

EESTI MURDE- JA FOLKLOORIALADE PIIRITLEMISEST*

ARVO KRIKMANN

Käesoleva artikli aluseks olev veebipublikatsioon (järgnevas UU) jätkab keele- ja folkloorialade piiritlemise arutelu empiirilise statistilise vaatluse põhjal, mille aluseks on sisuüksuste geograafilise leviku arvandmestikud kolmes üsna suures eestikeelses korpuses:

- 1) murdesõnavara levikuandmed „Väikeses murdesõnastikus” (VMS I–II);
- 2) mõistatuste levikuandmed eesti mõistatuste andmebaasis (Krikmann, J., Krikmann, A. 2012);
- 3) eesti vanasõnade geograafilise leviku register (EV IV: 267–354); andmed on veidi redigeeritud ja uuendatud.

Statistikaisse on võetud tüpoloogilised üksused, mis on registreeritud kahest või enamast kihelkonnast. Selliseid murdesõnu selgus olevat 41 777, mõistatustüüpe 1416 ja vanasõnatüüpe 4911.

Geoüksusteks on 105 Eesti kihelkonda, sh kihelkonna õigustes esinev Setumaa. Vaatlusest on välja jäetud Vormsi, kust on kogutud vaid kaduvväike hulk eestikeelset murde- ja folklooriainest, samuti keelesaared (Lutsi, Leivu, Kraasna); vanasõnade andmestikust on eemaldatud ka paariteistkümne kirjapanekuga esindatud ning seega statistiliselt töövõimetus Lihula ja Kirbla. Murdesõnade jagunemisest kesk-, ida- ja rannikumurde vahel mõnedes Kirde- ja Ida-Eesti kihelkondades pole hoolitud (vt ka Krikmann, Pajusalu 2000: 135–136).

Nn inimgeograafias on ammu kasutatud väga mitmekesist statistilist arsenalit, viimase poolsajandi jooksul ka rafineeritumaid mitmemõõtmelise analüüsi meetodeid, nagu faktoranalüüs, klasteranalüüs, nn mitmemõõtmeline skaleerimine jt. Humanitaarias on need meetodid laiemat rakendust leidnud eelkõige murdeuurimises, sh on olemas ulatuslik kirjandus murdematerjali areaalse analüüsi kvantitatiivsete meetodite kohta. Kõnealuses veebipublikatsioonis ei püüta vähimalgi määral refereerida inimgeograafias kasutatud kvantitatiivseid meetodeid ega analüüsida neid teoreetilises plaanis, kuid kirjutise lõpupoolel on tehtud arglikke katseid faktor- ja klasteranalüüsi võtete tegelikuks rakendamiseks. Põhiliselt on siin, nagu ka mu eelmistes töödes, nähtuste seosetiheduse hinnangutena kasutatud kaht elementaarset parameetrit:

- 1) põikeid lineaarsete vm korrelatsiooniväljade normfunktsioonist (nn residuaale);
- 2) Zbigniew Pawłowski (1967: 38) nn kolligatsioonikoeffitsienti ehk λ -koeffitsienti, mis põhineb tõenäosuste korrutamise reeglil ning leitakse valemist

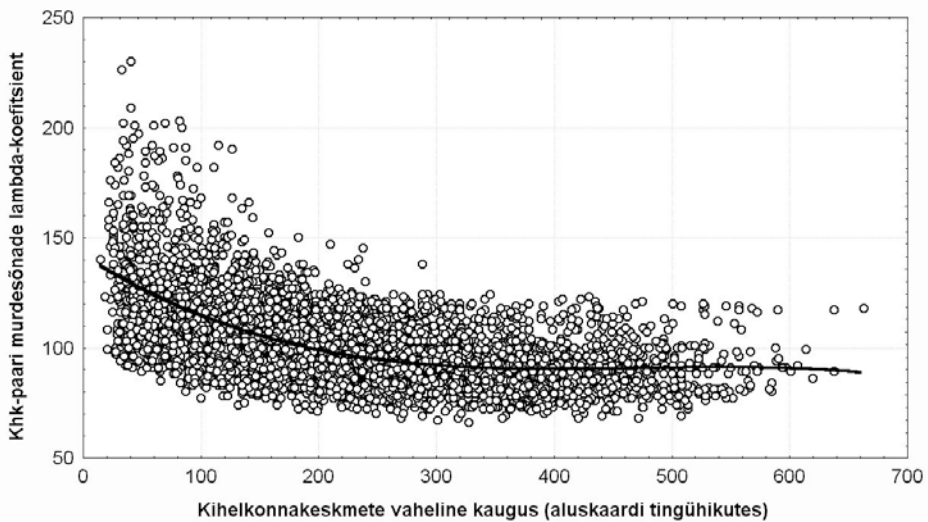
* Alljärgnev tekst on teejuht samateemalisse pikemasse e-publikatsiooni „Uusi unistusi eesti murde- ja folkloorialade piiritlemise teemal” (<http://www.folklore.ee/~kriku/TRANSPORT/Geotypo.pdf>; Krikmann 2014), mille värvikaarte pole trükitehnilistel põhjustel võimalik Keeles ja Kirjanduses avaldada.

$$\lambda = p(A_1 A_2) \div (p(A_1) \times p(A_2)),$$

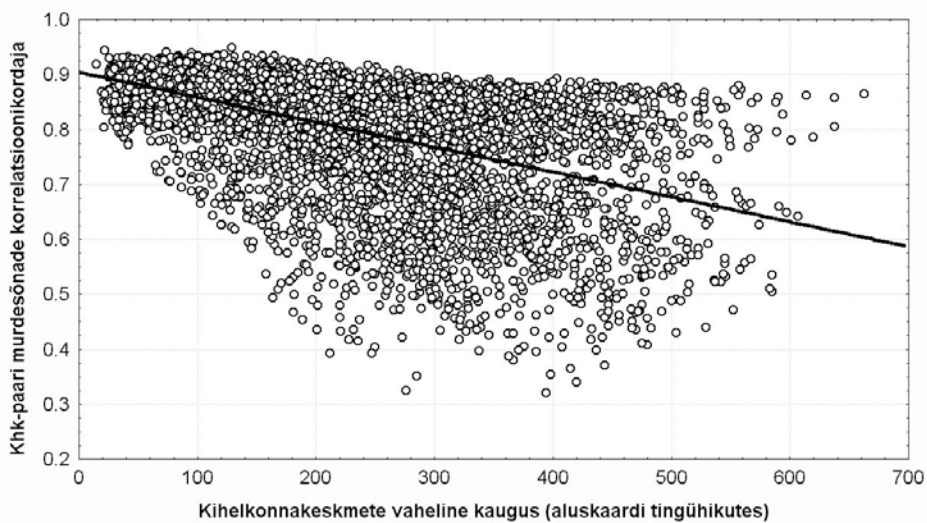
kus $p(A_1)$ ja $p(A_2)$ on mingite empiiriliste nähtuste summaarsed esinemissagedused ja $p(A_1 A_2)$ nende koosinemise sagedus; $\lambda = 1$ sõltumatute sündmuste korral, $\lambda > 1$ tähendab positiivset sõltuvust, $\lambda < 1$ negatiivset. Vaatluses on λ -koefitsiendid standardiseeritud korduvate iteratsioonide teel seni, kuni kõigi kihelkondade λ -summad piisava täpsuseni samastuvad. Andmete hoidmiseks ja töötlemiseks on kasutatud programmi Statistica 5.5 vahendeid. (Andmete menetlusviisist vt täpsemalt UU: 3.)

Nähtuste sarnasuse ~ kokkukuuluvuse ~ seosetiheduse lihtsamat järku hinnangutena on λ -koefitsiendist sagedamini kasutatud muid, praegustesse statistikaprogrammidesse sisse kodeeritud mõõdikuid, nt lineaarset korrelatsioonikordajat ja klasteranalüüsi aluseks olevat nn eukleidilist kaugust. UU: 4 esitab 9 graafi, mis kujutavad geoüksuste vahelise ühisosa suuruse ja nende kolme võimaliku seosemõõdiku suhteid kolmes nimetatud vaatlusmaterjalis. Kui ideaalis peaksid seosetugevused olema ühisosade absoluut-suurustest sõltumatud, siis näeme, et tegelikult pole seda neist ükski. Standardiseeritud lambda-väljade trendid on tõusev-koonduvad, korrelatsioonikordajate väljad ülespoole või horisontaalselt koonduvad ja eukleidiliste kauguste väljad üldiselt allapoole koonduvad. UU ei arutle lähemalt selle üle, kas ja millal tuleb nende jt võimalike seosemõõdikute suhtes rakendada n-ö ühtlaseks tempereerimise tehnikat, st püüda tingimata teisendada kõverjoonelise trendiga regressiooniväljad alul lineaarseteks, seejärel normeerida tõusva või alaneva trendiga lineaarsed väljad „horisontaalseteks”, seejärel normeerida ja standardiseerida ka paisuva või koonduva tendentsiga residuaaliväljad jne (vrd sellest ka Krikmann 1979: 34–36; Krikmann, Pajusalu 2000: 147–150). Samuti ei püüta leida kriteeriume otsustamiseks, kui sobivad on siinset laadi andmestike korral sisu- ja geoüksuste vaheliste seoste mõõdiku rolli täitma nt mainitud λ -koefitsient, korrelatsioonikordaja või eukleidiline kaugus. Provisoorsed katsetused on jätnud küll mulje, et λ -koefitsiendi korral on tekkiva „intuiitse müra” nivoo vähim, eukleidilise kauguse kasutamisel – vaatamata selle arvutusvalemi äärmisele lihtsusele – kõrgeim ja korrelatsioonikordaja korral kuskil nende kahe vahel.

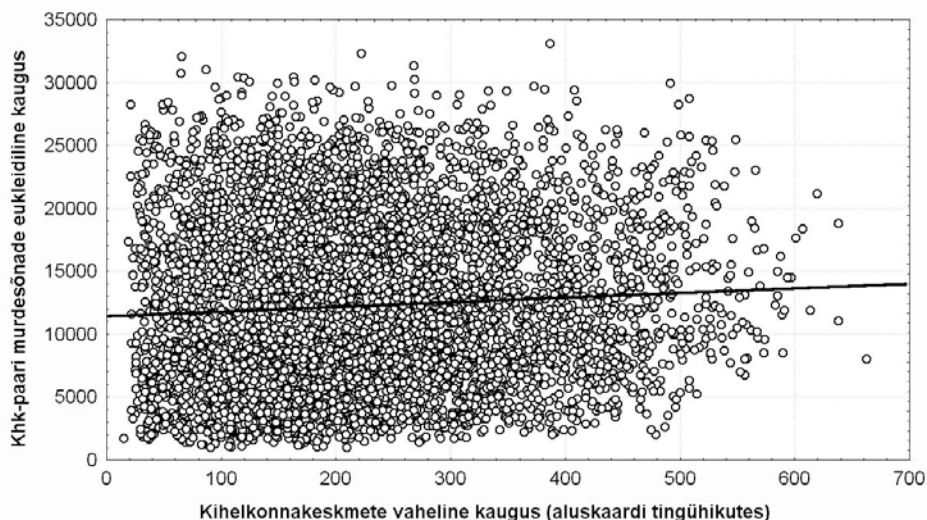
Paljud varasemad vaatlused on tõendanud, et kehtib kultuurigeograafiline triviaaltõde: m i d a l ä h e m a l, s e d a s a r n a s e m. See tõde peab paika ka eesti murdesõnade, mõistatuste ja vanasõnade andmestikes: kui sõltumatuks muutujaks võtta kihelkonnapaaride (keskmete) vaheline kaugus meil kasutatud aluskaardi tingühikutes ja sõltuvaks muutujaks nende paaride seosetihedust iseloomustavad λ -koefitsiendid, siis saame selgelt alaneva nõgusa kujuga regressiooniväljad; korrelatsioonikordajate väljad on alanev-hajuva lineaarse kujuga ning eukleidiliste kauguste korral hajusa puhmakujuliselt tõusva lineaarse trendiga (vt 9 graafi UU: 6; murdesõnade kolmik on esitatud joonistel 1–3).



Joonis 1. Murdesõnad: λ -koefitsientide sõltuvus kihelkondadevahelisest kaugusest.



Joonis 2. Murdesõnad: korrelatsioonikordajate sõltuvus kihelkondadevahelisest kaugusest.



Joonis 3. Murdesõnad: eukleidiliste kauguste sõltuvus kihelkondadevahelisest kaugusest.

Aluskirjutis UU näitab ära ka sisu- ja geooksuste vahelised sagedusjaotused, mis on omamoodi paradoksaalsed.

Sisuüksuste jagunemisel nii tekstilise kui ka geograafilise produktiivsuse plaanis näib lasuvat universaalne „Zipfi needus”: nende sagedusjaotus mõlemas plaanis on drastiliselt ebauhtlane, kusjuures kõige enam selgub ikka olevat tüpoloogiliselt või geograafiliselt „singulaarseid” üksusi, millel polegi seoseid teistega ning mida nt folklooriandmestikus justkui ei tohiks üldse olla, kuna folkloor on definitsioonikohaselt traditsioon, ja traditsiooni tuvastamiseks peaks meil olema dokumenteeritud andmeid vastavate sisuüksuste kordumise kohta. Teisalt on taoline improvisatoorsete „pürgijate” pilv kindlasti traditsiooni elujõu ja kestlikkuse sümptomiks ja garandiks. Logaritmilisel skaalal tavatsevad zipfilike sagedusjaotuste regressiooniväljad võtta lineaarse ilme; lineaarsete normeringute ebatäpsused vaadeldava kolme andmestiku logaritmilistel graafikutel võivad osalt johtuda ka sellest, et histogrammidelt puuduvad ühesainsas kihelkonnas esindatud sõnade ja folklooriüksuste tohutud hulgad (vt joonised UU: 7).

Geooksuste jagunemine neist registreeritud sisuüksuste hulga lõikes on hoopis teistsugune. Histogrammid UU: 8 meenutavad vasakule kallutatud raskuskeskmega normaaljaotust või mitme normaaljaotuse ühendit. Ka illustreerivad need graafikud kindlasti tööka, et murde- või keelearhiivist saadud sagedusandmetes on paratamatult pealistikku teave teatud sisuüksuste tegeliku tuntuse / mittetuntuse kohta teatud aladel kui ka teave neil aladel toimunud kogumistöö intensiivsuse / vähesuse kohta. Eriti sümptomaatiline on selles suhtes murdesõnade histogrammil nähtav kõrge „parempoolne küür”: selle ainuvõimalikuks seletuseks on, et murdeuurijail tekkisid millegi poolest esinduslikud või köitvad n-ö lemmikkihelkonnad, mida taas ja taas väisati. Ja pole ju loomulik uskuda, et mõne kihelkonna sõnavara on tegelikult suurusjärgu võrra vaesem kui teisel – et nt Kihelkonnal on olnud tegelikult

käibel 11 456 mittesingulaarsed murdesõna, Kuusalus 10 652, Kodaveres 10 547, Lüganusel 10 136 jne, seevastu Peetris kõigest 1518, Rakveres 1289, Noarootsis 928 ja Paides 728.

Murdeuurijad on tavatsenud murdepiire näha seal, kus ühtivad mitmete tähtsate isoglosside parmad. Meie kolmes aines on seevastu ülekaalus laiad, kuid koldelise ja katkendliku ilmega levikupildid, mis vastavad Mart Rummeli (1979: 18–19) levikutüübile „Something is everywhere”. Sedalaadi kaootiliste levikupiltide näiteid tuuakse S. H. Vestringi sõnaraamatus leiduvate sõnade murdetähtsustades (UU: 9) ning mitmetel kartogrammidel lõunaeesti mõistatus- te kogumikus „Tere tee, tere meele, tere egalõ talolõ” (Krikmann 2000a: 300 jj), mis jätavad mulje, nagu oleks tugev kagutuul puhunud üleskirjutusi nagu lehti Kagu- vm Lõuna-Eestis asuvast hunnikust üha hõrenevalt loode suunas laiali (vt UU: 10).

Kirjutises UU tuuakse ka Eesti kihelkondade murde- ja folklooriainese stereotüüpia / unikaalsuse andmed kolmes nimetatud vaatlusaines. UU: 11–14 leiduvad tabelid selgitavad stereotüüpshinnangute aluseks olnud parameetreid. Nendeks on kihelkondadest noteeritud murdesõnade, mõistatus- ja vana- sõnatüüpide hulk (t_K) ning neis t -komplektides esinevate sisuüksuste „kohtumiste” hulk teiste valimis esinevate geouksustega (r_K) (vrd ka Krikmann 1979: 28–29; 1980: 3–4; 1997: 225 jpt).

Arusaadavail looduslikel põhjustel pole kihelkonna keele- või folkloori- ainese t - ja r -näitude seos lineaarne, vaid kergelt kumer: kuna sisuüksuste sagedusjaotus mis tahes loomulikus ainevalimis on zipfilikult ebaühtlane (väike- si üksusi palju, suuri vähe, keskmisi keskmisel määral), siis saab pikem mittekorduvate sisuüksuste loend koosneda ainult suhteliselt unikaalsemast aimest, kuna sagedaste, stereotüüpsete üksuste varu on väike ja ammendub üsna peatselt.

Kuigi kõnealuselt kolmikvaatlusest oli välja lülitatud ainestike kõige unikaalsem osa, tõestas see jätkuvasti, et paika peab teine kultuurigeograafiline põhitõde: perifeeriates aines on unikaalne, südamaa aines stereotüüpne. Perifeeriatena tõusevad esile eelkõige Kagu-Eesti laias mõttes (Se, Võ, TaL) ja saared, nõrgemal määral ka Lõuna-Eesti kompleksi liituv Mulgi, saarte kompleksi liituv Läänemaa rannäärne mandriosa ning mõned Kirde-Eesti rannakihelkonnad (eriti Kuu, Hlj, VNg).

Mulle oli ammuilma teada juba üldisemaist loodusseadustest johtuv tõik, et stereotüüpse materjaliga Põhja- (ja eriti Süda-Eesti) geouksuste suhtlemispilt teiste geouksustega on loid ja üheülbaline, unikaalsema materjaliga alade oma aga dünaamiline ja reljeefne. Minu eelmainitud jm varasemais töödes leidub rohkesti siinsega sarnaste hindamise meetodite alusel saadud kartogramme Eesti kihelkondadest noteeritud murdesõnavara ning mõistatus- ja vanaõnaainese stereotüüpia/unikaalsuse kohta. Refereeritud kirjutises (UU: 15–16) on neid tõiku illustreeritud Anna, Rapla ja Pilistvere *versus* Viru-Nigula, Kihelkonna ja Urvaste kihelkonna suhtlemispiltidega saidil „Eesti kihelkondade murdesõnavaralisi suhteid” (Krikmann [2000b]) ning Türi, Vigala ja Pärnu-Jaagupi *versus* Nissi, Püha ja Vastseliina mõistatuste suhtlemispiltidega saidil „Eesti kihelkondade suhtlemistihedus mõistatus- tes” (Krikmann [1998]).

Oma varasemais töodes olen kihelkondadevahelise vanasõnade leviku- andmestiku seosetiheduse hindamiseks püüdnud välja töötada koguni oma- enda metoodikat, mis arvestaks mitte ainult nende ühisosa suurust, vaid ka suhtelist stereotüüpsust, olles arvanud, et küsimus pole mitte niivõrd „täpse lähendfunktsiooni leidmises, vaid sellise lähendfunktsiooni leidmises, millel oleks seesmine loogika” (Krikmann 1997: 224). Testides hiljem oma „vanasõnavalemi” võimalikku üldisemat kasutatavust, olen saanud kahetisi tule- musi: nt Eesti kihelkondade murdesõnavara stereotüüpsuse hindamiseks on see sobinud päris hästi (vt Krikmann, Pajusalu 2000: 141), H. Gösekeni, S. H. Vestringi ja A. Thor Helle vanasõnade geograafilise tausta leidmiseks aga mit- te, mistõttu selleks on kasutatud teistsugust arvutusviisi (vt Krikmann 1986: 112–114).

Kõnealuses kolmikvaatluses on *t*- ja *r*-vahekordi normeeriva lähendfunk- tsioonina kasutatud lihtsalt 2. astme polünoomi ja kihelkondlike sisukogumite hinnangutena normipõigete (nn residuaalide) suurusi (vt UU: 18 ülal) – ni- metagem neid siin *s*-koefitsientideks. Kolmel järgmisel kartogrammil (UU: 18 keskel) on toodud *s*-koefitsientide „toored” empiirilised jagunemispildid ning kolmel viimasel (UU: 18 all) silutud jagunemised. Silumised on toimunud liht- viisiliselt samamoodi nagu mu mitmetes eelmistes töodes: iga kihelkond saab igas järgmises iteratsiooniringis näiduks ta enda ja ta naabrite eelmise ringi näitude keskmised. Näiteks:

Ambla jaoks: (Amb+Kuu+Kad+VMr+JJn+JMd+Kos+HJn) ÷ 8;

Anna jaoks: (Ann+JMd+JJn+Pee+Pai+Tür+Kos) ÷ 7 jne.

Kui kolmikvaatluses olid iga kihelkonna λ -koefitsientide summad itereeri- tud võrdseteks, jäid nad siiski erinevaks oma hälbemäärade poolest: unikaal- sema ainega kihelkondadel on need eeldatavasti suuremad ning stereotüüp- sema ainega kihelkondadel väiksemad. Vaatlus näitaski, et λ -koefitsientide standardhälbed sobiksid samuti päris hästi geoüksuste stereotüüpsuse iseloo- mustajateks (vt kaardisarja UU: 19).

UU: 20 on toodud illustreeriv katke mõistatuste λ -tabelist, kus diagonaal- väärtused (s.o kihelkondade „ühisosad iseendaga”) on nullitud. Edasi tuletati iga sisukorpuse (murdesõnad, mõistatused, vanasõnad) λ -koefitsientide pool- tabeli põhjal kihelkonnapaaride koefitsientide (λ -de täisarvuliste 100-kordse- te) paremusjärjestuste tabel, mille pikkuseks on murdesõnade ja mõistatuste korral $(105 \times 104) \div 2 = 5460$ rida ja vanasõnade korral $(103 \times 102) \div 2 = 5253$ rida. Järgmine tabel (UU: 21) näitab katkendit mõistatuste λ -pingerea üla- otsast: kihelkonnapaaride lühendid, nende *x*- ja *y*-koordinaadid aluskaardil ning λ -koefitsiendid.

Edasi püüti jälgida, mis toimub nende λ -pingeridade sees: millised geo- üksused üksteist armastavad ja millised väldivad, millistes ruumisuhetes paarid domineerivad skaala eri osades jne.

Refereeritavas töös on toodud ulatuslikke väljavõtteid neist arengutest alul 100, siis 250 sammu pikkuste etappide kaupa murdesõnades (UU: 22–33), mõistatuses (UU: 34–45) ja vanasõnades (UU: 46–56):

1) iga tabelirea vasakpoolne kaart näitab antud vahemikus esindatud ki- helkondade seostumissagedusi teistega;

2) keskmine kaart annab vasakpoolsest silutud pildi (3 iteratsiooni);

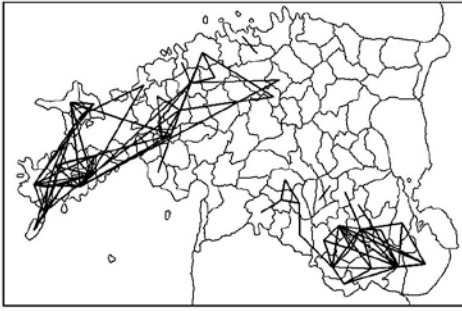
3) parempoolsetel kartogrammidel ühendavad mustad sirgjooned seosesse angažeeritud kihelkondade keskmeid, sinised kastikesed näitavad nende sirgete keskpunkte ja suuremad punased tähnid siniste joonepoolitajate keskmist asukohta.

Need UU leheküljevahemikud on meelega küljendatud nõnda, et pdf-faili kiirel lehitsemisel tekiks mõningane stroboskoobiefekt.

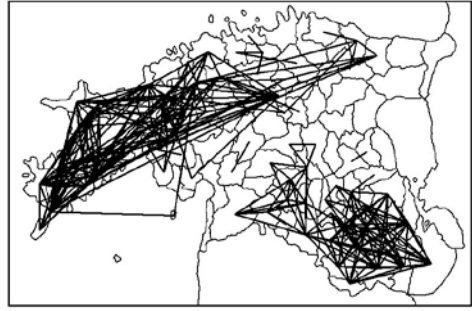
Vajalike arvutuste tegemiseks ja andmete parempoolsete must-valge põhjaga kartogrammid koostamiseks kasutasin möödunud sajandi lõpuaastail koostatud *offline*-režiimil töötavat kaardistajat, mis põhineb samuti Statistica 5.5 vahenditel; arvandmete ja graafika vahendamiseks oli vaja kirjutada mõningaid uusi skripte. Sini-kollase põhjaga kartogrammid on saadud eelnimetatud *online*-kaardimasina abil (Krikmann, J., Krikmann, A. 2012), mis laseb hõlpsasti teha ka silutud kaarte, kuid ei võimalda kanda kaartidele joongraafilisi objekte. Piiranguid tulenes ka sellest, et Statistica 5.5 ei luba oma .stg-laiendiga joonistele kanda rohkem kui 901 lisamärki; selle maksimumi ületamisega kaasnevad häälekad protestid.

Mind valdas hetkeline nukrus, kui taipasin, et mul pole mingit võimalust enamikule kaardiridadele juhuslikesse kohtadesse tekkivaist värvilaikudest ja praktiliselt kogu maad katvaist joonetihnikutest leida seaduspärasusi või tendentse, mis võiksid kehtida l-skaalade põhiosas. Selgituseks (UU: 58) on toodud kõrvutatult l-sageduste histogrammid ja neile vastavad „pinnalaotuslikud” variatsiooniread; nende vastavused teineteisele n-ö vasakult paremale on pöördpidised ja suhete neutraalsushetke näitavad punased kontrolljooned asuvad eri telgedel. Neid vaadates sugenes lohutav oletus: siit pole võib-olla põhjust mingeid seaduspärasusi otsidagi. Kõik tuvastatavad murde- või folkloorialad konstitueeritakse juba λ -astmike tajutavalt diskreetsete astakutega „paremikes”, esimese u 500 astme jooksul, kus nidusad võrgustikud tekivad kõigepealt Kagu-Eestisse ja/või saartele ning hargnevad sealt hajusamate ja hõredamate metastaasidena vastavalt Lõuna-Tartumaale ja Mulgisse ning Loode- ja Lääne-Eesti mandriossa, ning ühtlasi tekivad mõned uued, kuid nõrgemad ja hõredamad seostekolded Põhja-Eesti rannaalale, Järvamaale ja mujale. Suur osa kõigist seostumistest toimub aga järgneval pikkamisi lineaarselt alaneval kaldteel, mis katab λ -skaalade koguulatusest ainult kitsa, kõige neutraalsemaid suhteid hõlmava vahemiku (u 1,1–0,9; tehniliselt: 110–90). Põhiprotsessid toimuvad Põhja-Eestis, kuid põhimõtteliselt võivad igalt alalt pärit geouksused seostuda mis tahes muult alalt pärinevatega. Ning skaala lõpuosa, kus koefitsientide kahanemine taas kiireneb, haakuvad peamiselt unikaalsete alade esindajate (nõrgimad) seosed teistelt unikaalsetelt aladelt pärinevate geopartneritega.

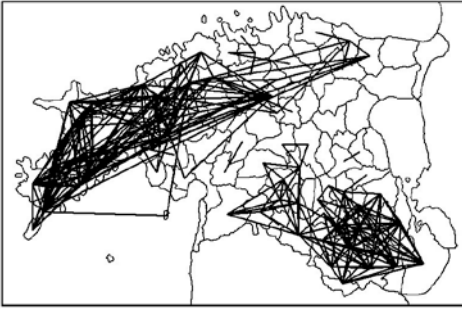
Kaardil 1 on toodud illustreeriv näide „mõistatusalade” tekkest: geouksuste seostevõrgud λ -pingerea vahemikes 1–100 (a), 1–200 (b), 1–300 (c) ja 1–400 (d) ning 1500 (e) seostumise järel tekkinud läbipaistmatu joonterägastik.



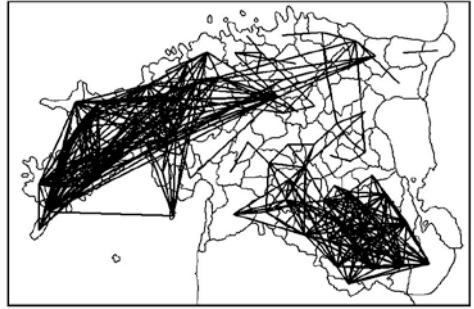
a) Vahemik 1–100.



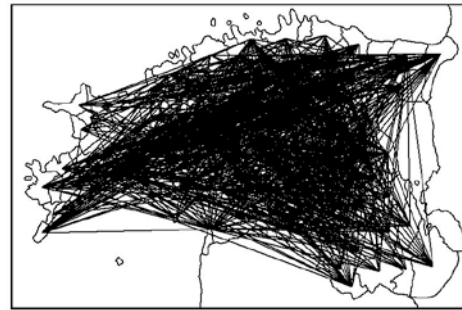
b) Vahemik 1–200.



c) Vahemik 1–300.



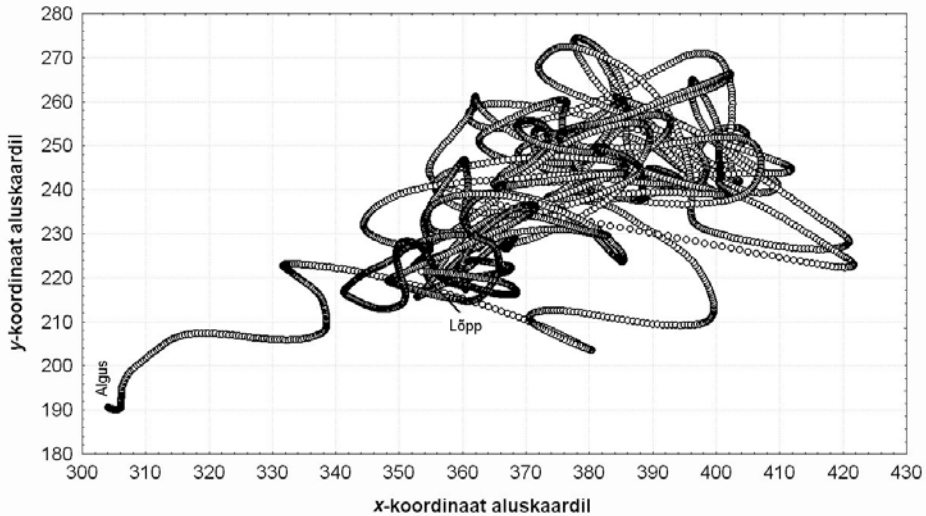
d) Vahemik 1–400.



e) Vahemik 1–1500.

K a a r t 1. Geotüksuste seostevõrgud mõistatuste λ -pingerea eri vahemikes.

UU: 58 alumisest graafikolmikust pärineb järgnev köiepuntra-kujund joonisel 4, mis näitab „siniste kastikeste” (st kihelkonnakeskmeid ühendavate sirgete keskpunktide) kaootilist liikumist läbi kogu mõistatuste λ -pingerea (näidud silutud libiseva keskmise meetodil: 3 iteratsiooni, silumisvahemik pingerea 50 astet).



Joonis 4. Köiepuntra-kujund: kihelkonnapaaride „silutud keskmise” asukoha muutumine läbi mõistatuste λ -pingerea.

Taipasin ühtlasi, et mu ammune katse „eesti vanasõnaraajoonide” leidmiseks (vt Krikmann 1980) oli olnud metoodiliselt väär, kuna kahepoolsed 1. astme rajoonid tekitati nii, et tugevaimatest seostest alates hakati maad täitma n a a b e r kihelkondi ühendavate graafidega ja jätkati seda seni, kuni iga kihelkond oli haaratud mõnesse graafi ning graafidevahelisi „velahkmeid” käsitati 1. astme rajoonide piiridena. Edasi hinnati 1. astme rajoonide vahelisi seose tihedusi ja sarnaseima sisuga rajoonid sulatati üheks suuremaks rajooniks ning jätkati seda seni, kuni maad läbis üksainus „viimne piir” (vt kaardinelik UU: 59).

Mu järgnevad kogemused, sh kõnealuse kolme andmestikuga, on siiski näidanud, et kuigi tugevaimad seosed tekivad tõesti just perifeersete naaber kihelkondade vahele, siiski üsna pea (ja eriti Põhja-Eestis) sugeneb rohkesti positiivseid suhteid ka mitte-n a a b e r kihelkondade vahel, mis muudavad sellise hierarhilise rajoneerimisviisi küsitavaks.

Kõik senised vaatlused jätavad igatahes mulje, et geouksuste perifeersus/südamaisuus, aine unikaalsus/stereotüüpsus ja geouksustest moodustuvate gruppide koherents / hajusus on sisuliselt ühe ja sama asja mõõtmed. Vaadeldud andmestikest nähtuvad kaks markantset ja üks vähem eenduv keele- ja traditsiooniala:

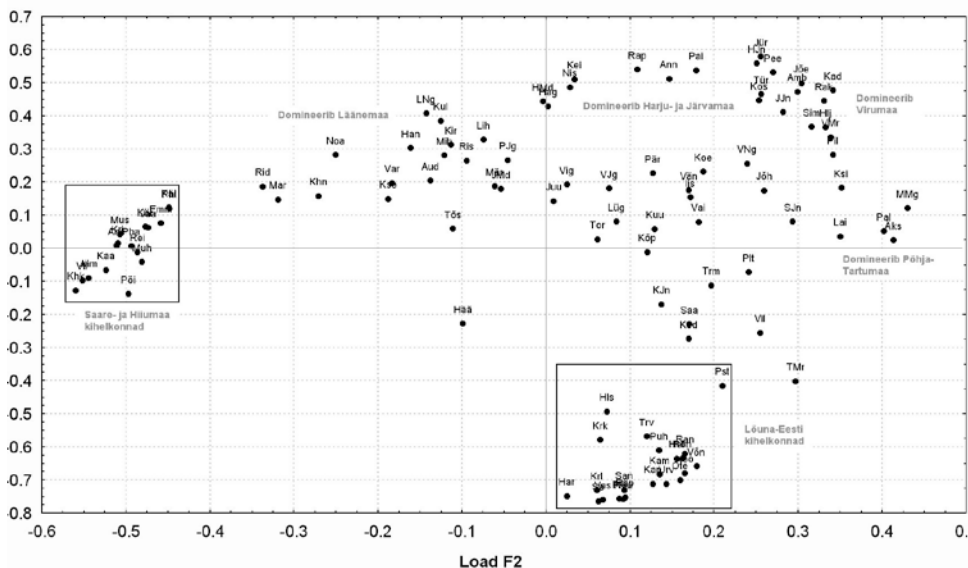
- 1) Kagu-Eesti laias mõttes (Se+Võ+TaL) koos nende külge haakuva haju-sama Mulgiga;
- 2) saared koos nende külge haakuva hajusama Lääne- ja Loode-Eestiga;
- 3) Eesti põhja- ja kirderannik.

Ka Põhja-Eesti keskosas kehtib üldiselt reegel „Mida ligemal, seda sarnasemad”, kuid mingeid suuremaid selgemalt koonduvaid georühmi ei nähtu.

Kontrollimaks neid oletusi suuremal proovitükil tegin järgmise lisavaatluse. Summeerisin samades positsioonides paiknevad λ -näidud läbi kõigi kolme maatriksi (murdesõnad + mõistatused + vanasõnad), käsitasin neid summasid algandmetena ja arvutasin nende põhjal lahtritele uued standardiseeritud λ -koefitsiendid. Valisin Statistica 5.5 faktoranalüüsi moodulist järgmise opt-sioonide jada:

Input file: Raw data. MD deletion: Casewise → Variables: Select All → Extraction method: Principal components. Maximum no. of factors: 3. Minimum eigenvalue: 1.000 → Factor rotation: Varimax raw → Plot of loadings: 2D → Factor 2 versus Factor 1.

Selle tulemusel tekkis graaf, mis meenutas äratuntavalt Eesti kihelkondade geograafilist konfiguratsiooni ning kihelkondade koonduvust ilmakaarte ja südamaisuse / perifeersuse lõikes. Faktorite 1 ja 2 vahel selgus olevat kerget positiivne lineaarne suhe; pärast selle normeerimist suurenes sarnasus Eesti kaardiga veelgi (vt joonist 5 ja UU: 61).



Joonis 5. Murdesõnade, mõistatuste ja vanasõnade koondandmestiku λ -koefitsientide faktoranalüüsi graafiline tulem.

Joonise 5 paremas allservas näeme Lõuna-Eesti klastrit tiheda Kagu-Eesti ja hajusama Mulgi komponendiga. Vasakul on teine ilmne tihend – saarte rühm. Sellest paremal eristub suhteliselt selgesti peamiselt Lääne- ja Loode-

Eesti kihelkondadest moodustuv hajusam jada, mis edasi alla paremale liikudes teiseb peamiselt Harju- ja Järvamaa, siis Virumaa ja edasi Põhja-Tartumaa kihelkondadest koosnevaks kaareks. Selle kaare ja Lõuna-Eesti klasteri vahele graafi paremkeskmisse ossa jääb kõige hajusam piirkond, kus domineerivad Põhja-Pärnumaa ja Põhja-Viljandimaa kihelkonnad. On huvitav mainida, et Häädemeeste kihelkond, mille aines koosneb valdavalt legendaarse murde- ja rahvaluulekoguja Marta Mäesalu kirjapanekuist, paikneb uhkes üksinduses telgede ristumispunkti lähedal.

Lõuna-Eesti sülemi suurenemine (UU: 62) näitab selgesti Võru-Setu, Lõuna-Tartumaa ja Mulgi allosade lähedusi ja tihedusi; saarte sülemi suurenemisel (sealsamas) pole selgemaid allrühmi näha: nt Sõrve kihelkonnad asuvad teineteisest üsna kaugel; Käina ja Pühalepa on väga lähestikku, kuid Emmaste ja Reigi neist tükk maad eemal jne.

Kui kihelkondade faktorskoore (tehniliselt: nende täisarvulisi korrutisi 100-ga 2 korda silutuna) mõtestada mingite neid iseloomustavate arvparameetritena, siis saaksime kartograafilised interpretatsioonid, kus Faktor 1 tähendaks loodepoolsust, Faktor 2 „põhja-kirdepoolset südamaisust” ja nõrgim Faktor 3 võib-olla „loode-kagusuunalist perifeersust” (vt UU: 62, alumised kartogrammide).

Katsetasin analoogilisi menetlusi ka murdesõnade, mõistatuste ja vanasõnadega eraldi. Nende graafilised tulemused olid mitmekesisema väljanägemisega ega olnud nii heas vastavuses Eesti geouksuste tegeliku konfiguratsiooniga.

Murdesõnade faktoranalüüsi graafil (UU: 64) moodustus ümber telgede lõikepunkti sõritaoline formatsioon, kuid murdealases plaanis eristus siin selgesti vaid alumine „Lõuna-Eesti kaar”, kogu ülejäänud (ülalpoollises) Põhja-Eesti osas paiknes kaootiliselt läbisegi nii lääne-, põhja- ja idapoolseid kui ka keskeestilisi geouksusi.

Mõistatuste graafil (UU: 65) vastas Lõuna-Eesti ja saarte sülemite paiknemine enim geograafilisele tegelikkusele, kui F1-faktori laadungväärtused olid pööratud 180° ümber telje ($x = 0$) ja sõltuvaks muutujaks oli võetud faktorlaadung F2 algkujul. Kuid saarte sülemisse oli sattunud ka palju mandrieestilisi „võörkehi” ning tekkinud sülemite piirid olid üldse ebamäärased.

Vanasõnade menetlusvariant oli täpselt samane sellega, mida katsetati koondmaterjalil (vt joonist 5), ning ka tulem (UU: 66) sellega kõige sarnasem. Lõuna-Eesti sülem eristus selgesti, kuid sisaldas ka mõningaid lähikihelkondi (Vil, TMr). Hiiumaa kihelkonnad olid saarte sülemi küljest lahti veninud ning sulanud kokku Loode- ja Lääne-Eesti kihelkondadega. Hajumi põhja- ja idapoolsed servad olid täidetud peamiselt Harju-, Viru- ja Põhja-Tartumaa kihelkondadega, stereotüüpseimad Kesk-Eesti alad jäid taas kõigi ülejäänute keskele, kõige lähemale telgede lõikepunktile.

Programmiga Statistica 5.5 on ka mõningad klasteranalüüsi võimed. Valisin selle klasteranalüüsi moodulist järgmise optioonide jada:

K-means clustering → Variables: Select All → Number of clusters: 3. Number of iterations: 10. Missing data: Casewise deleted. Choose observations to maximize initial between-cluster distances.

murdesõnalisi suhtluspilte näitavates kuuikutes (Krikmann [2000b]). Siin si-
rutuvad selgesti ja ilmakaareliselt enam-vähem õigetes suundades kolm uni-
kaalsemat poolsaart: lääne-loodepoolne, põhja-kirdepoolne ja lõunaeeestiline,
ning nende ristumiskohta jääb stereotüüpne südamaa.

Unikaalsemate mõistatuste graaf (UU: 69), mille lähteandmed olid saa-
dud siin kasutatud tavalise λ -protseduuriga, eristas selgesti kaks saart:
Saaremaast ja mõnedest Hiiumaa kihelkondadest koosneva ning Lõuna-Eesti.
Peamiselt Läänemaa *resp.* Põhja-Tartumaa kihelkondadest koosnevad ribu-
rajad ühendasid neid parem-ülalpoolse hajusa Põhja-Eesti platsdarmiga, mille
kirdenurka hakkas tekkima unikaalsematest kihelkondadest koosnev tipp.

Unikaalsemate vanasõnade graaf (UU: 70, lähteandmed samuti siinse
tavalise λ -protseduuri tulem) näitas kaht eraldiseisvat tihedakoelist saart –
Saaremaad (v.a Jaani) ja Kagu-Eestit, ning kaht hajusamat – Mulgi ja Põh-
ja-Tartumaa sülemit. Geograafiliselt segase koostisega Põhja-Eesti hajumist
graafi parem-ülalpoolses osas küünitus lääne suunas kitsas Pärnu-Lääne-
Hiiumaa kihelkondadest koosnev poolsaar.

Seega näib, et faktoranalüüsi võimendused Eesti murde- ja folkloorialade kaar-
distamiseks küünivad kaugele üle metafoorsuse piiride. Hulganiisti Eesti
kihelkondade murde- ja folklooriainese sarnasust mõõtvaid λ -hindeid, mis ei
kätkenud mingit geograafilist informatsiooni, sisestati analüsaatorisse, mis
toimis samuti vaid puhtabstraktsete ruumide tasandil, ja saadi vastuseks hul-
ganisti graafilisi väljundeid, mis meenutasid Eesti geograafilist kaarti ning
andsid ühtlasi kvaasigeograafilises mõõdukus teavet sellel paiknevate ob-
jektide lähedussuhete kohta.

Kirjandus

- EV IV = Eesti vanasõnad IV. Lisad. Toim Arvo Krikmann, Ingrid Sarv. Tallinn:
Eesti Raamat, 1988.
- Krikmann, Arvo 1979. Some aspects of proverb distribution. – Mathematical
Processing of Cartographic Data: Summaries of the Symposium. Tallinn: Academy
of Sciences of the Estonian S.S.R., lk 28–44.
- Krikmann, Arvo 1980. Towards the typology of Estonian folklore regions.
(Preprint KKI-16.) Tallinn: Academy of Sciences of the Estonian S.S.R.
- Krikmann, Arvo 1986. Fraseoloogiline aines eesti vanimais grammatikates ja
sõnastikes. Tallinn: Valgus.
- Krikmann, Arvo 1997. Sissevaateid folkloori lühivormidesse I: põhimõisteid,
žanrisuhteid, üldprobleeme. Tartu: Tartu Ülikooli Kirjastus.
- Krikmann, Arvo [1998]. Eesti kihelkondade suhtlemistihedus mõistatustes.
Tartu: EKM FO. <http://www.folklore.ee/~kriku/LEVIK/coefs.htm> (19. VI 2014).
- Krikmann, Arvo (koost) 2000a. Tere tee, tere mee, tere egalõ talolõ. Valik
lõunaeeesti mõistatusi. Tartu: [Eesti Kirjandusmuuseum].
- Krikmann, Arvo [2000b]. Eesti kihelkondade murdesõnavaralisi suhteid. Tar-
tu: EKM FO. <http://www.folklore.ee/~kriku/MURRE/Index.htm> (19. VI 2014).
- Krikmann, Arvo, Pajusalu, Karl 2000. Kus on keskmurde keskpunkt. – In-
ter dialectos nominaque. Pühendusteos Mari Mustale 11. novembril 2000. (Eesti
Keele Instituudi toimetised 7.) Koost ja toim Jüri Viikberg. Tallinn: Eesti Keele
Sihtasutus, lk 131–172.

- K r i k m a n n, Jaak, K r i k m a n n, Arvo (koost) 2012. Eesti mõistatused. Andmebaas. Tartu: EKM FO. <http://www.folklore.ee/moistatused> (19. VI 2014).
- P a w ł o w s k i 1967 = Збигнев Павловский, Введение в математическую статистику. Перевод с польского В. Д. Меникера; под ред. Ф. Д. Лившица. Москва: Статистика.
- R e m m e l, Mart 1979. Large Data Bases in Quantitative History: Some Handling Techniques. (Preprint KKI-12.) Tallinn: Academy of Sciences of the Estonian S.S.R.
- UU = Krikmann, Arvo 2014. Uusi unistusi eesti murde- ja folkloorialade piiritlemise teemal. Tartu: EKM FO. <http://www.folklore.ee/~kriku/TRANSPORT/Geotypo.pdf> (19. VI 2014).
- VMS = Väike murdesõnastik I–II. Toim Valdek Pall. Tallinn: Valgus, 1982, 1989.

On the areal division of Estonia according to dialect and folklore material

Keywords: dialect vocabulary, riddles, proverbs, areal distribution, dialect areas, folklore areas, proverb regions, statistics, strength of relationships, coefficient of colligation, correlation coefficient, euclidean distance, regression analysis, factor analysis, cluster analysis

The article draws on the author's web publication *Uusi unistusi eesti murde- ja folkloorialade piiritlemise teemal* („More dreams on the areal division of Estonia as based on dialect words, riddles and proverbs”, <http://www.folklore.ee/~kriku/TRANSPORT/Geotypo.pdf>), which studies the corpora of Estonian dialect words, riddles and proverbs seeking for arguments to support the following assumption:

The peripheral/central position of geographic units, the rare/common character of the material and agglomeration/dispersal of groups formed of the geographic units are, essentially, dimensions of the same thing.

Our analysis of the Estonian data of three domains reveals two well-defined areas of language and culture and one less salient:

- (1) South-Eastern Estonia in abroad sense (Se+Võ+TaL) together with the less concentrated Mulgi region;
- (2) the West-Estonian Islands together with the less concentrated Western and North-Western Estonia;
- (3) the Northern and North-Eastern Coast of Estonia.

For the central part of Northern Estonia, the closer the more similar is also the general rule, without revealing, however, any well-defined areal groups. The material similarity of the geographic units is measured by average standardized coefficients of colligation (λ -coefficients). Graphic identification of the dialect and folklore areas relies on the descending λ -rankings of the strength of relationships between pairs of parishes, complemented by some elements of factor and cluster analysis.

The article is introduced by a brief overview of the frequency distributions of cultural versus geographic units, relationships between λ -coefficients, correlation coefficients, and euclidean distances and the distances between geographic units, as well as of the geographical distribution of rare versus common material.

Arvo Krikmann (b. 1939), PhD, Member of the Estonian Academy of Sciences, Estonian Literary Museum, Department of Folklore, Senior Researcher, kriku@folklore.ee