valittuihin paikanlajeihin).

Leksikaalisten viimeisten osien käsittelyn jälkeen olen nimien kulumisen tarkastelua varten erottanut viimeistä osaa edeltäviä merkkejä. Luokittelin kuluneiksi tapauksiksi sellaiset, joissa viimeistä leksikaalista osaa edeltävä merkki on suomen fonotaksin kannalta täysin epäodotuksenmukainen (*d*, *g*, *h*, *j*, *k*, *m*, *p*, *v*). Lisäksi kävin läpi kaikki erilaiset *l*- ja *r*-tapaukset, joista poimin erilleen kulumattomat (*sammal*, *viher*) ja kuluneet (*käpä*[ä],

mäyr[ä]) muodot. Tapauksia, joissa viimeistä leksikaalista osaa edeltää *n*, *s* tai *t* on niin paljon, että niiden luokittelu ei tässä tutkimuksessa valitettavasti ole mahdollista. Pienen määrän kuluneita tapauksia saa näistäkin tunnistettua, kun etsii viimeistä leksikaalista osaa edeltävät konsonanttiklusterit (esim. *ht* nimessä *Reht|suo*). Valtaosa *n*-loppuisista lienee genetiivimuotoisia, mutta joukossa on myös nominatiiveja (esim. *ahven* nimessä *Ahven|lampi*) ja kuluneita tapauksia (esim. *hein* nimessä *Hein|suo*). *s*- ja *t*-loppuisten jakauma saattaisi erottaa itää ja länttä toisistaan (ks. SPNK, 258), mikäli mallinimeämiseen liittyvät, yleiskieliset ja kuluneet tapaukset saisi eroteltua. Mallintamista varten tapaukset on luokiteltu kahteen luokkaan: erikoiset ja mahdolliset. Erikoiset ovat täysin epäodotuksenmukaisia ja mahdolliset niitä, jotka voivat yleiskielessä esiintyä sanan lopussa. Tämä jättää luokittelun ulkopuolelle siis ne *n*-, *s*- ja *t*-tapaukset, jotka eivät ole osa konsonanttiyhtymää.

Nimenalut olen saanut erottamalla kunkin viimeistä leksikaalista osaa edeltävän nimen alkuosan alusta tietyn määrän merkkejä säännöllisellä lausekkeella omaan muuttujaansa. En erottanut nimenalkua koko nimestä, sillä tällä tavalla se on erillään leksikaalisesta viimeisestä osasta. Olen erottanut nimestä ensimmäisen, kaksi ensimmäistä, kolme ensimmäistä, neljä ensimmäistä ja viisi ensimmäistä merkkiä. Esimerkiksi nimestä *Honkapirtti* taa tällä tavoin nimenalut *H*, *Ho*, *Hon*, *Honk* ja *Honka*. Nimestä *Lankrunni* saa nimenalut *L*, *La* ja *Lan*, mutta neljä merkkiä ja viisi merkkiä jäävät tyhjäksi, sillä *krunki* eli 'matalikko, kari, luoto' on viimeinen leksikaalinen osa (SMS s.v. *krunki*).

3.4 Sijainnin ennustaminen ja muuttujat

Käytän tässä tutkimuksessa *Satunnaismetsä*-nimistä menetelmää, joka on päätöspuihin perustuva ohjatun koneoppimisen menetelmä. Ohjattu koneoppiminen eroaa luokittelutehtävän kannalta ohjaamattomasta siten, että ohjatussa algoritmillemme annetaan valmiit kategoriat sisältävä ennustettava muuttuja. Algoritmi luo sitten ennustettavan muuttujan kategorioista ennustavissa muuttujissa annettujen tietojen perusteella mallin, jolla havaintoja luokitellaan. Ohjaamattomassa koneoppimisessä algoritmi luo ennustavien muuttujien perusteella luokittelumallin vapaasti, ilman ennako-oletuksia ennustettavan muuttujan kategorioista. (Kyröläinen–Laippala 2020, 498–499; 503.) Tässä tutkimuksessa ennustettavan muuttujan kategoriat ovat murrealueita. Ennustavina muuttujina käytän paikanlajeittain ryhmiteltyjä viimeisiä osia, nimenalkuja ja nimien kulumista. Käytetyt aineistot ja R-kielellä kirjoitetut suomeksi kommentoidut skriptit ovat saatavilla verkossa (ks. Virtanen 2025).

Satunnaismetsä-luokittelussa algoritmi siis jakaa päätöspuumenetelmän tavoin aineistoa annetun ennustettavan muuttujan (engl. *response variable*) mukaisiin luokkiin annettujen ennustavien muuttujien (engl. *predicting variable*) perusteella siten, että erot muodostuvat aina mahdollisimman suuriksi aineiston osien välillä (algoritmista tarkemmin ks. Breiman 2001). Käytössäni on R-ohjelmointiympäristön ranger-nimisen kirjaston versio 0.17.0, ja mallin luomiseen käytän ranger-funktiota (ks. Wright–Wager–Probst 2024). Algoritmi jättää osan aineistosta (ns. *out-of-bag*-aineisto tai testiaineisto) sivuun ja luo valitun määrän päätöspuita, joista kukin on muodostettu satunnaisesti valituilla muuttujilla siten, että muuttujien valintahetkellä jokaisella muuttujalla on aina yhtä suuri todennäköisyys tulla valituksi. Lopullinen malli on kokoelma päätöspuita eli eräänlainen metsä. Alussa sivuun jätetty testiaineisto luokitellaan tällä harjoitusaineistolla koulutetulla mallilla, ja lopuksi verrataan mallin tekemää luokittelua ja varsinaisia luokkia. Menetelmä pystyy käsittelemään myös aineistosta puuttuvia arvoja. (Tagliamonte–Baayen 2012, 159–163.) Menetelmän vahvuus on siinä, että se ei ”yilisovitu” (engl. *overfit*), eli yhdellä aineistolla luotu malli toimii pääpiirteittäin yhtä hyvin täysin uuteen samankaltaiseen aineistoon (Breiman 2001, 29). Ranger-funktiossa testiaineisto on oletuksena noin kolmannes aineistosta (Wright ym. 2024, 20). Testiaineiston osuuden voi myös itse määrittellä, mutta olen tässä tutkimuksessa käyttänyt oletusarvoja.

Satunnaismetsämallin *out-of-bag*-ennustusvirhe (engl. *out-of-bag prediction error*) on alussa sivuun jätettyjen havaintojen todellisten luokkien ja mallin väärin ennustamien luokkien suhde (Wright ym. 2024, 24). Tarkkuutta voidaan havainnollistaa tarkemmin niin sanotulla sekaannusmatriisilla (engl. *confusion matrix*). Matriisi on ristitaulukko, josta näkyy, mihin ennustettavan muuttujan luokkiin malli on ennustanut havainnot. Alla olevaan kuvaan (ks. kuva 2) on havainnollistettu, miten sekaannusmatriisia tulkitaan: todelliset luokat kuvataan pystyakselilla ja ennustetut vaaka-akselilla. Sekaannusmatriisista siis selviää

1. montako kohdetta malli on luokitellut oikeisiin luokkiin
2. montako kohdetta malli on luokitellut muihin luokkiin eli tässä tapauksessa, mitkä alueet ovat mallissa keskenään siinä mielessä samanlaisia, että malli on sekoittanut ne keskenään.

ennustetut havainnot →	ennustettu luokkaan 1	ennustettu luokkaan 2
todelliset havainnot ↓		
todellinen luokka 1	luokkaan 1 oikein ennustetut havainnot	luokasta 1 luokkaan 2 ennustetut havainnot
	luokasta 2 luokkaan 1 ennustetut havainnot	luokkaan 2 oikein ennustetut havainnot
todellinen luokka 2		

Kuva 2 Sekaannusmatriisin tulkitseminen

Toinen mallista havainnoitava tieto on se, kuinka tärkeitä eri muuttujat ovat mallin tekemän ennustuksen kannalta (engl. *variable importance*). Tärkeimpien muuttujien tarkastelua on käytetty fennistiikassa ensin suomenoppijoiden akateemisten tekstien piirteiden tutkimiseen (ks. Ivaska 2014) ja ainakin puhutussa kielessä vaihtelevien ilmiöiden taustalla olevien tekijöiden tutkimiseen (ks. Priiki 2016; Väänänen 2015). Selvitän tässä tutkimuksessa muuttujien järjestyksen mutta tarkastelen tarkemmin tärkeimpiä muuttujia koko aineistosta, sillä mallien sisäinen tarkastelu vaatisi paljon laskentatehoa.

Tässä tutkimuksessa pääasiallisia tarkastelun kohteita ovat siis erilaiset viimeiset leksikaaliset osat ja niiden esiintymät eri paikanlajeissa ja eri murrealueilla. Rajaan aineistoni mallinnuksen tarpeisiin vielä leksikaalisesti moniosaisiin maasto- ja vesistönimien ja sellaisiin paikanlajeihin, joita esiintyy kaikilla murrealueilla. Tällaisia nimiä on yhteensä 429 461. Olen yhdistänyt liitteessä 3 mainituista paikanlajeista 1) KOHOUMA-muuttujaan kohoumaryhmän ja kohouman, 2) VILJELYSMAA-muuttujaan pellon ja niityn tai laitumen, 3) SAARI-muuttujaan saari- tai luotoryhmän ja saaren tai luodon sekä 4) VEDENKOKOUMA-muuttujaan järvi- tai lampiryhmän, järven tai lammen ja vesikuopan. Tämä noudattelee pääpiirteittäin esimerkiksi Kiviniemen (1990, 50–52) nimettyjen paikkojen luokittelua. Aineistossani on yhdistetyissä paikanlajeissa myös paljon samoja viimeisiä leksikaalisia osia. Paikanlajimuuttujia on siis yhteensä yhdeksätoista: kahdeksan maastopaikanlajia ja yksitoista vesistöpaikanlajia.

Toinen tarkasteltava kohde ovat nimenalut. Nimenalkumuuttujat korreloivat keskenään hyvin vahvasti: pidemmät nimenalut selittävät kaiken lyhyemmässä tapahtuvan vaihtelun (esim. nimenalku *Suuri* sisältää kaikki lyhyemmät saman nimen nimenalut *Suur*, *Suu*, *Su* ja *S*) ja lyhyempikin selittää osan pidemmissä tapahtuvasta vaihtelusta (esim. jos nimi alkaa *S*, kaikki saman nimen pidemmät nimenalut alkavat aina *S*). Seuraavassa luvussa käsiteltävässä nimenalkumallissa on mukana kaikki nimenalkumuuttujat, jolloin niiden merkitys korostuu. Nimien kuluminen on mukana nimenalkumallissa muuttujana, mutta se ei osoittautunut kovin tärkeäksi.

Päätöspuiden luokittelu perustuu suurimpiin eroihin ennustettavien luokkien välillä, joten havaintomäärien ero luokkien välillä vaikuttaa tuloksiin. Murrealueiden käyttö suoraan ennustettavana muuttujana ei siis ole paras vaihtoehto, sillä nimien määrä vaihtelee lounaismurteiden 13 748 nimestä savolaismurteiden 216 160 nimeen. Koko aineistolla luodussa mallissa ennustukset painottuvat savolaismurteiden alueelle. Menetelmällä saatavia tuloksia on myös vaivattomampaa tulkita, kun kussakin ennustettavan muuttujan luokassa on suunnilleen yhtä monta havaintoa. Olen tästä syystä poiminut ensimmäiseen osa-aineistoon (myöhemmin *pääryhmäaineisto*) jokaiselta murrealueelta 13 500 nimen satunnaisotoksen, ja käyttänyt tätä 108 000 nimen aineistoa kahden erilaisen mallin luomiseen. Otos kattaa tällöin 6,2 prosenttia kaikista savolaismurteiden alueen nimistä, noin neljäsosan hämäläismurteiden, Peräpohjolan murteiden ja Keski- ja Pohjois-Pohjanmaan murteiden alueen nimistä, kaksi kolmasosaa kaakkoismurteiden alueen nimistä, lähes 90 prosenttia lounaisten välimurteiden alueen nimistä ja melkein kaikki Etelä-Pohjanmaan ja lounaismurteiden alueen nimet.

Toiseen osa-aineistoon (myöhemmin *alaryhmäaineisto*) olen valikoinut murrealueiden alaryhmistä sellaiset joukot, joista 13 500 havaintoa kattaa suuremman osan ryhmän nimistä kuin esimerkiksi ensimmäisen osa-aineiston savolaismurteiden 6,2 prosenttia. Tällä tavalla eri luokat ovat paremmin vertailtavissa. Valitsin nimimäärältään suurimpien murrealueiden alaryhmistä sellaiset, jotka ovat lähimpänä 13 500 nimen rajaa. Lounaismurteiden, lounaisten välimurteiden, Etelä-Pohjanmaan murteiden ja kaakkoismurteiden alueen nimet on jo hyvin edustettuna, sillä otoksessa on 68,4–98,2 prosenttia alueiden nimistä. Hämäläismurteiden alueelta poimin Kaakkoishämäläisten murteiden Porvoon ryhmän ja Kaakkoishämäläisten murteiden Kymenlaakson eli Iitin ryhmän (13 500 nimen otos kattaa 96,3 %), Keski- ja Pohjois-Pohjanmaan murteista Keski-Pohjanmaan murteet (otos kattaa 66,8 %), Peräpohjolan murteista Kemin murteet (otos kattaa 47,5 %) ja savolaismurteista Päijät-Hämeen murteet (otos kattaa 89,0 %). Valintaperusteena oli lisäksi saman murrealueen sisäisten alaluokkien

maantieteellinen lähekkäisyys. Ennustettavia luokkia on alaryhmäaineistossakin siis kahdeksan, mutta suurimpien alueiden otos on pienemmältä alueelta. Käytän jatkossa tarkemmin rajatuista alueista murrealueiden alaryhmien nimiä, mutta hämäläismurteiden rajattua aluetta kutsun Hämeen kaakkoismurteiksi kaakkoismurteisiin sekaantumisen välttämiseksi. Sekä Pääryhmäaineiston että alaryhmäaineiston murrealueittaiset ja paikanlajeittaiset nimimäärät käyvät ilmi liitteenä olevasta taulukosta (ks. liite 4).

Tarkastelen tässä tutkimuksessa siis kolmea erilaista mallia. Alla olevassa taulukossa on yhteenveto erilaisista malleista (ks. taulukko 1). Yksinkertaisimmassa mallissa murrealueet ovat ennustettava muuttuja ja paikanlajimuuttujat ennustavia muuttujia. Toiseen malliin on paikanlajien lisäksi otettu ennustaviksi muuttujiksi nimenalut sekä nimen kuluminen. Kolmas malli on luotu alaryhmäaineistolla siten, että murrealueiden ja alaryhmien yhdistelmä on ennustettava muuttuja ja paikanlajit ja tärkein nimenalkumuuttuja ovat ennustavia muuttujia. Käyn läpi kaikkien mallien ennustustarkkuutta, selvitän tärkeimmät muuttujat ja kerron mallien välisistä eroista luvussa 4.2. Liitteenä on pieni aineistoesimerkki (ks. liite 5).

Taulukko 1 Yhteenveto ennustusmalleista ja muuttujista

Malli	Ennustettava muuttuja	Ennustava(t) muuttuja(t) ja muuttujien luokat
Paikanlajimalli	Murrealueiden pääryhmät (8 luokkaa)	Paikanlajimuuttujat: 19 muuttujaa, luokat viimeisiä leksikaalisia osia
Nimenalkumalli	Murrealueiden pääryhmät (8 luokkaa)	Paikanlajimuuttujat: 19 muuttujaa, luokat viimeisiä leksikaalisia osia Nimenalkumuuttujat: viisi muuttujaa, luokat erilaisia nimenalkuja Nimen kuluminen: yksi muuttuja, luokat erikoinen (sis. <i>d, g, h, j, k, m, p, v</i>) ~ mahdollinen (sis. <i>l, n, r</i> sekä <i>s, t</i> osana konsonanttiyhtymää)
Alaryhmäaineisto-malli	Murrealueiden pääryhmät tai alaryhmät (8 luokkaa)	Paikanlajimuuttujat (19 muuttujaa, luokat viimeisiä leksikaalisia osia) Nimenalkumuuttuja (yksi muuttuja, luokat erilaisia nimenalkuja)

Olen toistanut kunkin mallinnuksen viisi kertaa erilaisella jyvityksellä (engl. *seeding*, suomenkielinen käsite Ivaska 2014) eli asettanut satunnaistamisen aloitusarvoksi eri arvon. Aloitusarvon asettamisen idea on siinä, että vaikka malli valikoi muuttujia satunnaisesti, satunnaistus alkaa kerta kerralta samasta kohdasta. Ennustustarkkuudet eivät erilaisilla jyvityksillä eronneet toisistaan kuin korkeintaan prosenttiyksikön kymmenyksillä ja muutamalla kymmenellä osumalla suuntaan tai toiseen, joten käytin arvona kaikissa malleissa lukusarjaa 1234.

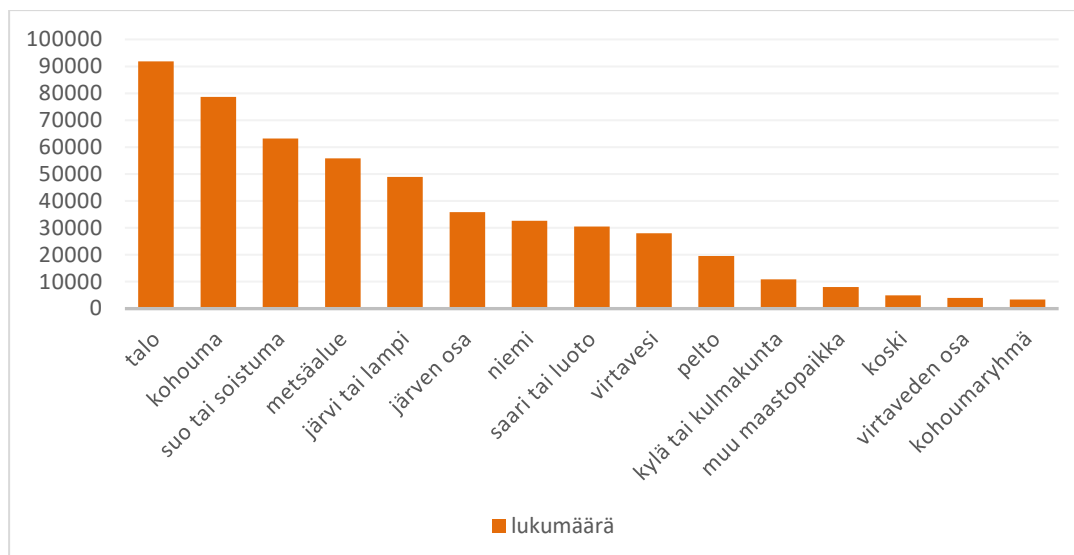
4 Paikannimien alueellinen variaatio

4.1 Paikannimet paikanlajikategorioittain ja murrealueittain

4.1.1 Koko paikannimi

Kaikilla murrealueilla esiintyviin paikannimiin rajatussa 705 712 paikannimen aineistossa on uusien paikanlajikategorioiden (ks. liite 3) mukaisia maastokohteiden nimiä 247 379 (35,5 %), vesistöjen ja vesistöihin rajautuvien kohteiden nimiä 207 110 (29,4 %) ja asutuskohteiden nimiä 250 290 (35,1 %). Mikäli erottaa nimistöntutkimuksen perinteen mukaan peltojen ja niittyjen tai laitumien nimet vielä omaan viljelysnimien kategoriaansa, maasto- ja vesistökohteiden nimien määrä tasoittuu entisestään. Maastokohteiden nimiä on tällöin siis 213 294 (30,3 %) ja viljelyskohteiden nimiä 34 085 (4,8 %). Kiviniemen (1990, 48) nimiarkistosta poimitussa aineistossa nimimäärät ovat maasto-, vesistö-, asutus- ja viljelysnimien kesken hyvin tasaiset: kussakin kategoriassa on noin neljäsosa aineiston nimistä. Kategorisointi tosin eroaa tässä tutkimuksessa käytetystä, sillä tässä tutkimuksessa vesistönimiin on laskettu myös kaikki vesistöjen äärellä sijaitsevien paikkojen nimet. Varsinaisia vesistöjen (vedenkokoumien ja virtavesien) nimiä on tämän tutkimuksen aineistossa 142 867 eli noin viidennes nimistä. Kotuksen vuoden 2010 tilastotietoihin verrattuna etenkin viljelysnimien määrä on laskenut (ks. Kotus, Peruskartan paikannimet; karttanimistön muutoksista ks. Kotus, Karttanimistön tarkistus).

Maastokohteiden kategoriassa on nimetty eniten kohoumia (78 697) ja soita tai soistumia (63 154), vesistökohteiden kategoriassa järviä tai lampia (48 896) ja järven osia (35 792), asutuskohteiden kategoriassa taas taloja (91 899) ja kyliä tai kulmakuntia (10 829). Jäljempänä olevassa kuviossa (ks. kuvio 1) näkyy viidentoista suurimman luokan nimimäärät. Maanmittauslaitoksen paikanlajiluokituksessa ei eroteta toisistaan esimerkiksi tiloja, taloja ja torppia, toisin kuin on tehty esimerkiksi Nimiarkistoon kerättyjen nimien yhteydessä. Taloja ja kyliä tai kulmakuntia on kuitenkin nimetty paikannimirekisteriin suunnilleen samassa suhteessa (karkeasti yhdeksää taloa kohden aina yksi kylä) kuin Kiviniemen (1990, 52) aineistossa asutuksia ja alueita, jotka hänen luokituksessaan ovat yläkategorioita.



Kuvio 1 Viidentoista nimimäärältään suurimman paikanlajiluokan nimimäärät

Nimien määrä vaihtelee paitsi paikanlajeittain myös murrealueittain. Alla olevaan taulukkoon on listattu kunkin murrealueen nimimäärä, nimien osuus kaikkien alueiden nimistä, nimitiheys sekä alueen pinta-ala (ks. taulukko 2). Murrealueittaiset pinta-ala tiedot ja asukasluvutiedot on saatu yhdistämällä Maanmittauslaitoksen tuorein kunnittainen pinta-ala tilasto ja Tilastokeskuksen tuorein kunnittainen asukasluvut tilasto aineistoon (ks. Maanmittauslaitos, Tilastot; Tilastokeskus 2025). Etenkin savolaismurteiden alue erottuu joukosta yli 336 000 nimellä, mikä on lähes puolet koko aineistosta. Savolaismurteiden alue on maantieteellisesti hyvin laaja, ja sen nimitiheys (2,7 nimeä/km²) on hyvin lähellä mediaania (2,8 nimeä/km²). Toiseksi eniten nimiä on hämäläismurteiden alueella, ja siellä nimitiheyskin on piirun verran savolaismurteiden aluetta korkeampi. Lounaismurteiden alueella, kaakkoismurteiden alueella ja Etelä-Pohjanmaan murteiden alueella on muihin alueisiin verrattuna melko vähän nimiä. Maantieteellisesti kaakkoismurteiden alue on kaikkein pienin, ja sen nimitiheys on kaikkein suurin. Lounaismurteiden ja Etelä-Pohjanmaan murteiden alueet taas ovat sen verran suuria, että nimitiheys on lähellä mediaania.

Taulukko 2 Paikannimien murrealueittainen määrä ja alueen nimien osuus kaikista nimistä sekä nimitys ja alueen pinta-ala

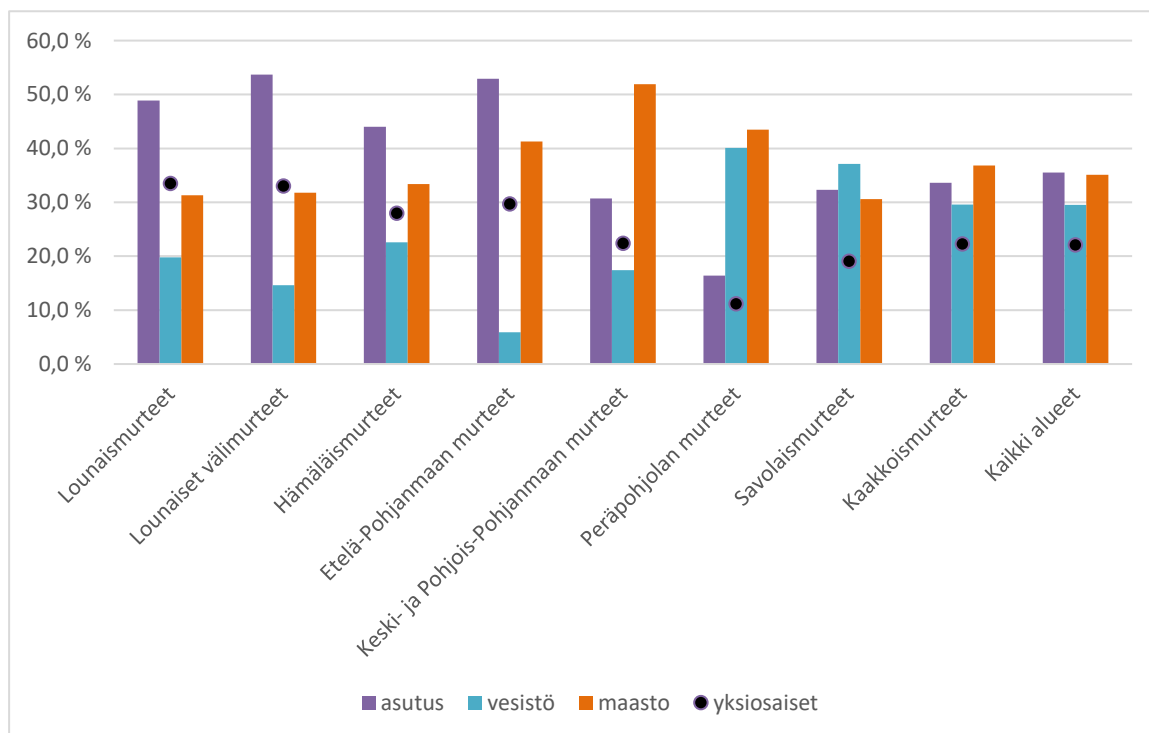
Murrealue	Nimien määrä	Osuus kaikista nimistä (%)	Nimitys (nimeä/km ²)	Alueen pinta-ala (km ²)
1 Lounaismurteet	30 630	4,3	2,7	12 206,68
2 Lounaiset välimurteet	37 117	5,3	3,2	11 989,03
3 Hämmäläismurteet	109 865	15,6	3,0	37 633,97
4 Etelä-Pohjanmaan murteet	31 617	4,5	2,9	11 079,05
5 Keski- ja Pohjois-Pohjanmaan murteet	69 893	9,9	1,6	44 327,53
6 Peräpohjan murteet	58 685	8,3	0,6	93 127,45
7 Savolaismurteet	336 707	47,7	2,7	124 340,11
8 Kaakkoismurteet	31 198	4,4	4,0	7 871,47
Yhteensä	705 712	100	(Md) 2,8	34 2575,29

Kiviniemi (1990, 49) esittää, että asutukseltaan tiheillä viljelyalueilla nimitys on korkeampi kuin harvaan asutetuilla alueilla. Murrealueet ovat kuitenkin sen verran laajoja, että käytännössä ainoastaan Peräpohjolan murteiden alue ja Keski- ja Pohjois-Pohjanmaan murteiden alue erottuvat alhaisella nimitysasteella muista alueista. Peräpohjolan murteiden alueen nimitysasteeseen vaikuttanee se, että pohjoinen Suomi on kokonaisuudessaan hyvin harvaan asuttua. Lisäksi alueen maisemassa on nähtävästi nimetty hyvin laajoja kohteita. Alueella on toki myös jonkin verran inarin-, koltan- ja pohjoissaamenkielisiä nimiä, mikä edelleen vähentää suomenkielisten nimien määrää. Keski- ja Pohjois-pohjanmaan murteiden alue on myös harvaan asuttua seutua, jonka maisemassa on laajoja alavia alueita.

Aineistoni nimistä 76,8 prosenttia (534 914 nimeä) on leksikaalisesti moniosaisia. Osuus on lähes sama kuin Salmenkiven ja kumppaneiden (2006, 195) raportoima 77,2 prosenttia. Moniosaisuus on siis huomattavasti tavallisempaa kuin yksiosaisuus. Syntaktisesti yksiosaisia nimiä on toki suurempi osuus, sillä leksikaalinen viimeinen osa ei aina viittaa paikan lajiin. Esimerkiksi talojen nimissä yleisimmät viimeiset leksikaaliset osat ovat *mäki*, *niemi*, *aho* ja *ranta*, joiden jälkeen järjestyksessään viidentenä on vasta asumukseen viittaava *talo*.

Syntaktisesti moniosaisia nimiä ovat nimittävän osan sisältävät paikannimet, joista osan tunnistaa paikannimirekisterissä kirjoitusasusta (epeksegeettisistä nimittävän osan sisältävistä nimistä ks. Kiviniemi 1990, 102; 120). Yleisnimiseen viimeiseen leksikaaliseen osaan jakamattoman erisnimisen viimeisen leksikaalisen osan sisältäviä nimiä (esim. puumalalainen saari *Pieni-Hansanen*) on aineistossa 14 705. Kirjoitusasun perusteella yhteensä 29 570 paikannimessä (0,4 % kaikista nimistä) on viimeisenä osana erisnimi. Tällaisia ovat esimerkiksi lohjalainen vedenkokouma *Pikku Kolmpersjärvi*, jossa viimeinen leksikaalinen osa on *järvi*, ja uusikaupunkilainen saari *Vähä Harmaaletto*, jossa viimeinen leksikaalinen osa on *letto*.

Alueen nimimäärään suhteutettuna eniten leksikaalisesti yksiosaisia nimiä on lounaismurteiden (33,5 %) ja lounaisten välimurteiden (33,0 %) alueella. Alla olevasta kuvioista (ks. kuvio 2) näkyy, että myös hämäläismurteiden (28,0 %) ja Etelä-Pohjanmaan murteiden (29,7 %) alueen nimistä noin kolmannes on yksiosaisia. Syynä lienee ainakin asutusnimien suuri osuus, sillä alueiden paikannimistä noin puolet on asutusnimiä ja aineistoni kaikista asutusnimistä yli puolet (53,9 %) on leksikaalisesti yksiosaisia.



Kuvio 2 Murrealueittainen ja kaikkien alueiden asutus-, vesistö- ja maastonimien sekä yksiosaisen nimien osuus

Reilu neljäsosa (26,6 %) kaikista nimistä eli 187 766 nimeä on uniikkeja, eli nimet esiintyvät aineistossa ainoastaan kerran. Loput nimet toistuvat rekisterissä 2–2 169 kertaa siten, että

erilaisia kirjoitusasuja on 61 955. Yhteensä erilaisia asuja on siis 247 689. Asutusnimissä uniikkeja nimiä on vain 14,8 prosenttia (36 984 nimeä), mutta maasto- ja vesistönimissä niitä on enemmän, jopa selvästi yli kolmannes (37,1 % eli 168 736 nimeä). Yleisimmät toistuvat nimet ovat *LA*-johtimisia asutusnimiä. Viiden kärjessä ovat *Koivula*, *Mäkelä*, *Peltola*, *Rantala* ja *Ahola*, kaikki yli 1 500 osumalla. Vasta viidennellätoista sijalla on kaksiosainen *Uusitalo*, joka sekin on asutusnimi. Valtaosa sijalla 19 olevista kohoumien nimiltä näyttävistä *Koivumäki*-tapauksistakin on asutusnimiä. Ensimmäinen varsinainen luonnonpaikan nimi on *Isosuo* 27. sijalla ja 714 osumalla ja toinen *Palokangas* 35. sijalla ja 616 osumalla.

Alueittain tarkasteltuna eniten uniikkeja nimiä on Peräpohjolan murteiden alueella: jopa 65 prosenttia alueen nimistä on uniikkeja. Syynä Peräpohjolan murteiden alueen suurelle uniikkien nimien määrälle lienee nimettyjen luonnonpaikkojen suuri määrä, mutta kontaktit kolmeen eri saamen kieleen ja ruotsiin voivat myös vaikuttaa uniikkeihin nimiin (ks. myös Torniojoki- ja Kemijokilaakson asuttamisesta Vahtola 1980). Kaakkoismurteiden ja lounaismurteiden alueellakin on suhteessa alueen nimimäärään paljon vain kerran esiintyviä nimiasuja: kaakossa uniikkeja nimiä on 58,8 prosenttia ja lounaassa 55,3 prosenttia. Osittain määrät selittynevät kielikontakteilla: lounaassa on vaikuttanut etenkin ruotsi mutta myös viro (ks. Pitkänen 2007, 9–10; Wiik 2004, 262–263) ja kaakossa ainakin ruotsi ja karjala (ks. Wiik 2004, 350–352). Uniikkien nimiasujen määrään voi kaakkoismurteiden alueella vaikuttaa myös maasto- ja vesistönimien suuri osuus. Toisaalta savolaismurteiden alueella on kaikkein vähiten uniikkeja nimiä (35,8 %), vaikka siellä maasto- ja vesistönimiä on paljon. Syynä voi olla alueen valtavan koon ja nimitiheyden suhde: kaikki murrealueet kattavassa aineistossakaan ei ole kuin reilu neljännes uniikkeja asuja.

4.1.2 Yleisimmät viimeiset leksikaaliset osat

Kaikilla murrealueilla esiintyviin paikanlajeihin rajatussa moniosaisissa maasto- ja vesistönimissä kattavassa 429 461 nimen aineistossa on yhteensä 2 357 erilaista viimeistä leksikaalista osaa. Näistä lähes puolet (1 053) on uniikkeja, eli ne esiintyvät aineistossa vain kerran. Uniikit viimeiset leksikaaliset osat ovat monet hyvin erikoistuneita paikanlajin tai ominaisuuden ilmauksia, kuten *heinistö*, *liimo* 'nuori lehtipuumetsä' (ks. SMS) tai *kierre*. Ne voivat myös olla yleisempien ilmausten äänteellisiä muunnelmia: esimerkiksi 'saarta' merkitsevän *karta*-sanana (ks. SMS) varianttina esiintyy kerran *karto* ja *koura*-sanana varianttina *kopra*. Useampaan kertaan esiintyvät viimeiset leksikaaliset osat esiintyvät aineistossa 2–40 765 kertaa. Viiden kärjessä ovat *suo* (40 765), *lampi* (32 502), *saari* (32 316), *niemi* (29 237) ja

mäki (28 332). Viisi yleisintä viimeistä leksikaalista osaa kattaa kaikista havainnoista 38 prosenttia. Liitteessä 3 mainitun kategorisoinnin asutusnimet sisältävän 705 712 nimen aineiston viisi yleisintä viimeistä leksikaalista osaa ovat järjestyksessään *mäki* (46 708), *suo* (42 208), *niemi* (37 888), *saari* (34 083) ja *lampi* (33 539). Osat ovat siis samat, mutta keskinäinen järjestys erilainen, sillä asutusnimiä on nimetty paljon etenkin mäiksi: esimerkiksi *Koivumäki* on tavallisemmin talo kuin kohouma.

Mainitut viimeiset leksikaaliset osat ovat myös yleisimmät luonnonpaikkoja luokittelevat sanat Kiviniemen (1990, 108–109) aineistossa ja yleisimmät nimenloppuiset lekseemit Salmenkiven ja kumppaneiden (2006, 195) aineistossa. Salmenkivi ja kumppanit (2006, 195–198) ovat jo selvittäneet lähes samalla aineistolla hyvin kattavasti, mitkä nimenloppuiset lekseemit ovat yleisimpiä kussakin paikanlajissa ja miten maakunnat eroavat toisistaan. Tarkastelen tässä tutkimuksessa kuitenkin viittä yleisintä viimeistä leksikaalista osaa ensin kaikkia esiintymiä paikanlajeittain ja sen jälkeen murrealueittaisia esiintymiä niin ikään paikanlajeittain. Käsittelen jatkossa siis vain liitteessä 3 mainitut maasto- ja vesistönimien paikanlajit sisältävää aineistoa.

Kolmannella sijalla oleva *saari* on jakautunut kaikkein tasaisimmin niiden paikanlajiluokkien kesken, joissa sillä on suurin esiintymistaajuus. Useimmin (23 512 kertaa) *saari* esiintyy saaren tai luodon nimessä eli paikan lajiin viittaavana, mutta se on viimeinen leksikaalinen osa myös lähes 5 000 metsän nimessä mahdollisesti metaforisessa 'saarekkeen' merkityksessä, ellei kyseessä sitten ole siirrynnäinen. Myös kohoumien ja niemien nimissä on usein *saari* viimeisenä leksikaalisena osana. Sen sijaan *suo* ja *lampi* ovat tiukemmin sidoksissa yhteen paikanlajiin: *suo* esiintyy suoalueiden nimissä 34 487 kertaa (81,7 % tapauksista) ja *lampi* vedenkokoumien nimissä 30 917 kertaa (92,2 % tapauksista). Toiseksi eniten *suo* esiintyy viljelysmaiden nimissä (4 079 kertaa): mahdollisesti siirrynnäisenä tai muistona aiemmasta paikanlajista. Se on joka tapauksessa viljelysmaiden nimissä yleisin viimeinen leksikaalinen osa. *lampi* taas esiintyy toiseksi eniten (812 kertaa) viimeisenä leksikaalisena osana suoalueiden nimissä, mahdollisesti samasta syystä kuin pelloissa esiintyy *suo*. *mäki* ja *niemi* ovat (huom. asutusnimiä lukuun ottamatta) rajautuneet melko tarkasti ilmaisemaan paikan lajia: *mäki* kohoumaa 27 618 esiintymällä ja *niemi* niemeä 27 394 esiintymällä.

Murrealueittain tarkasteltuna *mäki* on erityisen harvinainen Peräpohjolan murteiden ja Keski- ja Pohjois-Pohjanmaan murteiden alueella. Alla olevaan taulukkoon (ks. taulukko 3) on

korostettu alimmalle riville *mäki*-tapausten osuus kunkin murrealueen nimistä suurimmasta pienimpään eli sinisestä valkoisen kautta punaiseen. Peräpohjolan murteiden alueella kohoumien viimeinen leksikaalinen osa on yleisimmin *vaara* tai *kumpu*. Keski- ja Pohjois-Pohjanmaan murteiden alueellakin yleisempiä ovat *kangas* ja *saari*. Peräpohjolan murteiden alueella yleisimmät viimeiset leksikaaliset osat ovat siis paikan lajiin viittaavia, mutta Keski- ja Pohjois-Pohjanmaan murteiden alueella yleisimmät osat ovatkin paikkaa kuvailevia. Toki järjestyksessään seuraavaksi yleisimmät *kallio*, *harju* ja *vaara* ovat Keski- ja Pohjois-Pohjanmaan murteiden alueellakin paikan lajiin viittaavia, vaikka etenkin *kallio* voidaan ajatella myös maa-aineksen laatua kuvaavana. Vaikka kaikista *mäki*-tapauksista yli kolmannes on savolaismurteiden alueella, kaikkiin alueen moniosaisiin maasto- ja vesistönimisiin suhteutettuna *mäki* on alueella melko harvinainen. Etelä-Pohjanmaan murteiden alueella taas on alueen nimien määrään suhteutettuna eniten *mäki*-osaan loppuvia nimiä, vaikka kaikista *mäki*-tapauksista alueelle osuu alle kymmenesosa.

Taulukko 3 Viimeisen leksikaalisen osan *mäki* sisältävät nimet murrealueittain

Paikanlaji ↓ Murrealue →	Lounaismurteet	Lounaiset välimurteet	Hämäläismurteet	Etelä-Pohjanmaan murteet	Keski- ja Pohjois-Pohjanmaan murteet	Peräpohjolan murteet	Savolaismurteet	Kaakkoismurteet	Yhteensä
kohouma	1 825	2 335	7 448	2 312	449	49	10 057	3 143	27 618
metsäalue	74	61	55	201	13	6	50	16	476
viljelysmaa	23	18	41	27	5	0	55	7	176
muu maastopaikka	7	6	8	10	2	0	14	4	167
niemi	1	1	3	0	0	0	1	1	7
saari tai luoto	0	1	1	2	0	0	0	0	4
yhteensä	1 930	2 422	7 556	2 552	469	55	10 177	3 171	28 332
osuus <i>mäki</i> -tapauksista (%)	6,8	8,5	26,7	9,0	1,7	0,2	35,9	11,2	100
osuus alueen nimistä (%)	14,3	16,0	13,5	18,6	1,0	0,1	4,8	16,6	

Alla olevasta taulukosta (ks. taulukko 4) näkyy, että Etelä-Pohjanmaan ja Peräpohjolan murteiden alueella on hyvin vähän *suo*-loppuisia paikannimiä. Etelä-Pohjanmaan murteiden alueella suon tai soistuman viimeisenä leksikaalisena osana on yleisimmin *neva*, *rämäkkä*, *keidas* tai *korpi* ja Peräpohjolan murteiden alueella *jänkä*, *aapa*, *vuoma* tai *jänkkä*. Savolaismurteiden alueella *suo* sen sijaan on valtavan yleinen sekä lukumääräisesti että alueen nimimäärään suhteutettuna. Kaakkoismurteiden alueella tapauksia on lukumääräisesti

arvioiden melko vähän, mutta suhteellisesti lähes yhtä paljon kuin savolaismurteiden alueella. Kaikilla alueilla *suo* esiintyy toiseksi useiten viimeisenä leksikaalisena osana viljelysten nimissä. Etenkin lounaismurteiden, lounaisten välimurteiden ja kaakkoismurteiden alueella viljelysten osuus *suo*-tapauksista on suuri. Lukumäärät eivät ole kovin suuria (lounaisten välimurteiden 258 – kaakkoismurteiden 320), mutta reilusti yli kymmenesosa kaikista alueiden *suo*-tapauksista on viljelysnimiä.

Taulukko 4 Viimeisen leksikaalisen osan *suo* sisältävät nimet murrealueittain

Paikanlaji ↓ Murrealue →	Lounaismurteet	Lounaiset välimurteet	Hämäläismurteet	Etelä-Pohjanmaan murteet	Keski- ja Pohjois-Pohjanmaan murteet	Peräpohjan murteet	Savolaismurteet	Kaakkoismurteet	Yhteensä
suo tai soistuma	1 152	1 178	4 480	17	3 534	220	22 051	1 855	34 487
viljelysmaa	258	226	628	3	282	7	2 355	320	4 079
muu maastopaikka	28	28	134	0	49	5	1 252	205	1 701
metsäalue	13	19	200	0	31	1	132	10	406
painanne	2	2	9	0	0	0	62	10	85
vedenkokouma	4	1	0	0	0	0	1	0	6
ranta-alue	0	1	0	0	0	0	0	0	1
yhteensä	1 457	1 455	5 451	20	3 896	233	25 853	2 400	40 765
osuus <i>suo</i> -tapauksista (%)	3,6	3,6	13,4	0,0	9,6	0,6	63,4	5,9	100,0
osuus alueen nimistä (%)	10,6	9,4	9,5	0,1	8,6	0,5	11,6	11,5	

Etelä-Pohjanmaan murteet ja lounaiset välimurteet erottuvat alla olevaa taulukkoa (ks. taulukko 5) tarkasteltaessa muista alueista sillä, että paikannimien viimeisenä leksikaalisena osana on *niemi* selvästi harvemmin kuin muualla. Vaikka lounaismurteiden alueellakin on nimimäärään suhteutettuna vähän *niemi*-tapauksia, se on edelleen yleisin viimeinen leksikaalinen osa niemien nimissä, joskin hyvin yleisiä ovat myös *kari* ja *nokka*. Lounaisten välimurteiden alueella *nokka* on jopa hieman yleisempi kuin *niemi*. Etelä-Pohjanmaan murteiden alueella taas on ylipäänsä todella vähän nimettyjä niemiä, ja *niemi* onkin viimeisenä leksikaalisena osana neljäsosassa tapauksista metsän nimessä ja kymmenesosassa tapauksista kohouman nimessä. Keski- ja Pohjois-Pohjanmaan alueellakin *niemi* on muihin alueisiin verrattuna harvemmin niemien viimeinen leksikaalinen osa, ja siellä yli viidesosa *niemi*-tapauksista on metsien nimissä. Vesistörikkailta savolaismurteiden ja

kaakkoismurteiden alueilla *niemi* esiintyy viimeisenä leksikaalisena osana alueen nimimäärään suhteutettuna useammin kuin muualla: osuus on lähes kymmenen prosenttia.

Taulukko 5 Viimeisen leksikaalisen osan *niemi* sisältävät nimet murrealueittain

Paikanlaji ↓ Murrealue →	Lounaismurteet	Lounaiset välimurteet	Hämäläismurteet	Etelä-Pohjanmaan murteet	Keski- ja Pohjois-Pohjanmaan murteet	Peräpohjolan murteet	Savolaismurteet	Kaakkoismurteet	Yhteensä
niemi	367	219	2 344	127	960	1 875	19 725	1 777	27 394
metsäalue	5	30	67	58	307	65	312	79	923
kohouma	7	9	28	25	124	14	333	12	552
viljelysmaa	8	18	26	18	38	2	46	12	168
saari tai luoto	1	0	5	2	8	40	72	6	134
muu maastopaikka	0	0	8	0	7	3	20	2	40
suo tai soistuma	0	2	0	0	6	3	6	0	17
matalikko	0	0	0	0	1	5	1	0	7
louhikko tai kivikko	0	0	0	0	1	0	0	0	1
ranta-alue	0	0	0	0	0	1	0	0	1
yhteensä	388	278	2 478	230	1 452	2 008	20 515	1 888	29 237
osuus <i>niemi</i> -tapauksista (%)	1,3	1,0	8,5	0,8	5,0	6,9	70,2	6,5	100,0
osuus alueen nimistä (%)	2,9	1,8	4,4	1,7	3,2	4,2	9,7	9,9	

Edellä todettiin jo, että *saari* on jakautunut yleisimmistä viimeisistä leksikaalisista osista kaikkein tasaisimmin sen yleisimpiin paikanlajeihin. Alla olevasta taulukosta (ks. taulukko 6) voidaan nähdä, että murrealueittain tarkasteltuna Etelä-Pohjanmaan ja Keski- ja Pohjois-Pohjanmaan murteiden alueella *saari* esiintyy kohoumien ja metsäalueiden nimissä jopa useammin kuin saarien tai luotojen nimissä. Etelä-Pohjanmaan alueella tosin on ylipäänsä nimetty hyvin vähän saaria tai luotoja. Etelä-Pohjanmaalla metsäalueissa esiintyy *saari* leksikaalisena viimeisenä osana 568 kertaa, mikä on lähes puolet kaikista alueen *saari*-tapauksista. Lounaismurteiden alueella saarien tai luotojen nimien viimeinen leksikaalinen osa on *saarta* yleisemmin *kari* tai *luoto*, ja lounaisten välimurteiden alueellakin *kari* on hieman *saarta* yleisempi. Paikan lajiin viittaavana *saari* on kaikkein yleisimmän järjestyksessään Peräpohjolan murteiden, savolaismurteiden, hämäläismurteiden ja kaakkoismurteiden alueella. Näillä alueilla on myös eniten *saari*-esiintymiä.

Taulukko 6 Viimeisen leksikaalisen osan *saari* sisältävät nimet murrealueittain

Paikanlaji ↓ Murrealue →	Lounaismurteet	Lounaiset välimurteet	Hämäläismurteet	Etelä-Pohjanmaan murteet	Keski- ja Pohjois-Pohjanmaan murteet	Peräpohjolan murteet	Savolaismurteet	Kaakkoismurteet	Yhteensä
saari tai luoto	130	282	3 592	113	726	1 540	15 926	1 203	23 512
metsäalue	91	94	382	568	2 199	161	1 285	199	4 979
kohouma	51	39	174	242	1 014	27	602	76	2 225
niemi	10	30	353	13	57	20	710	105	1 298
viljelysmaa	10	4	25	31	31	0	23	4	128
muu maastopaikka	3	3	14	12	34	15	17	2	100
matalikko	0	0	0	0	3	26	14	1	44
suo tai soistuma	1	0	2	2	1	0	10	0	16
vesikivi tai -kivikko	0	0	0	0	2	3	8	0	13
louhikko tai kivikko	0	0	0	0	1	0	0	0	1
yhteensä	296	452	4 542	981	4 068	1 792	18 595	1 590	32 316
osuus <i>saari</i> -tapauksista (%)	0,9	1,4	14,1	3,0	12,6	5,5	57,5	4,9	100,0
osuus alueen nimistä (%)	2,2	3,0	8,1	7,1	9,0	3,8	8,8	8,3	

Viidenneksi yleisin viimeinen leksikaalinen osa *lampi* esiintyy kaikilla murrealueilla melko johdonmukaisesti lähinnä vedenkokoumien nimissä. Alla olevasta taulukosta (ks. taulukko 7) näkyy, että kaikkein johdonmukaisimmin *lampi* on paikan lajiin viittaavana Peräpohjolan murteiden, savolaismurteiden ja kaakkoismurteiden alueella. Lounaismurteiden alueella ja Etelä-Pohjanmaan murteiden alueella *lampi* on kuitenkin hyvin harvinainen viimeinen leksikaalinen osa. Lounaismurteiden, lounaisten välimurteiden ja hämäläismurteiden alueella vedenkokouman viimeinen leksikaalinen osa on *lampea* useammin *järvi* tai *lammi*. Lounaismurteiden alueella jopa leksikaalisesti yksiosaisia vedenkokoumien nimiä on enemmän kuin *lampi*-loppuisia. Etelä-Pohjanmaan murteiden ja Peräpohjolan murteiden alueella vedenkokouman viimeinen leksikaalinen osa on yleisimmin *järvi*, joka on kaikkien viimeisten leksikaalisten osien yleisyyslistalla kahdeksantena. Savolaismurteiden alueella on kuitenkin valtava määrä *lampi*-loppuisia nimiä (23 166) ja verrattain vähän *järvi*-loppuisia (5 083), joten *lampi* kipuaa yleisyyslistalla hyvin korkealle. Joka tapauksessa kaikilla alueilla pieni osa *lampi*-tapauksista on soiden tai soistumien nimien viimeisenä leksikaalisena osana.

Taulukko 7 Viimeisen leksikaalisen osan *lampi* sisältävät nimet murrealueittain

Paikanlaji ↓ Murrealue →	Lounaismurteet	Lounaiset välimurteet	Hämäläismurteet	Etelä-Pohjanmaan murteet	Keski- ja Pohjois-Pohjanmaan murteet	Peräpohjolan murteet	Savolaismurteet	Kaakkoismurteet	Yhteensä
vedenkokouma	26	166	1 208	137	1 862	3 046	23 166	1 306	30 917
suo tai soistuma	5	9	71	15	148	33	468	63	812
vedenkokouman osa	0	2	17	0	11	15	363	12	420
virtaveden osa	0	2	6	4	104	27	103	1	247
muu maastopaikka	0	0	4	1	17	1	23	5	51
viljelysmaa	0	0	2	1	8	1	18	4	34
metsäalue	0	0	3	0	5	0	0	1	9
painanne	0	0	0	0	1	0	3	1	5
virtavesi	0	0	0	1	4	0	0	0	5
kohouma	0	0	0	0	0	0	1	0	1
yhteensä	31	179	1 312	159	2 160	3 123	24 145	1 393	32 502
osuus <i>lampi</i> -tapauksista (%)	0,1	0,6	4,0	0,5	6,6	9,6	74,3	4,3	100,0
osuus alueen nimistä (%)	0,2	1,2	2,3	1,2	4,8	3,8	8,8	8,3	

Murrealueet siis näyttäisivät toisaalta eroavan toisistaan ja toisaalta yhtenevän toistensa kanssa. Monet yhtenevät seikat esiintyvät alueilla, jotka rajautuvat toisiinsa: Etelä-Pohjanmaan murteiden alueella on yhteisiä piirteitä ainakin Keski- ja Pohjois-Pohjanmaan murteiden alueen kanssa (esim. *saari* esiintyy usein metsäalueiden nimissä), lounaismurteiden ja lounaisten välimurteiden alueet näyttäisivät jakavan ilmiöitä (esim. *kari* yleinen saarten tai luotojen nimissä), ja Peräpohjolan ja Keski- ja Pohjois-Pohjanmaan murteiden alueetkin ovat joissain tapauksissa samanlaisia (esim. *mäki* ei ole yleisin kohouman nimissä). Kielen leksikaaliseen vaihteluun liittyvä ero on esimerkiksi se, että *suo* esiintyy ani harvoin Etelä-Pohjanmaan murteiden ja Peräpohjolan murteiden alueella, koska siellä soissa ja soistumissa on muita paikan lajiin viittaavia viimeisiä osia. Lounaismurteiden ja lounaisten välimurteiden alueen *lammi*-sanon yleisyys *lampi*-sanon kustannuksella taas on morfologista vaihtelua. Joidenkin piirteiden esiintyminen liittyy asutustiheyteen ja luonnonmaantieteeseen: esimerkiksi tiheään asutetuilla lounaismurteiden, lounaisten välimurteiden ja kaakkoismurteiden alueilla *suo* esiintyy usein viljelysten nimissä ja vesistörikkaalla savolaismurteiden alueella taas *lampi* on yleinen viimeinen leksikaalinen osa. Jotkut ilmiöt rajautuvat yksittäisille alueille johtuen niiden luonnonmaantieteestä: Etelä-Pohjanmaan

murteiden alueen vähävetisyys esimerkiksi vaikuttaa *saarien ja niemien* esiintymistäajuuteen yleensä ja myös jakautumiseen paikanlajeittain.

4.1.3 Yleisimmät kolmen merkin nimenalut

Seuraavassa luvussa (ks. luku 4.2) käsiteltävän nimenalkumallin tulosten perusteella tärkein nimenalkumuuttuja on nimen kolme ensimmäistä merkkiä. Tarkastelen tässä luvussa yleisimpien kolmen merkin nimenalkujen suhteellisen esiintyvyyden jakautumista murrealueittain. Malli perustuu absoluuttisiin esiintymistäajuuksiin, mutta se on luotu murrealueittain tasatulla aineistolla. Koko aineistossa suhteellinen esiintyvyys havainnollistaa alueita paremmin, sillä nimimäärät vaihtelevat alueittain. Kaikkien moniosaisten maasto- ja vesistönimien tarkastelussa yleisimpiä kolmen merkin nimenalkuja ovat *Iso-*, *Kor-*, *Pie-*, *Pit-*, *Mus-*, *Kar-*, *Kiv-*, *Kal-*, *Pal-* ja *Koi-*. Murrealueittain yleisimmissä kolmen merkin nimenaluissa ovat lisäksi *Haa-*, *Hau-*, *Hei-*, *Kot-*, *Kuu-*, *Mus-*, *Myl-*, *Per-*, *Pik-*, *Suo-*, *Suu-*, *Väh-* ja *Yli-*. Yleisimmät nimenalut kattavat aineiston nimistä noin viidenneksen.

Alla olevaan taulukkoon (ks. taulukko 8) on listattu aakkosjärjestyksessä yleisimmät nimenalut ja niiden suhteellinen esiintymistäajuus (nimenalkujen määrä / murrealueen nimien määrä * 100) murrealueittain. Taulukon oikeanpuoleinen sarake esittää kunkin nimenalun yleisyyden suhteessa kaikkien alueiden nimiin. Alimmalla rivillä näkyy, kuinka suuren osan kunkin murrealueen nimistä listatut nimenalut kattavat. Taulukossa on korostettu esiintymistäajuudet punaisesta valkoisen kautta siniseen siten, että tummin punainen merkitsee pienintä (0,03 kaakkoismurteiden *Väh*) ja tummin sininen suurinta (3,43 lounaismurteiden *Iso*) esiintymistäajuutta ja valkoinen on mediaanin kohdalla (0,88 esim. kaakkoismurteiden *Kal* ja lounaisten välimurteiden *Koi*). Kaikki alueet -sarake on korostettu itsenäisenä vaaleasta tummaan siten, että tummempi arvo on aina suurempi.

Taulukko 8 Yleisimpien kolmen merkin nimenalkujen suhteellinen yleisyys murrealueittain

Murrealue → Nimenalku ↓	Lounaismurteet	Lounaiset välimurteet	Hämäläismurteet	Etelä-Pohjanmaan murteet	Keski- ja Pohjois-Pohjanmaan murteet	Peräpohjojan murteet	Savolaismurteet	Kaakkoismurteet	Kaikki alueet
<i>Haa</i>	0,94	0,74	0,83	0,95	1,06	0,59	0,76	0,70	0,79
<i>Hau</i>	0,55	0,81	0,90	2,44	1,57	0,61	0,87	0,84	0,96
<i>Hei</i>	0,45	0,60	0,70	0,49	0,59	0,60	0,93	0,64	0,77
<i>Iso</i>	3,43	2,85	2,27	2,40	3,26	2,12	1,77	0,21	2,07
<i>Kal</i>	0,97	1,00	0,99	0,85	1,04	1,06	1,12	0,88	1,06
<i>Kar</i>	1,21	1,21	1,30	1,07	1,17	1,32	1,07	0,99	1,15
<i>Kiv</i>	0,64	0,96	1,06	1,64	1,53	0,87	1,04	1,06	1,08
<i>Koi</i>	0,95	0,88	0,98	0,78	1,01	0,77	1,00	1,05	0,96
<i>Kor</i>	1,21	1,48	1,53	2,51	1,95	1,31	1,32	1,27	1,45
<i>Kot</i>	0,36	0,71	0,66	1,11	1,12	1,05	0,79	0,71	0,83
<i>Kuu</i>	0,65	0,56	0,32	0,44	0,63	1,18	0,27	0,36	0,45
<i>Mus</i>	1,40	1,25	1,17	0,71	0,98	0,80	1,26	1,50	1,17
<i>My</i>	1,43	1,12	1,02	0,58	0,71	0,45	0,82	0,89	0,82
<i>Pal</i>	1,29	1,20	0,88	1,27	1,23	1,57	0,89	0,95	1,04
<i>Per</i>	0,73	0,71	0,70	1,37	0,47	0,49	0,35	0,52	0,49
<i>Pie</i>	0,26	0,16	0,32	0,22	0,65	0,70	1,89	1,35	1,22
<i>Pik</i>	0,55	1,04	0,78	1,01	1,75	1,89	0,48	0,11	0,83
<i>Pit</i>	1,07	1,29	1,23	1,30	1,19	1,01	1,21	1,12	1,18
<i>Suo</i>	0,79	0,60	0,54	0,24	0,36	0,39	0,57	1,06	0,54
<i>Suu</i>	0,37	0,17	0,46	0,22	0,09	0,20	0,66	1,94	0,54
<i>Väh</i>	1,89	0,78	0,92	0,83	0,27	0,26	0,11	0,03	0,35
<i>Yi</i>	0,06	0,26	0,24	0,38	0,55	0,96	0,15	0,05	0,30
Yhteensä	21,18	20,38	19,82	22,83	23,17	20,19	19,34	18,21	20,06

Suurin ero pienimmän ja toiseksi pienimmän esiintymistäajuuden väliltä löytyy koko aineistossa yleisimmästä *Iso*-nimenalusta: se on kaakkoismurteiden alueella huomattavan harvinainen. *Suu*-nimenalun esiintymistäajuus taas on kaakkoismurteiden alueella muita alueita suurempi. Ero muihin alueisiin selittyy murteistakin tunnetulla *iso~suuri*-vaihtelulla, sillä valtaosa *Suu*-alkuisista nimistä jatkuu kaakkoismurteiden alueella *Suur*-alkuisiksi. Myös savolaismurteiden ja hämäläismurteiden alueella *Suu*-alkuiset nimet jatkuvat usein *Suur*-alkuisiksi. Lounaismurteiden alueella taas on muista alueista poiketen taajaan *Väh*-alkuisia nimiä. *Väh*-alkuiset nimet jatkuvat lähes poikkeuksetta *Vähä*-alkuisiksi. Pohjoisessa ja idässä *Väh*-alkuisuus on harvinaista, koska siellä vähäiseen kokoon viitataan *pieni*- ja *pikku*-

ilmauksilla. *Pie-* ja *Pik-*alkuisten nimien taajuudet täydentävät kuvaa, sillä ne ovat yleisempiä pohjoisessa ja idässä. Hämmäläismurteiden ja Etelä-Pohjanmaan murteiden alueella *Väh-* ja *Pik-*alkuisuus on lähes yhtä yleistä, mutta vain hämmäläismurteiden alueella *Väh-* on kymmenen yleisimmän joukossa.

Etelä-Pohjanmaan murteiden alueetta erottaa muista *Hau-*, *Per-* ja *Kor-*alkuisten nimien yleisyys. *Hau-*alkuiset nimet näyttäisivät seuraavaa merkkiä tarkastelemalla jakaantuvan kahtaalle: toisaalta *Haut-* tai *Haud-*alkuisiin ja toisaalta *Hauk-*alkuisiin. Sekä *hauta* että *hauki* ja *haukka* ovat aiemman tutkimuksen perusteella yleisiä määriteosia ja alkuelementtejä (Kiviniemi 1990, 129–141; Salmenkivi ym. 2006, 202–203). Etelä-Pohjanmaan *Hau-*nimistä valtaosa lankeaa *hauta*-luokkaan, joka on Keski- ja Pohjois-Pohjanmaan murteiden aluetta lukuun ottamatta kaikkialla muualla harvinaisempi. *Hauki* ja *haukka* taas jakautuvat eri tavoin eri alueilla: Etenkin hämmäläismurteiden ja Keski- ja Pohjois-Pohjanmaan murteiden alueella mutta myös lounaisten välimurteiden alueella luokat ovat melko tasaväkiset. Vesistörikkaalla savolaismurteiden alueella *hauki* on huomattavasti yleisempi. Lounaismurteiden ja Etelä-Pohjanmaan murteiden alueella *haukka* on jonkin verran yleisempi. Kaakkoismurteiden alueellakin *haukka* näyttäisi olevan yleisempi, mutta pelkän *hauk-*nimielementin sisältäviä nimiä on siellä kuitenkin jokunen enemmän kuin *hauki-*nimielementin nimiä, joten täyttä varmuutta ei pelkän nimenalun perusteella asiasta saa.

*Per-*nimenalku taas jatkuu useimmin *Perä-* tai *Perk-*alkuiseksi. *Perä-*alkuun liittyvät sekä *Peri-* että *Peru-*alkuiset nimet: kaikilla viitataan johonkin perukkaan, perimmäiseen tai perällä olevaan. Kiviniemi (1990, 156) on luokitellut nämä lokaalipostpositioiden yhdyssanamuodoiksi ja todennut, että vaikka esimerkiksi juuri *perä-*sanalla on myös homonyyminen substantiivi *perä*, paikannimissä ja etenkin nimen alussa määriteosana ilmaus on useimmiten tulkittavissa juuri sijaintia ilmaisevaksi. Näin näyttäisi olevan myös paikannimirekisterissä kaikkien *Perä-*alkuisten kohdalla, vaikka joukosta löytyy myös jokunen *perätön-*ilmauksen sisältävä monitulkintaisempi poikkeus. *Perk-*alkuiset nimet taas liittyvät suurimmaksi osaksi viljelyksiin ja niiden raivaamiseen: listalta löytyy erilaisia *perkkiöitä*, *perkauksia* ja *perkkoja*. Etelä-Pohjanmaan murteiden alue erottuu kaikista muista sekä *Perä-* että *Perk-*alkuisten nimien suhteen. Muilla alueilla *Per-*alkuisissa nimissä on huomattavasti enemmän sijainninilmauksia.

Kaikkien kolmimerkkisten nimenalkujen yleisyyslistan toinen eli *Kor-* jakautuu aiempaa useammalle taholle. Suurimmat ryhmät ovat *korkea* (ja *korkia*), *korpi* ja *korte*. *Korppi-*,

Korpin- ja *Korva-*alkuisiakin nimiä on joitain, mutta määrä on vähäinen ja jakauma melko tasainen. Etelä-Pohjanmaan murteiden alueella sekä *korpi* että *korkea* (ja *korkia*) ovat yleisempiä kuin muualla. Peräpohjolan murteiden alueella sen sijaan *korpi* (siis sekä *Korve-* että *Korpi-*alut) on poikkeuksellisen harvinainen nimielementti. *Kort-*alkuiset nimet taas ovat lähes yhtä yleisiä Peräpohjolassa ja Etelä-Pohjanmaalla. Lounaismurteiden ja kaakkoismurteiden alueella kaikkien luokkien esiintymistaajuus on hieman alhaisempi kuin muilla, joten kokonaisuudessaankin taajuus jää alhaisimmaksi.

Peräpohjolan murteiden aluetta erottaa muista vielä *Kuu-* ja *Yli-*alkuisten nimien taaja esiintyminen. *Kuu-*alkuisista nimistä valtaosa liittyy sanaan *kuusi*. Useimmiten kyseessä lienee puunnimitys, sillä kuusi on aiemman tutkimuksen valossa muiden puiden ohella yksi yleisimmistä nimiaiheista (ks. esim. Kiviniemi 1990, 130–134). Sekä Kiviniemen (1990, 134) että Salmenkiven ja kumppaneiden (2006, 226) aineistossa *kuusi* on yleisin nimenomaan pohjoisessa. Kaakkoismurteiden, savolaismurteiden ja hämäläismurteiden alueella *kuusi* on nimen alussa verrattain harvinainen. Peräpohjolan murteiden alueelta löytyy myös jokunen kuukkelin mukaan nimetty paikka (esim. Kittilän *Kuukkeliautto*) sekä valikoima saamen ’pitkää’ merkitsevistä *guhki-*sanasta (ks. esim. SPNK 2007, 200) juontuvia nimiä (esim. Ranuan *Kuukaslampi*), joita on muualla vain ani harvassa ja vaihtelevammassa asuissa (esim. *Kukka-*alkuisina; ks. Aikio 2003, 100).

*Yli-*alkuisten nimien esiintyminen listalla johtuu yksinomaan Peräpohjolan murteiden alueen *Yli-*alkuisista nimistä. Esimerkiksi Kiviniemi (1990, 156) raportoi *Ala-* ja *Ali-*kategorian hieman *Ylä-* ja *Yli-*kategoriaa yleisemmäksi. Salmenkiven ja kumppanien (2006, 202) mukaankin *ala* on yleisempi nimen alkuelementti kuin *yli*. Kaikkien alueiden kolmen merkin nimenaluissa *Ala* onkin 52. sijalla 1 820 osumalla ja *Yli* taas 94. sijalla 1 284 osumalla. *Ali-* ja *Ylä-*alkuisia nimiä taas on koko aineistossa kumpiakin noin 500. Pelkkä *Yli* ennen nimen viimeistä leksikaalista osaa on piirun verran yleisempi Keski- ja Pohjois-Pohjanmaan murteiden alueella, mutta kaiken kaikkiaan (esim. ilmauksilla *ylinen* tai *ylimmäinen*) ylempänä olemista ilmaistaan nimissä siis useiten Peräpohjolan murteiden alueella. Lounaismurteiden ja kaakkoismurteiden alueella *Yli-* taas on tarkastelluista nimenaluista harvinaisimpia.

Loput nimenalut esiintyvät nimissä alueen nimimäärään suhteutettuna hyvin tasaisesti. Hajaannusta syntyy osittain siksi, että eri ilmauksilla on sama alku. *Koi-*alku esimerkiksi liittyy etenkin *koivu-*sanaan, joka on yksittäisistä puunnimityksistä yleisin paikannimissä

(Kiviniemi 1990, 131). Kiviniemen (1990, 131) listalla yleisemmäksi nimeämisaiheeksi nousee mänty, mutta käsitteeseen onkin niputettu kaikki 'mäntyä' merkitsevät ilmaukset, siis paitsi *mänty* myös *honka* ja *petäjä*. Tällä laskutavalla *mänty*-nimet peittoavat *koivu*-nimet myös tässä aineistossa. *Haa*-alkuisista nimistä suurin osa liittyy haapaan, joka on niin ikään yksi yleisimmistä paikannimiaiheista (Kiviniemi 1990, 131–132; Salmenkivi ym. 2006, 202). *Koi*-alkuisten nimien toiseksi suurin ryhmä ovat *koira*-nimet. *Koira*a yleisempi elementti nimissä on aiemman tutkimuksenkin perusteella kuitenkin *karhu* (ks. Kiviniemi 1990, 137; 142). *Kar*-nimenalun jakavat kuitenkin myös esimerkiksi monet nimet, joissa on ensimmäisenä nimielementtinä *karja*, *kari*, *karva* tai *karhea*.

Mus-alkuisista nimistä valtaosa liittyy *musta*-nimielementtiin, joka on aikaisemman tutkimuksen perusteella hyvin yleinen nimenomaan maasto- ja vesistönimissä (ks. Kiviniemi 1990, 150–152; Salmenkivi ym. 2006, 210). Koska tarkastelun kohteena ovat nyt ainoastaan maasto- ja vesistönimet, *musta* näyttäisi olevan tasaisen yleinen kaikkialla. Mikäli tarkasteltaisiin myös asutusnimiä, tilanne saattaisi olla toinen, sillä asutusnimissä *musta* on paljon harvinaisempi nimielementti (ks. Salmenkivi ym. 2006, 207). Toinen adjektiivisiin nimielementteihin liittyvä nimenalku on *Pit*-, joka on lähes aina alku *pitkä*-nimielementille. *Pitkä*-elementillä alkaa noin prosentti kaikkien alueiden maasto- ja vesistönimistä.

4.2 Paikannimen viittauskohteen sijaintialueen ennustaminen

Tässä luvussa käsiteltävissä luokittelumalleissa on käytetty kunkin murrealueen tai murrealueen alaluokan leksikaalisesti moniosaisista nimistä poimituista 13 500 nimen satunnaisotoksista koostettuja 108 000 nimen aineistoja. Kaikissa malleissa ennustettavia luokkia on kahdeksan, ja jokaisessa luokassa on 13 500 havaintoa. Käsittelen ensin mallia, jossa on käytetty ennustavina muuttujina paikanlajeja. Tällöin havaintoina ovat viimeiset leksikaaliset osat (ks. tarkempi kuvaus piirteistä ja muuttujista luvuissa 3.3 ja 3.4). Toisessa mallissa ennustavina muuttujina ovat paikanlajien lisäksi nimenalut (1–5 merkkiä) ja nimen kuluminen. Kummassakin mallissa ennustettava muuttuja ovat murrealueiden pääryhmät. Kolmas malli on luotu alaryhmäaineistolla siten, että ennustavina muuttujina ovat kaikki paikanlajimuuttujat ja tärkein nimenalkumuuttuja ja ennustettavana muuttujana yhdistelmä murrealueiden pääryhmiä ja alaryhmiä (ks. tarkempi kuvaus luvussa 3.4).

Paikanlajimalli ennustaa oikeaan luokkaan keskimäärin 31,6 prosenttia havainnoista. Satunnaisella jaolla oikeisiin luokkiin osuisi 12,5 prosenttia, eli malli ennustaa luokkia yli kaksi kertaa sattumaa paremmin. Erilaisilla jyvityksillä poimituilla otoksilla luodut mallit

eivät kokonaistarkkuudessa eroa toisistaan kuin prosentin kymmenyksiä, mutta olen valinnut tähän viidestä mallista tarkimman. Mallin valintaan vaikutti myös se, että luokat sekaantuivat toisiinsa useimmissa malleissa samoin kuin tässä mallissa. On silti hyvä huomioida, että kaakkoismurteiden ja hämäläismurteiden ennustustarkkuuteen erilaisilla otoksilla on jonkin verran vaikutusta: hämäläismurteiden luokkaan erilaiset mallit ennustavat oikein 500–1470 havaintoa ja kaakkoismurteiden luokkaan 733–2078. Jäljempänä on sekaannusmatriisi (ks. kuvio 3), johon on korostettu vihreällä kaikki oikein ennustetut luokat ja lisäksi sinisellä suurimmat määrät väriin luokkiin ennustettuja havaintoja kultakin alueelta. Parhaiten paikanlajimalli ennustaa havaintoja oikein Etelä-Pohjanmaan murteiden (64,5 %), Peräpohjolan murteiden (61,0 %) ja savolaismurteiden luokkaan (53,9 %). Edelleen melko hyvin malli ennustaa havaintoja lounaismurteiden luokkaan (33,0 %). Järjestyksessään heikoimmin se ennustaa havaintoja lounaisten välimurteiden (1,6 %), kaakkoismurteiden (5,7 %), hämäläismurteiden (10,9 %) ja Keski- ja Pohjois-Pohjanmaan murteiden luokkaan (23,1 %). Edellä suluissa on esitetty oikein ennustettujen havaintojen osuus luokkiin todellisesti kuuluvien havaintojen määrästä, joka on 13 500.

ennustettu luokka → todellinen luokka ↓	Lounaismurteet	Lounaiset välimurteet	Hämäläis- murteet	Etelä- Pohjanmaan murteet	Keski- ja Pohjois- Pohjanmaan murteet	Peräpohjolan murteet	Savolaismurteet	Kaakkois- murteet
Lounaismurteet	4449	102	799	3839	355	1743	1878	335
Lounaiset välimurteet	2332	211	438	5275	574	2112	2129	429
Hämäläismurteet	990	114	1470	4052	640	1896	3897	441
Etelä-Pohjanmaan murteet	500	70	311	8703	1254	1902	577	183
Keski- ja Pohjois- Pohjanmaan murteet	646	73	112	4471	3117	2496	2380	205
Peräpohjolan murteet	164	49	157	1345	788	8231	2661	105
Savolaismurteet	512	45	544	1841	1368	1507	7275	408
Kaakkoismurteet	1104	54	849	3248	705	1035	5741	764

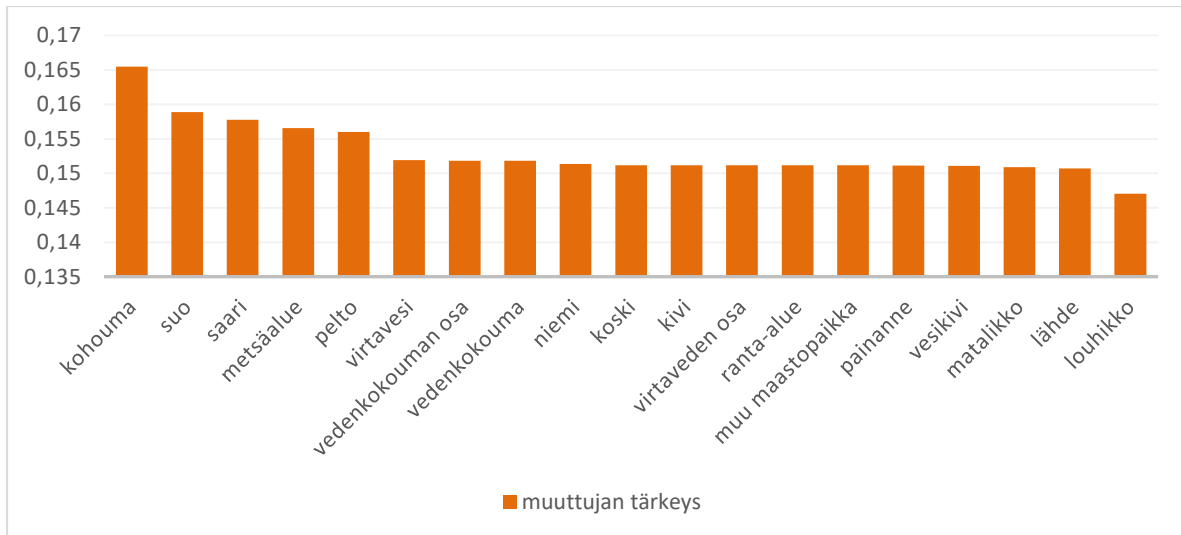
Kuvio 3 Paikanlajimallin sekaannusmatriisi eli oikeisiin ja väriin luokkiin ennustetut havainnot

Edellä mainituista neljästä paikanlajimallin heikoimmin ennustamasta luokasta malli on ennustanut havaintoja järjestelmällisemmin yhteen tai useampaan väärään luokkaan kuin oikeisiin luokkiin. Keski- ja Pohjois-Pohjanmaan murteiden luokasta malli on ennustanut enemmän havaintoja Etelä-Pohjanmaan luokkaan (33,1 %) kuin oikeaan luokkaan.

Kaakkoismurteiden luokasta malli taas on ennustanut havaintoja järjestelmällisimmin savolaismurteiden luokkaan (42,5 %) ja Etelä-Pohjanmaan murteiden luokkaan (24,1 %). Lounaisten välimurteiden luokasta malli on ennustanut havaintoja järjestelmällisemmin kaikkiin muihin kuin oikeaan luokkaan, eniten Etelä-Pohjanmaan murteiden (38,8 %) ja lounaismurteiden (17,3 %) luokkaan. Mallissa on siis näiden neljän heikoimmin oikein ennustetun ryhmän sisällä hyvin vähän sellaista vain niille ominaista säännönmukaisuutta, joka eroaa muista luokista ja jonka perusteella havaintoja voisi luokitella oikeisiin luokkiin.

Malli on luokitellut ylipäänsä kaikkein eniten havaintoja muista luokista väärin Etelä-Pohjanmaan murteiden luokkaan (kaikkiaan 32 774 havaintoa). Ainoastaan Peräpohjolan murteiden ja kaakkoismurteiden luokkaan kuuluvia havaintoja malli ennustaa väärin suuremman määrän johonkin muualle kuin Etelä-Pohjanmaan murteiden luokkaan. Sekä Peräpohjolan murteiden että kaakkoismurteiden luokan havaintoja malli ennustaa eniten savolaismurteiden luokkaan silloin, kun ennustus ei osu oikeaan. Kaakkoismurteiden havainnoista malli ennustaa silti Etelä-Pohjanmaan murteiden luokkaan lähes viidenneksen (24,1 %) ja Peräpohjolan murteiden havainnoistakin kymmenesosan. Silloin, kun ennustus on väärässä, malli ennustaa Etelä-Pohjanmaan murteiden havaintoja eniten Peräpohjolan murteiden luokkaan (13,6 %).

Paikanlajimallin ennustuksen kannalta viisi tärkeintä muuttujaa ovat 1) KOHOUMA, 2) SUO, 3) SAARI, 4) METSÄALUE sekä 5) VILJELYSMAA. Viiden tärkeimmän muuttujan suurimmat luokat ovat *mäki* (10 756), *suo* (7 039), *saari* (3 779), *kangas* (3 715) ja *suo* (1 078). Alla olevasta kuvioista (ks. kuvio 4) näkyy, että etenkin viiden mainitun muuttujan jälkeen seuraavaksi tärkeimmät kolmetoista muuttujaa eivät numeerisilta arvoiltaan juurikaan eroa toisistaan. LOUHIKKO-muuttujassa on hyvin vähän havaintoja, mikä voi selittää sen sijoittumisen viimeiseksi.



Kuvio 4 Paikanlajimallin muuttujien tärkeys

Toisessa ennustusmallissa eli nimenalkumallissa ennustavina muuttujina on paikanlajimuuttujien lisäksi viisi nimenalkumuuttujaa (1–5 ensimmäistä merkkiä) ja nimen kulumiseen liittyvä muuttuja (muuttujista tarkemmin ks. luku 3.4). Nimenalkumalli ennustaa oikeisiin luokkiin havainnoista 34,2 prosenttia, mikä on 2,5 prosenttiyksikköä enemmän kuin paikanlajimalli. Nimenalkumalli kuitenkin ennustaa luokkia aiemmin tarkasteltua paikanlajimallia tasaisemmin oikeisiin luokkiin. Tässä tarkasteltavassa mallissa on mukana kaikki nimenalkumuuttujat, mutta yksittäisillä nimenalkumuuttujilla luotujen mallien ennustusvirhe on paikanlajimallin kanssa lähes sama.

Jäljempänä olevaan sekaannusmatriisiin (ks. kuvio 5) on korostettu vihreällä oikein ennustetut havainnot ja sinisellä suurin väärin ennustettujen havaintojen rypäs. Kaakkoismurteiden ja savolaismurteiden luokista malli on ennustanut kahteen luokkaan havaintoja melko tasaisesti, joten alimmilla riveillä on siksi korostettuna sinisellä kaksi luokkaa. Etelä-Pohjanmaan murteiden luokkaan malli ennustaa edelleen eniten havaintoja oikein (62,5 %). Malli ennustaa Etelä-Pohjanmaan murteiden luokkaan myös edelleen selvästi eniten havaintoja, joiden todellinen luokka on lounaismurteiden, lounaisten välimurteiden, hämäläismurteiden tai Keski- ja Pohjois-Pohjanmaan murteiden luokka. Nimenalkumalli ennustaa paikanlajimallin tavoin toiseksi tarkimmin havaintoja Peräpohjolan murteiden luokkaan. Kolmanneksi tarkimmin nimenalkumalli kuitenkin ennustaa havaintoja lounaismurteiden luokkaan. Lounaismurteiden luokkaan oikein luokiteltujen havaintojen määrä on 4,6 prosenttiyksikköä enemmän kuin paikanlajimallissa. Savolaismurteiden alueelle nimenalkumalli ennustaa havaintoja heikommin (36,0 %) kuin paikanlajimalli, joka ennusti oikein yli puolet luokan

havainnoista. Luokassa on kuitenkin edelleen enemmän oikein ennustettuja havaintoja kuin yksittäisiin väriin luokkiin ennustettuja havaintoja.

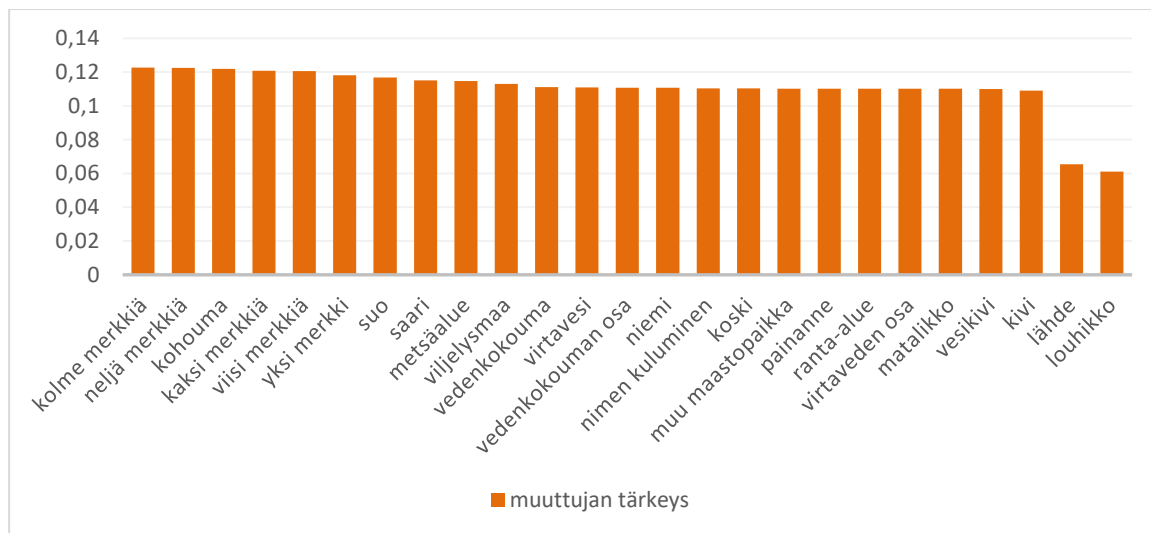
ennustettu luokka → todellinen luokka ↓	Lounaismurteet	Lounaiset välimurteet	Hämäläis- murteet	Etelä- Pohjanmaan murteet	Keski- ja Pohjois- Pohjanmaan murteet	Peräpohjolan murteet	Savolaismurteet	Kaakkois- murteet
Lounaismurteet	5078	984	712	3248	645	1050	720	1063
Lounaiset välimurteet	2414	1442	849	4667	997	1204	959	968
Hämäläismurteet	1303	803	1656	3701	1074	1364	1953	1646
Etelä-Pohjanmaan murteet	578	468	436	8436	1437	1102	584	459
Keski- ja Pohjois- Pohjanmaan murteet	672	400	492	3620	4252	1888	1561	615
Peräpohjolan murteet	472	307	512	1251	1297	7362	1655	644
Savolaismurteet	630	374	887	1715	1816	1393	4855	1830
Kaakkoismurteet	1071	412	828	2769	942	831	2772	3875

Kuvio 5 Nimenalkumallin sekaannusmatriisi eli oikeisiin ja väriin luokkiin ennustetut havainnot

Nimenalkumalli siis ennustaa ainoastaan lounaisten välimurteiden ja hämäläismurteiden havaintoja systemaattisemmin johonkin muualle kuin oikeaan luokkaan. Lounaisten välimurteiden luokasta malli ennustaa eniten havaintoja Etelä-Pohjanmaan murteiden (34,6 %) ja lounaismurteiden (17,9 %) luokkaan kuten paikanlajimallikin. Nimenalkumalli ennustaa kuitenkin kokonaisuudessaan tarkemmin myös näiden luokkien havaintoja: lounaisten välimurteiden luokkaan malli ennustaa oikein 10,7 prosenttia ja hämäläismurteiden luokkaan 12,3 prosenttia havainnoista. Kaikkein eniten parannusta on kaakkoismurteiden luokan ennustustarkkuudessa, sillä nimenalkumalli ennustaa oikeaan luokkaan yli neljänneksen, kun paikanlajimallin ennustustarkkuus on vain 5,6 prosenttia. Näyttäisi siis siltä, että etenkin näissä kolmessa luokassa ja erityisesti kaakkoismurteiden luokassa nimenaluissa on enemmän luokat muista erottavaa systemaattisuutta kuin eri paikanlajeissa esiintyvissä viimeisissä leksikaalisissa osissa.

Nimenalkumallin tekemän ennustuksen kannalta viisi tärkeintä muuttujaa ovat 1) KOLMEN MERKIN NIMENALKU, 2) NELJÄN MERKIN NIMENALKU, 3) KOHOUMA, 4) KAHDEN MERKIN NIMENALKU sekä 5) VIIDEN MERKIN NIMENALKU. Alla olevaan kuvioon (ks. kuvio 6) on havainnollistettu muuttujien tärkeysjärjestys. Muiden paitsi kahden vähiten tärkeän muuttujan

tärkeydessä ei tässäkään mallissa ole suurta eroa. Nimenalkumuuttujat ovat mallissa kärkipaikoilla. Tämä voi johtua siitä, että ne ovat vahvasti yhteydessä toisiinsa (ks. luku 3.4). Yksittäisillä nimenalkumuuttujilla luoduissa malleissa nimenalkumuuttujien sijoitus vaihteli seitsemänneistä kymmenenteen. Muutoin muuttujien tärkeysjärjestys on kaikissa malleissa sama kuin paikanlajimallissa. Nimen kulumiseen liittyvä muuttuja (kuviossa nimellä *nimen kuluminen*) sijoittuu tärkeydessä keskivaiheille.



Kuvio 6 Nimenalkumallin muuttujien tärkeys

Nimenalkumallin ennustustarkkuus on siis hieman parempi mutta selvästi tasaisempi kuin paikanlajimallin. Tarkimmin ennustettujen luokkien välillä ei nimenalkumallin ja paikanlajimallin välillä ole valtavaa eroa savolaismurteiden luokkaa lukuun ottamatta. Nimenalkumuuttujien mukaan ottaminen parantaa kuitenkin paikanlajimallilla heikoimmin ennustettujen luokkien tarkkuutta.

Yksi tekijä siinä, että nimenalkumalli ennustaa paikanlajimallia heikommin havaintoja savolaismurteiden alueelle, voi olla otoskoko. Savolaismurteiden alue on hyvin laaja, ja satunnaisesti koko alueelta poimittu 6,2 prosentin otos ei kata yhtä suurta osaa alueen nimistä eikä ole yhtä rajatulta alueelta kuin vaikkapa lounaismurteiden tai Etelä-Pohjanmaan murteiden alueen nimet aineistossa. Etenkin jos nimenalkujen muodostamat ryhmittymät liittyvät kantanimen ja liitynnäisnimien muodostamiin rypäisiin, kattavammassa otannassa on enemmän systemaattisuutta. Savolaismurteiden alueen nimitiheys on lähellä mediaania (ks. luku 4.1.1). Nimet muodostavat Kiviniemen (1990, 35–36) mukaan maantieteellisesti laajalle levittyviä ketjuja silloin, kun alue on nimetty harvaan. Loin siis vielä kolmannen

mallin, jonka aineistoon on valittu satunnaisesti nimiä maantieteellisesti pienemmiltä alueilta eli murrealueiden alaryhmistä (ks. luku 3.4).

Erilaisilla jyvityksillä poimituilla otoksilla on alaryhmäaineistoissa merkitystä paitsi mallin ennustusvirheeseen myös yksittäisten luokkien ennustustarkkuuteen. Parhaimmillaan ennustusvirhe on 52,0 prosenttia, eli malli ennustaa oikein jopa 48,0 prosenttia luokista. Viidestä otoksesta suurimman osan ennustusvirhe on kuitenkin vähän yli 68 prosenttia, eli mallit ennustavat oikein hieman alle 32 prosenttia luokista kuten paikanlajimallikin. Olen valinnut tähän tarkasteltavaksi suurinta osaa edustavan mallin. Eri savolaismurteiden alaryhmistä otetut otokset sen sijaan eivät vaikuta mallin ennustusvirheeseen prosenttiyksikköä enempää, mutta etenkin hämäläismurteiden, kunkin savolaismurteiden alaryhmän ja kaakkoismurteiden ennustustarkkuus vaihtelee eri otoksilla. Kainuun murteiden alueen nimet esimerkiksi näyttäisivät sekaantuvan useammin Kemin murteiden ja kaakkoismurteiden kanssa kuin vaikkapa Päijät-Hämen murteiden alueen nimet. Keski-Suomen murteiden alueen nimet taas sekaantuvat muita useammin Etelä-Pohjanmaan nimiin.

Alla olevaan sekaannusmatriisiin (ks. kuvio 7) on havainnollistettu luvussa 3.4 kuvatulla aineistolla luodun mallin eri luokkien ennustustarkkuus. Erilaisiin alaryhmäaineistoihin (joissa siis savolaismurteiden otos on eri alaryhmästä) perustuvat mallit ennustavat aina tarkimmin havaintoja Kemin murteiden luokkaan. Oikein ennustettuja havaintoja on tässä tarkasteltavassa alaryhmäaineistomallissakin yli 60 prosenttia. Kaakkoismurteiden luokka on tässä mallissa myös paikanlajimalliin verrattuna tarkemmin ennustettu luokkia, sillä sen ennustustarkkuus on lähes 16 prosenttia. Etelä-Pohjanmaan luokan tarkkuus on sekä paikanlajimalliin että nimenalkumalliin verrattuna parempi, yli 65 prosenttia. Etelä-Pohjanmaan luokka on kaikissa malleissa kolmen kärjessä ennustustarkkuudessa. Melkein kaikissa malleissa luokkaan ennustetaan myös eniten havaintoja väärin muista luokista. Lounaisten välimurteiden ennustustarkkuus on samaa luokkaa paikanlajimallin kanssa eli melko heikko. Hämeen kaakkoismurteiden luokassa on eri otoksilla paljon vaihtelua. Tässä mallissa Hämeen kaakkoismurteiden havaintoja ennustetaan erityisen paljon paitsi Etelä-Pohjanmaan luokkaan myös itämurteita edustaviin luokkiin ja Kemin murteiden luokkaan. Päijät-Hämeen ennustustarkkuus paranee nimenalkumallin savolaismurteisiin nähden, mutta paikanlajimallin tasolle tarkkuus ei aivan yllä. Lounaismurteiden luokan ennustustarkkuus ei juurikaan eroa kahdesta edellisestä mallista, mutta tarkalleen ottaen tarkkuus on parempi kuin paikanlajimallissa ja heikompi kuin nimenalkumallissa. Lounaismurteiden ennustustarkkuuteen erilaisilla otoksilla ja malleilla näyttäisi ylipäänsä olevan melko vähän

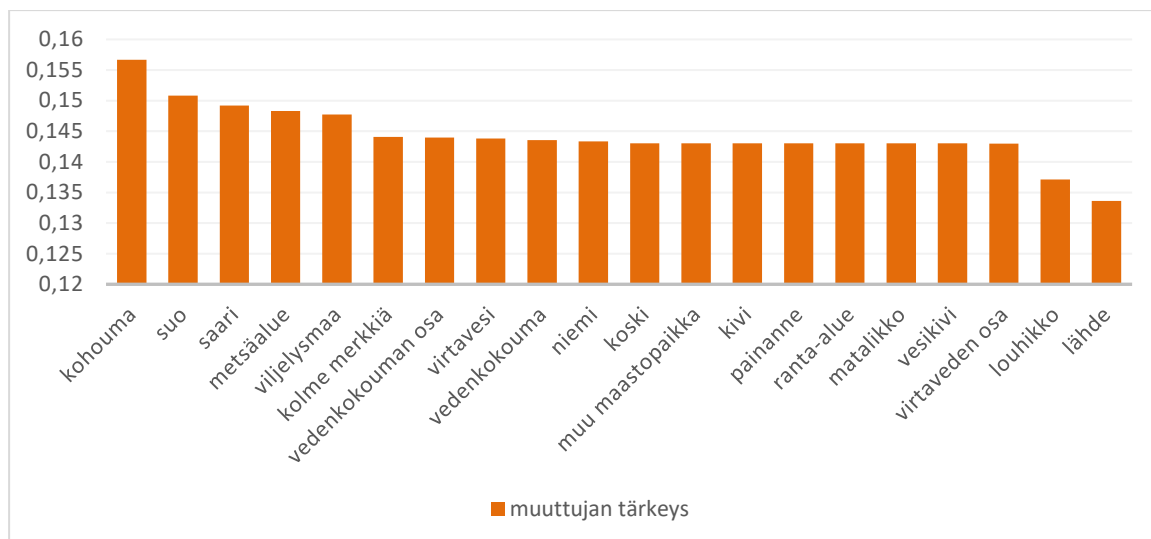
vaikutusta. Keski-Pohjanmaan murteiden alueen oikein ennustettujen havaintojen määrässä ei ole juuri muutosta paikanlajimallin Keski- ja Pohjois-Pohjanmaan murteiden alueeseen verrattuna. Alaryhmäaineistomalli ennustaa kuitenkin Keski- ja Pohjois-Pohjanmaan murteiden alueelta hieman enemmän havaintoja kaakkoismurteiden alueelle ja vähemmän savolaismurteiden (tässä siis Päijät-Hämeen murteet) alueelle.

ennustettu luokka → todellinen luokka ↓	Lounais- murteet	Lounaiset välimurteet	Hämeen kaakkois- murteet	Etelä- Pohjanmaan murteet	Keski- Pohjanmaan murteet	Kemin murteet	Päijät- Hämeen murteet	Kaakkois- murteet
Lounaismurteet	4730	159	380	3866	324	1734	1368	939
Lounaiset välimurteet	2433	216	320	5304	583	2107	1733	804
Hämeen kaakkoismurteet	1379	131	821	4191	630	1924	3208	1216
Etelä-Pohjanmaan murteet	641	91	156	8805	1204	1840	485	278
Keski-Pohjanmaan murteet	632	131	114	4534	2970	2572	1929	618
Kemin murteet	224	61	145	1456	668	8352	2094	500
Päijät-Hämeen murteet	670	68	328	1821	1358	1607	5958	1690
Kaakkoismurteet	1464	87	340	3286	727	1093	4355	2148

Kuvio 7 Alaryhmäaineistolla luodun mallin sekaannusmatriisi

Alaryhmäaineistomallin tärkeimmät muuttujat ovat samat kuin paikanlajimallinkin.

Nimenalkumuuttuja on järjestyksessään kuudenneksi tärkein. Alla olevasta kuviosta (ks. kuvio 8) näkyy, että paikanlajimuuttujien tärkeysjärjestys on lähes sama kuin paikanlajimallissa (ks. kuvio 4) ja nimenalkumallissa (ks. kuvio 6).



Kuvio 8 Alaryhmäaineistolla luodun mallin muuttujien tärkeys

Esimerkiksi alaryhmäaineistomallin muuttujien luokkien tarkempi tarkastelu selittäisi paremmin mallien tekemiä ennustuksia, mutta käytettävissäni ei ole tarpeeksi tehokkaita työkaluja mallin sisäisen rakenteen selvittämiseen. Tulkinnat on siis tehtävä peilaamalla malleja siihen aineistoon, johon ne perustuvat.

5 Päätelmät

Asetin johdannossa kaksi tutkimuskysymystä:

1. Millaisia eri murrealueiden paikannimet ovat, ja millaisia paikannimien yleisimmät leksikaaliset viimeiset osat ja yleisimmät kolmen merkin nimenalut ovat?
2. Millaisen ennustuksen paikannimen viittaushohteen sijaintialueesta voi tehdä ohjatun koneoppimisen menetelmin eri paikanlajeissa esiintyvien leksikaalisten viimeisten osien sekä nimenalkujen ja nimien kulumisen perusteella?

Tulosten perusteella eri murrealueilla on erilainen määrä leksikaalisesti yksi- ja moniosaisia paikannimiä. Salmenkivi ja kumppanitkin (2006, 193) raportoivat eroja yksi- ja kaksisanaisten nimien esiintymisessä maakunnittain. Maakunnat ovat kooltaan ja nimimäärältään tasaisempia, joten erot eri alueiden välillä ovat murrealueissa suurempia. Suurin selittävä tekijä leksikaalisesti yksiosaisien ja moniosaisien nimien esiintymisessä lienee nimettyjen kohteiden määrä: esimerkiksi taloilla on usein johdettu, leksikaalisesti yksiosainen nimi, kuten *Nummila*, *Koivisto* tai *Hakanen*, ja tiheämmän asutuksen alueilla on enemmän nimettyjä asumuksia.

Toinen selitys yksiosaisien nimien esiintymiseen voi olla nimen ikä, sillä vanhat nimet ovat useammin läpinäkymättömiä. Tällöin nimi hahmottuu paitsi syntaktisesti yksiosaiseksi myös leksikaalisesti yksiosaiseksi, koska leksikaalisia osia ei enää erota toisistaan: esimerkiksi *Mustio* on alun perin päättynyt *joki*-sanaan, joka on kulunut lopusta (ks. SPNK, 277). Vanhoja nimiä on eniten paikkakunnilla, joissa on ollut kauimmin pysyvää asutusta, ja vanhojen reittien varrella (ks. Kiviniemi 1990, 39). Etenkin lounaismurteiden ja lounaisten välimurteiden alueiden leksikaalisesti yksiosaisissa nimissä lienee paljon vanhoja nimiä. Myös merkittävien kohteiden (esim. suuret vedenkokoumat ja vesireitit) nimet ovat usein vanhoja ja semanttisesti läpinäkymättömiä sijainnista riippumatta. Ihmisillä on kuitenkin taipumus selittää erilaisia kansanetymologioita, mikä saattaa muuttaa nimiä siten, että hämärtyneet nimet muuttuvat läpinäkyvämmiksi (ks. luku 2.3). Nimiin on voitu lisätä merkitystä selventäviä osia silloin, kun ne ovat hämärtyneitä (ilmiöstä ruotsin kontaktialueilla ks. Pitkänen 2007). Tällä tavalla jotkin hyvin yleiskieliset viimeiset leksikaaliset osat, kuten *mäki*, *suo*, *joki* ja *lampi*, ovat voineet yleistyä paikallisten ilmausten kustannuksella. Tässä tutkimuksessa ilmiötä ei voi suoraan todeta, mutta sitä voi pitää aiemman tutkimuksen (ks. Ainiala 1997) valossa mahdollisena.

Uniikkien nimien määrän murrealueittainen vaihtelu liittyy toisaalta siihen, paljonko erilaisia paikkoja on nimetty, ja toisaalta siihen, paljonko alueella on ollut kontakteja eri kielten kanssa. Peräpohjolan murteiden alueella yhdistyy sekä luonnonpaikkojen suuri määrä että runsaat kontaktit monen eri kielen kanssa, kun taas lounaismurteiden alueella kielikontaktit voivat olla suurin selittävä tekijä (havainnollistus ruotsin ja suomen kontaktialueista ks. Kotus, Kielestä kieleen).

Viimeisten leksikaalisten osien jakautuminen murrealueittain samalla tavalla eri paikanlajeihin heijastelee alueiden luonnonmaantiedettä. Esimerkiksi asutukseltaan tiheillä viljelysvaltaisilla alueilla luonnonpaikkoihin viittaavia sanoja on usein viimeisenä leksikaalisena osana viljelysnimissä. Samaan paikanlajiin viittaavissa sanoissa on aineiston perusteella näkyvissä murrealueittain sekä leksikaalista (esim. suohon viittaavien *aapa-* ja *neva-*sanojen vaihtelu) että morfologista vaihtelua (esim. *lammi-* ja *lampi-*vaihtelu). Tulokset ovat linjassa aiemman tutkimuksen kanssa (ks. Kiviniemi 1990; Salmenkivi ym. 2006).

Kolmen merkin nimenaluissa näkyy etenkin aiemmasta tutkimuksesta tuttu *iso~suur(i)-* ja *pieni~pikku~vähä-*vaihtelu lännen ja idän välillä. Myös nimeämisasiheiden vaihtelu murrealueittain käy ilmi yleisimpiä nimenalkuja tarkastelemalla. Aiheet liittyvät usein alueen luonnonmaantieteeseen, kuten vesistöjen määrään tai tietyn puulajin esiintymiseen alueella: jos kuusia on melko harvassa, kuuset ovat todennäköisemmin paikkoja muista erottavia, ja jos alueella on paljon vesistöjä, siellä esiintyy oletettavasti enemmän eri kalalajeja kuvaamassa paikkoja. Kielikontaktitkin vaikuttavat alueittain yleisimpien nimenalkujen esiintyvyyteen, kuten Peräpohjolan *Kuu-*alkuisten nimien yleisyys todistaa. Harvinaisempien nimenalkujen, kuten erilaisten konsonanttiklusterien, tarkastelu olisi ehkä paljastanut kielikontaktivaikutuksia myös muualla, oletettavasti ainakin ruotsin kielikontaktialueilla lounaassa ja etelä- ja länsirannikolla.

Satunnaismetsä-menetelmällä luotu ennustusmalli kykenee pelkkien paikanlajeittain ryhmiteltyjen viimeisten osien perusteella ennustamaan sijaintialueen parhaiten Etelä-Pohjanmaan murteiden, Peräpohjolan murteiden ja savolaismurteiden luokkaan. Peräpohjolan murteiden alueen ennustustarkkuus läpi kaikkien mallien ei ole yllättävää, sillä aikaisemmassa tutkimuksessa on jo todettu pohjoisimman Suomen eroavan nimenloppuisilta lekseemeiltään muusta Suomesta (ks. Salmenkivi ym. 2006, 216–217; SPNK, 14–15). Etelä-Pohjanmaan murteiden alueellakin on tärkeimmässä muuttujissa muista alueista erottuvia luokkia, kuten soissa *neva* ja kohoumissa *mäki*. Toisaalta samalla esimerkiksi juuri *mäki* yhdistää monia

alueita ja *neva* etenkin Etelä-Pohjanmaata ja Keski- ja Pohjois-Pohjanmaata.

Savolaismurteiden alueella varsinkin vedenkokoumissa esiintyy koko aineiston tarkastelun perusteella viimeisenä leksikaalisena osana hyvin usein *lampi*, kun muualla viimeisissä leksikaalisissa osissa näyttäisi olevan enemmän vaihtelevuutta. Koska malli perustuu merkittävimmille eroille, johdonmukaisuus yhden leksikaalisen viimeisen osan käytössä tietyssä paikanlajissa ohjaa ennustukset yhdelle alueelle.

Kaikkien nimenalkumuuttujien ja nimen kulumista kuvaavan muuttujan ottaminen mukaan malliin parantaa ennustusta kokonaisuudessaan. Yksittäisillä nimenalkumuuttujilla luodut mallit eivät kuitenkaan kokonaistarkkuudeltaan eroa paikanlajimallista. Etenkin kaakkoismurteiden alueen ennustustarkkuus näyttäisi kuitenkin tarkentuvan, kun nimenalkumuuttujia otetaan mukaan malliin. Savolaismurteiden luokan ennustustarkkuus sen sijaan heikkenee. Nimenalkuja kannattaisi jatkossa luokitella tarkemmin semanttisiin ryhmiin esimerkiksi Kiviniemen (1990, 122) mallin mukaan. Tällä tavalla voisi paremmin huomioida sen, että yhteen käsitteeseen voidaan viitata eri tavoin: esimerkiksi mänty voi olla *männyn* lisäksi *honka* tai *petäjä* ja käärme *käärmeen* lisäksi *kärme* tai *mato* tai ruotsista mukautettu *ormi*.

Nimen kuluminen ei ole suuressa osassa nimenalkumallin tekemässä ennustuksessa. Se voi johtua siitä, että kaikkia tapauksia ei ollut tässä tutkimuksessa mahdollista saada aineistosta esiin. Havainnot painottuvat lounaismurteiden, lounaisten välimurteiden ja kaakkoismurteiden alueelle, ja muuttujan jättäminen pois heikentää hieman lounaismurteiden mutta parantaa jonkin verran kaakkoismurteiden ennustustarkkuutta. Nimien kuluminen voi kuitenkin liittyä koko nimisysteemiin ja siihen, että nimistä on ajan saatossa tullut taloudellisempia (ks. Kiviniemi ym. 1974, 22).

Erilaiset alaryhmäaineistomallit voisivat olla mielenkiintoinen jatkoselvityksen kohde. Vaikka kokonaistarkkuus ei eri savolaismurteiden alarymistä poimituilla otoksilla juurikaan vaihtelee, alueiden sekaantuminen toisiinsa muuttuu jonkin verran. Esimerkiksi silloin, kun aineistossa on otokset Pohjois-Pohjanmaan murteiden alueelta, Kemin murteiden alueelta ja Kainuun murteiden alueelta, malli ennustaa havaintoja väärin nimenomaan näiden alueiden kesken. Alueiden maantieteellinen läheikkäisyys voi olla yksi selittävä tekijä, mutta ihan kaikki maantieteellisesti lähekkäiset alueet eivät samalla tavalla sekaannu keskenään.

Ohjatun koneoppimisen menetelmällä saadut tulokset vahvistavat etenkin aikaisemmassa tutkimuksessa havaittua eroa pohjoisimman Suomen ja muiden alueiden välillä. Erot

näyttäisivät liittyvän sekä leksikaalisiin viimeisiin osiin että nimenalkuihin. Leksikaalisten viimeisten osien esiintymisen itä–länsi-akselin ääripäinä näyttäisivät olevan Etelä-Pohjanmaan murteiden alue ja savolaismurteiden alue. Etelä-Pohjanmaan murteiden alueella on yhteisiä piirteitä hyvin monen alueen kanssa, sillä kaikki mallit ennustavat muilta alueilta paljon havaintoja Etelä-Pohjanmaan murteiden alueelle.

6 Lopuksi

Tässä tutkimuksessa on tarkasteltu paikannimien viimeisiä leksikaalisia osia ja nimenalkuja ja pyritty selvittämään, millaisen ennustuksen paikannimen sijaintialueesta voi tehdä ohjatun koneoppimisen menetelmin. Paikannimisysteemissä on paljon muutakin säännönmukaisuutta, jota voisi tutkia nyky menetelmillä. Esimerkiksi erilaisten johdinainesten tai suffiksien ottaminen mukaan piirteiksi voisi valaista alueellista vaihtelua vielä pidemmälle. Silloin voisi ottaa mukaan myös leksikaalisesti yksiosaiset nimet.

Aluejaonkin voisi toteuttaa hyvin monella eri tavalla: esimerkiksi siten, että se vastaisi paremmin Leinon ja kumppaneiden (2006) murre sanojen levikkiä koskevien havaintojen perusteella tehtyä jakoa viiteentoista alueeseen. Se olisi voinut sanoihin pohjautuvana jakona sopia jopa paremmin viimeisten leksikaalisten osien tarkasteluun. Mallinnuksessa voisi jatkossa käyttää ennustettavan muuttujan luokkina myös murrealueiden alaryhmiä tai tutkia sitä, miten yhteen alaryhmään perustuva malli luokittelee muita saman pääryhmän paikannimiä. Vielä tarkemmin mitä tahansa kohdealueita vastaavan jaon saisi toteutettua käyttämällä kuntien sijaan vertailukohtana koordinaattipisteitä.

Murrealueiden koko ja siten myös nimimäärät ja nimitys sekä maasto ovat erilaisia. Tämä aiheuttaa vinoumia aineistoon. Esimerkiksi Leino (2006, 3–4) on pyrkinyt minimoimaan alueiden koosta kumpuavia vinoumia jakamalla Suomen tasakokoisiin alueisiin. Hänen vertailukohtanaan olivat kunnat. Alueen jakaminen maantieteellisesti tasan ei kuitenkaan poista vaihtelevuutta nimityksessä. Suuremmissa alueissa ero tasoittunee, mutta maastoon ja alueiden käyttöön liittyvät erot näkyvät edelleen tuloksissa. Toisaalta perinnäisissä paikannimissä on aina mukana historiaa, joten nimiä on hyvä tarkastella aina niiden esiintymisalue huomioiden.

Työkaluja ajatellen automaattiset lauseenjäsentäjät ovat kyllä kehittyneet, mutta paikannimiaineistoon automaattinen lauseenjäsennys ei edelleenkään oikein toimi. Kokeilin 1 000 nimen satunnaisotoksen leksikaalisten osien erotteluun kolmea lauseenjäsentäjää: R-ohjelmointiympäristöön saatavilla olevaa udpipen annotointia sekä Python-ohjelmointikieleen perustuvien Uralic NLP:n (ks. Hämäläinen 2018) ja Trankitin (ks. Nguyen ym. 2021) työkaluja. Uralic NLP ja Trankit olivat huomattavasti tarkempia kuin udpipen. Uralic NLP:tä voisi kuvailla jopa tyydyttäväksi, vaikka esimerkiksi *Teerimaa* olikin sen analysoimana *Tee|rimaa* (yksikön nominatiivi sanasta *tee* ja yksikön partitiivi sanasta *rima*). Ongelmat

Trankitin ja Uralic NLP:n välillä olivat erilaisia, joten ehkä yhdistelemällä ja ajan kanssa olisi päätyntä hyväksyttävään lopputulokseen. Rekisterimuotoisten paikannimiaineistojen käsittelyssä olisi hyvä esimerkiksi määrittää, missä sijamuodoissa yksittäiset osat voivat olla, koska mahdollisuudet ovat paljon rajatummalla kuin pidemmissä teksteissä (paikannimissä leksikaalinen viimeinen osa on lähes poikkeuksetta nominatiivissa ja sitä edeltävä osa joko nominatiivissa tai genetiivissä). Epäodotuksenmukaiset muodot toki saa suodatettua pois tuloksista.

Menetelmällisesti mielenkiintoinen vaihtoehto olisi yrittää muodostaa paikannimistä erotettujen piirteiden perusteella ryhmiä klusterianalyysilla, eli käyttää paikannimien ryhmittelyyn ohjaamatonta koneoppimista. Klusterianalyysilla muodostettuja alueita voisi valituista piirteistä riippuen verrata vaikkapa juuri Leinon ja kumppaneiden (2006) murreanastotutkimuksen alueisiin tai perinteisiin murrealueisiin.

Aineistoakin voisi vielä laajentaa niillä Venäjän, Ruotsin ja Norjan suomenkielisillä paikannimillä, jotka lankeavat murrealueille. Etenkin jos pidättäytyy perinteisessä murrealuejaottelussa. Ainakin kaakkoismurteiden alue on maantieteellisesti hyvin pieni, ja sitä voisi laajentaa itärajan toiselle puolelle. Ongelmana saattaa kuitenkin olla se, että aineisto ei ole samanlaista kuin paikannimirekisterin aineisto, jos sen poimii paikallisista kartoista. Nimiarkistostakin voisi poimia nimiä nykyisen Suomen rajojen ulkopuolelta ottaen huomioon arkiston ja paikannimirekisterin eroavaisuudet. Rekisterin sijaan nimiä voisi tarkastella myös kielenkäytössä: paikannimiä voisi poimia esimerkiksi Lauseopin arkiston murrekorpukselta, johon on annotoitu erikseen henkilönnimet ja muut erisnimet.

Paikannimissä ja paikannimirekisterissäkin riittää siis vielä paljon tutkittavaa, etenkin kun rekisteriä päivitetään jatkuvasti. Tätä tutkimusta varten käsitellyt aineistot ja satunnaismetsien luonnissa käytetyt R-ohjelmointiympäristölle kirjoitetut skriptit ovat saatavilla jatkotutkimuksia varten verkossa (ks. Virtanen 2025).

Lähteet

Aineslähteet

Nimistörekisterin Paikannimet-tiedosto. Maanmittauslaitos, Helsinki. Päivitetty 4.6.2024.

Noudettu 10.10.2024. Päivitetty versio saatavissa

<https://asiointi.maanmittauslaitos.fi/karttapaikka/tiedostopalvelu/nimisto?lang=fi>.

Virtanen, Emilia 2025. Paikannimisatunnaismetsät. Saatavissa <https://osf.io/xmqdp/>.

Muut lähteet

Aarikka, Lotta 2023. *Murre ja sen tutkimus. Näkökulmia fennistisen murteentutkimuksen historiaan ja kieli-ideologioihin 1871–2017*. Väitöskirja. Turun yliopisto, Turku. URN:ISBN 978-951-29-9301-7.

Aikio, Ante 2003. Suomen saamelaisperäisistä paikannimistä. *Virittäjä* 107(1), s. 99–106.

Kotikielen seura, Helsinki. Saatavissa

<https://journal.fi/virittaja/article/view/40243/9670>.

— 2007. The Study of Saami Substrate Toponyms in Finland. *Onomastica Uralica* 4, s. 159–197. Toim. Ritva Liisa Pitkänen ja Janne Saarikivi. Saatavissa

<https://mnytud.arts.unideb.hu/onomural/kotetek/ou4.html>.

Ainiala, Terhi 1997. *Muuttuva paikannimistö*. Väitöskirja, suomen kielen laitos. SKS, Helsinki.

— 2018. Paikannimien kerrostumat maaseudulla ja kaupungissa nimistöntutkijan kohteina.

Elore 25(1), s. 96–106. Suomen kansantietouden tutkijain seura, Joensuu. DOI:

<https://doi.org/10.30666/elore.72817>.

Ainiala, Terhi – Sjöblom, Paula 2021. Nimistöntutkimuksen metodit *Virittäjän* kirjoituksissa.

Virittäjä 125(4), s. 521–546. Kotikielen seura, Helsinki. DOI:

<https://doi.org/10.23982/vir.98030>.

Breiman, L 2001. Random forests. s. 5–32. *Mach Learn* 45. DOI:

<https://doi.org/10.1023/A:1010933404324>.

Hyvän nimistön opas. [Verkojulkaisu.] Toim. Ulla Onkamo. Kotimaisten kielten keskuksen verkkojulkaisuja 78. Kotus, Helsinki. Viitattu 17.1.2025. URN:ISBN 978-952-7359-18-1.

Itkonen, Terho 1989. *Nurmijärven murrekirja*. Kotiseudun murrekirjoja 10. SKST 498. SKS, Helsinki.

- 1997. *Nimestäjän opas*. 3., uud. p. Apuneuvoja suomalais-ugrialaisten kielten opintoja varten 13. Suomalais-ugrilainen seura, Helsinki.
- Ivaska, Ilmari 2014. Edistyneen oppijansuomen avainrakenteita. Korpusnäkökulma kahden kielimuodon tyypillisiin rakenteellisiin eroihin. *Virittäjä* 118(2), s. 161–193. Kotikielen seura, Helsinki. Saatavissa <https://journal.fi/virittaja/article/view/9368>.
- Ivaska, Ilmari – Johnson, Mirva – Kurki, Tommi 2023. Identifying the dialectal background of American Finnish speakers using a supervised machine-learning model. *Nordic Journal of Linguistics* 47(3), s. 1–25. DOI: <https://doi.org/10.1017/S0332586523000057>.
- Ivaska, Ilmari – Siitonen, Kirsti 2017. Lerner language morphology as a window to crosslinguistic influences: A key structure analysis. *Nordic Journal of Linguistics* 40(2), s. 225–253. DOI: <https://doi.org/10.1017/S0332586517000129>.
- Häkkinen, Kaisa 1994. *Agricolasta nykykieleen: Suomen kirjakielen historia*. WSOY, Porvoo.
- Hämäläinen, Mika 2019. UralicNLP: An NLP Library for Uralic Languages. *Journal of open source software*, 4(37). DOI: <https://doi.org/10.21105/joss.01345>.
- Kettunen, Lauri 1940. *Suomen murteet. 3 B, Selityksiä murrekartastoon*. SKS, Helsinki.
- KS = *Kielitoimiston sanakirja*. [Päivitettävä verkkojulkaisu.] Kotimaisten kielten keskuksen verkkojulkaisuja 35. Kotus, Helsinki. Päivitetty 19.3.2024. URN: NBN:fi:kotus-201433.
- Kiviniemi, Eero 1971. *Suomen partisiippinimistöä: ensimmäisen partisiipin sisältävät henkilön- ja paikannimet*. Väitöskirja, suomen kielen laitos. SKST 295. SKS, Helsinki.
- 1975. *Paikannimien rakennetyypeistä*. Suomi 118:2. SKS, Helsinki.
- 1977. *Väärät vedet: tutkimus mallien osuudesta nimenmuodostuksessa*. SKST 377. SKS, Helsinki.
- 1978. Paikannimistö systeeminä. *Nimistöntutkimus ja paikallishistoria*, s. 52–72. Toim. Eero Kiviniemi. Paikallishistoriallisen toimiston julkaisuja 2. Paikallishistoriallinen toimisto ja Kotus, Helsinki.
- 1990. *Perustietoa paikannimistä*. Suomi 418. SKST 516. SKS, Helsinki.
- Kiviniemi, Eero – Pitkänen, Ritva Liisa – Zilliacus, Kurt 1974. *Nimistöntutkimuksen terminologia. Terminologin inom namnforskningen*. Castreanuksen toimitteita 8. Helsingin yliopisto, Helsinki.
- Klemettinen, Riina 2010. Suurella sydämellä – tai isolla. *Kielikello* 2/2010. [Verkkoversio.] Kotus, Helsinki. Saatavissa <https://kielikello.fi/suurella-sydamalla-tai-isolla/>.

- Kotus = Kotimaisten kielten keskus. Karttanimistön tarkistus. [Verkkosivu.] Viitattu 26.2.2025. <https://kotus.fi/nimistonhuolto/nimisto-ja-sen-huoltaminen/karttanimiston-tarkistus/>.
- Kielestä kieleen. Paikannimen tarina -verkkonäyttely. [Verkkojulkaisu.] Viitattu 22.3.2025. Saatavissa <https://kotus.fi/nimistonhuolto/nimijulkaisut/paikannimen-tarina/kielesta-kieleen/>.
- Kieliaineistojen käyttö. Arkistojen käytösäännöt. Sähköisten aineistojen käyttö. [Verkkosivu.] Viitattu 15.1.2025. Saatavissa <https://kotus.fi/kotus/kieliaineistot/kieliaineistojen-kaytto/>.
- Peruskartan paikannimet. Paikannimen tarina -verkkonäyttely. [Verkkojulkaisu.] Viitattu 20.3.2025. Saatavissa <https://kotus.fi/nimistonhuolto/nimijulkaisut/paikannimen-tarina/peruskartan-paikannimet/>.
- Suomen murteet. Murrejako. [Verkkosivu.] Viitattu 23.2.2025. Saatavissa <https://kotus.fi/sanakirjat/suomen-murteiden-sanakirja/suomen-murteet/#murrejako>.
- Kyröläinen, Aki-Juhani – Laippala, Veronika. Määrällinen korpuslingvistiikka. *Kielentutkimuksen menetelmiä I-IV*. s. 487–524. Toim. Minna Luodonpää-Manni, Markus Hamunen, Reetta Konstenius, Matti Miestamo, Urpo Nikanne & Kaius Sinnemäki. SKST 1457. SKS, Helsinki. DOI: <https://doi.org/10.21435/skst.1457>.
- Leino, Antti 2004. Computational Overview of Finnish Hydronyms. *Onomastica Lettica II*, s. 239–268. Saatavissa <https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=c2499213011188923365ab9102e9b0bb5383098f>.
- 2006. Regional variation in Finnish lake and hill names. *Nordiske navnes centralitet og regionalitet. Rapport fra NORNA 35. symposium på Bornholm 4(7)*. Saatavissa <https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=d1fc2f200a49bb25af4dee5b1edcac5f3e240c9e>.
- Leino, Antti – Hyvönen, Saara – Salmenkivi, Marko 2006. Mitä murteita suomessa onkaan? Murreosanaston levikin kvantitatiivista analyysiä. *Virittäjä* 110(1), s. 26–45. Kotikielen seura, Helsinki. Saatavissa <https://journal.fi/virittaja/article/view/40455>.
- Lehtinen, Tapani 2007. *Kielen vuosituhanneet. Suomen kielen kehitys kantauralasta varhaisuomeen*. Tietolipas 215. SKS, Helsinki.
- Maanmittauslaitos 2022. Maanmittauslaitoksen paikannimirekisterin paikkaluokitus. [Verkkodokumentti.] Viitattu 10.3.2025. Saatavissa

https://www.maanmittauslaitos.fi/sites/maanmittauslaitos.fi/files/attachments/2023/09/Paikannimirekisterin_paikkaluokitus.pdf.

Maanmittauslaitos. Karttapaikka. [Verkkosivu.] Viitattu 15.1.2025. Saatavissa

<https://asiointi.maanmittauslaitos.fi/karttapaikka/?lang=fi>.

— Nimistö. [Verkkosivu.] Viitattu 15.1.2025. Saatavissa

<https://www.maanmittauslaitos.fi/kartat-ja-paikkatieto/aineistot-ja-rajapinnat/tuotekuvaukset/nimisto>.

— Nimistön kyselypalvelu. [Verkkosivu.] Viitattu 10.3.2025. Saatavissa

<https://www.maanmittauslaitos.fi/nimiston-kyselypalvelu-ogc-api>.

— Tilastot. [Verkkosivu.] Viitattu 21.1.2025. Saatavissa

<https://www.maanmittauslaitos.fi/tietoa-maanmittauslaitoksesta/organisaatio/tilastot>.

Nguyen, Minh Van – Lai, Viet – Veyseh, Amir Pouran Ben – Nguyen, Thien Huu 2021.

Trankit: A Light-Weight Transformer-based Toolkit for Multilingual Natural Language Processing. *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations*. Saatavissa <https://arxiv.org/pdf/2101.03289>.

Nimiarkisto. [Verkkotietokanta.] Kotus, Helsinki. Päivitetty 14.4.2022. Viitattu 17.1.2025.

Saatavissa <https://nimiarkisto.fi/>.

— Aarni, Juhani 1956. *Ärästi*. Saatavissa <https://nimiarkisto.fi/wiki/Q7227934>.

— Nuutinen, Olavi 1960. *Ahonpää*. Saatavissa <https://nimiarkisto.fi/wiki/Q6224996>.

— Kiiski, Marjatta 1965. *Uudismaa*. Saatavissa <https://nimiarkisto.fi/wiki/Q6736662>.

Nimisampo 2019. [Verkkopalvelu.] Helsingin yliopisto ja Aalto yliopisto, Helsinki. Viitattu 22.1.2025. Saatavissa <https://nimisampo.fi/fi/>.

Nimistön tietokortit. [Verkkosivusto.] Maanmittauslaitos, Helsinki. Viitattu 16.1.2025.

Saatavissa <https://tietokortit.maanmittauslaitos.fi/nimisto>.

Närhi, Eeva Maria 1978. Suomen nimitoimiston kokoelmat tutkimuksen lähteistönä.

Nimistöntutkimus ja paikallishistoria, s. 52–72. Toim. Eero Kiviniemi.

Paikallishistoriallisen toimiston julkaisuja 2. Paikallishistoriallinen toimisto ja Kotus, Helsinki.

Pitkänen, Ritva Liisa 1985. *Turunmaan saariston suomalainen lainanimistö*. SKS, Helsinki.

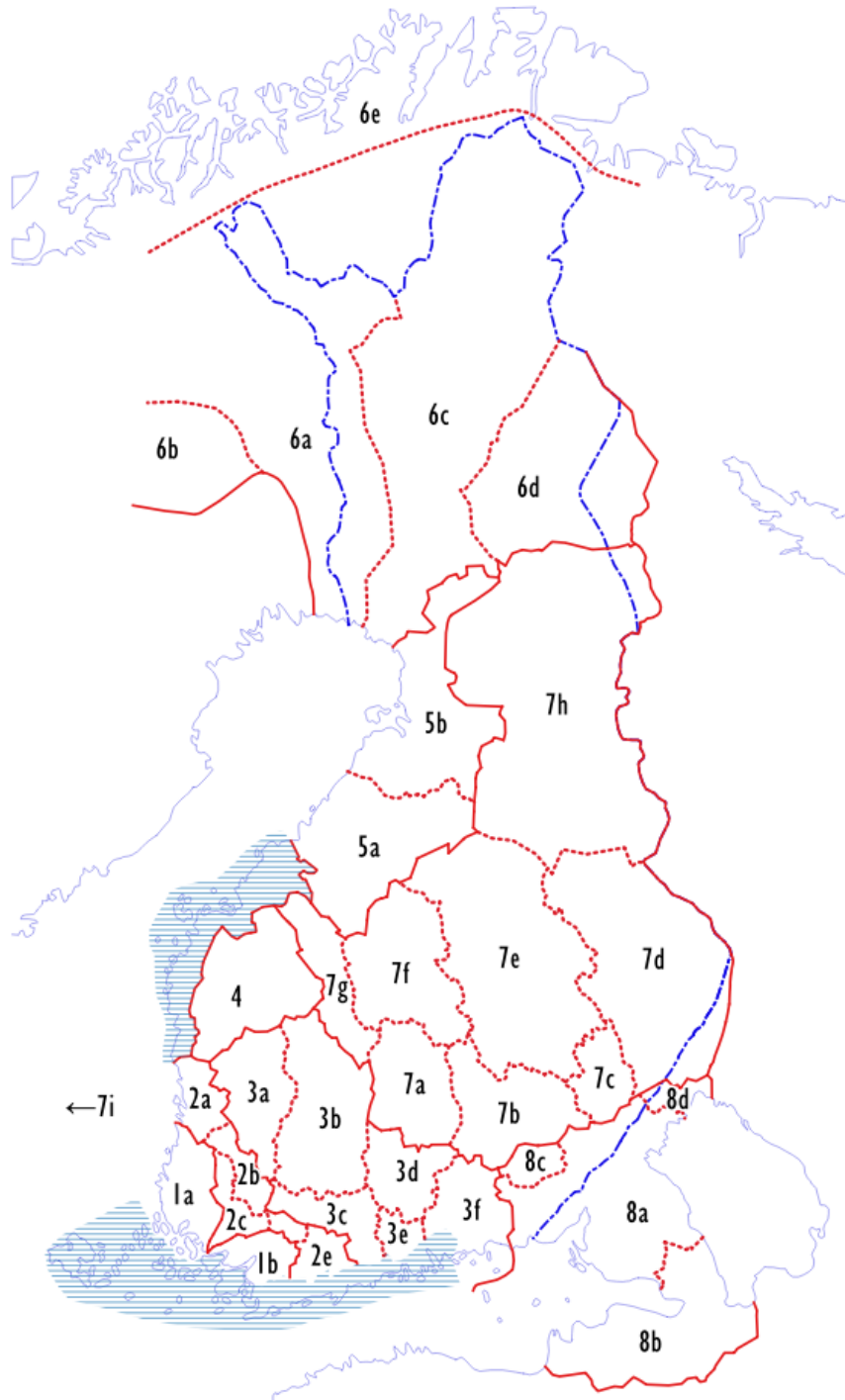
— 2007. Finnish-Swedish Contacts in Finnish Nomenclature. s. 9–26. *Borrowing of Place Names in the Uralian Languages*. Toim. Ritva Liisa Pitkänen ja Janne Saarikivi. Onomastica Uralica, Helsinki.

- Priiki, Katri 2016. Puhutun suomen kielioppia ja yksilöllistä vaihtelua. Kvantitatiivinen tutkimus *hän*-pronominista Kaakkois-Satakunnan nykypuhekielessä. *Sananjalka* 58(1), s.112–135. Suomen kielen seura, Turku. DOI: <https://doi.org/10.30673/sja.86748>.
- Rapola, Martti 1969. *Johdatus suomen murteisiin*. SKS, Helsinki.
- Räisänen, Alpo 2005. Saamelaisjäljet Kuusamon ja Posion paikannimistössä. *Virittäjä* 109(3), s. 336–356. Kotikielen seura, Helsinki. Saatavissa <https://journal.fi/virittaja/article/view/40417/9843>.
- Salmenkivi, Marko – Pulkkinen, Risto – Tuominen, Heikki 2006. Leksikaalisten ja syntaktisten nimenosien yleisyydestä ja levinneisyydestä peruskartan paikannimistössä. *Virittäjä*, 110(2), s. 190–228. Kotikielen seura, Helsinki. Saatavissa <https://journal.fi/virittaja/article/view/40476/9902>.
- Skinnarila, Niina. 1994. *Hirvijokilaakson ja Maskunjokilaakson kylännimet*. Pro gradu -tutkielma, suomen kieli. Turun yliopisto, Turku.
- Sjöblom, Paula – Ainala, Terhi 2020. Nimistöntutkimus. *Kielentutkimuksen menetelmiä I-IV*. s. 800–830. Toim. Minna Luodonpää-Manni, Markus Hamunen, Reetta Konstenius, Matti Miestamo, Urpo Nikanne & Kaius Sinnemäki. SKST 1457. SKS, Helsinki. DOI: <https://doi.org/10.21435/skst.1457>.
- SKS = Suomalaisen Kirjallisuuden Seura.
- SKST = Suomalaisen Kirjallisuuden Seuran toimituksia.
- SMS = *Suomen murteiden sanakirja* 2024. Kotimaisten kielten keskus, Helsinki. Kotimaisten kielten keskuksen verkkojulkaisuja 30. [Päivitettävä verkkojulkaisu.] Päivitetty 5.6.2024. Viitattu 15.1.2025. URN:NBN:fi:kotus-201110.
- SPNK = *Suomalainen paikannimikirja* 2007. Päätoimittaja: Sirkka Paikkala. Karttakeskus, Helsinki. URN:ISBN 978-952-5446-96-8.
- Tagliamonte, Sali – Baayen, Harald 2012. Models, forests, and trees of York English: *Was/were* variation as a case study for statistical practice. *Language Variation and Change* 24(2), s. 135–178. DOI: <https://doi.org/10.1017/S0954394512000129>.
- Tilastokeskus 2025. Luokitukset. [Verkkosivu.] Viitattu 18.2.2025. Saatavissa <https://stat.fi/fi/luokitukset>.
- Törmälä, Outi 2020. *Perseen paikka. Hävyttömät paikannimet nimistössä ja 2010-luvun mediassa*. Pro gradu -tutkielma, suomen kieli. Helsingin yliopisto, Helsinki. Saatavissa <https://helda.helsinki.fi/server/api/core/bitstreams/0890841c-f895-4757-bfd0-1ae1d8f75c78/content>.

- Vahtola, Jouko 1980. *Tornionjoki- ja Kemijokilaakson asutuksen synty: Nimistötieteellinen ja historiallinen tutkimus*. Pohjois-Suomen historiallinen yhdistys, Rovaniemi.
- Virrankoski, Pentti 1978. Mitä historiantutkija odottaa nimistöntutkimukselta. *Nimistöntutkimus ja paikallishistoria*, s. 9–23. Toim. Eero Kiviniemi. Paikallishistoriallisen toimiston julkaisuja 2. Paikallishistoriallinen toimisto ja Kotus, Helsinki.
- VISK = Hakulinen, Auli – Vilkuna, Maria – Korhonen, Riitta – Koivisto, Vesa – Heinonen, Tarja Riitta – Alho, Irja 2004. Iso suomen kielioppi. [Verkkoversio.] Suomalaisen kirjallisuuden seura, Helsinki. URN:ISBN:978-952-5446-35-7.
- Väänänen, Milja 2015. Subjektin ilmaiseminen persoonapäätteellä ja pronomiinilla. *Sananjalka* 57(1), s. 7–28. Suomen kielen seura, Turku. DOI: <https://doi.org/10.30673/sja.86735>.
- Wickham, Hadley. 2023. *stringr: Simple, Consistent Wrappers for Common String Operations*. R-project. [Verkkodokumentti.] DOI: doi.org/10.32614/CRAN.package.stringr.
- Wiik, Kalevi 2004. *Suomalaisten juuret*. Atena, Jyväskylä.
- Wright, Marvin N. – Wager, Stefan – Probst, Philipp 2024. *Package 'ranger': A Fast Implementation of Random Forests*. R-project. [Verkkodokumentti.] DOI: <https://doi.org/10.32614/CRAN.package.ranger>.
- Zilliacus, Kurt 1966. *Ortnamnen i Houtskär : en översikt av namnförrådets sammansättning*. Skrifter utgivna av Svenska litteratursällskapet i Finland, 416. Svenska litteratursällskapet i Finland, Helsingfors.

Liitteet

Liite 1. Murrealueet



Murrealueiden pääluokat ja alaryhmät ja kirjoitushetkellä niihin yhdistettävät pitäjäjaosta poikkeavat kunnat on selvitetty alla. Suomen murteiden sanakirjan pitäjälistauksen perusteella kokoamani luokitusavain on verkossa vapaasti saatavilla (ks. Virtanen 2025). Alla olevassa listassa on mainittu vain ne murrealueiden alaryhmät, joilta on aineistoa. Viimeisenä on mainittu ruotsin- ja kaksikielisiin alueisiin luettavat kunnat, joiden nimet on rajattu tämän tutkimuksen ulkopuolelle.

Länsimurteet

1 Lounaismurteet

1a Lounaismurteiden pohjoisryhmä: Uusikaupunki (ent. Pyhämaa, Kalanti, Lokalahti)

1b Lounaismurteiden itäryhmä: Salo (ent. Angelniemi, Halikko, Kuusjoki, Kiikala, Uskela, Pertteli, Muurla, Kisko, Suomusjärvi, Perniö)

2 Lounaiset välimurteet

2a Porin seudun murteet

2b Ala-Satakunnan murteet: Keikyä nyk. osa Sastamalaa (Perihämäläiset murteet)

2c Turun ylämaan murteet: Maaria ja Paattinen nyk. osa Turkua
(Lounaismurteiden pohjoisryhmä)

2d Someron murre

2e Länsi-Uudenmaan murteet: Karkkila (ent. Pyhäjärvi Ul.)

3 Hämmäläismurteet

3a Ylä-Satakunnan murteet: Sastamala (ent. Suodenniemi, Mouhijärvi, Suoniemi, Karkku, Kiikoinen, Kiikka, Keikyä, Tyrvää)

3b Perihämäläiset murteet: Hämeenlinna (ent. Hauho, Tuulos, Kalvola, Vanaja), Mänttä-Vilppula (ent. Pohjaslahti, Vilppula), Valkeakoski (ent. Sääksmäki), Jämsä kokonaan Päijät-Hämeen murteiden ryhmässä

3c Etelähämäläiset murteet: Järvenpää ja Kerava (ent. Tuusula), Renko nyk. osa Hämeenlinnaa (Perihämäläiset murteet), Riihimäki (ent. Hausjärvi)

3d Kaakkoishämäläisten murteiden Hollolan ryhmä: Lahti (ent. Hollola)

3e Kaakkoishämäläisten murteiden Porvoon ryhmä

3f Kaakkoishämäläisten murteiden Kymenlaakson eli Iitin ryhmä: Hamina (ent. Vehkalahti), Kotka (ent. Kymi), Kouvola (ent. Anjala, Elimäki, Iitti, Jaala, Sippola, Valkeala)

4 Etelä-Pohjanmaan murteet

5 Keski- ja Pohjois-Pohjanmaan murteet

5a Keski-Pohjanmaan murteet: Alaveteli ja Teerijärvi nyk. osa Kruunupyötä (Ruotsin- ja kaksikieliset alueet), Kokkola (ent. Kälviä, Lohtaja, Ullava), Siikalatva (ent. Kestilä, Piippola, Pulkkila)

5b Pohjois-Pohjanmaan murteet: Rantsila ja Temmes nyk. osa Siikalatvaa (Keski-Pohjanmaan murteet), Vaala kokonaan Kainuun murteiden ryhmässä

6 Peräpohjan murteet

6a Tornion murteet: Tornio (ent. Alatornio)

6c Kemin murteet: Keminmaa (ent. Kemi)

6d Kemijärven murteet

Itämurteet

7 Savolaismurteet

7a Päijät-Hämeen murteet

7b Etelä-Savon murteet

7c Savonlinnan seudun välimurteet: Savonlinna (ent. Savonranta, Kerimäki, Punkaharju, Sääminki)

7d Pohjois-Karjalan murteet: Lieksa (ent. Pielisjärvi), Outokumpu (ent. Kuusjärvi)

7e Pohjois-Savon murteet: Varkaus (ent. Kangaslampi)

7f Keski-Suomen murteet

7g Keuruun–Evijärven seudun välimurteet: Kauhava kokonaan Etelä-Pohjanmaan murteiden ryhmässä

7h Kainuun murteet: Ranua kokonaan Pohjois-Pohjanmaan murteiden ryhmässä

8 Kaakkoismurteet

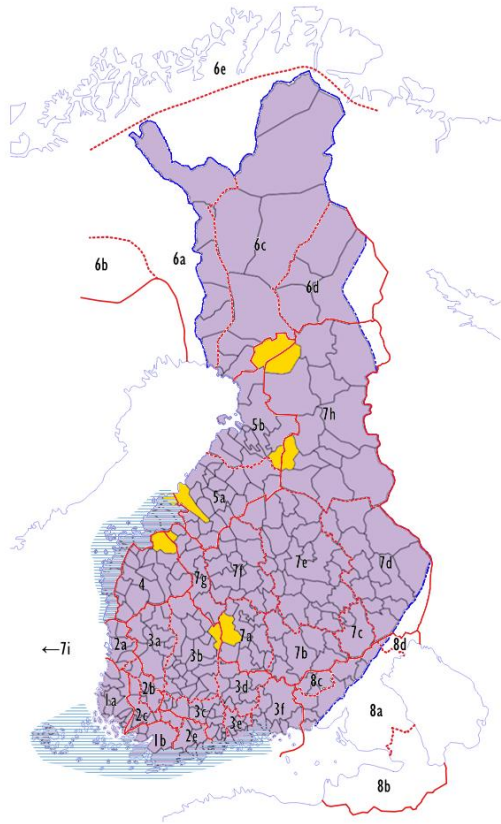
8a Varsinaiset kaakkoismurteet: Imatra (ent. Nuijamaa), Lappeenranta (ent. Lappee ja Ylämaa)

8c Lemin seudun välimurteet

Ruotsin- ja kaksikieliset alueet

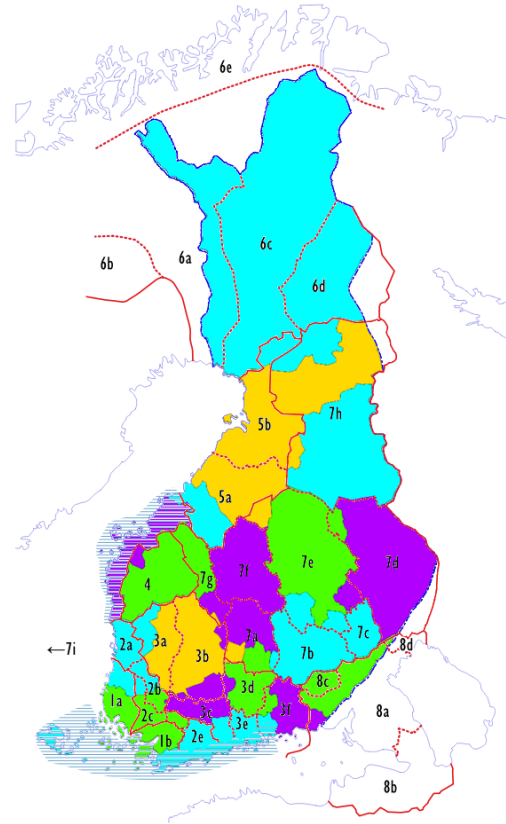
Brändö, Eckerö, Espoo, Finström, Föglö, Geta, Hammarland, Hanko, Helsinki, Inkoo, Jomala, Kaskinen, Kauniainen, Kemiönsaari, Kirkkonummi, Korsnäs, Kristiinankaupunki, Kruunupyy, Kumlinge, Kökar, Lemland, Loviisa, Lumparland, Luoto, Maalahti, Maarianhamina, Mustasaari, Närpiö, Parainen, Pedersören kunta, Pietarsaari, Raasepori, Saltvik, Sipoo, Siuntio, Sottunga, Sund, Uusikaarlepyy, Vaasa, Vårdö, Vöyri

Liite 2. Maakunnat, kunnat ja murrealueet kartalla



Murrekartta: © Kotus

Kuntakartta: © Maanmittauslaitos



Murrekartta: © Kotus

Maakuntakartta: © Maanmittauslaitos

Murrealueita havainnollistava kartta on Kotimaisten kielten keskuksen ja perustuu Itkosen (1989, 342–343) käyttämään murrejaotteluun. Yhtenäiset punaiset viivat kuvaavat pääalueiden rajoja ja katkoviivat alaryhmien rajoja. Aluekoodit on selitetty liitteessä 1. Viivoitetut alueet ovat ruotsin- tai kaksikielisiä. Taustakartat on haettu Maanmittauslaitoksen Paikkatietoikkunasta tammikuussa 2025. Vasemmanpuoleiseen kuntakarttaan on keltaisella merkitty tekstissä mainitut kunnat (pohjoisesta lukien: Ranua, Vaala, Kokkola, Kauhava ja Jämsä), jotka ovat kahden murrealueen alueella. Oikeanpuoleisen maakuntakartan värit auttavat hahmottamaan eri maakuntia, jotta murrealueiden ja maakuntien rajoja on helpompi vertailla.

Liite 3. Maasto-, vesistö- ja asutuskohteiden luokitus

Kustakin paikanlajista on esitetty ensin paikanlajikoodi ja paikanlaji ja lopuksi vielä paikanlajin kuvaus. Paikanlajien nimitykset ja kuvaukset on otettu suoraan Maanmittauslaitoksen luokituksesta (Maanmittauslaitos 2022). Vain ne paikanlajit on listattu, joita on käsitelty osassa tätä tutkielmaa. Nimenosien tarkastelussa ja mallinnuksen aineistossa on vain maastokohteiden ja vesistökohteiden nimiä. Paikanlajikoodin ensimmäinen merkki osoittaa alkuperäisen paikkakategorian (1 maasto, 2 vedet ja 3 asutus).

Maastokohteet

- 1010205 Kohoumaryhmä. Kahden tai useamman kohouman muodostama ryhmä.
- 1010210 Kohouma. Ympäristöään korkeammalla oleva maaston kohta tai alue; tunturi, vaara, mäki, harju tms.
- 1010305 Painanne. Ympäristöään matalammalla oleva maaston kohta tai alue; laakso, sola, rotko, kuru tms.
- 1020105 Metsäalue. Laajahko tai ympäristöstään erottuva metsää kasvava alue.
- 1020205 Suo tai soistuma. Vetinen tai kostea suokasveja kasvava maa-alue; kostea notkelma tai muu maastokohta.
- 1020305 Pelto. Hyötykasvien viljelyä varten muokattu laajahko maa-alue.
- 1020405 Niitty tai laidun. Puuton tai lähes puuton, ruohoa ja heinää kasvava maa-alue; alue, jossa kotieläimet laiduntavat.
- 1020505 Louhikko tai kivikko. Kivenlohkareiden tai kivien peittämä maa-alue.
- 1030105 Kivi. Kivenlohkare maalla.
- 1999999 Muu maastopaikka. Muihin paikanlajeihin kuulumaton maastopaikka; luola, hiidenkirnu tms.; tervahauta; entinen pelto.

Vesistökohteet

- 1010105 Saari- tai luotoryhmä. Kahden tai useamman saaren tai luodon muodostama ryhmä.
- 1010110 Saari tai luoto. Veden kokonaan tai lähes kokonaan ympäröimä maa-alue meressä, järvessä tai joessa.
- 1010115 Vesikivi tai -kivikko. Vedessä oleva kivi tai kivikko; vedenpäällinen tai vedenalainen kari.
- 1010120 Niemi. Meren-, järven- tai joenrannan ulkonema.
- 1010130 Ranta-alue. Meren, järven tai joen rantamaa-alue.
- 2010205 Järvi- tai lampiryhmä. Kahden tai useamman järven tai lammen muodostama ryhmä.
- 2010210 Järvi tai lampi. Vedenkorkeudeltaan yhtenäinen makean veden luonnonallas.
- 2010215 Järven osa. Järven tai lammen osa; selkä, salmi, lahti tms.

2020105 Virtavesi. Luonnon muovaama tai maahan kaivettu uoma tai ura, jossa vesi virtaa; joki, puro tai oja.

2020110 Virtaveden osa. Joen osa joka ei ole järvi, lampi tai koski; suvanto, kaarre, salmi, lahti tms.

2020115 Koski. Vuolas, aaltoileva ja kohiseva, yleensä kivikkoinen joen osa; putous.

2030105 Lähde. Paikka, missä pohjavettä purkautuu maan pinnalle.

2030110 Vesikuoppa. Lampea pienempi luonnon vesiallas.

Asutuskohteet

1999905 Hävinnyt asuinpaikka. Hävinneen talon tai kylän mukaan nimensä saanut maastopaikka.

3010110 Kylä tai kulmakunta. Alueellisen kokonaisuuden muodostava ryhmä lähekkäin sijaitsevia maa- tai pientaloja, varsinkin maaseudulla.

3030105 Talo. Asuinrakennuksen, pihapiirin ja mahdollisten läheisten talous- ym. rakennusten muodostama kokonaisuus.

Liite 5. Esimerkkiotos mallinnusaineistosta

Esimerkkiotokseen on poimittu mallinnukseen käytetystä aineistosta satunnaisesti yksi kutakin Maanmittauslaitoksen valmiiden koodien mukaista paikanlajia (huomaa, että esimerkiksi saari-sarakkeessa on paikanlajit 1010105 Saari- tai luotoryhmä ja 1010110 Saari tai luoto). Kategorisointi käy ilmi sarakeotsikoista (esim. vedenkokoumissa on kolme paikanlajiluokkaa ja kohoumissa kaksi). Paikanlajimallissa ennustettavana muuttujana on murrealuekoodi (8 luokkaa) ja ennustavina muuttujina paikanlajit (yhteensä 19 muuttujaa). Nimenalkumallissa on ennustavina muuttujina lisäksi nimenalut (yhteensä 5, ensimmäinen merkki – viisi merkkiä) ja nimen kuluminen (taulukossa nimellä *erikoinen merkki*).

Paikannimi	Paikanlajikoodi	Vilmeinen leksikaalinen osa	Murrealuekoodi	Saari	Vesikivi	Niemi	Ranta-alue	Kohouma	Painanne	Matalikko	Metsäalue	Suo	Viljelysmaa	Louhikko	Kivi	Muu maastopaikka	Vedenkokouma	Vedenkokouman osa	Virtavesi	Virtaveden osa	Koski	Lähde	Ensimmäinen merkki	Kaksi merkkiä	Kolme merkkiä	Neljä merkkiä	Viisi merkkiä	Erikoinen merkki
Eliisanniemi	niemi	1010120	8			niemi																F	El	Eli	Elii	Eliis		
Hakojärvi	järvi	2010210	4														järvi						H	Ha	Hak	Hako		
Hörskönvuoret	vuori	1010205	3					vuori															H	Hö	Hör	Hörs	Hörsk	
Kaksoskarit	kari	1010105	2	kari																			K	Ka	Kak	Kaks	Kakso	
Kannakkeenlahti	lahti	2010215	8															lahti					K	Ka	Kan	Kann	Kanna	
Karhukivi	kivi	1030105	2												kivi								K	Ka	Kar	Karh	Karhu	
Kertunlähde	lähde	2030105	2																			lähde	K	Ke	Ker	Kert	Kertu	
Kunniinsanta	santa	1010130	1			santa																	K	Ku	Kun	Kunn	Kunni	
Kydönlammi	lammi	2030110	3														lammi						K	Ky	Kyd	Kydö	Kydön	
Kylänalusta	alusta	1020305	2										alusta										K	Ky	Kyl	Kylä	Kytän	
Köröttäjälammit	lammi	2010205	5														lammi						K	Kö	Kör	Körö	Köröt	
Lotikonkellari	kellari	1010305	2					kellari															L	Lo	Lot	Loti	Lotik	
Läiskyvätkivet	kivi	1010405	1							kivi													L	Lä	Läi	Läis	Läisk	
Mansikkakarvo	karvo	1010110	5	karvo																			M	Ma	Man	Mans	Mansi	
Matarmäki	mäki	1010210	1				mäki																M	Ma	Mat	Mata	Matar	mahdollinen
Metsönpahikko	pahikko	1020505	6											pahikko									M	Me	Met	Mets	Metso	
Musta-alho	alho	1020205	1									alho											M	Mu	Mus	Must	Musta	
Muuramsuot	suo	1020205	1									suo											M	Mu	Muu	Muur	Muura	erikoinen
Myllykoski	koski	2020115	2																	koski			M	My	Myl	Myll	Myly	
Palonhaka	haka	1020105	4							haka													P	Pa	Pal	Palo	Palon	
Poosuvanto	suvanto	2020110	6																suvanto				P	Po	Poo			
Punkinoja	oja	2020105	5																oja				P	Pu	Pun	Punk	Punki	
Raumanhohde	hohde	1010115	1		hohde																		R	Ra	Rau	Raum	Rauma	
Satoalho	alho	1999999	1												alho								S	Sa	Sat	Sato		

Kuva 3 Esimerkkiotos mallinnusaineistosta