

TARTU ÜLIKOOL
HUMANITAARTEADUSTE JA KUNSTIDE VALDKOND
EESTI JA ÜLDKEELETEADUSE INSTITUUT

Mirjam Ruutma

Kaassõnad eesti murretes

Magistritöö

Juhendajad PhD Kristel Uiboaed ja MA Maarja-Liisa Pilvik

Tartu 2016

Sisukord

Sissejuhatus	3
1. Teoreetiline taust	5
1.1. Kaassõnad eesti keeles	5
1.1.1. Kaassõna mõiste	6
1.1.2. Kaassõnade tähendustest	7
1.1.3. Lauseliikmete vaheliste suhete väljendamine analüütiliselt ja sünteetiliselt ..	9
1.1.4. Piirid kaassõnade ja teiste sõnaliikide vahel	11
1.2. Korpuspõhine dialektoloogia	13
2. Andmestiku kirjeldus	15
2.1. Lähtematerjal	15
2.2. Kodeerimisskeem	18
3. Meetod	21
3.1. Korrespondentsanalüüsi põhimõtted	21
3.2. Klasteranalüüsi põhimõtted	22
3.3. Sagedusandmete normaliseerimine	24
4. Tulemused	25
4.1. Pre- ja postpositsioonide kasutus	25
4.2. Erinevused lekseemide kasutuses	28
4.3. Komplementi käänded	32
4.4. Kaassõnade tähendused	38
4.5. Tulemuste kokkuvõte	43
4.6. Edasised uurimissuunad	45
Kokkuvõte	46
Kirjandus	48
Adpositions in Estonian Dialects. Summary	52
Lisa 1. Lekseemide normaliseeritud sagedusandmestik	55

Sissejuhatus

Käesolevas magistritöös vaatlen kaassõnade kasutust eesti murretes. Kaassõnu on eesti keeles küll palju uuritud, kuid põhjalikke murretel põhinevaid töid on vähe. Varasemalt on kaassõnade kasutusest murretes andnud deskriptiivse ülevaate Harald Ploom (1955). Kvantitatiivses valguses on kahe paiknemisvõimalusega kaassõnu uurinud Ruutma jt (2016) ning sama teemat on käsitlenud ka selle magistritöö autor oma bakalaureusetöös (Ruutma 2013). Mõningaid murdepäraseid jooni kaassõnade kasutuses on välja toonud ka Paula Palmeos (1982) ning tähelepanekuid kaassõnade kohta on esitatud sarjas „Eesti murded“ (Tanning 1961; Must 1995; Keem, Käsi 2002; Lonn, Niit 2002; Juhkam, Sepp 2000). Selle magistritöö laiem eesmärk ongi väheuuritud teemat avada ning selgitada välja, kuidas kaassõnu murretes kasutatakse – kas esile tõuseb murretevahelisi erinevusi ning kas teatud murded moodustavad kaassõnade kasutuse järgi sarnasemalt käituvaid rühmi. Keskendun kaassõnade lekseemidele, kaassõnafaasi morfoloogilisele poolele ja kaassõnade semantikale. Täpsemateks uurimisküsimusteks on:

- kas ees- ja tagasõnade kasutuses on murrete vahel erinevusi,
- millised kaassõnad on eri murretele iseloomulikud,
- kas ja missugune on seos murrete ning kaassõna laiendava noomeni käänete vahel,
- kas erinevate kaassõnade tähenduste kasutuses on murdeti erinevusi.

Eelduseks on, et adpositsioonide lekseemitasandil võiksid murded jaguneda traditsiooniliselt põhja- ning lõunaeesti murderühmaks. Lisaks on varasemalt täheldatud (nt Pajusalu 2009; Palmeos 1982), et läänepoolsetele murretele on iseloomulikum selliste kaassõnade kasutus, millel on olemas ka käändelised vasted, st et need murded võiksid eelistada analüütilist väljendusviisi sünteetilisele. Kaassõna laiendava noomeni käänete kasutuses võib eeldada murrete vahel varieeruvust, sest suulises keeles üldisemaltki võib käänete valikus esineda varieerumist ning käändekasutust võiks mõjutada murdetaust.

Kaassõnadega on võimalik väljendada hulgaliselt tähendusi, kuid eeldada võib, et kõige tavalisem on kohatähenduse märkimine.

Uurimisküsimustele vastamiseks viin läbi kvantitatiivsed analüüsid, mis põhinevad mahukal andmestikul. Töös analüüsitav andmestik on kogutud eesti murrete korpuse morfoloogiliselt märgendatud murdetekstidest ning kaasatud on kõik adpositsioonideks märgendatud sõned ja nende lähikontekst. Materjal on käsitsi üle kontrollitud ja nõnda sisaldab lõplik andmestik võimalikke märgendusvigu minimaalselt. Kvantitatiivse andmeanalüüsi eelduste tõttu on andmestikust välja jäetud kaassõnad, mida esines vähem kui 20 korda. Töö lõplik andmestik sisaldab 12 785 kaassõnade kasutuskorda.

Järgin magistritöös tänapäevase korpuspõhise dialektoloogia põhimõtteid ehk olen analüüsinud empiirilist materjali peamiselt kvantitatiivselt. Uurimisküsimustele vastamiseks olen viinud läbi neli analüüsi: Pearsoni χ^2 -test, lihtne ja mitmene korrespondentsanalüüs ning klasteranalüüs (täpsemad kirjeldused peatükis 3). Lisaks toetun materjali analüüsis geograafilistele sagedusandmetele. Analüüsi tulemusi illustreerin murdekorpuse lausenäidetega.

Magistriöö koosneb neljast peatükist, mis omakorda jagunevad alapeatükkideks. Esimeses peatükis tutvustan töö teoreetilist tausta. Kirjeldan eesti kaassõnade süsteemi üldisemalt ning annan ülevaate korpuspõhisest dialektoloogiast, millele olen töös tuginenud. Kaassõnade ülevaate puhul keskendun peamiselt kaassõnade väljendatud suhete, kaassõnu laiendavate noomenite käänete, käänete ning kaassõnade paralleelse kasutuse ning sõnaliigi määramise problemaatika kirjeldamisele.

Töö teises peatükis annan ülevaate töö analüüsis kasutatud andmestikust. Tutvustan andmestiku lähtematerjali, materjali kogumise kriteeriumeid ja andmestiku kodeerimise põhimõtteid. Kolmandas peatükis kirjaldan töö analüüsis rakendatud meetodeid: lihtsat ja mitmest korrespondentsanalüüsi, klasteranalüüsi ning kvantitatiivsete analüüside läbiviimiseks vajaliku andmete normaliseerimise põhimõtteid. Töö viimases, neljandas peatükis, esitan analüüside tulemused ning tähelepanekud.

1. Teoreetiline taust

Käesolevas peatükis annan ülevaate magistritöö teoreetilisest raamistikust. Esmalt kirjeldan kaassõnade süsteemi eesti keeles – määratlen, mis on kaassõna, esitan erinevad kaassõnade väljendatavad suhted ning kaassõnu laiendavate noomenite käänded, vaatlen kaassõnade ning käänete paralleelset kasutust ja esitan peamised kaassõnade ja teiste sõnaliikide eristamise probleemid. Lisaks annan ülevaate korpuspõhise dialektoloogia põhimõtetest, millele olen töös tuginenud.

1.1. Kaassõnad eesti keeles

Selle peatüki alapeatükkides kirjeldan täpsemalt kaassõnade süsteemi eesti keeles. Oma töös vaatlen kaassõnade kasutust murretes, kuid töö teoreetiliseks taustaks on siiski põhiliselt uurimused, mis tuginevad eesti kirjakeelele, sest murretel põhinevaid töid ei leidu palju. Üldiselt on kaassõnu uuritud eesti keeles üsna erinevatest lähtepunktidest, näiteks on kaassõnu kirjeldatud põhjalikult mitmetes grammatikakäsitlustes (nt EKG I 1995; EKG II 1993; Palmeos 1982). Lisaks on kaassõnade kujunemist vaadeldud nii läänemeresoome keeltes üldisemalt kui ka grammatiseerumisteooriast lähtudes (nt Grünthal 2003; Habicht 2000; Habicht, Penjam 2007; Jürine, Habicht 2013). Üksikuid kaassõnu on uuritud ka kognitiivsemantilise vaatenurgast (nt Veismann 2009; Klavan 2012). Ometi ei ole kaassõnadest palju murretel põhinevaid uurimusi. Kaassõnade kasutuse kohta murretes on tehtud mõningaid tähelepanekuid sarjas „Eesti murded“ (Tanning 1961; Must 1995; Keem, Käsi 2002; Lonn, Niit 2002; Juhkam, Sepp 2000) ning deskriptiivse ülevaate eri murretes kasutatavate kaassõnade lekseemidest on andnud oma diplomitöös Harald Ploom (1955). Kaassõnafraasi sõnajärje varieerumist murretes on vaadeldud täpsemalt Ruutma jt (2016) ning seda sama teemat käsitles käesoleva magistritöö autor ka oma bakalaureusetöös (Ruutma 2014).

1.1.1. Kaassõna mõiste

Kaassõnad ehk adpositsioonid on muutumatud sõnad, mis esinevad alati koos üksiku noomeni või noomenifraasiga (EKG I 1995: 33–34). Adpositsioonidel on lauses morfosüntaktilised ülesanded – süntaktiliselt väljendavad adpositsioonid koos laiendava

noomeniga lauseliikmete vahel erinevaid seoseid ning morfoloogiliselt kasutatakse adpositsioone sarnaselt käänetega käändefunktsioonide väljendamiseks (Palmeos 1982: 7). Käänetega võrreldes on kaassõnadel siiski mitmekesisemad tähendusvarjundid ning neid erinevaid tähendusi kirjeldan täpsemalt alapeatükis 1.1.2.

Nagu käänetelgi, puudub kaassõnadel iseseisev leksikaalne tähendus ning eraldiseisvalt ei saa kaassõna olla omaette lauseliige. Niisiis selgub kaassõna grammatiline tähendus ainult konkreetse kontekstis koos noomeniga. (EKG I 1995: 33–34; Palmeos 1982: 7) Koos kohustusliku laiendava noomeniga ehk komplemendiga moodustab kaassõna kaassõnafraasi (Karlsson 2002: 164). Kaassõnafraasi põhjaks ei ole mitte adpositsiooni komplement vaid adpositsioon ise, sest see määrab komplemendi kääne (EKG II 1993: 137). Adpositsioon võib seejuures oma komplemendile eelneda (*keset teed*) või järgneda (*tee peal*), olles seega vastavalt pre- või postpositsioon. Eesti keeles esineb ka mõlema paiknemisvõimalusega kaassõnu ehk ambipositsioone¹ (*mööda teed ~ teed mööda*). (EKG I 1995: 37)

Ajalooliselt on postpositsiooniline sõnajärg kujunenud varem kui prepositsiooniline ning eesti keeles, nagu ka teistes soome-ugri keeltes, on postpositsiooniline fraas ka tavalisem (Palmeos 1982: 69–72). Grünthal (2003: 36; 2005: 40) on selgitanud postpositsioonirohkust rektsioonireeglite kaudu: kaassõnad on algselt kujunenud substantiividest, mis on moodustanud mõne teise substantiiviga possessiivse konstruktsiooni ning sellise possessiivkonstruktsiooni genitiivne täiend peab soome-ugri keelte reegleid järgides eelnema oma põhisõnale. Tüüpiliselt ongi postpositsioonifraasid genitiivse ning prepositsioonifraasid partitiivse rektsiooniga (EKG I 1995: 37–38; Grünthal 2003: 84). Ambipositsioonide puhul on murdematerjali põhjal leitud, et nende pre- ja postpositsiooniline paiknemine ei sõltu ainult genitiivsest ja partitiivsest rektsioonist, vaid eelkõige konkreetsest ambipositsioonist endast. Kui ambipositsiooni komplemendi kääne varieerub, on eessõnad seotud partitiiviga ning tagasõnad genitiiviga. Lisaks mõjutab ambipositsiooni paiknemist sellega väljendatav suhe ning teatud kaassõnade puhul ka murdeala. (Ruutma 2014; Ruutma jt 2016)

¹ Terminit *ambipositsioon* olen kasutanud Claude Hagège'i (2010: 114) eeskujul.

Kaassõnad nõuavad lisaks genitiivile ja partitiivile oma komplemendilt teisigi käändeid ning EKG I osas (1995: 37–38) on nendeks märgitud näiteks nimetavat, seestütlevat ning kaasütlevat käänat. Grünthal (2003: 58–59) on esitanud oma monograafias „Eesti keele seletava sõnaraamatu“ järgi 183 adpositsiooni ning neist adpositsioonidest nõuavad mõnda muud käänat kui genitiivi või partitiivi vaid 18 adpositsiooni. Nende 18 adpositsiooni puhul on komplemendilt nõutav kääne nomatiiv (*paitsi, otsa, ringi*), elatiiv (*väljas, saadik, saati, hoolimata, tagapool*), allatiiv (*tänu, vaatamata*), adessiiv (*pool*), ablatiiv (*poolt*), terminatiiv (*kuni*), abessiiv (*ilma*) või komitatiiv (*koos, kõige, käsikäes, seltsis*) (Grünthal 2003: 58–59).

Lisaks sellele, et prepositsioonid nõuavad tavaliselt partitiivi ning postpositsioonid genitiivi, sõltub kaassõna nõutav kääne ka kaassõnaga väljendatavast tähendusest (EKG II 1993: 138–139). Erinevate tähendustega seostuvaid käändeid kirjeldan põhjalikumalt järgmises alapeatükis (1.1.2).

1.1.2. Kaassõnade tähendustest

Kaassõnadel on lauses üsna sarnased ülesanded nagu käänatelgi ning tihtipeale langevadki kaassõnade ja käänete väljendatud suhted kokku. Kaassõnad on tähenduselt siiski tihtipeale spetsiifilisemad ning väljendavad käänetega võrreldes mitmekesisemaid tähendusnüansse. (EKG I 1995: 34–36; Palmeos 1982) EKG I osas (1995: 34–36) on esitatud 16 erinevat kaassõnade väljendatud suhet ehk tähendust, mille järgi jagunevad kaassõnad järgmiselt:

1. **kohakaassõnad**, mis jagunevad siht-, asu- ja lähtekoha kaassõnadeks (*pall veeres laua alla, pall on laua all, võta pall laua alt üles*). Lisaks on veel orienteerimata kohakaassõnu, mis väljendavad mingil pinnal liikumist või millegi katmist ilma kindla sihtkohata (*lennuk lendas üle puude*), ning prolatiivseid kohakaassõnu, mis väljendavad midagi mööda liikumist (*liikusime piki jõekallast*);
2. **ajakaassõnad** (*ta saabus mõne päeva eest*);
3. **põhjускаassõnad** (*teen sulle sünnipäeva puhul välja*);
4. **viisi- ja seisundikaassõnad** (*tiidruk langes teiste põlguse alla*);

5. **vahendit või abinõu väljendavad kaassõnad** (*kelle abil sa selle töö lõpuks valmis said*);
6. **suhtumist või suhet väljendavad kaassõnad** (*ta on oma ea kohta arukas*);
7. **hulka, mõõtu ja määra väljendavad kaassõnad** (*seal oli üle viieteistkümne inimese*);
8. **ligikaudsust väljendavad kaassõnad** (*õunu oli viie kilo ümber*);
9. **kaasasolu väljendavad kaassõnad** (*Jüri läks teistega ühes*);
10. **ilmaolu väljendavad kaassõnad** (*ilma tööta läheb ta kohe närviliseks*);
11. **asendust väljendavad kaassõnad** (*sa pead kellegi enda asemel saatma*);
12. **piiri väljendavad kaassõnad** (*läksime kuni suure teeni koos*);
13. **eksklusiivsust väljendavad kaassõnad** (*peale Jüri ei tea sellest otsusest veel keegi*);
14. **tegijat väljendavad kaassõnad** (*tuule poolt murtud puud jäidki seisma*);
15. **valdajat väljendavad kaassõnad** (*kelle käes on võti*);
16. **objekti väljendavad kaassõnad** (*mees lõi kirvega jala pihta*).

Kaassõnade algseks väljendatud tähenduseks peetakse ruumisuhteid. Noomenid, millest kaassõnad on välja kujunenud, ongi tähistanud koos teise noomeniga mõnda ruumilist suhet: näiteks fraasis *maja ees* on *ees* olnud kohakäändega substantiiv **edes-nä* ja *maja* seda laiendav genitiivne täiend. Fraas **majan edesnä* on tähendanud 'maja ees oleval alal'. Põhisõna **edesnä* tähendus aga tuhmus ning see muutus postpositsiooniks, mis ei oma ilma laiendava komplementita tähendust. Sellele, et kaassõnad on algselt tähistanud ruumisuhteid, viitab ka ruumisuhteid väljendavate kaassõnade rohkus. (Palmeos 1982: 69–70) Adpositsioonide grammatiseerumisprotsessis, mida kirjeldan täpsemalt alapeatükis 1.1.4, ongi tõenäoliselt esimeseks kaassõnade kujunemise etapiks kohatähenduse tähistamine ning aja- ja muud abstraktsemad suhted on lisandunud alles hiljem (Habicht 2000: 21; Palmeos 1982: 69–72). Samuti on kognitiivsemantilises valguses leitud, et ka keelekasutaja peab kaassõna tähendusvõrgustiku keskmeks ruumisuhteid, mis teeb sellest tähendusest primaarse, ning kaassõnade areng on olnud konkreetsemast abstraktsema poole (Veismann 2009).

Nagu selgus alapeatükis 1.1.1, nõuavad kaassõnad üldiselt prepositsioonina partitiivi ning postpositsioonina genitiivi ning lisaks sellele võib adpositsiooni komplement esineda muudiski kääntes. Samas seostuvad kaassõnade komplementide käänded ka kaassõnade tähendusega ning komplementi kääne võib olla mõjutatud kaassõna tähendusest (EKG I 1995: 37). Postpositsioone, mille komplement on genitiivis, leidub siinses peatükis esitatud kõikides tähendusrühmades (EKG II 1993: 138–139). EKG II osa (1993: 138–139) põhjal nõuavad postpositsioonid oma komplementilt nominatiivi või partitiivi vaid koha- ja ajasuhteid väljendades. Lisaks võivad postpositsioonide komplementid olla elatiivis koha- ning ajatähendust väljendades ning komitatiivis kaasasolu väljendades.

Prepositsioone laiendav noomen on aga koha- ja ajasuhteid tähistades tavaliselt partitiivis. Samas võivad prepositsioonid koha- ning määrakaassõnadena nõuda komplementilt ka genitiivi. Lisaks esinevad prepositsioonide komplementid põhjускаassõnadena allatiivis, koha- ja ajakaassõnadena terminatiivis, ilmaolu väljendavana abessiivis ning kaasasolu väljendavana komitatiivis. (EKG II 1993: 138–139)

Murretes on kaassõna komplementi käänete kasutus mõnel juhul kirjakeelest erinev. Näiteks on kirderanniku murretes aja- ja kohasuhteid väljendades prepositsioonide *allpool, enne, kesk, mööda, piki, pealt, pärast* ja *vastu* komplement genitiivis (Pajusalu jt 2009: 113) ning kui täpsustatakse mõnda ajalõiku, on tavaline adessiivi kasutus (*mennel aastal, enne vanal aiga*) (Must 1995: 38).

1.1.3. Lauseliikmete vaheliste suhete väljendamine analüütiliselt ja sünteetiliselt

Kuigi kaassõnad annavad sageli käänetega võrreldes edasi täpsemaid ja mitmekesisemaid tähendusvarjundeid, võivad adpositsioonid olla mõnel juhul ka käänetega asendatavad. Analüütiline, kaassõnaga väljendatud vorm ei ole tingimata sünteetilisest, käändega väljendatud vormist tähenduslikult erinev. Näiteks on analüütiline *laua pealt* võrdne sünteetilise variandiga *laualt*. (EKG II 1993: 137, Palmeos 1982: 17)

Tänapäeva eesti kirjakeeles paralleelselt kasutatavaid alalütlevat ning kaassõnaga *peal* konstruktsioone on oma doktoritöös uurinud Jane Klavan (2012). Korpusuuringute ja katsete tulemused viitasid, et mida komplekssem on fraas, seda enam peaks esinema

süntheetilist väljendust ehk kohasuhete markeerimist väliskohakäänetega. Ometi tõi Klavan (2012: 274) välja, et alati see siiski paika ei pea – põhjuseks võib olla, et kaassõnaga saab kohasuhet rõhutada enam kui käände abil. Analüütilist väljendusviisi kasutataksegi tihtipeale selguse mõttes. Näiteks võib mõne substantiivi käändeline vorm langeda vormiliselt kokku kaassõnaga: substantiiv *pea* on väliskohakäänetes vormiliselt kokkulangev kaassõnadega *peale*, *peal* ja *pealt*. Nõnda on selle substantiivi puhul kohasuheteid segaduse vältimiseks selgem väljendada analüütiliselt: *pea peal on suur juuksepuudar*. (Palmeos 1982: 18)

Üldjoontes ilmnes Klavani (2012) väitekirjas kirjakeele korpusandmete põhjal, et süntheetilisi vorme kasutatakse mõnevõrra rohkem kui analüütilisi, kuid katsete põhjal jällegi eelistati käänetele analüütilist väljendust. Ka vanas eesti kirjakeeles näiteks on autorid eelistanud üldiselt rohkem analüütilisust (Palmeos 1982: 16). Seda seostatakse indoeuroopa keelte mõjuga: vana kirjakeele tekstid on otsetõlked saksa keelest, mis on analüütiline keel, ning saksa keele prepositsioonid on lihtsam tõlkida adpositsioonideks kui käändelõppudeks (Habicht 2000: 23). Üldiselt on soome-ugri keeled arenenud küll suurema süntheetilise suunas, kuid eesti keel, nagu ka indoeuroopa keeled, näib arenevat just analüütilisemaks (Rätsep 1977: 31–32).

Mõnedel murdealadel on täheldatud analüütiliste väljendusvõimaluste kasutamise eelistamist selgemalt. Näiteks on Lääne-Eestis saarte ning läänemurdes eelistatud sise- ja väliskohakäänete asemel samu funktsioone täitvaid kaassõnu. (Pajusalu jt 2009: 113) Analüütilisus on seejuures eriti tavaline kohasuhete väljendamisel: näiteks illatiivi puhul (GEN+*sisse/peale*) ning inessiivi puhul (GEN+*sees/peal*) (Juhkam, Sepp 2000: 29; Lonn, Niit 2002: 49). Ka Võru murdes on toodud esile kohakäänete asemel analüütilise väljendusviisi eelistamist (Keem, Käsi 2002: 41). Kirderanniku murrete puhul on aga märgatud, et mitmete suhete väljendamisel on tavalisem just süntheetiline väljendus. Näiteks kasutatakse kohasuheteid tähistades palju sisekohakäänedeid (*viede kalu mantri*) ja kellegi juures viibimise puhul on tavaline väliskohakäänete kasutus (*lähän arstile*). (Must 1995: 38)

Lääne- ning lõunaestis murrete analüütilisuse eelistust on täheldatud ka kaassõna *peal* kasutuse puhul (Klavan jt 2015). Nendes piirkondades, eriti aga lääne-, Setu, Tartu ja Võru murdealadel, on kohatähenduses eelistatud analüütilist *peal* konstruktsiooni

süntheetilisele adessiivi vormile. Ranna- ja kirdemurdes on võrreldes teiste murretega aga kaassõna *peal* asemel vastupidiselt lääne- ning lõunaeesti murretele just adessiivi kasutus tavalisem. Tendentsi on selgitatud keelekontaktidega – ranna- ja kirdemurdes on olnud palju kontakte soome keelega, kus on sünteetilisus harilikum, ning läänepoolsetes ja lõunaeesti murretes on olnud tihedamad kontaktid germaani või balti keeltega, kus on jällegi analüütilisus tavaline. (Klavan jt 2015)

Samas on murretes toodud esile mõne analüütilise vormi rohkust ka näiteks sel juhul, kui kaassõnale vastavat käänat murdes lihtsalt ei esine. Näiteks kasutatakse Mulgi murdes essiivi väga harva ning seetõttu on selle käände funktsioonid asendunud analüütiliste vastetega nagu *tervest peräst, toorest taol* ('toorena, toorelt'). Lisaks on Mulgi murdes tavaline kasutada pronoomenitega seoses inessiivi ning ablatiivi analüütilisi vasteid. (Tanning 1961: 44–45)

1.1.4. Piirid kaassõnade ja teiste sõnaliikide vahel

Adpositsioonide ajalooline seotus substantiividega tekitab kaassõnade määramisel probleemi – tihtipeale on substantiivi ja adpositsiooni sõnaliigi piirid hägusad. Lisaks on mõned kaassõnad arenenud verbivormidest (*hoolimata, vaatamata*). Nimelt on kaassõnad tekkinud grammatiseerudes – algsed täistähenduslikud sõnavormid on kaotanud oma leksikaalse sisu ning on hakanud väljendama grammatilisi abstraktseid suhteid. (EKG I 1995: 38–39)

Lehmann (1985: 303–318) esitab oma grammatiseerumismudelil järgmised adpositsioonide arenemisjärgud:

- (substantiiv)
 - suhet väljendav substantiiv
 - sekundaarne adpositsioon
 - primaarne adpositsioon
 - aglutinatiivne afiks
 - fusiivne afiks.

Sekundaarsed adpositsioonid on alles kujunemisjärgus ning nende seos lähtevormiga on läbipaistev. Primaarsed kaassõnad on juba juurdunud ning nende seos substantiiviga

ei pruugi olla enam nähtav. (Hopper, Traugott 1994: 107–108) Näiteks on sellise arenguahela teel tekkinud kaasäitlev kääne. Algset substantiivi **kansa* hakati kasutama käändelises vormis koos teatud substantiiviga lauseliikmete vaheliste suhete väljendamiseks. Sõna muutus oma tähenduselt üha läbipaistmatumaks ning hakkas täitma grammatilist ülesannet. Koos genitiivis komplemendiga muutuski sõna postpositsiooniks *käden kansak*. See primaarne kaassõna kujunes mitmete häälikumuutuste tagajärjel hiljem tänapäeval kasutatavaks komitatiivi käändelõpuks *–Ga*. (Rätsep 1979: 77–78) Niisiis on kaassõnade süsteem pidevas muutumises ning piiri sõnaliikide ning ka grammatiseerumisahela eri etappide vahele on keeruline tõmmata.

Rudolf Karelson (1972) on käsitlenud kaassõnade sõnaliigi jaotusaluste problemaatikat põhjalikumalt ning on esitanud mõningad kriteeriumid, mis aitavad eristada kaassõnu teistest sõnaliikidest. Osalt kattuvad need kriteeriumid EKG I osas (1995) esitatutega. Esiteks on Karelson (1972) toonud EKG-ga (1995) sarnaselt välja, et kaassõna nõuab komplemendilt teatud rektsiooni – kaassõna mõjutab alati mingil moel oma laiendavat noomenit ning ei saa esineda iseseisvalt. Noomeni vormi sõltumine mõnest teisest moodustajast ei viita aga alati sellele, et viimase näol oleks tegemist adpositsiooniga. Näiteks on substantiiv rektsiooniseoses nii fraasis *mindi mööda teekallast* kui ka *ta püüdis meist mööda minna*. Viimase puhul on tegemist siiski afiksaaladverbiga – sõna *mööda* seostub siin süntaktiliselt verbiga, mitte noomeniga. Seega peab sõna adpositsiooniks määrates veenduma, et seos oleks just nimelt noomeniga. (Karelson 1972: 74–75) Eesti keeles kattuvadki paljude kaassõnade vormid nii iseseisvate adverbide kui ka afiksaaladverbidega ning kaassõnalist ja määrsõnalist kasutust saab eristada vaid kaassõna funktsiooni ja ümbritseva konteksti abil (Habicht, Penjam 2007: 52–54).

Lisaks on kaassõnal võrreldes substantiiviga, adverbiga ning verbiga erinev sisu (Karelson 1975: 74). EKG-stki (1995) selgub, et kaassõna tähendus on peamiselt grammatiline ning kaassõna väljendab mõnda seost komplemendiga. Adpositsiooni eristab sageli substantiivi käändevormist ka mingil moel muutunud vorm – näiteks on sõna *küüs* mitmuse seesütlev vorm *küüntes*. Kaassõnastunud vormi puhul tarvitatakse aga *i*-tüvelist mitmust *küüsis*. Samuti ei tohiks kaassõna ning komplemendi vahel olla täiendeid (kui tegemist pole prepositsiooniga). (Karelson 1972: 71–77) Seega on näiteks

kaasaütlevat käänet nõudvad kaassõnad Karelsoni (1972: 77) järgi piiripealsed juhtumid, sest näiteks lauses *laps läks emaga kaasa* võiks lisada vahele ka täiendeid *laps läks emaga linna kaasa*. Sellist juhtumit võiks ehk pidada kaassõnale lähenevaks adverbiks – sõna on veel lauses üsna iseseisev, kuid kindlasti väljendab ta grammatilisi suhteid lauseliikmete vahel (Karelson 1972: 77). Kaassõnade tüüpilise kujunemisprotsessi järgi omandabki kaassõna esmalt muutumatu vormi, siirdub avatumast klassist suletud klassi, kinnistub süntaktiliselt ning ei saa viimaks esineda iseseisvalt (Jürine, Habicht 2013: 740).

Karelson (1972: 79) on toonud välja ka asjaolu, et grammatikates lähtutakse tihtipeale sõnaliigi määramisel ortograafiast. Näiteks liigitatakse *alussärgi väel* kaassõnafrasiks, aga *särgiväel* liitadverbiks. Ortograafia alusel, mis on paljuski kokkuleppeline, pole aga sõnaliike võimalik määrata suulises kõnes ning siinseski suulisel murdekõnel põhinevas töös kerkib esile sarnane probleem.

Kokkuvõtvalt on Karelson (1972) nentunud, et sõnaliigi määramisel kaassõnaks peab eelkõige lähtuma konkreetsest lausest ning arvestama laiema kontekstiga. Siinse töö seisukohalt on kaassõnade eristamine teisest sõnaliikidest oluline materjali valikus – murdekorpuse märgendajad on küll samadest põhimõtetest lähtunud, kuid korpuses on ebahütlusi ning seega olen Karelsoni (1972) ja EKG (1995) põhimõtete järgi materjali üle kontrollinud.

1.2. Korpuspõhine dialektoloogia

Murdeuurimuste lähtematerjaliks on olnud traditsiooniliselt murdeatlased, mis toetuvad peamiselt n-ö esile kutsutud keelematerjalile – informantidele on ette antud küsitluskavad ning murdeainestiku kogujad ja atlase koostajad on saanud materjali kohandanud. Korpuspõhises dialektoloogias kasutatavad korpused sisaldavad seevastu muutmata keelematerjali, spontaanset kõnet. See pakub keelekasutuse kohta otsest ja tegelikku materjali ning annab keelenähtusest objektiivsema ja usaldusväärsema pildi. (Szmrecsanyi, Wolk 2011: 564–565) Szmrecsanyi (2014: 3–4) järgi ongi atlasepõhise ja korpuspõhise dialektoloogia suurimaks erinevuseks just see, et traditsioonilises dialektoloogias ei kajastu loomulik keelekasutus kõige paremini. Esitatakse vaid murdele tüüpilised juhtumid – kas mingi nähtus esineb või mitte. Kuna korpusmaterjal sisaldab aga muutmata loomulikku keelekasutust, võimaldab see vaadelda ka uuritava

keelenähtuse paralleelvorme ning kaasata analüüsi nähtuse sageduse. Niisiis peegeldub korpusmaterjalis ka nähtuste varieerumine ning esinemus. Korpuspõhist dialektoloogiat võibki nimetada ka sageduspõhiseks. (Szmrecsanyi 2014)

Seminaritöös olen lähtunud korpuspõhise dialektoloogia põhimõtetest. Korpuspõhine murdeuurimus peaks Szmrecsanyi ja Wolki (2011) järgi vastama järgmistele põhimõtetele:

- Uurima peaks mitmeid keelelisi nähtuseid korraga, sest see vähendab üksikute irrelevantsete tegurite mõju analüüsile.
- Analüüsis võiks kasutada erinevaid arvutuslikke mitmemõõtmelise statistika meetodeid.
- Uurimistulemusi peaks visualiseerima erinevaid tehnilisi võimalusi rakendades.

Siinses töös uurin korraga ainult üht keelelist nähtust, kaassõnu, ja teisi keelenähtusi oma analüüsi ei kaasa. Muus osas toetub töö korpuspõhise dialektoloogia põhimõtetele ning töö analüüs põhineb peamiselt korrespondentsanalüüsil, mida kirjeldan peatükis 3.1.

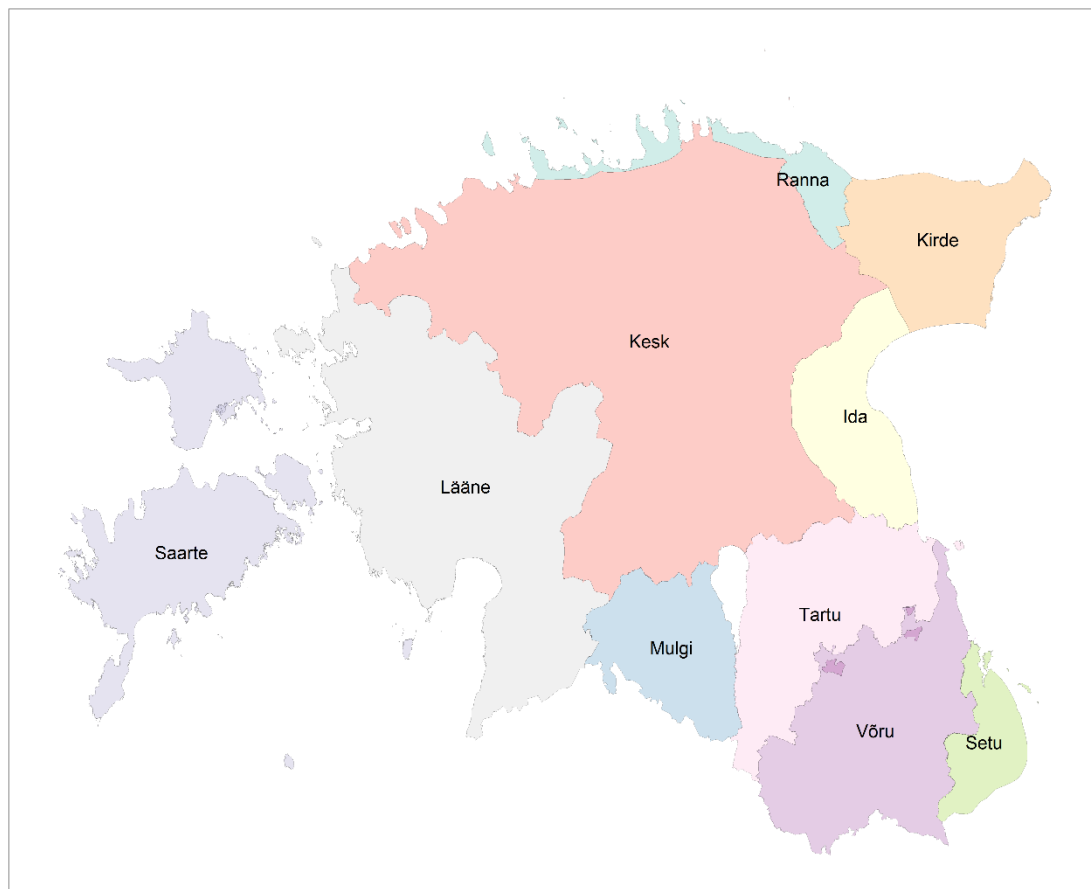
2. Andmestiku kirjeldus

Selles peatükis annan ülevaate töö analüüsis kasutatud andmestikust. Kirjeldan andmestiku lähtematerjali ning toon välja, milliste kriteeriumite alusel on uurimismaterjal kogutud. Seejärel esitan andmete kodeerimispõhimõtted ja -skeemi.

2.1. Lähtematerjal

Töö uurimismaterjal on pärit eesti murrete korpusest. Eesti murrete korpus on elektrooniline andmebaas, mis sisaldab kõigi Eesti murdealade keelematerjali: murdelindistusi ning murdetekste nii foneetilises ja lihtsustatud transkriptsioonis kui ka morfoloogilise märgenduse ning süntaktilise analüüsiga. Lisaks on korpuses andmed keelejuhtide, lindistuste ja litereeringute kohta. Murdekorpuse aluseks olevad lindistused on tehtud põhiliselt 1960.–1970. aastatel. (EMK) Keelejuhtideks on peamiselt 70–90-aastased ning enamuses on naiskeelejuhid. Salvestiste temaatika on üpris sarnane ja jututeemad puudutavad igapäevaelu ehk töid, kombeid, pere jne. Ka suhtlussituatsioon on murdelindistustel samasugune – lindistused on pigem institutsionaalsed, sest intervjuerijad ja keelejuhid ei ole varasemalt tuttavad. See tähendab, et murdekorpuse tekstid on omavaheliseks võrdluseks piisavalt ühetaolised. (Uiboed 2013: 49–52)

Eesti murrete korpuses eristatakse kümme murdeala: põhjaeesti murderühma moodustavad ida-, kesk-, lääne- ja saarte murre, lõunaeesti murderühma kuuluvad Mulgi, Võru, Tartu ja Setu murre ning kirderanniku murderühm jaguneb kirde- ja rannamurdeks. (EMK) Setu murret ei ole varasemates traditsioonilistes murdejaotustes eraldi murdeks peetud, vaid seda on vaadeldud kui üht Võru murrakurühma (nt Pajusalu jt 2009: 184). Murdekorpuse lühitutvustuses (EMK) on aga välja toodud, et lisaks kultuurilisele ja ajaloolisele eripärale on Setu murdeala osutunud Võru murdest eristuvaks ka mitmetes murdesüntaksit vaatlevas uurimustes (nt Uiboed 2013; Lindström jt 2015). Seetõttu võib korpuses kasutatavat jaotust pidada põhjendatuks. Joonisel 1 on kaardil esitatud murdekorpuse murdejaotus.



Joonis 1. Eesti murdealad.²

Eesti murrete korpust on koostatud Tartu Ülikooli eesti ja üldkeeleteaduse instituudi ja Eesti Keele Instituudi koostöös alates 1998. aastast ning seda täiendatakse seni. (EMK) Tabelis 1 on esitatud murdekorpuse morfoloogiliselt märgendatud sõnade arv igas murdes käesoleva töö uurimismaterjali kogumise hetkel mais 2015. Korpuses pole kõik murded võrdselt esindatud ning enim tekste on saarte, lääne- ja keskmurdest. Setu, ida- ja kirdemurde tekste on korpuses kõige vähem.

² Kaart Uiboaed 2016.

Tabel 1. Murdekorpuse morfoloogiliselt märgendatud sõnade arv murdeti

	<u>Sõnu</u>
Idamurre	45280
Keskmurre	130086
Kirdemurre	47660
Läänemurre	154400
Mulgi murre	63516
Rannamurre	51667
Saarte murre	166898
Setu murre	39175
Tartu murre	65591
Võru murre	<u>70038</u>
Kokku	<u>834311</u>

Käesoleva töö andmestikku on kaasatud murdekorpuse morfoloogiliselt märgendatud tekstidest kõik lausungid, mis sisaldavad mõnda korpuses adpositsiooniks märgendatud sõne. Kuna murdekorpuse tekstid on käsitsi märgendatud, esineb seal märgendusvigu. Nagu peatükis 1.1.4 mainitud, võib eesti keeles sama sõna esineda nii kaassõna kui ka adverbina – sõnaliik sõltub laiemast kontekstist. Adpositsiooni ja adverbi piir pole aga alati kergesti tuvastatav ning nõnda on murdekorpuses märgendatud adverbe ekslikult kaassõnadeks ning vastupidi. Sarnast probleemi esineb ka adpositsioonide ja substantiivide eristamisel. Kattuvate vormide ning hägusate sõnaliigipiiride tõttu võib seega olla murdekorpuses kaassõnadeks märgendatud adverbe või afiksaaladverbe, üksikuid kaassõnu on omakorda märgendatud (afiksaal)adverbideks ning substantiivi ja adpositsiooni piirimail olevaid sõnesid on kord peetud substantiivideks ning teinekord adpositsioonideks.

Murdekorpuse märgenduse ebahühtlusest tuleneb ka siinse töö kitsaskoht: välja on jäänud need kaassõnad, mis on märgendatud murdekorpuses vääralt või ebajärjepidevalt. Tuginen niisiis vaid murdekorpuses kaassõnadeks märgendatud materjalile. Lausungid, mis sisaldasid kaassõnadeks märgendatud sõnesid, olen üle kontrollinud ning välja jätnud juhtumid, kus kaassõna asemel on tegemist iseseisva adverbi või verbiga seotud afiksaaladverbiga. Materjalist jätsin välja ka juhtumid, mille puhul polnud lähikonteksti põhjal võimalik määrata, kas tegemist on adverbi või kaassõnaga.

Kontrollitud materjal on kokku 215 erinevat kaassõna. Siinses töös keskendun vaid sagedamini esinenud vormidele ning olen tööst välja jätnud kõik murdekorpuses vähem

kui 20 korda esinevad kaassõnad, kuna harvaesinevate vaatluste kaasamine kvantitatiivsesse analüüsi oleks andmeanalüüsi seisukohalt problemaatiline. Selliseid kaassõnu, mis esinevad murdekorpuses rohkem kui 20 korda, on kokku 68.

2.2. Kodeerimisskeem

Kokku on töö andmestikus 12 785 mõnda adpositsiooni sisaldavat lausungit. Iga lausung on käsitsi üle kontrollitud, tuginedes peatükis 1.1.4 esitatud kaassõnade piiritlese põhimõtetele, ning lausungitele on andmestikus juurde lisatud teave komplemendi käände ja adpositsiooni väljendatava suhte kohta. Järgnevalt tutvustangi neid andmestikus kodeeritud tegureid täpsemalt ning esitan kodeerimise põhimõtted.

Murre. Tähistab murdeala, kust analüüsitav lausung pärit on. Töös kasutan murdekorpuse jaotust, kus eristatakse 10 murdeala: idamurre (ida), keskmurre (kes), kirdemurre (kir), läänemurre (lää), Mulgi murre (mul), rannamurre (ran), saarte murre (saa), Setu murre (set), Tartu murre (tar) ja Võru murre (võr).

Adpositsiooni lekseem. Võib arvata, et erinevates murretes eelistatakse erinevate kaassõnade kasutamist, sest paljud kaassõnad on sünonüümsed. Kuna kaassõnade funktsioonid sarnanevad paljuski käänete omadega, võib lisaks oletada, et kui mõnes murdes esineb mõnda muidu sagedast kaassõna teiste murretega võrreldes palju vähem, on võimalik, et selle kaassõna funktsiooni täidab kääne. Peatükis 1.1.3 selguski varasemate uurimuste põhjal, et läänepoolsemates ning lõunaesti murretes peaks analüütiline väljendus olema sagedasem kui sünteetiline. Seetõttu ongi töösse kaasatud lemmavormid, et selgitada võimalikke erinevusi üksikute kaassõnade kasutamises erinevates murretes. Andmestikus on kokku 68 erinevat kaassõna mis, esinesid korpuses rohkem kui 20 korda.

Kaassõnafraasi sõnajärg. Nagu tuli esile ka siinse töö esimeses peatükis, on soome-ugri keeltele tavaline postpositsioonide rohkus. Sõnajärg võib samas olla seotud ka teatud kaassõna väljendatud tähendustega ning samuti võivad olla seotud ka komplemendilt nõutav kääne ning pre- ja postpositsioonilisus. Vaatan töös, kas pre- ja postpositsioonide valikus on murrete vahel (ka statistiliselt olulisi) erinevusi ning kaassõnafraasi sõnajärge uurin ka korrespondentsanalüüsi abil.

Kääne. Peatükis 1.1.1 selgus, et erinevad adpositsioonid võivad nõuda oma komplemendilt erinevaid käändeid. Lisaks võib üks ja sama adpositsioon erinevat tähendust kandes esineda eri käändes oleva komplemendiga. Murdematerjali puhul võib eeldada komplemendi käändekasutuses palju varieerumist, seda nii kirjakeelega võrreldes kui ka murdeid kõrvutades. Näiteks ilmes ka siinse töö peatükis 1.1.2, et teatud kaassõnad nõuavad mõnes murdes kirjakeelega võrreldes oma komplemendilt erinevat käänet. Kodeeritud tunnuse eesmärk ongi leida, kas murrete ning komplementide käänete vahel on seos. Siinses materjalis esinevad kaassõnade komplemendid nimetavas (nom), omastavas (gen), osastavas (par), seestütlevas (ela), alaleütlevas (all), alalütlevas (ade), alaltütlevas (abl), ilmaütlevas (abe) ning kaasaütlevas (kom) käändes.

Adpositsiooni väljendatav tähendus. Erinevaid tähendusi, mida kaassõnad väljendavad, olen määranud EKG I osa (1995: 34–36) eeskujul. EKGs (1995: 34–36) on kokku esitatud 16 eri kaassõnade tähendusrühma (vt täpsemalt ptk 1.1.2). Siinses materjalis esines eri tähendusi 14, materjalis ei olnud ainsana EKG (1995: 36) esitatud piiri väljendavaid kaassõnu ja objekti väljendavaid kaassõnu. Tunnust kodeerides pidasin silmas, kas tähendusrühmi võiks olla lisaks EKGs esitatule rohkemgi, kuid kõik kaassõnad sobitusid siiski EKG jaotusesse. Nõnda on selles töös jagatud kaassõnad järgmistesse rühmadesse:

- kohakaassõnad ehk ruumikaassõnad (lok),
- ajakaassõnad (tem),
- põhjuskaassõnad (kau),
- viisi- ja seisundikaassõnad (vii),
- vahendit või abinõu väljendavad kaassõnad (vah),
- suhtumist või suhet väljendavad kaassõnad (suh),
- hulka, mõõtu ja määra väljendavad kaassõnad (kva),
- ligikaudsust väljendavad kaassõnad (lig),
- kaasasolu väljendavad kaassõnad (kaa),
- ilmaolu väljendavad kaassõnad (ilm),
- asendust väljendavad kaassõnad (ase),
- eksklusiivsust väljendavad kaassõnad (eks),

- tegijat väljendavad kaassõnad (teg),
- valdajat väljendavad kaassõnad (val).

Tunnuse abil selgitan, kas kaassõnade erinevate tähenduste kasutuses on murdeti erinevusi.

3. Meetod

Tuginen oma magistritöös tänapäevase murdesüntaksi uurimise põhimõtetele, mida olen täpsemalt tutvustanud peatükis 1.2. Toetun niisiis töös murdekorpuse materjalile ning empiirilise matejali analüüsis kasutan mitmemõõtmelise statistika meetodeid. Töö on seega olemuselt peamiselt kvantitatiivne. Uurimistulemuste visualiseerimiseks olen kasutanud erinevaid tehnilisi võimalusi. Andmestiku analüüsis rakendan tavalist ning mitmest korrespondentsanalüüsi ja klasteranalüüsi, mille viin läbi vabavaralise statistikaprogrammiga R (R Core Team 2013). Lihtsa korrespondentsanalüüsi tegemiseks kasutasin paketti *ca* (Nenadic, Greenacre 2007) ja mitmese korrespondentsanalüüsi läbiviimiseks paketti *FactorMineR* (Husson jt 2016). Klasteranalüüsid viisin läbi paketiga *pvclust* (Suzuki, Shimodaira 2015). Järgnevalt tutvustan töös rakendatud meetodeid lähemalt.

3.1. Korrespondentsanalüüsi põhimõtted

Korrespondentsanalüüs (edaspidi viitan terminile lühendiga CA ingliskeelse nimetuse *Correspondence Analysis* järgi) võimaldab tuvastada mahukatest sagedusandmestikest seaduspärasid ja kategooriliste tunnuste vahelisi seoseid, mida lihtsalt vaatluse teel poleks võimalik leida. CA peamine ülesanne on andmeid visualiseerida ning analüüsi üks väljunditest ongi graafik. Analüüs põhineb tabelikujulisel sagedusandmestikul, mille ridades ning veergudes olevate tunnuste vahelisi seoseid võrreldakse. (Greenacre 2007; Lebart jt 1998; viidatud Uiboaed 2013: 66–67 kaudu; Uiboaed 2013: 67–71; Levshina 2015: 369–370) Samal põhimõttel toimivad nii lihtne kui ka mitmene korrespondentsanalüüs (edaspidi MCA ingliskeelse termini *Multiple Correspondence Analysis* järgi), mida mõlemat ka siinses töös kasutatan. Võrreldes CA-ga on MCA eeliseks võimalus vaadelda mitme tunnuse vahelisi seoseid korraga (Abdi, Valentin 2016). Samas ei võimalda MCA kasutada sisendina normaliseeritud sagedusandmeid ning analüüsi sisendiks on ilma sagedusandmeteta koondandmestik kodeeritud kujul.

Siinses magistritöös vaatlen murrete ning erinevate andmestikus kodeeritud tunnuste omavahelisi sageduspõhiseid seoseid. CA sisendiks on seega sagedustabel, kus ridades on ühe tunnuse eri tasemed ning tulpades teise omad. Siin töös rakendan CA-d lekseemide sageduste ja murrete seose tuvastamiseks, seega on CA sisendiks kaassõnade lemmade normaliseeritud sagedusandmed murrete kaupa. Nõnda paiknevad sagedustabeli lahtrites lemmade sagedused murretes. Töö mõlema MCA graafiku sisendiks on kolm tunnust. Mõlemasse MCA analüüsi on kaasatud tunnused murre ja adpositsioonifraasi sõnajärg ning lisaks sellele on kahes erinevas MCA analüüsis veel kas kaassõna tähendus või komplemendi kääne. Lihtsa CA andmed on normaliseeritud. Normaliseerimise põhimõtteid kirjeldan täpsemalt peatükis 3.3.

Sagedusandmetel põhineva korrespondentsanalüüsi graafik illustreerib tunnustevahelisi seoseid kahemõõtmeliselt. Sisestatud andmed (ehk uuritavad tunnused) on graafikus esindatud punktidenä, mis paiknevad graafiliselt kahemõõtmelises ruumis. Graafiku keskel on hallid horisontaalne ja vertikaalne joon, mis tähistavad mõlema mõõtme n-ö keskmist murret ehk olukorda, kui kõikides murretes oleks kaassõnade kasutus sarnane. Keskmine murre esitab seega uuritavate tunnuste sageduste keskmist väärtust. Mida kaugemal on tunnuse punkt joonest, seda rohkem erineb ta keskmisest või "tavalisest" selle mõõtme suhtes ja mida kaugemal punktid paiknevad teineteisest, seda erinevamad nad on. Kahe joone ristumiskohta nimetan töös keskpunktiks.

Kuna CA põhineb kauguste mõõtmisel andmestiku ridade ning veergude vahel, siis graafikul esitatavad üksustevahelised kaugused on tõlgendatavad sama tüüpi tunnuste omavaheliste sarnasuste ja erinevustena või eri tüüpi tunnuste tõmbumise ja tõukumisena. CA graafik illustreerib tulemusi siiski ligikaudselt ning seda tuleb tulemuste tõlgendamisel arvesse võtta (Lebart jt 1998, Greenacre 2007, viidatud Uiboaed 2013: 66–67 kaudu). Analüüsis kasutatava hii-ruut-kauguse abil on tulemus siiski nii täpne üldistus ridade ja veergude korrelatsioonist kui võimalik (Baayen 2008: 139–146).

3.2. Klasteranalüüsi põhimõtted

Korrespondentsanalüüsi tulemuste täiendamiseks kasutan töös klasteranalüüsi. Klasteranalüüsis on mingid uuritavad objektid (siin töös murded) jagatud nn klastripuus omaette üksusteks, st et alustatakse olukorrast, kus iga murre on omaette rühm. Seejärel

hakatakse neid uuritava tunnuse põhjal (siin töös lekseemide, käänete või tähenduste põhjal) liitma rühmadesse põhimõttel, et üksteisele kõige sarnasemad murded kuuluksid samasse rühma (Levshina 2015: 309–311). Nõnda tekivad klasteranalüüsis tunnuste põhjal murretest nii homogeenised rühmad kui võimalik. Rühmade liitmine toimub seni, kuni kõik murded on liidetud üheks rühmaks. Analüüsi tulemused esitatakse puulaadse graafikuna ehk dendogrammina ning moodustatud klastrid esitavad vähimate erinevustega objekte. Sama klatri objektide erinevust tähistab klatri kõrgus – mida kõrgemalt klaster on liidetud, seda erinevamad objektid on. (Baayen 2008: 148; Levshina 2015: 309)

Selles töös põhineb klasteranalüüs murretevaheliste eukleidiliste kauguste mõõtmisel, kus murded moodustavad rühmi vastavalt sellele, kui lähestikku paiknevad nad mitmemõõtmelises ruumis. Seejuures toimub suurtemateks rühmadeks liitmine Wardi algoritmiga. Algoritm otsitakse liitmiseks sellised lähestikku paiknevad objektid, mille liitmisel oleks klatri objektide omavahelise kauguse hajuvus minimaalne. Meetodi abil tekivad väiksemad klastrid, mis on üsna sama suurusega ning meetod sobib kasutamiseks erinevat tüüpi materjali puhul. (Levshina 2015: 306–311; Gries 2013a: 347–348)

Selle töö dendogrammides liidetakse klastriteks kolme erineva tunnuse (lekseemide, käänete ja tähenduste) põhjal sarnasemalt käituvaid murdeid. Klasteranalüüsi sisendiks on seejuures sarnaselt CA analüüsile tabelikujuline sagedusandmestik, mille veergudes on sedapuhku klasterdamise aluseks oleva tunnuse tasemed, ridades murded ning tabeli lahtrites normaliseeritud sagedusandmed murrete kaupa.

Dendogrammis on esitatud ka suurima silueti laiusega ehk analüüsitava andmestiku seisukohalt optimaalne klastite arv. See tähendab, et minimaalne arv klastreid, mis oleksid parimaks murrete erinevuste kirjeldamiseks vajalikud, on joonistel ümbritsetud roheline kastiga. Kuna murdeid on kümme, võib optimaalseks olla 2–9 klatri, sest kõik murded ühes klatri või omaette kümnes klatri ei oleks huvipakkuv. Igale võimalikule klatri arvu antakse silueti laiuse väärtus, ning see klatri arv, mille laiuse väärtus on suurim, on ka optimaalseks klatri arvaks. (Levshina 2015: 312)

3.3. Sagedusandmete normaliseerimine

Siinse töö lihtsas korrespondentsanalüüsis ja klasteranalüüsis on kasutatud sagedusandmed normaliseeritud. Märgendatud sõnade arv ei ole murdekorpuses ühtlane: näiteks nagu selgus tablis 1, on saarte murre esindatud 166 898 sõnaga ning Setu murre vaid 39 175 sõnaga. Seega ei ole ka kaassõnade absoluutsagedused võrreldavad, sest mida suurem on üksiku murde materjalimaht, seda rohkem esineb seal tõenäoliselt ka kaassõnu. Normaliseerimise abil aga muudetakse sagedusandmed omavahel võrreldavaks.

Magistritöös olen teinud normaliseerimise korpuse murdetekstide keskmise sõnade arvu põhjal, milleks on 83 431 sõna. Näiteks on Setu murdes kaassõna *man* absoluutsagedus 22 ning kokku on Setu murdetekstides 39 175 sõna. Adpositsiooni *man* normaliseeritud sagedus saadakse arvutusega $22 \times 83431/39175 = 47$. Kaassõna absoluutsagedus murdes korrutatakse seega korpuse keskmise sõnade arvuga ja jagatakse omakorda murde sõnade arvuga korpuses. See tähendab, et kui Setu murdetekstide sõnaarv oleks murdekorpuse keskmine 83 431, esineks kaassõna *man* Setu tekstides 47 korda. Töö kõik andmed on samal viisil normaliseeritud ning seega on eri tekstimahuga esindatud murrete sagedusandmed omavahel võrreldavad. Need normaliseeritud sagedusandmed on aluseks magistritöö lihtsatele korrespondentsanalüüsidele ja klasteranalüüsidele.

4. Tulemused

Selles peatükis esitan kvantitatiivsete analüüside tulemused. Analüüsid põhinevad peatükis 2 tutvustatud andmestikul ning andmestiku normaliseeritud sagedustabelitel. Analüüside eesmärk on tuvastada murretevahelisi erinevusi ja sarnasusi kaassõnade kasutuses. Selle selgitamiseks keskendun kaassõnade lekseemitasandile ning morfosüntaktilistele ja semantilistele teguritele. Vaatlen täpsemalt,

- kas ees- ja tagasõnade kasutuses on murrete vahel erinevusi,
- millised kaassõnad on eri murretele omasemad,
- kas ja missugune on seos murrete ning kaassõna komplemendi käänete vahel,
- kas erinevate kaassõnade tähenduste kasutuses on murdeti erinevusi.

Vastavalt uurimisküsimustele viin läbi neli põhilist analüüsi. Esimeses neist keskendun vaid pre- ja postpositsioonide ning murrete statistilise seose tuvastamisele Pearsoni χ^2 -testiga. Teises analüüsis vaatlen lihtsa korrespondentsanalüüsi ja klasteranalüüsi abil kaassõna lekseemide varieerumist murretes. Kolmas analüüs tuvastab komplemendi käänete ning murrete seosed mitmese korrespondentsanalüüsi, klasteranalüüsi ning ka geograafiliste sagedusandemete põhjal. Neljandas analüüsis vaatlen mitmese korrespondentsanalüüsi, klasteranalüüsi ning geograafiliste sagedusandmete abil, kuidas on murretes kasutatud erinevaid kaassõnade tähendusi. Rakendatavatest meetoditest oli täpsemalt juttu peatükis 3 ning sageduskaartide esitamiseks olen kasutanud selleks väljatöötatud R-i skripte (Uiboed 2016).

4.1. Pre- ja postpositsioonide kasutus

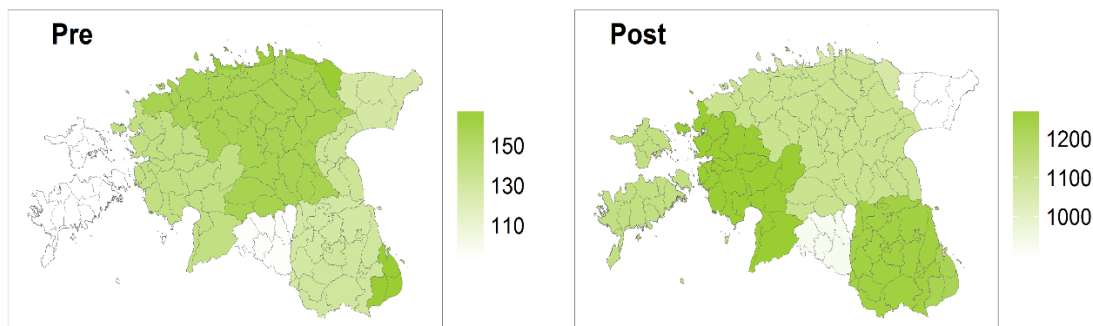
Postpositsioonifraasi peetakse eesti keeles ning ka teistes soome-ugri keeltes vanemaks kui prepositsioonifraasi (Palmeos 1982: 69– 72; Grünthal 2003: 36). Valdav osa adpositsioonidest ongi kirjakeeles postpositsioonid ning prepositsioone leidub vähe (Palmeos 1982: 4–5). Postpositsioonirohkus ilmneb ka murdematerjalis. Tabelis 2

esitatud normaliseeritud sagedusandmete põhjal on postpositsioone kasutatud prepositsioonidega võrreldes oluliselt rohkem. Sagedusandmetes on murrete vahel väiksemaid erinevusi, kuid postpositsiooniline sõnajärg on siiski kõikjal selgelt eelistatum.

Tabel 2. Pre- ja postpositsioonide normaliseeritud sagedused

	pre	post
IDA	133	1118
KESK	159	1113
KIRDE	128	895
LÄÄNE	141	1277
MULGI	95	926
RANNA	168	1071
SAARTE	93	1140
SETU	168	1227
TARTU	128	1250
VÕRU	131	1256

Joonisel 2 on kaassõnafraasi sõnajärje sagedusandmed kujutatud geograafiliselt ning nii jooniselt kui ka tabelist 2 ilmneb, et prepositsioone on kõige enam kasutatud kesk-, ranna- ja Setu murdes ning postpositsioone lääne-, Tartu, Võru ja Setu murdes. Kõige vähem on prepositsioone kasutatud saarte ja Mulgi murdes ning kõige vähem postpositsioone esineb Mulgi ja kirdemurdes. See, et mõnes murdes on kasutatud nii pre- kui ka postpositsioone võrreldes teiste murretega rohkem, võib viidata, et nendes murretes eelistatakse kaassõnu käänetele. Prepositsioonirohkus taas võib viidata indoeuroopa keelte mõjule. Seda, kas aga murre ning kaassõna paiknemine on ka tegelikult omavahel statistiliselt seotud, testin Pearsoni χ^2 -testiga.



Joonis 2. Pre- ja postpositsioonide jagunemine geograafiliselt

Testi tulemusena selgus, et murretevahelised erinevused kaassõna paiknemises on statistiliselt olulised ($p < 0,001$). Samas on Pearsoni jäägid (tabel 3) enamjaolt kõik nullilähedased, mis tähendab, et tunnuste seosed pole tugevad. Pearsoni jäägid, mis on väiksemad kui -3.841 või suuremad kui 3.841 , viitavad väga tugevale seosele (Gries 2013a: 369), ning siinsete andmete põhjal on tugevaim seos murde ning kaassõna positsiooni vahel ranna- ning saarte murdel, millest esimeses on prepositsioonide esinemise tõenäosus seose juhuslikkuse eeldusel oodatust suurem ning teises oodatust väiksem. Prepositsioone esineb oodatust mõnevõrra rohkem ka kesk-, kirde- ja Setu murdes ning vähem Mulgi, Tartu ja Võru murdes. Postpositsioone on lisaks saarte murdele kasutatud oodatust mõnevõrra rohkem ida-, lääne-, Mulgi, Tartu ja Võru murdes. Oodatust väiksem postpositsioonide kasutus on lisaks rannamurdele kesk-, kirde- ja Setu murdes.

Tabel 3. Pre- ja postpositsioonide normaliseeritud sagedusandmete Pearsoni jäägid

	Pre	Post
IDA	-0.051	0.018
KESK	2.022	-0.698
KIRDE	1.812	-0.626
LÄÄNE	-0.817	0.282
MULGI	-1.355	0.468
RANNA	3.137	-1.083
SAARTE	-3.348	1.156
SETU	1.614	-0.557
TARTU	-1.517	0.524
VÕRU	-1.370	0.473

Craméri V väärtus, mis näitab murrete ning kaassõnade omavahelise seose tugevust, on 0.058. See ei viita samuti tugevale statistilisele korrelatsioonile. Craméri V väärtus jääb alati 0 ja 1 vahele, ning mida lähemal on tulemus ühele, seda tugevam on korrelatsioon tunnuste vahel (Gries 2013a: 370–371). Seega viitavad tulemused, et seos murrete ning kaassõna paiknemise vahel on küll statistiliselt oluline, ent seos on nõrk ning selle põhjal oleks ennatlik teha järeldusi murretevaheliste erinevuste kohta.

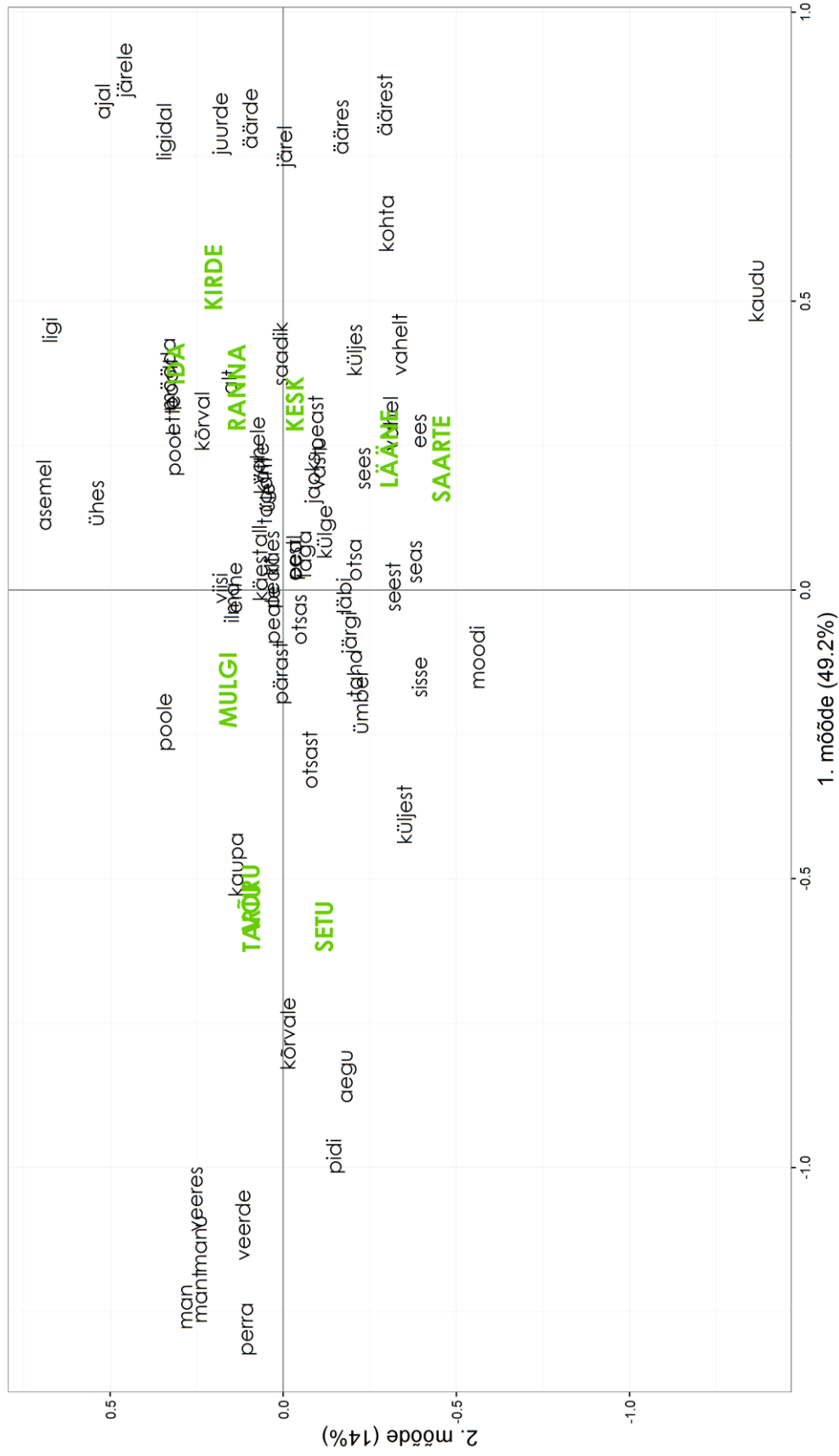
4.2. Erinevused lekseemide kasutuses

Joonis 3 illustreerib korrespondentsanalüüsi tulemusi kaassõnade lekseemide normaliseeritud sageduste põhjal. Analüüsis kasutatud lekseemide normaliseeritud sagedusandmestik on esitatud lisas 1. Graafiku esimene, horisontaalne mõõde kirjeldab 49% ning teine, vertikaalne mõõde 14% andmestiku variatiivsusest (kokku 63%). Joonisel 4 on esitatud sama sagedusandmestiku põhjal klasteranalüüs.

Joonisel 3 eristuvad horisontaalse mõõtme põhjal kaks suuremat kaassõnu sarnasemalt kasutatavat rühma: Mulgi, Tartu, Võru ja Setu murde rühm ehk lõunaeesti rühm ning ida-, kirde-, ranna-, kesk-, lääne- ja saarte murde rühm ehk põhjajaeesti rühm. Vertikaalse mõõtme põhjal näeb väiksemaid erinevusi. Lõunaeesti rühmas on Mulgi murre tõmbunud teistest kaugemale ja on seega võrreldes teiste rühma kuuluvate murretega põhjajaeesti rühmale kõige lähemal. Seejuures on huvitav, et teise mõõtme põhjal osutuvad kaassõnade kasutuse osas väga sarnasteks Tartu ja Võru murre ning Setu murre paikneb graafikul mõnevõrra neist murretest eemal. See viitab taas, et Setu murre erineb mõnevõrra Võru murdealast. Lekseemi tasandil oleks võinud eeldada just Võru ja Setu sarnasust.

CA analüüs kaassõnade lemmadest

a kaassõna lemma a murre



Joonis 3. Kaassõnade lemmade korrespondentsanalüüsi graafik

Vertikaalse mõõtme põhjal osutuvad põhjaeesti rühmas teistest murretes eristuvaks lääne- ning saarte murre, mis viitab, et lekseemitasandil kasutatakse nendes kahes murdes sagedamini sarnaseid kaassõnu.

Lekseemide puhul on paljud kaassõnad koondunud graafiku keskpunkti ning see viitab, et neid kaassõnu kasutatakse kõigis murretes üsna sarnaselt ehk nad vastavad n-ö keskmisele murdele. Ometi ilmneb teatud adpositsioonide puhul selge lekseemi ja murde seos. Kaassõnad *man*, *manu*, *mant*, *veeres*, *veerde*, *perra* ning ka *pidi*, *aegu*, *kõrvale* ja *kaupa* on graafiku põhjal iseloomulikumat lõunaeesti murretele (näited 1–10).

- (1) *ja syss kool'i+darō **mann** oli jälle* (TAR)
- (2) *`tõije `Vahtsõ+liina vai sinnäq liina **kõr'd's'i mano*** (VÕR)
- (3) *ul'l' (siija) hobõsõga kodu **kerikku mant*** (VÕR)
- (4) *ku ma **järve veeren** `nuutta vedäsi* (MUL)
- (5) *syss veime **järve `viirde*** (SET)
- (6) *ol sõss nii vana **kalendri `perrä*** (VÕR)
- (7) *a sannaq ol'l'iva kyik' tuut **uja viirt pit't'i** alahh # jaq* (SET)
- (8) *ss `tul'lime **kellä ühe aigu*** (TAR)
- (9) *viis ällü einä+maale `pantti `ikki sinna **küüini `kõrva*** (MUL)
- (10) *jälle **nädälide `kauppa*** (TAR)

Neist kaassõnadest teises äärmuses on esimese mõõtme järgi kaasõnad *ääres*, *järel*, *juurde* jt, mis on seega seotud põhjaeesti murretega (näited 11–13)

- (11) *krahv üksibäini söönud **lauwa ääres*** (IDA)
- (12) *ja siiss **selle lahingu järel** jäi sial see liin `seisma* (LÄÄ)
- (13) **läksin Talina siin **oma *sõprade *juure*** (RAN)

Nende selgelt kas lõunaeesti või põhjaeesti murretega seotud kaassõnade puhul ongi peamiselt tegemist erinevate sünonüümipaaridega. Näiteks vastab lõunaeesti murrete juures paiknevale kaassõnale *manu* (14) põhjaeesti murrete läheduses olev kaassõna *juurde* (15).

- (14) *`min't't'i **ves'k'i mano*** (SET)

(15) *no kus sa läksid läksid kirikku juure* (LÄÄ)

Üksteisele vastavad ka paarid *veeres* (16) ja *ääres* (17) ning *perra* (18) ja *järele* (19).

(16) *tuu `pes't'i `nahku ynnõ uja veerehh* (SET)

(17) **seitse+dõist+*kümmedell *jõulu iel siis pesima pesu jõe ääres* (KIR)

(18) *ol sõss nii vana kalendri perrä* (VÕR)

(19) *sie õli *nüütse ka*lenderi järele* (KIR)

Huvitava rühma moodustavad kaassõnad *sisse* (20), *sees* (21), *seest* (22), mis näivad graafiku põhjal olevat omasemad lääne- ja saarte murdele.

(20) *ehk `jälle paneme kot'ti sisse* (LÄÄ)

(21) *taa mä^ngab sääl liiva sehes* (SAA)

(22) *vana vana `suurde metsade seest korjat männi koored* (SAA)

Nende adpositsioonide suur kasutussagedus võib viidata sellele, et nendes murretes eelistatakse sisekohasuhete väljendamisel analüütilisi vorme – sisekohakäänete asemel on kasutatud nendele vastavaid kohasuhete väljendavaid kaassõnu. Sellist tendentsi on Lääne-Eesti murretes täheldanud ka Palmeos (1982: 15) ja Pajusalu jt (2009: 113) ning sarnaste tulemusteni on jõudnud ka Klavan jt (2015).

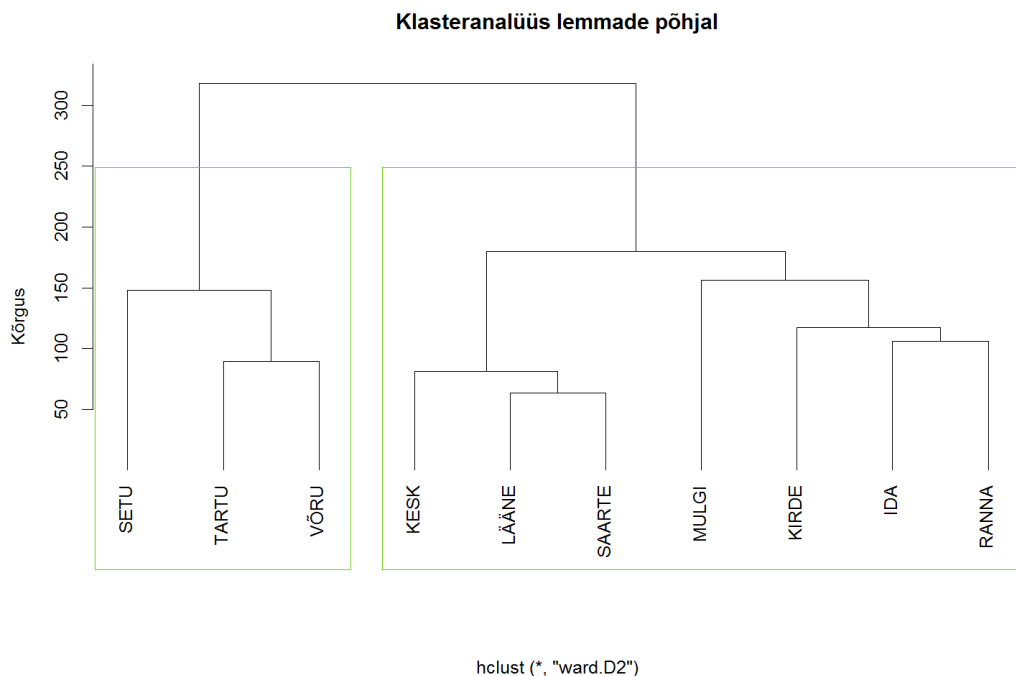
Üldjoontes aga osutavad kaassõna lekseemi ning murrete korrespondentsanalüüsi tulemused harilikule põhja- ja lõunaeeesti murrete eristusele, mis lekseemitasandil on ootuspärane. Sarnane põhja- ja lõunaeeesti jaotus ilmneb ka joonise 4 klasteranalüüsis, kus on moodustunud kaks suuremat rühma: Setu, Tartu ja Võru murdega lõunaeeesti rühm ning kesk-, lääne-, saarte, Mulgi, kirde-, ida- ja rannamurdega põhjaeeesti rühm. Need kaks suuremat rühma on ka analüüsi põhjal optimaalsed (silueti väärtusega 0.36) ning neid kahte rühma võib pidada klasteranalüüsi kõige olulistemaks.

Mulgi murre kuulub seejuures üllatuslikult põhjaeeesti rühma, kuid tegelikult ka CA graafiku põhjal paiknes Mulgi murre teistest lõunaeeesti murretest kaugemal ja oli üsna lähedal just ida- ja rannamurdele.

Suurem põhjaeeesti rühm jaguneb kaheks: kesk-, lääne- ja saarte murde rühmaks ning Mulgi, kirde-, ida- ja rannamurde rühmaks. Mulgi murre ei moodusta seejuures kirde-,

ida- ja rannamurdega ühtset klastrit, vaid eraldub neist varem. Ka kirdemurre on ida- ja rannamurdest varem eraldunud, kuid korrespondentsanalüüsi järgi olid kõik need kolm murret üksteisele sama lähestikku. CA graafiku põhjal on lääne- ning saarte murre kaassõnade kasutuse poolest sarnased ning ka klasteranalüüsis moodustavad need kaks murret keskmurdest eraldi klastri.

Lõunaeeesti murrete rühmas eraldub esmalt teistest Setu murre. Tartu ja Võru murrre moodustavad niisiis omaette klastri ning ka CA graafikul paiknesid need kaks murret kohakuti. See viitab, et Tartu ja Võru murdes kasutatakse kaassõnu üsna sarnaselt.



Joonis 4. Klasteranalüüs lekseemide normaliseeritud sageduste põhjal

4.3. Komplementide käänded

Komplementide käänete kasutust murretes olen vaadelnud mitmest korrespondentsanalüüsi ja klasteranalüüsi rakendades ning sagedusandmeid kõrvutades. MCA graafikus (joonis 5) on esindatud murdealad, komplementide käänded ning kaassõnafraasi sõnajärg. Klasteranalüüs (joonis 7) põhineb ainult komplementide käänetel.

MCA analüüs komplementide käänetest

a KÄÄNE a MURRE a POS



Joonis 5. Kaasõnade komplementide käänete mitmese korrespondentsanalüüsi graafik.

Joonisel 5 ilmneb esimese ning teise mõõtme põhjal, et murded on koondunud graafiku keskpunkti lähiümbrusesse. Teise mõõtme põhjal saab eristada kahete rühma: kesk-, lääne-, ida- ja Mulgi murrete rühm ning saarte, Tartu, Võru, ranna-, kirde- ja Setu murrete rühm, millest Setu paikneb mõnevõrra teistest eemal. Seejuures väärib märkimist, et üksteisega kattuvad ranna- ning kirdemurde punktid ning samasse rühma kuuluvateks võib pidada ka lõunaesti Tartu, Võru ja Setu murret. Seega viitavad tulemused komplementide käänete kasutuse puhul mõnevõrra murrete jagunemisele omamoodi kesk- ja äärealade rühmadeks. Ometi paikneb Mulgi murre keskala murrete rühmas. Mõlemad teise mõõtme põhjal moodustunud rühmad paiknevadki tegelikult väga keskpunkti lähedal ning seega on neid keeruline eristada. Üksikute käänete puhul ilmneb aga selgelt seos pre- või postpositsioonilise sõnajärgjega. Genitiiv on ootuspäraselt seotud postpositsioonidega ning paikneb ka üsna lähedal graafiku keskpunktile (23).

(23) `valge linane riije `pandi sis seena `peale (SAA)

Genitiivse konstruktsiooni graafiku keskpunkti juures paiknemine tähendab, et genitiivi kasutus on kõikides murretes ühtlane. Peatükis 1.1 on välja toodud, et postpositsioonifraas, kus komplement on genitiivis, ongi eesti keeles harilik ning ka ajalooliselt algne konstruktsioon (EKG I 1995: 37–38; Grünthal 2003; Palmeos 1982). Niisiis selgub siingi, et murretes on postpositsioonidega koos genitiivi kasutus kõige sagedasem. Lisaks on EKG II osas (1993: 138–139) esitatud, et näiteks elatiivi, nominatiivi ning adessiivi nõuavad kaassõnad oma komplementidilt vaid postpositsioonina ning seegi kajastub graafikul ja mainitud käänded paiknevad prepositsioonidest kaugel (näited 24–26).

(24) et kuiss timä nii pal'l'o `vaiva latsõst saa'ni nägi (VOR)

(25) talvõd läbi tas'si `einu puid (SAA)

(26) ja siis `öösse ühel ajal (SAA)

Nii nagu esineb genitiiv ootuspäraselt enim postpositsioonina, esineb graafiku põhjal komplement partitiivis kõige sagedamini koos prepositsioonidega (27).

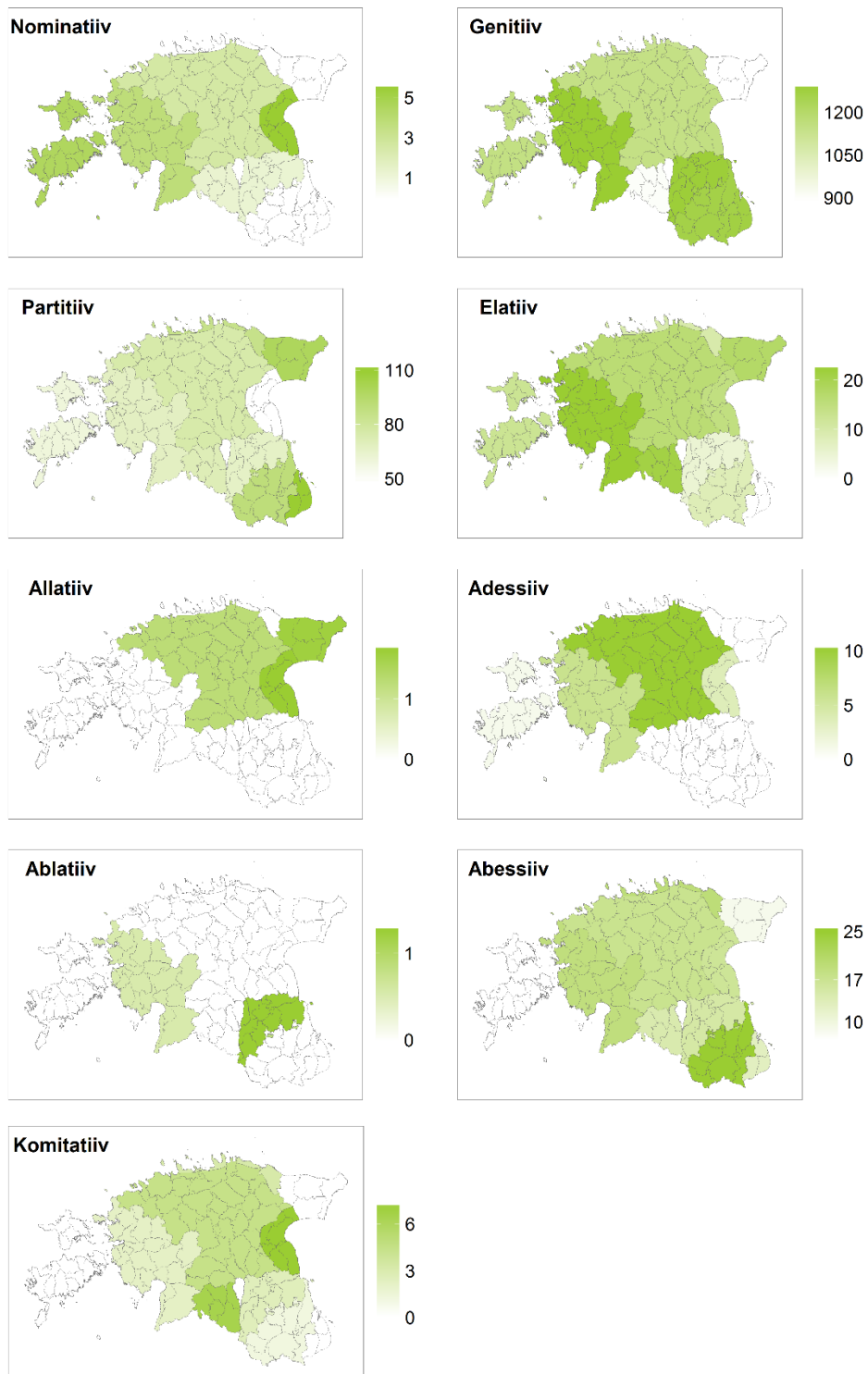
(27) *ülget vedada *müöda paljast jääd (KIR)

Graafiku keskmest paiknevad partitiiv ning prepositsioonid siiski eemal – prepositsioonid ongi murdekorpuses vähe esindatud ning seega on ka nendega seostuvat partitiivi vähe. Murretest paiknevad partitiivile ning seega ka prepositsioonidele kõige lähemal ranna-, kirde- ja Setu murre ning neist kolmest murdest eemal, kuid siiski prepositsioonidele ning partitiivile kõige lähemal on ka keskmurre. Tabelis 4 on esitatud komplemendi käänete normaliseeritud sagedused ning siingi selgub, et genitiivi kasutus on valdav ning teisi komplemendi käändeid esineb murdekorpuses vähe. Seejuures on huvitav, et elatiiv, mida kasutatakse graafiku põhjal enim postpositsioonidega, on kõigis murretes üsna ühtlase kasutussagedusega, kuid Tartu, Võru ning Setu murdes esineb seda käänat oluliselt vähem. Ilmselt on elatiiv seotud kaassõnadega *peast* ja *saadik*, mis joonise 3 CA analüüsi põhjal ongi omasem pigem põhjaeesti murretele. Mulgi murre taas on elatiivi kasutussageduse põhjal murretest teisel kohal.

Tabel 4. Komplemendi käänete normaliseeritud sagedusandmestik

	nom	gen	par	ela	all	ade	abl	abe	kom
IDA	6	1152	48	15	2	4	0	17	7
KESK	3	1145	75	16	1	10	0	17	4
KIRDE	0	895	100	18	2	0	0	9	0
LÄÄNE	4	1297	68	23	0	6	1	19	2
MULGI	1	906	71	21	0	0	0	14	7
RANNA	0	1121	86	10	0	0	0	19	3
SAARTE	4	1149	59	12	0	1	0	7	0
SETU	0	1271	111	0	0	0	0	13	0
TARTU	1	1287	66	5	0	0	1	15	3
VÕRU	0	1262	91	8	0	0	0	25	1

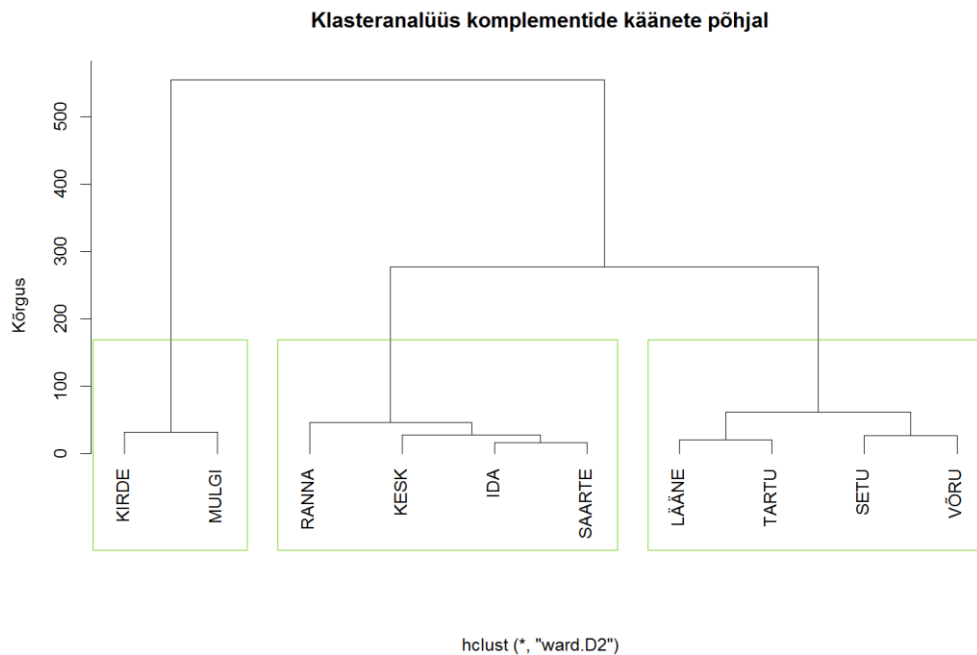
Samad tabelis 4 esitatud sagedusandmed on esitatud ka geograafiliselt murdekaartidel (joonis 6).



Joonis 6. Komplementide käänete jagunemine geograafiliselt

Klasteranalüüsi (joonis 7) põhjal tekivad selgemad murderühmad kui mitmese korrespondentsanalüüsi (joonis 5) järgi. Optimaalselt jaguneksid murded kolmeks

rühmaks (silueti väärtus 0,77) ning kolme rohelisega märgitud klasteri järgi selgelt põhja ja lõuna või ida ja lääne eristust ei ilmne. Sagedusandmetes on ülekaalus genitiivis komplement ning klasteranalüüs ongi põhiliselt selle käände kasutusjuhtudest mõjutatud.



Joonis 7. Klasteranalüüs komplementide käänete normaliseeritud sageduste põhjal

Ka joonisel 6 on näha, et geograafiliselt on genitiivi kasutatud enim lääne-, Tartu, Võru ja Setu murdes. Need neli murret ongi moodustanud dendogrammis omaette klasteri, mis sisaldab veel eraldi lääne- ja Tartu murde rühma ning Setu ja Võru rühma. Selle taustal võib olla nende murrete kalduvus analüütilisusele: läänemurde kohta on toodud esile kohasuhete väljendamisel analüütilise konstruktsiooni eelistamine, milles komplement on enamasti just genitiivis (Juhkam, Sepp 2000: 29; Lonn, Niit 2002: 49) ning sama on täheldatud ka Võru murde kohta (Keem, Käsi 2002: 41). Materjalis esinebki kohasuhteid palju (sellest täpsemalt peatükis 4.4) ning selle väljendatud tähendusega on ajalooliselt seotud genitiivikonstruktsiooni kasutamine, sest algselt on kaassõnad tähistanud vaid ruumilisi suhteid ning nende komplement on olnud genitiivis (nt Grünthal 2003; Grünthal 2005).

Klasteranalüüsis on teisteks moodustunud rühmaks kirde- ja Mulgi murde rühm. Nendes murretes on genitiivi kasutus kasutus kõige vähesem (joonis 6). Kirderanniku

murrete ülevaates (Juhkam, Sepp 2000: 29) on mainitud, et kirde- ja rannamurdes eelistatakse kohasuhete väljendamisel kaassõnadele vastavaid kohakäändeid ning taas, kuna enamjaolt on andmestikus kasutatud genitiivi seoses kohta väljendavate kaassõnadega, on see mõjutanud siinseid tulemusi. Rannamurre paikneb eraldi kirde- ja Mulgi klastrist, kuid on eraldunud seejuures ülejäänud sama rühma murretest, kesk-, ida- ja saarte murdest.

4.4. Kaassõnade tähendused

Kaassõnade erinevate tähenduste väljendamist murretes vaatlen mitmese korrespondentsanalüüsi, klasteranalüüsi ning sagedusandmete põhjal. MCA analüüsis (joonis 8) on esitatud eri murdealad, kaassõnade tähendused ning kaassõnafraasi sõnajärg. Klasteranalüüs (joonis 10) põhineb erinevatel kaassõnade tähendustel.

MCA graafikus (joonis 8) ilmneb, et murded on koondunud keskpunkti ümber. Seejuures kattuvad ranna-, Setu, Võru ning läänemurre pea täielikult ning ülejäänud murdedki ei paikne tihedamast tuumikust kaugel. See tähendab, et kaassõnade eri tähendusi kasutatakse kõikides murretes üpris ühtlaselt ning sarnase kasutussagedusega. Erinevate tähenduste ühtlane kasutussagedus murretes ilmneb ka sagedusandmete järgi tabelis 5 ning sama on nähtav ka geograafiliselt joonisel 9.

Kui aga võrrelda kasutussagedusi tähenduste kaupa, tuleb tabelis 5 iga murde puhul esile kohasuhete (28) rohke esinemine.

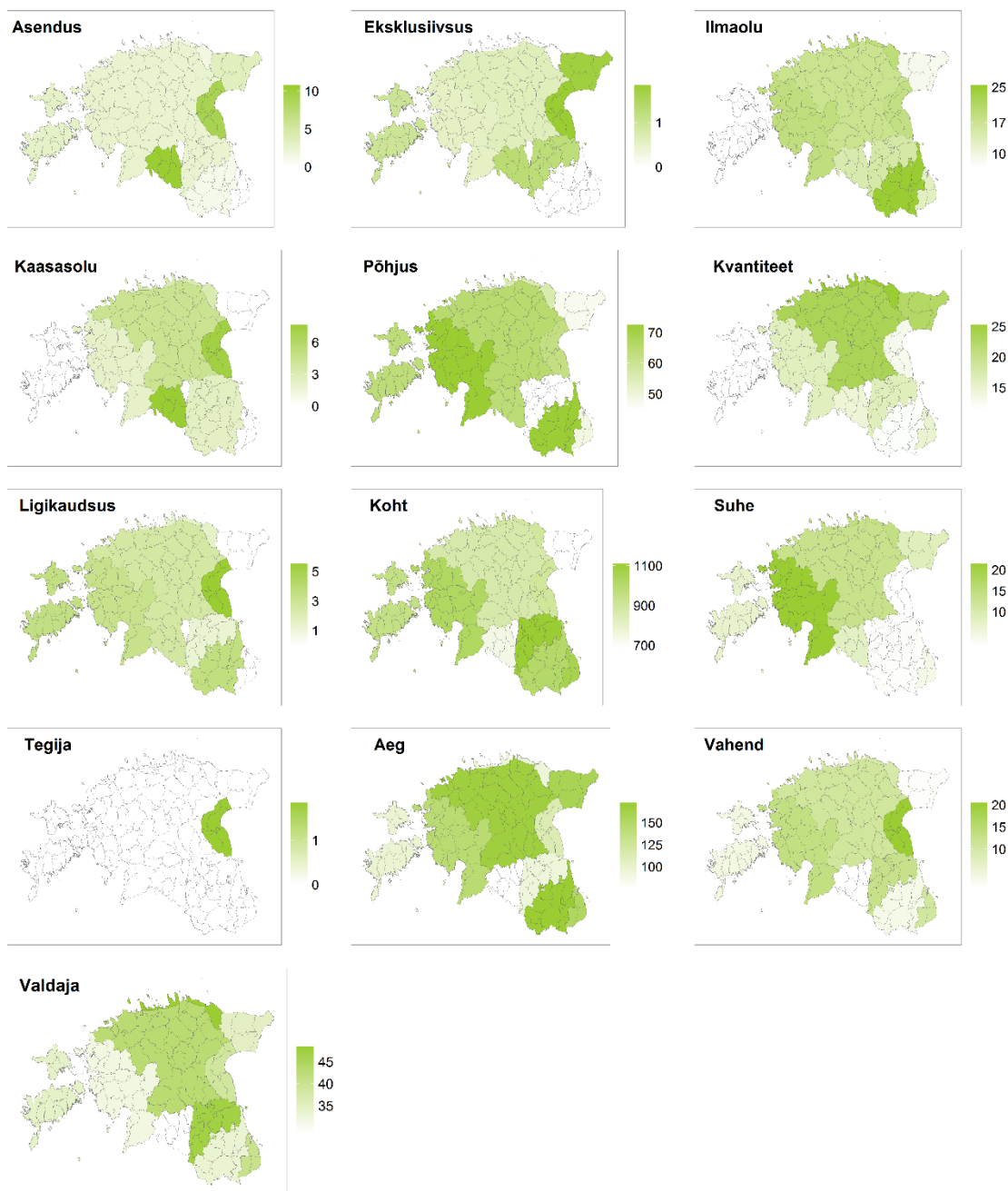
(28) *koloosi tüö+goda onn sial uovi pial* (KES)

Kohasuhe paikneb ka MCA graafikul keskpunkti läheduses ning teised tähendused on väikese esinemissageduse tõttu keskpunktist eemaldunud. Nagu selgus ka töö esimeses peatükis, ongi kaassõnad algselt väljendanud kohasuhteid ning teised abstraktsemad suhted on välja arenenud just kohasuhetest (Habicht 2000: 21; Palmeos 1982: 69–72). Samuti on tänapäeva eesti keele uurimustes leitud (Veismann 2009), et kõnelejale on kaassõnade puhul ruumisuhete väljendamine primaarne ning üldisemaltki on see tähendus kaassõnadele prototüüpne. Seega võib kohasuhete suur esinemissagedus ning graafiku keskpunktis esinemine viidata sellele, et ka suulise murdekõne põhjal on kohasuhete väljendamine kaassõnade peamine ülesanne.

Tabel 4. Kaassõnade tähenduste normaliseeritud sagedusandmestik

	ase	eks	ilm	kaa	kau	kva	lig	lok	suh	teg	tem	vah	val
IDA	9	2	18	7	63	13	6	905	2	2	114	20	39
KESK	3	1	17	4	65	24	3	872	13	0	167	11	43
KIRDE	4	2	9	0	47	23	0	688	9	0	159	2	35
LÄÄNE	3	1	19	2	73	16	3	1030	22	0	145	14	32
MULGI	11	1	14	8	64	14	3	743	8	0	76	1	29
RANNA	2	0	19	3	57	26	2	904	15	0	95	10	48
SAARTE	3	1	7	0	64	11	3	951	7	0	98	4	34
SETU	0	0	13	0	49	15	0	1067	4	0	153	11	40
TARTU	3	1	15	3	45	17	1	1110	3	0	94	13	47
VÕRU	1	0	25	2	73	12	4	1013	2	0	172	4	33

Geograafiliselt (joonis 9) ilmneb, et kohasuhteid on kõige enam kasutatud lääne-, saarte, Tartu, Võru ning Setu murdes. Taas võib taustal olla analüütilisuse eelistamine, mida on nendes murretes just nimelt kohatähenduste puhul märgatud (Juhkam, Sepp 2000: 29; Lonn, Niit 2002: 49; Keem, Käsi 2002: 41). Kirdemurdes on seejuures jällegi teiste murretega võrreldes kohasuhete sagedus madalam ning üldjoones kehtib sama ka muude tähenduste kohta. Ometi on kirdemurdes üsna kõrge ajasuhete kasutus, seega ei saa öelda, et kõigis tähendustes võiks kirdemurdes olla eelistatud sünteetilised konstruktsioonid.



Joonis 9. Kaassõnade väljendatud tähenduste jagunemine geograafiliselt

Kaassõnafraasi sõnajärje ning erinevate tähenduste puhul on märkimisväärne, et taaskord paiknevad postpositsioonid ootuspäraselt graafiku (joonis 8) keskpunkti läheduses. Prepositsioonide puhul aga näib olevat seos spetsiifilistemate tähendustega, mis liituvad tegelikult vaid mõne teatud lekseemiga. Näiteks kaasasolu väljendavateks

kaassõnadeks on andmestikus kaassõna *ühes*, eksklusiivsust väljendatakse peamiselt kaassõnaga *peale* ning ilmaolu kaassõnaga *ilma* (29–31).

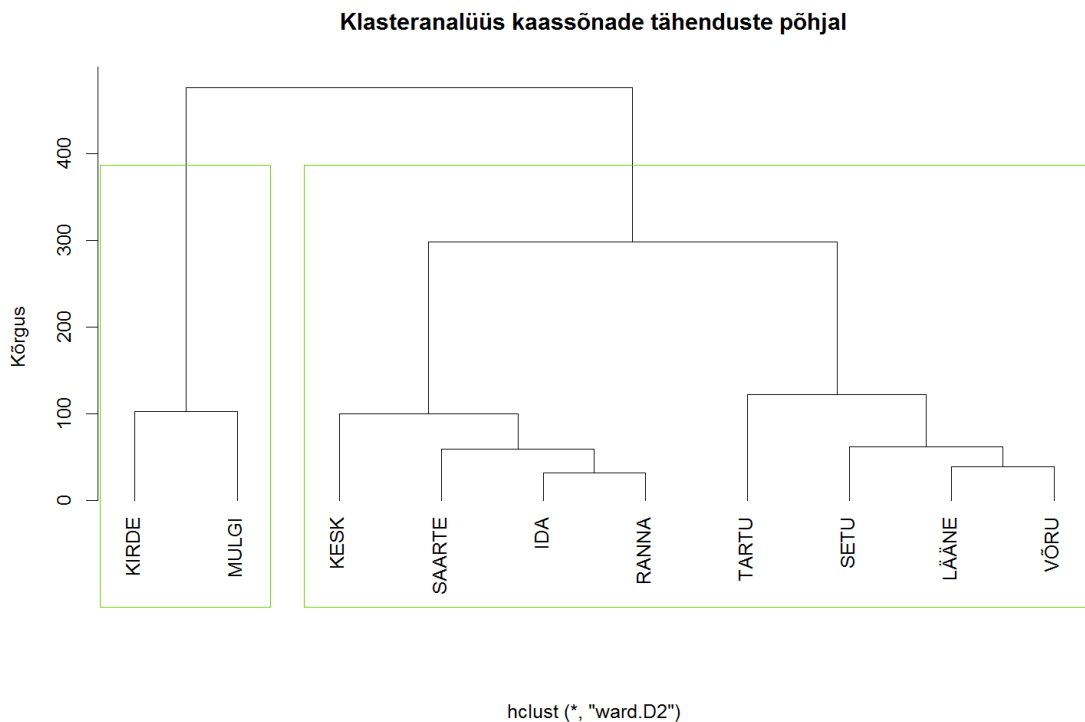
(29) *need `ol'lid koa **ühes** minu vanemattega* (IDA)

(30) *põle änam seal kedagid mtte **piale** minu* (SAA)

(31) *siis ta on `piale seda **ilma** `mõistusetta veel tükk `aega* (LAA)

Prepositsioonide punkt ning seetõttu ka prepositsioonidega seostuvad tähendused paiknevad MCA graafikul kõige lähemal ranna-, kesk- ja kirdemurdele. Joonise 5 graafiku järgi, kus oli vaadeldud komplementide käändeid, olid need samad murded samuti prepositsioonidele kõige lähemal.

Klasteranalüüsis (joonis 10) on jagunemine taas enim mõjutatud kohasuhetest, mida esineb materjalis kõige rohkem. Dendogrammis märgitud optimaalsed, kaks rühma (silueti väärtusega 0,54), moodustuvad kirde- ja Mulgi murdest ning teise rühma moodustavad kõik ülejäänud murded. Kirde- ja Mulgi murdes ongi kohasuhteid ning kaassõnu kasutatud kõige vähem. Teised murded moodustavad kaks suuremat klastrit: kesk-, saarte, ida- ja rannamurde ning Tartu, Setu, lääne- ja Võru murde rühmad. Neist esimeses rühmas on klasteranalüüsi põhjal sarnasemad ida- ja rannamurre ning teises rühmas taas lääne- ja Võru murre. Lääne-, Tartu, Setu ja Võru murdes ongi analüütilisus tavalisem, kuid varasemate tähelepanekute põhjal (nt Pajusalu jt 2009: 113) võiks eeldada ka saarte murde kuulumist samasse rühma.



Joonis 10. Klasteranalüüs tähenduste normaliseeritud sageduste põhjal

4.5. Tulemuste kokkuvõte

Analüüside eesmärgiks oli tuvastada murretevahelisi erinevusi kaassõnade kasutuses ning leida eristuvaid rühmi. Erinevate analüüside põhjal kokku ei ilmnenud küll selgelt eristuvaid murderühmi, kuid kaassõnade kasutuses on siiski murretevahelised erinevused olemas. Näiteks selgus, et morfosüntaktiliste ning semantiliste tegurite põhjal ei jagunenud murded traditsioonilise põhja- ja lõunaesti murdejaotuse järgi, vaid pigem genitiivi ning kohasuhete sagedusandmete järgi, mille taustal võib olla analüütilisuse või sünteetilisuse eelistamine. Lekseemide analüüsis aga eristusid murded traditsiooniliselt põhja- ning lõunaesti murderühmadeks. Lõunaesti murretest moodustasid küll Tartu, Võru ja Setu murre kõikides analüüsid ühtsema rühma, kuid erinevates analüüsid eristus teistest lõunaesti murretest alati Mulgi murre. Lekseemide kasutuse poolest kuulus Mulgi murre siiski korrespondentsanalüüsis (joonis 3) pigem lõunaesti murretega samasse rühma, kuid klasteranalüüsi järgi sarnaneb Mulgi murre rohkem põhjaesti murretega. See osutab, et Mulgi murre on kaassõnade kasutuse poolest vahepealsem ning

ei kuulu kaassõnade lekseemide, morfosüntaksi ning semantika poolest selgelt teiste lõunaesti murrete juurde.

Iga üksiku analüüsi põhjal ilmesid üsna selged murrete grupid. Lekseemide korrespondentsanalüüsis selgus, et eristada võib kolme homogeensemata murderühma: esiteks Tartu, Võru ja Setu murre, teiseks ida-, ranna, kirde- ja keskmurre ning kolmandaks lääne- ja saarte murdest moodustuv rühm. Nagu mainitud, polnud Mulgi murre seejuures selgelt seotud ühegi mainitud rühmaga ning paiknes kahe esimese vahel. Lekseemide puhul olid klasteranalüüsi tulemused sarnased korrespondentsanalüüsiga ning Mulgi murre kuulus seal põhjaesti murrete rühma. Üldjoontes järgisid lekseemide tulemused siiski traditsioonilist murdejaotust põhja- ning lõunaesti murreteks.

Lekseemide puhul tuli lääne- ning saartemurdes esile sisekohakäänetele vastavaid tähendusi väljendavate kaassõnade suur kasutussagedus. See võib viidata analüütilisemate väljendusvõimaluste kasutamise eelistamisele nendes murretes. Näiteks ongi osutatud, et nendes murretes kasutatakse kohatähenduses pigem kaassõnu kui nendega (lähi)sünonüümseid käändeid (Pajusalu 2009: 113–114; Palmeos 1982: 15). Lekseemide analüüsis ilmnis ootuspäraselt, et üksikud kaassõnad on rohkem seotud mõne murdealaga. Teatud sünonüümsete kaassõnapaaride puhul, nagu näiteks *man* ja *juures*, kasutatakse üht varianti lõunaesti ning teist varianti põhjaesti murretes.

Komplementi käänete kasutuse põhjal eristusid keskalade rühm kesk-, ida-, lääne- ja Mulgi murretega ning äärealade rühm saarte, Tartu, Võru, ranna-, kirde- ja Setu murdega. Genitiiv komplementi käändena on materjalis ülekaalus ning selle kasutus on ka kõikides murretes üsna ühtlane. Vorm on kõige enam seotud eesti keelele omasema postpositsioonilise sõnajärgjega. Partitiivi kasutus on harudasem ning seda kasutatakse kõige sagedamini just prepositsioonidega. Prepositsioonide ning postpositsioonide kasutuses murrete vahel tugevat statistilist seost ei tuvastatud. Suurim seos kaassõna paiknemisega on rannamurdel, kus prepositsioone esineb oodatust rohkem ning saarte murdel, kus prepositsioone esineb oodatust oluliselt vähem. Ka mitmest korrespondentsanalüüsides paiknes rannamurre võrreldes saarte murdega lähemal prepositsioonidele ning saarte murre paiknes lähemal taas postpositsioonidele.

Tähenduste puhul ilmnis kohasuhete olulisus – seda suhet kasutatakse suhteliselt ühtlaselt ning sama sagedusega kõigis murretes. Teisi tähendusi on palju vähem ning see

kinnitab varasemaid kirjakeele uurimuste tulemusi, mille järgi kaassõnade põhiülesandeks on väljendada kohasuhteid ning muud abstraktsemad suhted on lisandunud alles hiljem (nt Habicht 2000: 21; Veismann 2009). Enim oli kohasuhteid kasutatud lääne-, saarte, Tartu, Võru ning Setu murdes, ning see viitab taas varem täheldatud tendentsile, et kohasuhteid väljendades on nendes murretes eelistatud analüütiline väljendusviis (Juhkam, Sepp 2000: 29; Lonn, Niit 2002: 49; Keem, Käsi 2002: 41). Kirdemurdes oli pea kõiki tähendusi teiste murretega võrreldes vähem ning see võib viidata sünteetilisuse eelistamisele, mida on täheldatud kirdemurde puhul varemgi (nt Must 1995: 38; Klavan jt 2015).

4.6. Edasised uurimissuunad

Käesolev magistritöö keskendus kaassõnade kasutusele üldisemalt ning andis uudset empiirilist infot varem põhjalikult uurimata teema kohta. Töö keelematerjali põhjal saaks keskenduda veel täpsemalt näiteks ühele kaassõna sagedaale tähendusele korraga. Materjali kogudes ilmnes ka kaassõnade sõnaliigi piiri hägususe problemaatika ning sellega seoses võiks murdekorpusest välja otsida ebahühtlaselt märgendatud kaassõnade tüüpjuhte ning uurida täpsemalt nende grammatiseerumisprotsessi. Töö tulemused osutavad eri murrete tendentsidele eelistada sünteetilisust või analüütilisust ning see teema vääraks kindlasti edasiuurimist. Täpsemalt võikski vaadelda üksikute käänete ning nendele vastavate adpositsioonifraaside kasutust, nagu seda on varasemalt uurinud adessiivi ja kaassõna *peal* puhul Klavan jt (2015).

Kokkuvõte

Käesolevas magistritöös uurisin kaassõnade kasutust eesti murretes. Töö põhieesmärgiks oli välja selgitada, kas kaassõnade kasutuses on murretevahelisi erinevusi ning leida kaassõnade kasutuse poolest sarnasemaid murderühmi. Kitsamalt olid uurimisküsimusteks, 1) kas ees- ja tagasõnade kasutuses on murrete vahel erinevusi, 2) millised kaassõnad on eri murretele iseloomulikud, 3) mis seos on murrete ning komplementide käänete vahel ning 4) missuguseid kaassõnade tähendusi eri murretes eelistatakse.

Uurimisküsimustele vastamiseks viisin läbi neli kvantitatiivset analüüsi. Analüüsid põhinesid eesti murrete korpuse morfoloogiliselt märgendatud murdetekstidel. Andmestik sisaldas vaid selliseid kaassõnu, mida esines korpuses vähemalt 20 korda ning lõplik andmestik koosnes 12 785-st kaassõna sisaldavast lausungist ning neljast kodeeritud kaassõnade kasutust potentsiaalselt mõjutavast tunnusest (murre, kaassõna lemma, komplementi kääne, kaassõna väljendatav suhe). Andmete analüüsiks kasutasin erinevaid statistilisi meetodeid, nagu lihtne ja mitmene korrespondentsanalüüs, klasteranalüüs ning Pearsoni χ^2 -test.

Murrete ning ees- ja tagasõnade kasutuse vahel ei ilmnenud tugevat statistilist seost. Esile tõusis rannamurre, kus prepositsioone esineb oodatust rohkem, ning saarte murre, kus prepositsioone esineb oodatust oluliselt vähem.

Lekseemide analüüsides eristusid kolm põhilist rühma: Tartu, Võru ja Setu murderühm; ida-, ranna, kirde- ja keskmurde rühm ning lääne- ja saarte murderühm. Mulgi murre polnud seejuures selgelt ühegi rühmaga seotud. Peamiselt jagunesid murded kaassõnade lekseemide kasutuse poolest siiski traditsiooniliselt põhja- ning lõunaeesti murreteks. Lekseemide kasutuses ilmnes ka, et mõned kaassõnad on seotud teatud murretega või suuremate murdealadega. Näiteks on lääne- ning saarte murdes kasutatud palju sisekohakäänetele vastavaid kaassõnu, mis võib viidata analüütilisuse eelistamisele neis murretes. Paljude sünonüümsete kaassõnapaaride puhul, nagu näiteks *man* ja *juures*,

veeres ja *ääres* ning *perra* ja *järele*, eelistati lõuna- ning põhjaeesti murretes vastavalt kas esimest või teist varianti.

Komplemendi käände kasutuse põhjal eristusid nn keskalade rühm, mis sisaldas kesk-, ida-, lääne- ja Mulgi murret, ning äärealade rühm, mis sisaldas saarte, Tartu, Võru, ranna-, kirde- ja Setu murret. Kõikides murretes oli enim kasutatud kääne genitiiv ning selle kasutus oli ka kõikjal ühtlane. Partitiivi kasutus oli oluliselt väiksem ning peamiselt seotud prepositsioonidega. Analüüside põhjal oli partitiiv enim seotud ranna-, kirde-, Setu ja keskmurdega ning nendes murretes esines ka normaliseeritud sageduste järgi kõige enam prepositsioone.

Kaassõnade tähendustega tõusis esile kohasuhete rohkus. Kohasuhteid kasutati ülekaalukalt kõikides murretes ning teisi tähendusi esines vähe. See tulemus kinnitab varasemate kirjakeele uurimuste tulemusi (nt Habicht 2000: 21; Veismann 2009), mille järgi kaassõnade keskseks ülesandeks on just kohasuhete väljendamine. Enim esines kohasuhteid lääne-, saarte, Tartu, Võru ning Setu murdes, mis ühtib varasemate tähelepanekutega, et kohasuhteid tähistades on nendes murretes eelistatud analüütiline väljendusviis (Juhkam, Sepp 2000: 29; Lonn, Niit 2002: 49; Keem, Käsi 2002: 41). Kirdemurdes aga oli pea kõiki erinevaid väljendatuid tähendusi teiste murretega võrreldes kõige vähem, mis võib viidata sünteetilisuse eelistamisele.

Üldjoontes ei osutanud erinevate analüüside tulemused aga ühte selget suundumust, kuidas murded võiksid kaassõnade kasutuse poolest rühmadesse jaguneda. Lekseemide, morfosüntaktiliste ning semantiliste tegurite analüüsid viitasid üsna erinevale murdejaotusele. Kõikides analüüsides moodustasid Tartu, Võru ja Setu murre küll ühtsema lõunaeesti rühma, kuhu võis kuuluda ka teisi põhjaeesti murdeid. Analüüsides eristus aga muidu ühtsemast lõunaeesti rühmast alati Mulgi murre, millest järeldub, et Mulgi murde kaassõnade kasutus erineb ülejäänud lõunaeesti murrete omast.

Kirjandus

- Abdi, Hervé ja Dominique Valentin 2016. *Multiple Correspondence Analysis*.
<https://www.utdallas.edu/~herve/Abdi-MCA2007-Pretty.pdf>. Vaadatud
29.04.2016
- Baayen, R. Harald 2008. *Analyzing linguistic data: a practical introduction to statistics using R*. Cambridge: Cambridge University Press.
- EMK = Eesti murrete korpus 2014. <http://www.keel.ut.ee/et/keelekogud/murdekorpus>.
Vaadatud 03.03.2016
- EKG I = Ereht, Mati, Reet Kasik, Helle Metslang, Henno Rajandi, Kristiina Ross, Henn Saari, Kaja Tael ja Silvi Vare 1995. *Eesti keele grammatika I*. Tallinn: Eesti Teaduste Akadeemia Eesti Keele Instituut.
- EKG II = Ereht, Mati, Reet Kasik, Henno Rajandi, Kristiina Ross, Henn Saari, Kaja Tael ja Silvi Vare 1993. *Eesti keele grammatika II*. Tallinn: Eesti Teaduste Akadeemia Keele ja Kirjanduse Instituut.
- Greenacre, Michael 2007. *Correspondence analysis in practice. 2nd edition*. Boca Raton Fla.: CRC Press.
- Gries, Stefan Th. 2013a. *Elementary statistical testing testing with R*. In Manfred Krug & Julia Schlüter (eds.). *Research methods in language variation and change*. New York: Cambridge University Press, 361–381.
- Gries, Stefan Th. 2013b. *Statistics for Linguistics with R: A Practical Introduction*. Berlin/Boston: Walter de Gruyter GmbH.
- Grünthal, Riho 2003. *Finnic Adpositions and Cases in Change*. – *Suomalais-Ugrilaisen Seuran Toimituksia* 244. Helsinki: Société Finno-Ougrienne.
- Grünthal, Riho 2005. *Miksi itämerensuomessa on prepositioita?*. – *Virittäjä* 1, 28–51.
- Habicht, Külli 2000. *Grammaticalization of adpositions in Old Literary Estonian*. *Estonian: Typological Studies IV*. Tartu: Tartu University Press, 19–58.
- Habicht, Külli ja Pille Penjam 2007. *Kaassõna keeleuurija ja -kasutaja käsituses*. – *Emakeele Seltsi Aastaraamat* 52 (2006). Tallinn: Emakeele Seltsi aastaraamat.

- Hagège, Claude 2010. *Adpositions*. Oxford: Oxford University Press.
- Hopper, Paul J. ja Elizabeth Closs Traugott 1994. *Grammaticalization*. Cambridge: Cambridge University Press.
- Husson, François, Julie Josse, Sebastien Lê ja Jeremy Mazet 2016. *FactoMineR: An R Package for Multivariate Analysis*. R Package Version 1.32.
- Juhkam, Evi ja Aldi Sepp 2000. *Eesti murded VIII. Läänemurde tekstid*. Tallinn: Eesti Keele Instituut.
- Jürine, Anni ja Külli Habicht 2013. *Kaassõnade tsükliline areng*. – Keel Ja Kirjandus 10, 51–56.
- Karelson, Rudolf 1972. *Märkmeid kaassõna piirimailt*. – Emakeele Seltsi Aastaraamat 18, 71–84.
- Karlsson, Fred 2002. *Üldkeeleteadus*. Tartu: Eesti Keele Sihtasutus.
- Keem, Helga ja Inge Käsi 2002. *Eesti murded VI. Võru murde tekstid*. Tallinn: Eesti Keele Instituut.
- Klavan, Jane 2012. *Evidence in Linguistics: Corpus-Linguistics and Experimental Methods for Studying Grammatical Synonymy*. Doktoritöö. Tartu: Tartu University Press.
- Klavan, Jane, Kristel Uiboaed ja Maarja-Liisa Pilvik 2015. *The Use of Multivariate Statistical Classification Models for Predicting Constructional Choice in Spoken, Non-Standard Varieties of Estonian*. – SKY Journal of Linguistics, 28, 187–224.
- Lebart, Ludovic, André Salem ja Lisette Berry 1998. *Exploring textual data*. Dordrecht: Kluwer Academic Publishers.
- Lehmann, Christian 1985. *Grammaticalization: Synchronic Variation and Diachronic Change*. – Lingua E Stile 20.
- Levshina, Natalia 2015. *How to do Linguistics with R*. Amsterdam/Philadelphia: John Benjamins Publishing Co.
- Lindström, Liina, Maarja-Liisa Pilvik, Mirjam Ruutma ja Kristel Uiboaed 2015. *Mineviku liitaegade kasutusest eesti murretes keelekontaktide valguses*. – Võro Instituudi Toimõndusõq 29, 39–70.
- Lonn, Varje ja Ellen Niit 2002. *Eesti murded VII. Saarte murde tekstid*. Tallinn: Eesti Keele Instituut.

- Must, Mari 1995. *Eesti murded V. Kirderannikumurde tekstid*. Tallinn: Eesti Teaduste Akadeemia Eesti Keele Instituut.
- Nenadic, Oleg ja Michael Greenacre 2007. *Correspondence Analysis in R, with Two and Three-dimensional Graphics: The ca Package*. – Journal of Statistical Software 20 (3), 2–13.
- Pajusalu, Karl, Tiit Hennoste, Ellen Niit, Peeter Päll ja Jüri Viikber 2009. *Eesti murded ja kohanimed*. Tallinn: Eesti Keele Sihtasutus.
- Palmeos, Paula 1982. *Eesti keele grammatika. II osa. Neljas vihik. Kaassõna*. Kolmas Trükk. Tartu: Tartu Riiklik Ülikool.
- Ploom, Harald 1955. *Kaassõnad eesti murdeis*. Diplomitöö: Käsikiri Tartu Ülikooli eesti keele õppetoolis.
- Rätsep, Huno 1977. *Eesti keele ajalooline morfoloogia I*. Tartu: Tartu Riiklik Ülikool. Eesti keele kateeder.
- Rätsep, Huno 1979. *Eesti keele ajalooline morfoloogia II*. Tartu Riiklik Ülikool. Eesti keele kateeder.
- R Core Team 2013. *R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria*. <http://www.R-project.org/>.
- Ruutma, Mirjam 2014. *Kaassõnade läbi, mööda, vastu, üle ja ümber pre- ja postpositsioonilisest varieerumisest eesti murretes*. Bakalaureusetöö: Käsikiri Tartu Ülikooli eesti keele õppetoolis.
- Ruutma, Mirjam, Aki-Juhani Kyröläinen, Maarja-Liisa Pilvik ja Kristel Uihoaed 2016. *Ambipositsioonide morfosüntaktilise varieerumise kirjeldusi kvantitatiivsete profiilide abil*. – Keel Ja Kirjandus, 2, 92–113.
- Suzuki, Ryota ja Hidetoshi Shimodaira 2015. *pvclust: Hierarchical Clustering with P-Values via Multiscale Bootstrap Resampling*. R Package Version 2.0-0.
- Szmrecsanyi, Benedikt 2014. *Methods and Objectives in Contemporary Dialectology*. – Современные Методы В Диалектологии. Ареал Северных, Северо-Западных Русских И Белорусских Говоров / Contemporary Approaches to Dialectology: The Area of North, Northwest Russian and Belarusian Vernaculars).

- Szmrecsanyi, Benedikt ja Cristoph Wolk 2011. *Holistic Corpus-Based Dialectology*. –
Revista Brasileira de Linguística Aplicada 2, 561–592.
- Tanning, Salme 1961. *Eesti murded I. Mulgi murdetekstid*. Tallinn: Eesti Riiklik Kirjastus.
- Uiboed, Kristel 2016. *Spatial Visualization with R*. spatial-visualization-with-r 1.0.
<http://dx.doi.org/10.5281/zenodo.51473>.
- Uiboed, Kristel 2013. *Verbiühendid eesti murretes*. Doktoritöö. Tartu: Tartu Ülikooli Kirjastus.
- Veismann, Ann 2009. *Eesti keele kaas- ja mäarsõnade semantika võimalusi*. Doktoritöö. Tartu: Tartu University Press.

Adpositions in Estonian Dialects. Summary

This thesis focuses on gaining insight into the use of adpositions in Estonian dialects. Adpositions have been extensively examined in written Estonian, but we lack research on non-standard language, e.g. dialects. One of the few previously conducted studies observes lexical characteristics of adpositions (e.g. Harald Ploom 1955) and single observations have been made about the use of some individual adposition or its morphosyntactic features (e.g. Palmeos 1982; Tanning 1961; Must 1995; Keem, Käsik 2002; Lonn, Niit 2002; Juhkam, Sepp 2000). Quantitative research on the subject has so far focused on the variation of prepositional and postpositional order in adpositional phrase (Ruutma 2013, Ruutma jt 2016), but there are no systematic works about the general use of adpositions in nonstandard varieties of Estonian.

The aim of this study is to contribute to the research of this understudied subject, to detect potential regional variation in the use of adpositions, and to discover dialects that are more similar to each other in terms of how they use adpositions. The main focus of this thesis is on the lexical, morphosyntactic, and semantic features of adpositions. The specific research questions are:

- How are pre- and postpositions used in dialects?
- Which lexemes are characteristic of different dialects?
- Is there any correlation between dialects and the case of the adpositions' complements,
- Do dialects differ in terms of the meanings the adpositions express?

The empirical data were obtained from the morphologically annotated Estonian Dialect Corpus (EMK). The corpus consists of 834 311 words from ten Estonian dialect areas (North Estonian dialects: Coastal, Eastern, Insular, Mid, Northeastern, and Western; and South Estonian dialects: Mulgi, Tartu, Võru, and Setu). The dataset was manually proofread and the cases that were incorrectly annotated as adpositions were removed. Due to the prerequisites of quantitative data analysis, the adpositions which occurred less than 20 times were left out. The final dataset contains 12 785 occurrences of 68 different

adpositions and their nearest context. Four variables which potentially affect the use of adpositions were chosen to be coded in the dataset.

The dialect data were analysed by following the principles of corpus-based dialectology (Szmrecsanyi 2014; Szmrecsanyi & Wolk 2011), which gives emphasis to the importance of frequency, statistical analysing techniques, and visualisation of the results. In this study, the quantitative analysis was carried out by applying four different methods: Pearson's χ^2 -test, Correspondence Analysis, Multivariate Correspondence Analysis, and agglomerative clustering. When interpreting the results, the normalised frequency tables and geographical distributions of the adpositions were observed as well. The R environment for statistical computing (R Core Team 2013) was used to perform all statistical analyses and compile the frequency maps.

The analyses revealed that the correlation between the use of pre- and postpositions and dialects is not statistically remarkable. However, it was noted that adpositional phrase order has the strongest correlation with Coastal dialect, where prepositions occur more than statistically expected, and with Insular dialect, where prepositions are less observed than statistically expected.

In the lexical analyses, three major dialect groups emerged: Tartu, Võru, and Setu group; Eastern, Coastal, Northeastern, and Mid group; Insular and Western group. Mulgi dialect was not attached to any of the groups. In general, the emerged groups followed the traditional division into northern and southern dialects. The study showed that adpositions which have the same functions as locative cases were used the most in Insular and Western dialects. This could indicate that analytic constructions are preferred when expressing locative function. In case of some adpositions that have synonymous counterparts (e.g. *veeres* and *ääres* 'beside, by'), one counterpart was strongly preferred in Northern dialects while the other in Southern dialects.

Two groups of dialects were formed based on the use of the complements' case: Mid, Eastern, Western, and Mulgi dialect; Insular, Tartu, Võru, Coastal, Northeastern, and Setu dialect. Genitive was the most used case in all dialects. The usage frequency of partitive was significantly lower and correlated with prepositional use. The results suggest that the use of partitive mostly correlates with Coastal, Northeastern, Setu, and Mid dialects. The use of prepositions was also the highest in these dialects.

Investigation of the use of adpositions' different meanings showed that the use of spatial relations was dominant and other expressed meanings were rare. This confirms previous results that have found spatial relations to be the core meaning of adpositions in written Estonian (e.g. Habicht 2000: 21; Veismann 2009). The spatial relation occurred the most in Western, Insular, Tartu, Võru, and Setu dialects and these results coincide with previous studies which have indicated, that in the mentioned dialects, analytical constructions are preferred when expressing spatial relations (Juhkam, Sepp 2000: 29; Lonn, Niit 2002: 49; Keem, Käsi 2002: 41). Compared to other dialects, the Northeastern dialect had the smallest number of occurrences of different meanings and this might indicate the preference of synthetic constructions.

In summary, the results of different analyses do not enable to make a definite conclusion about how dialects could be divided into groups based on their usage of adpositions. Each conducted analysis showed a slightly different grouping of dialects. However, in each analysis, the southern Tartu, Võru and Setu dialects formed a coherent group that could also contain some of the northern dialects. Mulgi dialect was always located apart from other southern dialects. This indicates that Mulgi dialect does not use adpositions quite like other southern dialects.

Lisa 1. Lekseemide normaliseeritud sagedusandmestik

	IDA	KESK	KIRDE	LÄÄNE	MULGI	RANNA	SAARTE	SETU	TARU	VÕRU
äärde	18	6	12	11	0	0	8	0	0	1
ääres	13	28	37	32	3	13	19	0	1	0
äärest	2	3	5	3	0	2	4	0	0	0
aegu	0	8	2	37	9	2	16	66	55	85
ajal	37	35	58	11	4	3	3	0	3	5
all	28	38	21	28	26	47	26	26	28	31
alt	20	12	11	10	5	24	12	6	8	8
asemel	7	1	4	2	9	0	0	0	3	1
ees	2	9	7	21	8	16	18	4	9	4
eest	26	33	19	29	37	31	33	30	24	24
enne	18	25	18	9	9	31	11	32	9	27
ette	24	10	11	7	7	16	8	6	13	4
ilma	18	17	9	19	14	19	7	13	15	25
jaoks	26	21	14	34	14	19	21	15	8	24
järel	2	20	14	12	4	34	9	0	0	0
järele	26	8	14	8	0	6	4	0	0	0
järgi	2	5	9	10	22	3	17	9	4	12
juurde	64	30	49	33	0	45	33	0	5	1
käes	15	14	5	8	3	8	8	13	8	10
käest	20	26	26	21	22	27	21	28	28	21
kätte	7	12	5	9	5	19	10	0	17	5
kaudu	2	1	2	13	1	0	24	0	1	0
kaupa	0	3	0	2	5	0	1	0	4	5
kohta	2	5	4	11	4	10	5	0	0	0
kõrval	7	0	5	4	4	5	4	0	5	2
kõrvale	0	1	0	1	1	2	1	4	5	6
külge	9	12	4	18	5	27	15	11	15	11
küljes	7	9	5	15	4	21	14	4	6	1
küljest	2	4	0	6	4	2	6	11	5	6
läbi	24	12	5	32	8	8	15	23	19	10
ligi	13	4	4	2	0	6	0	0	4	4
ligidal	6	4	4	2	1	5	1	0	0	0

man	0	0	0	0	24	0	0	47	122	81
mant	0	0	0	0	4	0	0	9	13	13
manu	0	0	0	0	62	0	0	75	85	70
mööda	17	30	35	22	42	31	5	0	10	13
moodi	0	17	7	23	7	3	18	30	3	12
otsa	9	28	5	35	7	18	13	11	25	14
otsas	6	17	7	17	11	15	11	6	22	15
otsast	2	3	0	2	1	0	1	2	4	2
pärast	9	19	16	19	45	13	23	19	22	27
peal	164	175	98	185	138	126	161	124	151	155
peale	182	162	149	162	116	197	153	266	173	200
pealt	53	46	26	46	39	40	40	45	52	37
peast	4	5	0	10	8	0	0	0	1	0
perra	0	0	0	0	1	0	0	13	5	18
pidi	2	3	0	5	8	5	8	64	19	36
pool	31	12	25	19	26	34	11	9	6	29
poole	64	14	16	19	29	36	8	51	38	63
poolt	33	26	32	16	25	34	14	2	10	25
saadik	6	8	18	10	9	10	12	0	4	8
seas	2	3	0	3	0	0	3	0	5	0
sees	50	63	53	85	32	19	82	26	46	44
seest	7	7	7	13	8	10	21	15	8	11
sisse	18	63	33	85	34	23	96	104	60	61
taga	18	19	9	14	8	8	18	13	18	11
tagant	6	3	7	4	3	3	4	6	5	0
taha	7	10	11	10	1	11	17	26	18	7
ühes	7	4	0	2	8	3	0	0	3	2
üle	28	47	37	41	29	55	30	26	39	29
ümber	28	13	5	29	11	8	30	40	20	29
vahel	13	11	7	23	13	16	29	2	14	5
vahele	17	11	7	13	1	18	8	9	10	7
vahelt	2	4	5	10	3	8	9	0	3	4
vastu	7	21	23	31	12	27	17	13	20	10
veerde	2	0	0	0	7	0	0	17	14	1
veeres	2	1	0	0	11	0	0	15	20	11
viisi	7	8	7	4	1	16	3	13	10	6

Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks

Mina MIRJAM RUUTMA

(*autori nimi*)

(sünnikuupäev: 10.11.1990)

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) enda loodud teose
„Kaassõnad eesti murretes”,
(*lõputöö pealkiri*)

mille juhendajad on Kristel Uiboaed ja Maarja-Liisa Pilvik,
(*juhendaja nimi*)

- 1.1. reprodutseerimiseks säilitamise ja üldsusele kättesaadavaks tegemise eesmärgil, sealhulgas digitaalarhiivi DSpace-is lisamise eesmärgil kuni autoriõiguse kehtivuse tähtaja lõppemiseni;
- 1.2. üldsusele kättesaadavaks tegemiseks Tartu Ülikooli veebikeskkonna kaudu, sealhulgas digitaalarhiivi DSpace'i kaudu kuni autoriõiguse kehtivuse tähtaja lõppemiseni.
2. olen teadlik, et punktis 1 nimetatud õigused jäävad alles ka autorile.
3. kinnitan, et lihtlitsentsi andmisega ei rikuta teiste isikute intellektuaalomandi ega isikuandmete kaitse seadusest tulenevaid õigusi.

Tartus, 19.05.2016