

ÜHENDVERBID EESTI MURRETES*

KRISTEL UIBOAED

Sissejuhatus

Püsiühendite¹ uurimisega on tegeldud palju eesti kirjakeele puhul. On uuritud püsiühendeid, sh ka ühendverbe ning nende automaatse tuvastamise võimalusi (nt Kaalep, Muischnek 2009). Keelekorpuste areng on muutnud võimalikuks keele uurimise suuremahulise materjali põhjal, aga samal ajal tekitanud vajaduse välja töötada erinevaid meetodeid, mis võimaldaksid just uurijat huvitavat materjali suurest tekstikogust kätte saada. Alates 1998. aastast on koostatud Tartu Ülikooli ja Eesti Keele Instituudi koostöös eesti murrete korpus,² mis on teinud võimalikuks korpusuuringud ka eesti murrete puhul. Murdeuurimused on siiani peaaesjalikult keskendunud sõnavara, sõnatuletuse, foneetika ja fonoloogia ning mõnevõrra sõnajärje uurimisele (vt täpsemalt Pajusalu 2002). Eesti murrete püsiühendeid pole seni uuritud. Artiklis tulebki juttu eesti murrete ühendverbidest ning nende automaatse tuvastamise võimalusest eesti murrete korpusse tekstide põhjal. Lisaks on see töö esmakordne suurem katse rakendada eesti murrete uurimisel arvutilingvistilisi meetodeid. Eesmärgiks oli välja töötada ühendverbide automaatse tuvastamise meetod murdetekstide jaoks ning sama meetodit kasutades leida kõigis murretes esinevad ühendverbid, mille põhjal omakorda oleks võimalik leida erinevusi ning sarnasusi nii murrete lõikes kui ka kirjakeelega võrrelduna. Kuigi eesmärgiks polnud ideaalse osalausestaja (s.t osalause piiride automaatse tuvastaja) väljatöötamine, eeldas ühendverbide automaatse tuvastamine siiski mõningast teksti eeltöötlemist, seega moodustas osalausestamismeetodi väljatöötamine kirjeldatavast uurimistööst üsna arvestatava osa.

Artiklis olen ühendverbi mõistet käsitanud väga laialt. Murretevahelisi sünonüüme käsitlen sagedusloendites erinevate ühendverbidena. Kindlaid mitteühendverbe olen püüdnud analüüsimaterjalist eemaldada (nt verbiga *olema* moodustunud ühendid ning mõningad modaalverbidega moodustunud ühendid). Samuti olen välja jätnud iseseisva tähendusega adverbi sisaldavad ühendid.

Ühendverbid ja neile iseloomulikke jooni

EKG järgi on ühendverbid perifrastilised verbid, mis koosnevad finiiitverbist ning afiksaaladverbist, mille üks komponent käitub sarnaselt afiksile ning muudab finiiitse verbi tähenduse keerulisemaks (EKG II: 19–20).

* Artikli valmimist on toetanud ETF-i grant nr 7464 ja riikliku programmi „Eesti keele keeletehnoloogiline tugi 2006” projekt EKKTT06-14.

¹ Püsiühend on kahe või enama sõna või sõnavormi ühend, mida mingi tähenduse väljendamiseks kasutatakse tavaliselt koos. Püsiühendi mõiste on väga lai, kuid selles artiklis tuleb juttu ainult ühendverbidest.

² <http://www.murre.ut.ee/triip/murdekorpus/>.

Huno Rätsep esitab „Eesti keele lihtlause tüüpides” eelnevalt tehtud uuri-
muste põhjal lisaks ortograafilisele veel kolm kriteeriumi, mille abil võiks eris-
tada ühendverbe muudest verbiühenditest.

Morfoloogilise või konstruktsioonilise kriteeriumi puhul
toetutakse väitele, et verbid koosnevad verbist ja afiksaaladverbist (Rätsep
1978: 26; vt ka Eesti keele süntaks 1964: 83; Rätsep 1969).

Süntakiline kriteerium tähendab, et ühendverbi mõlemad osad
kuuluvad lauses öeldise juurde (Eesti keele süntaks 1964: 83).

Semantilise kriteeriumi põhjal on ühendverbidel tähendus-
varjund, mida samal verbil ilma afiksaaladverbita poleks. Afiksaaladverb võib
anda verbile tavalisest erineva tähenduse, muuta verbi tegevuse lõplikuks
või anda sellele teatava suuna (*välja tulema, ette võtma, vastu tulema*) (Muuk
1938).

Produktiivsuse põhjal võib ühendverbid jagada kahte rühma. Osa afiksaal-
adverbe moodustavad seerialisi sõnaühendeid tervete verbirühmadega, s.t
üks afiksaaladverb võib kombineeruda paljude verbidega ning sellised ühend-
verbid moodustavad semantiliselt lähedasi verbide rühmi. Samuti võib teata-
va verbirühma iga liige koos esineda ühendverbina selle rühma kõigi adver-
bidega (Rätsep 1978: 28). Murdekorpuses esinevad näiteks *ees/ette/läbi/rin-
gi/taga/tagasi/välja/üle jooksmata/sõitma/tulema*. Teine rühm ühendverbe
on ainukordsed sõnaühendid, kus niisugune varieerumine pole võimalik: nad
esinevad koos ainult ühes teatavas väljendis (Rätsep 1978: 28). Näiteks: *juur-
de arvama, peale käima, välja tulema*.

Artiklis käsitlen ühendverbe ainult konstruktsioonilist kriteeriumi aluseks
võttes. Väljatöötatud automaatne meetod ei võimalda arvesse võtta süntak-
sit, seega pole võimalik automaatselt kindlaks teha, kas vastav ühend on lau-
se süntakiline kese või mitte.

Materjal

Artikli materjal pärineb eesti murrete korpusest³ (kokku umbes 400 000 teks-
tisõna 2008. aasta 2. veebruari seisuga). Eesti murrete korpus on elektrooni-
line andmebaas, mis sisaldab autentseid murdetekste kõikidest eesti murre-
test. Murdekorpus põhineb murdelindistusel, millest vanemad on pärit aast-
tast 1938, kuid suurem osa on lindistatud 1960.–1970. aastatel. Töö aluseks
on morfoloogiliselt märgendatud tekstid, mis on XML-formaadis.⁴ Märgenda-
miseks on kasutatud abiprogrammi Mark (vt Lindström 2009).

Murdetekstide automaatsest osalausestamisest

Ühendverbide leidmiseks oli eelnevalt vajalik murdetekstide osalausestami-
ne,⁵ et oleks võimalik leida ühendverbe ühe osalause piirides. Kui osalause
piire ei ole määratud, paneb programm ühendverbid kokku terves lauses (või
voorus) leiduvatest verbidest ja adverbidest, aga üks lause (või kõnevoor) võib

³ <http://www.murre.ut.ee/triip/murdekorpus/>.

⁴ XML ehk laiendatav märgistuskeel (ingl *Extensible Markup Language*) on lihtne, väga
pandlik tekstiformaat, mis võimaldab esitada infot teatud struktureeritud kujul, nii et tekst
on arvuti jaoks n-ö arusaadav, vt <http://www.w3.org/XML/>.

⁵ Edaspidi kasutan sünonüümselt mõisteid *lausestamine* ja *osalausestamine*.



koosneda mitmest osalausest, see omakorda muudaks tulemused ebausaldusväärseks.

Osalausestamine on ühendverbide leidmiseks palju olulisem kui muude verbikesksete püsiühendite tuvastamiseks, sest ühendverbe moodustavad sõnad on keeles sagedased, näiteks afiksaaladverbid võivad tekstis esineda ka kaassõnadena. Väljendverbi koosseisu kuuluvad käändsõnavormid pole nii sagedased (Muischnek, Kaalep 2006).

Kui ühendverbide leidmiseks kasutada ainult ühte kõnevooru (lausestamata tekstijuppi), mis sageli tähendab väga pikka tekstiosa, oleks tulemuseks liiga palju ühendverbide seisukohalt valesid ühendeid. Näites 1 on esitatud üks kõnevoor, kus esinevad ühendverbid *ära minema* ja *kokku kiskuma*.⁶

- (1) noohh (.) ega (.) eks ta ikke võtt kua puol `pääva (.) kui ol'i pikk ka~ngas `ikke puole pääva **läks ära** (...) vaest ol'i (.) viel kalevi ka~ngas sie `kangeste (.) ol'i paha (...) `kiarida (.) tema `akkas neid `kiar+buid `kiskuma nagu (...) villane `ikke (.) **kisub `kokku** (...) jahh (.) aga siss (...) jahh (...) mina olen ikke kot'ti=ka~ngast ja=ja labast ja kõikke olen (.) paelu teind (...) neid `kangaid (...)

Kui nüüd proovida leida ühendverbe sel viisil, et esmalt moodustada kaheliikmelised sõnakombinatsioonid kõigi kõnevooru sõnadega ning sellest võtta edasiseks analüüsiks ainult adverbi ja verbi ühendid (mida võimaldab korpuse morfoloogiline märgendus, sellest näitest on morfoloogilised märgendid loetavuse huvides eemaldatud), saaksime tulemuseks ka ühendid, nagu *ära võtma*, *ära minema*, *kokku võtma*, *kokku minema*, *kokku kiskuma*, *ära kiskuma*, *kokku tegema* jne, kuigi õiged on ainult kaks ühendverbi *ära minema* ja *kokku kiskuma*.

Lausepiiride määramiseks seadsin esmalt piirid erinevate kõnevoorude, pikkade pauside, side- ning küsisõnade järgi.⁷ Näide 2 illustreerib, kuidas programm tekitas lausepiirid küsi- ja sidesõnade järele (lausepiiri tähistab kaldkriips).

- (2) tema käis siss kool'is noorelt ja / pärast ta ol'i kirjanik no võe / mina änam läbi taga käesin ni+ba t
ma tsi / kuna ta suri
seda `tiadis / millas sa `sün'dind `ol'lid

Kui eelnevatest reeglitest lausestamiseks ei piisanud, võtsin aluseks verbide finiiivormid. Enne seda lausestatud järjenditest otsis programm verbi pöördelist vormi, ning kui „osalausest“ oli neid rohkem kui üks, tekitas programm lausepiiri iga teise, kolmanda jne finiiivverbi ette. Järgmises näites pole ühtegi side- ega küsisõna, samuti puudub pikka pausi tähistav (...), seega tuleb lausepiir seada finiiivverbi järgi:

⁶ Lihtsustatud transkriptsioonis kasutatavad märgid: (.) – lühike paus (kuni 0,2 sekundit); ` – III vältel sõnade ees graavis, v.a 1-silbilistes sõnades; ' – palatalisatsioon; ~ – ng, nk, nt *kangas*; * – kolmveerandpikkus, märgitud sõna ees; (...) – pikk paus; + – liitsõna; = – kokkuhääldus.

⁷ Murdekorpused on ka siduvad küsisõnad märgendatud lihtsalt küsisõnadeks.





(3) ta Palalt ärä läks läks As'sikvere `küllä `Puuspa `Jakpille sii

Rakendades finiiitverbi reeglit, seab programm osalause piiri lausestamata lõigu teise finiiitverbi ette ning tulemuseks on lausestatud tekst näites 4 esitatud kujul:

(4) 1) ta Palalt ärä läks

2) läks As'sikvere `küllä `Puuspa `Jakpille sii

Finiitverbi järgi lausepiiri seades võib mõne analüütilise verbivormi üks osa jääda eelmisesse osalauseesse, seetõttu korrastasin järgnevalt automaatselt veel selliseid osalauseid, nagu on viiendas näites.

(5) 1) ei sa `kõn'dida ei

2) saa se jala pale `toettatta

Esitatud näites jääb analüütilise verbivormi *ei saa* esimene osa eelmisesse osalauseesse, sest lausepiir on tekitatud lausestamata tekstiosa teise finiiitverbi ette, pärast osalause te korrastamist saab korrektse tulemuse, mis on esitatud näites 6.

(6) ei sa `kõn'dida

ei saa se jala pale `toettatta

Lõpptulemuseks on täielikult osalausestatud tekst, mis teeb võimalikuks ühendverbide automaatse tuvastamise ühe osalause piirides.

Näites 1 esitatud kõnevooru lausestamise tulemuseks on üheksa osalause:

(7) noohh (.) ega (.) eks ta ikke võtt kua puol `pääva (.)

kui ol'i pikk ka~ngas

`ikke puole pääva **läks ära** (...)

vaest ol'i (.) viel kalevi ka~ngas

sie `kangeste (.) ol'i paha (...) `kiarida (.)

tema `akkas neid `kiar+buid `kiskuma

nagu (...) villane `ikke (.) **kisub `kokku** (...)

jahh (.) aga siss (...) jahh (...) mina olen ikke kot'ti=ka~ngast ja=ja labast

ja kõikke olen (.) paelu teind (...) neid `kangaid (...)

Lausestamise järel moodustatakse ühendverbid ühe osalause piirides ning tulemuseks on ainult ühendverbid *ära minema* ja *kokku kiskuma*. Moodustuvad veel ühendid *kangesti kerima* neljandas osalause ning *palju tege-ma* üheksandas osalause, kuid need saab kergesti eemaldada stoppsõnade loendit kasutades (täpsemalt vt järgmisest alapeatükist).

Ühendverbide automaatse tuvastamise meetod

Nagu mainitud, on ühendverbide tuvastamise aluseks morfoloogiliselt märgendatud ning eelnevalt osalausestatud murdekorpuse tekstid. Ühendverbide leidmiseks moodustasin esmalt iga osalause piirides kaheliikmelised kom-



binatsioonid kõigi selle osalause sõnadega. Seejärel eraldasid tulemustest automaatselt verbi ja afiksaaladverbi ning verbi ja adverbi ühendid. Murdekorpuse märgendamisel ei ole täielikku ühtlust, millal sõna märgendatakse adverbiks või afiksaaladverbiks (ja sageli ongi piir väga hajus). Sellepärast ei piisanud vaid verbi ja afiksaaladverbi ühendite eraldamisest. Saadud loetelude puhul rakendasin adverbide stoppsõnade loendit, mis eemaldas sellised adverbühendid, mille adverbiline komponent reeglina ühendverbe ei moodusta, näiteks *vaikselt*, *koledasti* jmt. Veel eemaldasid ühendite loetelust kõik verbiga *olema* moodustatud ühendid. Samuti jätsin välja suurema osa modaalverbidega ühenditest, sest kontrolli käigus selgus, et enamjaolt niisugustes kombinatsioonides adverbiline komponent ühendverbi osa ei ole. Näiteks osalauses *mee võind ju mörra ede panna* moodustuksid automaatselt ühendid *ette panema* ja *ette võima*, millest viimane ei ole ühendverb. Modaalverbidega ühendid on eemaldatud käsitsi kontrolli käigus, sest modaalverbid ei ole murdekorpuses eraldi märgendatud sõnaklass.

Ühendverbide leidmiseks kasutasin sõnade algvorme kirjakeelestatud kujul, mis on teksti lisatud morfoloogilise märgendamise käigus. Samad sõnad võivad murdekorpuses esineda väga erineval kujul. Näiteks adverbi *laiali* kõik esinemisvormid murdekorpuses on: *laijali*, **laijali*, **laijale*, *laijale*, *laikka*, *lakjalõ*, *laikka*. Ühendverbi komponentide erinevad esinemiskujud on töö käigus salvestatud ning vajaduse korral saab neid kergesti kontrollida.

Töös ei ole kasutatud ühendverbide komponentide vahelise seose tugevuse mõõtmiseks nn hüpoteesi testimise meetodeid. Brigitte Krenn ja Stefan Evert on võrrelnud levinumaid sõnade omavahelise seose tuvastamise meetodeid ning jõudnud järeldusele, et ükski neist ei anna oluliselt paremaid tulemusi kui tavaline sageduse arvutamine (Krenn, Evert 2001). Seetõttu olengi piirdunud vaid sageduse leidmisega.

Ühendverbide leidmise meetodi tõhusust saab kontrollida saagise ja täpsuse väljaarvutamisega. Saagise ja täpsuse arvutamiseks koostasid testkorpuse, milles oli üks murdetekst igast murdest – kokku kümme murdeteksti.

Saagis näitab kõigi korpusest leitud ühendverbide (kõigi automaatse meetodiga tuvastatud ühendite) suhet kõigi võimalike õigete ühendverbidega (ühenditega, mis korpuses esinevad tõepoolest ühendverbidena) ning kirjeldab, kui suurt osa õigetest ühendverbidest õnnestus andmebaasist leida. Saagis jääb 0% ja 100% vahele ning null tähendab seda, et ei ole leitud ühtegi õiget ühendverbi, 100% seda, et on leitud kõik korpuses olevad ühendverbid.

Täpsus kirjeldab leitud õigete ühendverbide suhet kõigi leitud ühendverbide hulgaga ja näitab, kui suur osa leitud ühendverbidest on korrektsed ühendverbid. Täpsus jääb samuti 0% ja 100% vahele.

Testkorpusest leitud ühendverbide tulemus on esitatud tabelis 1. Tabel osutab, et kokku leiti testmaterjalist 206 ühendverbi, millest korrektsed on 160 ning 46 ei olnud ühendverbid, vaid verbi ja adverbi ühendid. Leidmata jäi kolm testmaterjalist esinenud ühendverbi. Seega osutavad leitud õiged ühendverbid ühenditele, mis esinevad tekstis tõepoolest ühendverbina, ning ka automaatne meetod tuvastas vastavad ühendid. Leidmata õigete ühendite arv osutab korpuses esinenud ühendverbidele, mida automaatne meetod ei tuvastanud. Leitud valed ühendid on ühendid, mida automaatne meetod pidas ühendverbideks, kuid mis tekstis tegelikult ühendverbidena ei esine.

Tabel 1.

Ühendverbi leidmise meetodi saagis ja täpsus

	ÕIGED	VALED	KOKKU
LEITUD	160	46	206
LEIDMATA	3	0	3
SAAGIS	98 %		
TÄPSUS	78 %		

Saagise arvutamiseks on kasutatud valemit *leitud õiged / (leitud õiged + leidmata õiged)*. Täpsus on arvatud järgnevalt: *leitud õiged / (leitud õiged + leitud valed)*.

Murrete ja kirjakeele sagedasemad ühendverbid

Eespool kirjeldatud meetod võimaldab leida ühendverbe murdekorpuses esineval kujul. Ühendverbid on (peale muude verbikesksete püsiühendite) märgendatud ka 300 000-sõnalises kirjakeele korpuses,⁸ lisaks on olemas eesti keele verbikesksete püsiühendite andmebaas, mis on kättesaadav aadressil <http://www.cl.ut.ee/ressursid/pysiyhendid/index.php?lang=et>. Need keeleressursid võimaldavad võrrelda ühendverbide esinemist murretes ja kirjakeeles.

Kokku vaatlesin 2876 erinevat ühendverbi: kõiki ühendverbe, mis esinevad uurimismaterjali põhjal nii kirjakeeles kui ka murretes. Murrete ühendverbid on leitud eespool kirjeldatud meetodiga. Kirjakeelseks võrdlusmaterjaliks on 300 000-sõnalises kirjakeele korpusest leitud ühendverbid. Korpus koosnes ilukirjandustekstide, ajalehetekstide ja populaarteaduslike tekstide sõnadest (kõiki umbes 100 000 sõna). Et korpuste mahud oleksid omavahel võrreldavad, olen kõigi ühendverbide esinemissagedused ümber arvutanud korpuse 100 000 sõna suhtes. 300 000-sõnalises korpusest ei tule välja paljud ühendverbidest, mis kirjakeeles tegelikult olemas on, seetõttu kasutasin ka eesti keele verbikesksete püsiühendite andmebaasi, et teada saada, millised murretes esinevad ühendverbid kirjakeelest puuduvad.

15 sagedasemat ühendverbi murretes ja kirjakeeles on esitatud tabelis 2. Veerg „sagedus” esitab ühendverbi esinemissageduse 100 000 sõna kohta korpuses.

Kõigepealt näeme, et murdetekstides kasutatakse sagedasemaid ühendverbe märgatavalt rohkem kui kirjakeeles: murrete kõige sagedasem *ära minema* esines 54 korral 100 000 sõna kohta, kirjakeele kõige sagedasem *vastu võtma* vaid 23 korral 100 000 sõna kohta.

Murdematerjalis esines 1589 sellist ühendverbi, mida ei olnud kirjakeelles uurimismaterjalis, ja selliseid ühendverbe, mida ei olnud kogu kirjakeele püsiühendite andmebaasis, oli tuhatkond.

Viie sagedasema hulgas on kirjakeeles ja murrete koosarvestuses ühine ainult *välja tulema*. Siiski esineb kirjakeele sagedasim *vastu võtma* suhteliselt tihti ka murretes. Kirjakeeles sagedased *ette võtma*, *kaasa tooma* on murretes küll olemas, kuid üsna madala esinemissagedusega.

⁸ <http://www.cl.ut.ee/korpused/morfkorpus/>; <http://www.cl.ut.ee/korpused/morfliides/>.

Tabel 2.

Murrete ja kirjakeele sagedasemad ühendverbid

MURDED		KIRJAKEEL	
ÜHENDVERB	SAGEDUS	ÜHENDVERB	SAGEDUS
ära minema	54	vastu võtma	23
ära võtma	50	välja tulema	19
ära saama	46	tagasi tulema	12
välja tulema	46	ette võtma	11
ära surema	40	kaasa tooma	11
peale panema	37	välja nägema	9
ära tulema	36	välja andma	8
sisse panema	35	ette kujutama	8
ära viima	33	ära võtma	8
välja minema	24	ette nägema	8
välja võtma	22	ära minema	7
ära kaduma	20	maha jätma	7
kinni panema	20	alla kirjutama	7
üles panema	19	läbi viima	7
sisse minema	19	esile kutsuma	7

Murrete sagedasimad *ära minema*, *ära võtma* esinevad suhteliselt tihti ka kirjakeeles, kuid pole siiski viie kõige sagedasema hulgas. Murretes väga sage *ära surema* esineb kirjakeelses uurimismaterjalis ainult kolm korda.

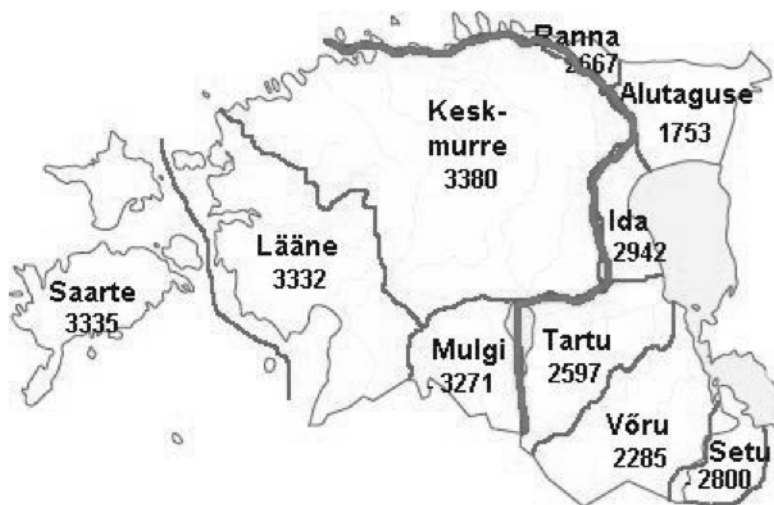
ära saama puudub kirjakeelses materjalis sootuks. Ühendverb *ära saama* esineb tavaliselt kontekstis *mingit tööd ära saama*, *kuhugi/kuskilt ära saama*, *vanaks ära saama* jne. Ühendverbiga *ära saama* tekib probleeme sageduste arvestamisel. Nimelt toodab ühendverbide leidmise meetod selliseid ühendeid rohkem, kui neid tegelikult ühendverbidena esineb. Probleem on suurem murretes, kus kasutatakse palju *saama*-passiivi (nt *töö sai ära tehtud*), mille tagajärjel moodustatakse ühendverb ka abiverbina kasutatud tegusõnaga *saama*. Et murdekorpuses on abiverbiks märgendatud ainult verbi *olema* vormid, siis verbi *saama* abiverbina või põhiverbina esinemist ei olnud lausestamismeetodit kasutades võimalik eristada. Eriti palju esineb uuritud materjali põhjal *saama*-passiivi kesk-, saarte- ja läänemurdes. Seevastu idamurdes ja Setu⁹ murdes on passiivi väljendamine sellisel kujul tunduvalt harvem. Alutaguse materjalis esines *ära saama* ainult üks kord ning see oli täiesti korrektne ühendverb. Eriti sageli esineb *ära saama* ühendverbina Tartu murdes.

Murrete sagedasemad ühendverbid

Joonisel 1 on esitatud ühendverbide kasutussagedused 100 000 sõna kohta igas murdes (murdekorpuses kasutatava murdejaotuse põhjal).

Nagu näha, on ühendverbide kasutamises suurem erinevus ida ja lääne vahel, lääne pool kasutatakse ühendverbe rohkem. Ala, kus ühendverbe kasu-

⁹ Murdekorpuses on traditsiooniliselt Võru murde üheks murrakuks peetud Setu liigitatud omaette murdeks.



Kaart 1. Eesti murrete ühendverbide esinemissagedused 100 000 sõna kohta.

tatakse vähem, moodustavad kirderanniku murderühm, idamurre ning Tartu, Võru ja Setu murre. Kõige vähem esineb analüütilist väljendusviisi ühendverbide põhjal otsustades Alutaguse murdes. Suure erinevusena torkab silma, et Mulgi murdes kasutatakse ühendverbe palju sagedamini kui teistes lõunaeesti murretes, nii et Mulgi murre sarnaneb selles osas pigem lääne, saarte ja keskmurdega. Tartu, Võru ja Setu murdest on kõige vähem analüütiline Võru murre. Kuigi saarte murdest tuli välja kõige rohkem erinevaid ühendverbe, on sageduste põhjal otsustades ühendverbide kasutamine kõige levinum keskmurdes.

Järgnev osa esitab kõige sagedasemad ühendverbid murderühmade ja murrete kaupa. Tabelites on esitatud vastava murde 15 kõige kõrgema esinemissagedusega ühendverbi ning iga ühendverbi juurde on lisatud näide, millele järgneb sulgudes kihelkond, kust see pärit on. Sagedused on ümber arvutatud korpuse 100 000 sõna kohta, et eri murrete materjal oleks omavahel võrreldav.

Põhjaeesti murderühm

Idamurre. Tabel 3 esitab 15 kõige kõrgema esinemissagedusega ühendverbi. Sagedasematest idamurde ühendverbidest ei mahu kõigi murrete koosarvestuses sagedasemate hulka *ära kuivatama* ja *ära keetma*, mis küll kõigi murrete koosarvestuses on esimese saja hulgas, kuid on 15 kõrgeima sagedusega ühendverbide hulgas vaid Tartu murdes. Idamurde suhteliselt sage *kokku laskma* ei ole sagedasemate loendis üheski teises murdes ning murrete koosarvestuses mahub see alles esimese 300 hulka. Ühendverb *kokku laskma* esineb näiteks kontekstis *loomi kokku laskma* või käsitöö kirjeldamise kontekstis (tavaliselt ketramisega seotud).

Tabel 3.

Idamurre

ÜHENDVERB	SAGEDUS	NÄIDE
ära minema	69	vot nüid lähmä ärä kodo (Kod)
ära viima	59	vei ära selle pois'ikkese `metsä jaa (Avi)
ära tulema	55	tul'ime saanast ära (Trm)
ära surema	50	vana+isä suri ära (Pal)
välja tulema	41	õrass alle õts tuleb `väl'jä (Trm)
ära võtma	37	Seppa pojalt õl'l siis `ikke mõtte seda `möldri tüttard ära `võtta (Trm)
välja minema	37	meije eij ole `väl'lä läind Kaare+bere vallast ega `kos'kil `ulkkumas (Pal)
ära kuivatama	37	`rukkid siis jällä sial partte pial kuivetteti ära (Avi)
sisse minema	32	lähme meije kaa `sisse (Trm)
sisse tulema	32	Jukko ei taha üle ukse `sauna mitte `sissegi tulla (Kod)
ära keetma	32	`keetsid jälle käbid padan ärä (Kod)
kokku laskma	32	`las'ti loomad jüri+bääva ommukku `kokku kuij=õl'i koa lumi moas aga las'ti ikke luomad `kokku (Avi)
ära saama	27	sain viina ära (Trm)
välja võtma	27	eks me `võtnud näid `kiissu võrguss `väl'jä niid (Kod)
ära tegema	27	ta Suomenn juba õma tüü no ärä õl'i tehnu (Kod)
ära käima	27	mee käis'ime silla päl ära (Trm)
ära korjama	27	õl'i kõik juba nied põhud kõik juo ära kor'jattud (Avi)
ära pesema	27	`lambit peset'ti `lambad ärä (Kod)
välja viima	27	kui se Nurga talu `väl'lä `viidi siäl (Pal)

Keskmurre. Kõigi murrete koosarvestuses ei mahu sagedasemate hulka keskmurdes tihti esinevad ühendverbid *ringi käima*, *tagasi tulema*, *üles saama*. Sagedasemate hulgas on need veel ainult rannamurdes, *üles saama* ka Alutaguse murdes ja *tagasi tulema* Võru murdes. Kõigi murrete koosarvestuses on need siiski suhteliselt sagedased ning kõik kolm mahuvad esimese 50 hulka (vt tabelit 4). Keskmurdes kasutatakse ühendverbe kõige rohkem (vt kaarti 1).

Läänemurre. Läänemurde sagedasemad ühendverbid on sagedased ka kõigi murrete koosarvestuses. Läänemurde suhteliselt sage *ära tooma* on sage ka kõigi murrete koosarvestuses, kuid sagedasimate hulgas esineb ainult veel Saarte murdes (vt tabelit 5).

Saarte murre. Saarte murde sagedasimatest ühendverbidest pole kõigi murrete koosarvestuses sagedusloendi tipus vaid *ära tooma* ja *läbi käima* (vt tabelit 6). Ühendverb *läbi käima* on suhteliselt sage kõigi murrete koosarvestuses, kuid ei esine kõige sagedasemate hulgas muudes murretes eraldi arvestatuna.

Tabel 4.

Keskmurre		
ÜHENDVERB	SAGEDUS	NÄIDE
ära võtma	63	noo siis `akkas't siis seda (.) alus+`kuube sel'last ää `võtma (VJg)
ära minema	55	kolm `pääva läks `pulmadega ära jahh (VJg)
ära saama	53	`ikke sai ära (VJg)
välja tulema	50	ühe korraga sada `uoppi tuleb `väl'la nohh (VJg)
ära surema	46	Venemale ta surigi ära kahh (VJg)
kinni panema	44	eks ik `viina norisid viina+norijad `jälle panid tied `kin'ni (Kei)
ära viima	37	noo `vargad viisid ää jälle (VMr)
välja minema	35	siit Pal'diski kala+med läksid `väl'la (HMD)
peale panema	35	`iasti kahh `senna `pan'di `rammu `piale (Pil)
sisse panema	33	rajuttud nii+baelu et paned kodara `sisse (Juu)
ära tulema	26	ss kui tul'i sie sõl'm ää (VMr)
ringi käima	26	sui `ringi käis'id `en'di `pakkidega (Pil)
üles saama	26	mina lehma alt ei saa ülesse (Pil)
tagasi tulema	26	`Ruelast tul'i tagasi jälle seije Kul'inalle tagasi (VJg)
välja võtma	24	nüüd vel sohooss sii=siss võttab oma jagu `väl'lä siit (Äks)

Lõunaeesti murderühm

Mulgi murre. Mulgi murde seitse sagedasemat ühendverbi kuuluvad kõige sagedasemate hulka kõigi murrete koosarvestuses (vt tabelit 7). Kõigi murrete koosarvestuses ei ole 15 sagedasima hulgas Mulgi murde tihti esinevatest ühendverbidest *ära nägema*, *ära käima*, *ära andma*, *ära sööma*, *kinni võtma*, *vastu tulema*, *ära tapma*. Teistest Mulgi murde sagedasematest ühendverbidest esineb idamurde sagedusloendi tipus *ära käima*. Ülejäänud, mis pole kõigi murrete koosarvestuses sagedasemad, pole ka muude murrete sagedusloendite esimeses osas. Ühe suurema erinevusena tuleb Mulgi murdes välja sagedasimate ühendverbide rohkearvulisem kasutamine (üle 100 korra 10 000 sõna kohta) võrreldes muude murretega ning ühendverbide tunduvalt kõrgem sagedus teiste lõunaeesti murretega võrreldes (vt ka kaarti 1), mis näitab, et Mulgi murdes kasutatakse analüütilist väljendusviisi ühendverbide kasutussageduste põhjal otsustades rohkem kui muudes sama murderühma murretes.

Tartu murre. Tartu murde sagedasim *ära koolema* (*ta latsõkkõnõ `kuuli ärä*) ei esine mujal kui lõunaeesti murretes (v.a Mulgi murre). Setu ja Võru murdes on see ühendverb alles vastavalt kuuendal ja viiendal kohal. Erinev kõigi murrete koosarvestuses on ühendverbi *ära elama* (*no ärä ole ma nüüd elänu seeni+aani*) kõrge sagedus Tartu murdes (vt tabelit 8). Rohkem esineb see ühendverb vaid Mulgi ja keskmurdes. Saarte, Setu ja Võru murdes on ühendverb küll olemas, kuid on väga madala esinemissagedusega.

Tabel 5.

Läänemurre

ÜHENDVERB	SAGEDUS	NÄIDE
välja tulema	86	lebad tulid ilosad `välja (Mih)
ära võtma	83	külm siss `võttis kõik marja+õied ära (Var)
ära minema	70	sii ku te ää läksit muul mettu `asja `meelis (Mih)
sisse panema	65	mõrrad need `pandi madalamale `sisse (Var)
peale panema	59	siis `pandi see kiut+kört siija `piäle (Var)
ära viima	59	ärä viib `jälle pisiksed mustad talu+lapsed nüid ära (Rid)
ära kaduma	51	siis kadus see ovuste `vargus `rohkem ära (Mih)
ära saama	46	takkust need said jämedad ned ää (Var)
ära tooma	43	`rohkem obu i jõun ää `tuuvagi (Var)
ette panema	38	siis taa pann kaks `musta `ärga `ette (Mih)
välja minema	38	`luht+ruudist lähme `väl'la (Han)
ära surema	35	alt `preester `surri ära (Han)
maha laskma	32	`jälle ärräle `kaibama et Künnisso vana Ants lasn luige maha (Var)
ära tulema	30	viiendamal `aastal (.) tulime säält ära Kiide+ma `mõisa (Rid)
läbi minema	22	selle talu talu maea asemelt lähäb tee läbi (Mih)
maha kukkuma	22	kui äkkeste teine tüttar+laps `kukkund maha (Mih)
maha lööma	22	ussi maha löön selle mürgitse ussi (Var)
ära teadma	22	meste+`rahvad ja laeva+mehed `tiäväd ärä jõe+obused (Var)
üles tõmbama	22	kus=see jalg=oli ammusted (.) `tõmmas selle naha tükki ülessi=ja (.) `tõmmas `nuaga selle (.) tükki seest `väl'la (Mih)

Võru murre. Võru murdes suhteliselt sage *ära ostma* (vt tabelit 9) on mõnevõrra kõrgema sagedusega veel ranna- ja Tartu murdes. Teistes murretes on see madala esinemissagedusega või puudub täielikult. Võru murde sage *ära eksima* on kõigi murrete koosarvestuses sageduselt teise saja hulgas. Paljudes murretes puudub see ühendverb sootuks ja üheski teises murdes pole see eriti sage.

Setu murre. Ühisosa on idamurdega (*ära keetma, ära pesema*). Ühendverb *ära koolesma* esineb sagedasemate hulgas lõunaeesti murretes, v.a Mulgi murdes, kus see ühendverb puudub täielikult. Suurema erinevusena torkab silma ühendverbi *ära tapma* väga kõrge esinemissagedus. Kõige sagedasemate hulgas on see veel vaid Mulgi murdes ning muudes murretes on see pigem madala esinemissagedusega ühendverb. Selline kõrge koht kõigi murrete koosarvestuses tulebki ennekõike kõrge sageduse tõttu Mulgi ja Setu murdes. Kõigi murrete koosarvestuses jäävad alles teise saja hulka Setu murde sagedased ühendverbid *peale, ära hõõruma, ära jahtuma* (vt tabelit 10).

Tabel 6.

Saarte murre

ÜHENDVERB	SAGEDUS	NÄIDE
ära saama	76	varasemini `ikka saunast sai ää (Khk)
ära võtma	73	see võttab ulga `aega ää (Käi)
ära tulema	64	`Peetter oli siis `jälle Kihelgonnalt ära tulnd (Khk)
välja tulema	57	tulid juba juba need vabrigu võrgud `välja (Khk)
peale panema	57	oleks ussi verd `peale `panna oln siis põleks midagid oln (Muh)
sisse panema	52	siss silgu- silgu dükkid said kaa `sisse `pandud (Phl)
ära surema	41	siiss ema Emmi ema suri kolme+gümne esimesel `aastal ära (Khk)
ära tooma	38	ära `toodi põllalt `ikka `korma (Phl)
ära viima	35	siis `viidi kohe se leib lava pealt ää (Muh)
läbi käima	35	nee `ollid neli viis+kümmend `pulma juba läbi köin (Muh)
ära kaduma	35	`Atla lahess siit jää ära kadus (Khk)
üles panema	35	see sari `pandi nii+`moodi üles ülese kusagil+`poole (Phl)
välja võtma	34	naised käisid ikka abis `lesti `välja `võtmas (Khk)
ära minema	31	üks kolm neli `päeva `ikka läks ära (Phl)
sisse minema	31	sisse `mindud valõ `võtmõttõgä uksõd olid `lukkis (Khn)

Tabel 7.

Mulgi murre

ÜHENDVERB	SAGEDUS	NÄIDE
ära surema	114	nüid suress küll ää meelege juba ärä (Hls)
ära minema	111	nüit tõise lääve puha `sanna ärä (Krk)
välja tulema	65	temä ol `Retsnigu `niitte pääle madale pääle `väl`lä tullu (Trv)
ära võtma	52	`tahten tõist sõss ärä `võtta (Krk)
peale panema	49	taa `panden säält ilusti `pääle (Krk)
ära saama	49	nevä ei saa mud`u ärä et pane obesed `ette (Hls)
välja minema	49	lännu kõrtsi manu `vällä sõnna lähpeppess (Krk)
ära viima	46	Lunna Mär`t `viidi Siberise ärä (Hls)
ära nägema	46	sedä Vinkläri poiga sedä näi mea kahh ärä (Krk)
ära käima	42	Arnu käis kesk+kool`in ärä (Trv)
sisse minema	39	suur paing läits sisse sõss (Trv)
ära andma	36	kõlvanu oma poja iist äräde soldattiss anda (Krk)
ära sööma	33	asi puha ik sedäsi `õhtu+kosti ärr sõönü (Krk)
sisse panema	29	ss pannime nooda `sissi (Trv)
ära tapma	29	miu käe omm ärä tappet (Trv)
kinni võtma	29	s võet sii Maimu `kinni (Krk)
üles panema	29	tüdrukku `pandsive sõss rehi üles (Hls)
vastu tulema	29	sõss oru viirt müüdä tullu kolme `aastene väikke tüdruguk seant `vastu (Krk)

Tabel 8.

Tartu murre

ÜHENDVERB	SAGEDUS	NÄIDE
ära koolema	87	ta latsökkõnõ `kuuli ära (Ote)
ära minema	84	viil palavik jälle nii ära läits (Rõn)
ära saama	79	et `ruttu ära saass (Kam)
ära surema	63	tõne orava+boig sures ära (Kam)
ära tegema	50	kui tu kanõpp ol'li kyik' ära `tettu (Rõn)
ära võtma	39	`vir'ttin `ütless võtta mino ära (Kam)
sisse panema	39	siss `pan'di tuu kanep'pi+jahu sinnä `sisse (Rõn)
peale panema	37	pan'nime obese `ette ja pan'ni latse `pääle (Rõn)
välja tulema	32	ma ss `tihka `väl'lä tulla (Võn)
ära tulema	26	tule ära `ruttu (Võn)
üles panema	26	nuu `pantti siss üless jäl'l' tarõ pääle (Ote)
ära elama	26	ära ole ma nüid elänu seeni+aani (Rõn)
kinni hoidma	24	tule `oija pen'i `kin'ni (Nõo)
üles tulema	21	siss tullu üits miiss üless (Rõn)
ära kuivatama	21	kanep'pi ss iloste kuivatted'i ära (Rõn)
ära lõikama	21	Mar'i `lõikka `meije pen'i ära (Nõo)
ära peksma	21	rüä ol'li siss ära pessettu (Nõo)
üles ajama	21	syss ta `aije rehele joba üless (Kam)
ühes võtma	21	ss `lõunõ+süük' `võet'ti ütten (Rõn)
sisse tulema	21	tul'li venelane `sisse (Kam)

Tabel 9.

Võru murre

ÜHENDVERB	SAGEDUS	NÄIDE
ära minema	65	neede hobõstõga ei saa `Tarttohe ärq minnãq (Plv)
ära saama	42	ku piim jäll ära `saije (Har)
peale panema	42	sääl pandass `auttul lihm `pääle (Har)
sisse panema	36	`pan't'ti mu arost noid humal'it ka `sisse (Vas)
ära koolema	29	taahh rehe saina ala Villakku per+nanõ tuu mm ärq koolnu (Vas)
välja tulema	26	`tul'l'igi `säärtse `mustri `vällä (Har)
ära viima	26	`paraegust tu tuul' kahh kurad'i kyik vei sinnaq ärq (Vas)
ära võtma	23	kohegi `tall'i `sisse mint hobenõ ära võet (Har)
tagasi tulema	23	laulattuse `aigu tagasi `tul'l'i (Har)
ära ostma	23	`lambaq kyik' ära ostõduq (Har)
ära tulema	20	nii tul'l' pin'i mu manu ärr (Vas)
kinni panema	20	siss silmäq `pan't'ti kin'n'iq (Vas)
välja minema	20	i tiijãq kos na Verolt ol'l' `väl'l'ä lännüq (Vas)
maha jääma	20	jäi noid `pääkkeise `maaha (Rãp)
maha tapma	20	naakka nüüt *tüühõ tappaq soe `maaha (Rãp)
kinni pidama	20	pei hobõst kin'n'iq (Vas)
ära eksima	20	olõ kaq paar `kõrda elun ärãq `essünü mõtsan (Har)

Tabel 10.

Setu murre

ÜHENDVERB	SAGEDUS	NÄIDE
ära minema	78	lät's'ivä siss kohegi `mõtsa `tüühü ärq (PSe)
ära tapma	68	`ark'k'i tap'p' pap'p'i tap'p' ärr (PSe)
ära võtma	58	a vot nüüd nuidõ kalos'k'idõga `võtk'i jalaq ärq (LSe)
välja tulema	49	tun'n'is't'iva esäle ka üleest tul'l' kyik' as'i aval'igust `väl'l'ä (PSe)
peale panema	49	jälq suuhh pant siss `kuurma `pääle jaaq lät's' Ees't'i+`maalõ siijä sise+`maalõ ärq (PSe)
ära koolema	49	`Tarttohõ ol'l' ärq kooluq (PSe)
välja võtma	44	ss `katlast võt't'i `väl'l'ä (LSe)
ära panema	44	sinnä panõt yks kas't'i ko ärr (LSe)
ära saama	39	nä ärq nädäl'i vanast jo saava (LSe)
sisse panema	39	pan'n'i `suk'r'it ka vähäkkene `sisse (LSe)
ära pesema	39	noo vana n'eo põllõq peses'i ärq (LSe)
peale ajama	39	ai keevä viiq `pääle salt (LSe)
ära keetma	34	ol'l' nuuq kalakkõsõ keededü ärr (LSe)
ära jahtuma	34	tumõ pruun' liim siss `jaahtu ärr (LSe)
ära hõõruma	34	kat's' `vok'k'e hõq- hõõr'i ma vanast arq (PSe)

Kirderanniku murderühm

Alutaguse murre. See murre erineb sagedasemate ühendverbide koosseisu poolest teistest murretest kõige rohkem. Murrete koosarvestuses mahuvad sagedasemate hulka ainult *ära tulema*, *välja tulema*, *ära surema*, *ära võtma*, *üles panema*. Harvem esinevad kõigi murrete koosarvestuses *kokku vedama* ja *läbi ajama*. Ühendverbid *läbi minema* ja *ette panema* esinevad sagedasemate hulgas veel läänemurdes ning *peale hakkama* rannamurdes. Ühendverbide kasutussageduse põhjal on Alutaguse murre kõige vähem analüütiline (vt tabelit 11).

Rannamurre. Sagedasim ühendverb *välja vedama* (vt tabelit 12) ei ole sagedasemate hulgas üheski teises murdes ning ka kõigi murrete koosarvestuses on see alles 77. kohal. Ühendverb *välja vedama* esineb ainult kontekstis *võrku (veest) välja vedama*. Muude murrete sagedasimate ühendverbidega on ühised *maha laskma* (läänemurre) ja *üles minema* (Võru murre).

Mõningaid tähelepanekuid

Eri murderühmade puhul ilmnes ettearvatavaid sarnasusi. Sagedasemad ühendverbid on sarnasemad Tartu, Setu ja Võru murdes. Näiteks ühendverb *ära koolema* esineb kõigis kolmes murdes viie sagedasema ühendverbi hulgas, samal ajal kui muudes murretes ei esine sellist ühendit üldse. Samatähenduslik ühendverb *ära surema* on siiski suhteliselt sage kõigis murretes, v.a Võru ja Setu, kus see küll esineb, kuid märgatavalt sagedasem on *ära koolema*.

Tabel 11.

Alutaguse murre

ÜHENDVERB	SAGEDUS	NÄIDE
ära tulema	45	*laulattamast ära tulima (Lüg)
ära võtma	39	sial onn *enne kõik põlev+givi ära *võettud (Jõh)
peale hakkama	27	*akkasin *jälle kaa *niiskese *jusku isä veri sies *ikke *kauplussega *jälle *peale (Lüg)
välja tulema	21	sialt oli *tulnud kaa siis *niisukkesi *pruunisi kivisi *välja (Jõh)
üles saama	21	ülesse ei saand ehk enamp (Jõh)
vastu võtma	21	*tohter *võttis *vasta (Jõh)
kokku saama	21	no eks me nüüd *kokko *saima (Lüg)
ette panema	21	pani *püssi *rauwad *ette (Lüg)
ära surema	18	kui ära oli surd juo (Jõh)
tagasi tulema	18	säl tuli tagasi *kuoli (Jõh)
läbi minema	18	jõgi läks Vabat+külä alt läbi (Lüg)
üles panema	18	sai sai sie peso ülesse *pandud (Lüg)
läbi saama	18	miss krun't *annab senega *piatta läbi *saama (Lüg)
läbi ajama	18	sääld aeasimma siis *luomad läbi (Lüg)
kokku vedama	18	*vihmasel aeal sai *einu *kokku vedada (Jõh)

Tabel 12.

Rannamurre

ÜHENDVERB	SAGEDUS	NÄIDE
välja vedama	55	vedasin siis e viest *vergud *väl'ja (Jõe)
ära tulema	50	tulime merelt ära (Jõe)
ära minema	44	tämal siit nied kaks *sõrme *läksivad ära (Hlj)
välja tulema	42	sie asi ei tuld *väljä (Kuu)
sisse panema	34	siis *sinna panime *ankru *sisse (Hlj)
sisse laskma	34	läks las'i *selle *sekli+*ket'ti *kiirest *sisse (Hlj)
ära võtma	31	saa *poisi+nadikkas ei *võttand *mütsi piast ära (Hlj)
välja minema	31	alt läks kol'inal *väl'ja (Jõe)
ära viima	31	siis *pukser vei *selle e sääld ära (Hlj)
ära saama	29	no kuhu mina enamb sain sääld ära (Hlj)
ära surema	26	*Kunda parun suri Kareppalle ära (Hlj)
tagasi tulema	26	sield ei *suida enamb tagasi *tulla (Jõe)
peale hakkama	26	*Suome lahest *akkab ko- Pietterbori just *piale (Jõe)
maha laskma	26	sis oleks maha *lastud *sakslane oleks *tõesti maha last (Hlj)
ringi käima	26	jää pial käid *ringi (Jõe)
edasi minema	24	paat läks purje all *täije *vaardiga edasi (Jõe)
üles minema	24	siit *läksime nii+*bal'ju ülesse *Nieme maa *puole (Jõe)
kinni pidama	24	mõni pere+mes *ikka pidas oma *kohta *kinni (Jõe)

Üllatuslikult on Mulgi sagedasemad ühendverbid sarnasemad keskmurdega kui Tartu ja Võru murdega.

Niisama üllatuslik on, et Setu murde sagedused on sarnasemad lääne- ja Mulgi murdega kui Võru murdega. Tartu murdes suhteliselt sage *ära tegema* ei ole viie sagedasema ühendverbi hulgas üheski teises murdes.

Kõige sagedasemate ühendverbide osas leidub erinevusi kõigis murretes. Sagedasim ühendverb on sama ainult Setu ja Võru murdes (*ära minema*). Teistest kõige erinevam on Alutaguse murre, kus viie kõige sagedasema ühendverbi hulgas on kõige rohkem selliseid ühendverbe, mida enamjaolt teistes murretes viie sagedasema hulgas ei ole, näiteks *peale hakkama*, *ette panema*, *kokku saama*, *vastu võtma* ja *üles saama*. Samuti ei kuulu osa selliseid ühendverbe kõigi murrete koosarvestuses sagedasemate hulka. Erinevuse põhjuseks on ilmselt asjaolu, et Alutaguse murdes moodustab produktiivne afiksaaladverb *ära* ühendverbe tunduvalt harvem kui muudes murretes.

Rannamurde sagedasim *välja vedama* pole üheski teises murdes viie sagedasema hulgas. Rannamurde sagedasemad ühendverbid erinevad enim lõunaeesti murrete ühendverbidest.

Eri murrete kohta põhjalikumate järelduste tegemiseks oleks kindlasti vaja korpuse mahtu suurendada. Kõige rohkem on korpuses esindatud saarte murre, mis kajastub ka tulemusel: enim tuli erinevaid ühendverbe välja just saarte murdes. Siiski ei saa öelda, et materjali suurus ja tulemus oleks alati üksüheses vastavuses. Kaks korpuses kõige rohkem esindatud murret (saarte ja keskmurre) toodavad küll mitmekülgseima tulemuse, kuid näiteks Alutaguse murre on korpuses tunduvalt rohkem esindatud kui idamurre, aga erinevaid ühendverbe on idamurdes rohkem, mis omakorda viitab jällegi Alutaguse murde väiksemale analüütilisusele. Mulgi murregi pole korpuses esindatuim, kuid genereeritud ühendverbide loend on suhteliselt mitmekülgne.

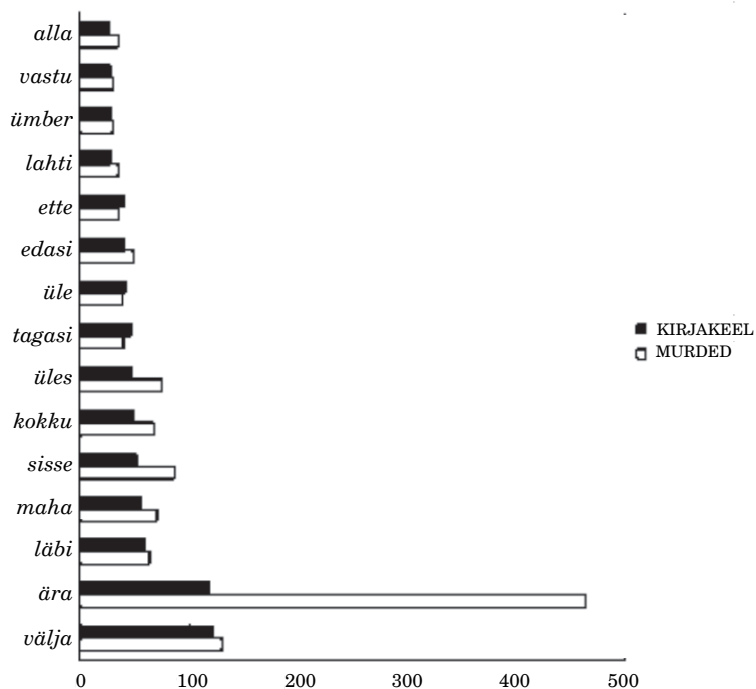
Kindlasti on kasutatud meetodiga leitud ühendverbide seas palju juhuslikke ühendeid, eriti just nende hulgas, mis esinevad loendis üks kord. Nende üle otsustamiseks oleks vaja rohkem materjali.

Kirjakeele püsiühendite andmebaasist puuduvad ühendverbid

Murdekorpusest tuvastati 1065 sellist ühendverbi, mida kirjakeele püsiühendite andmebaas ei sisaldanud. Neist 690 olid sellised, mida murdematerjalis esines ainult üks kord. Murrete sagedasemad ühendverbid, mis kirjakeele andmebaasist puuduvad, on näiteks *ära koolema*, *ära teadma*, *peale tegema*, *läbi panema*, *ära lagunema*, *ära niitma*, *ära kuduma*, *valla saama*, *ära jahvatama*, *ära leotama*, *ära hautama*, *manu panema*, *ära kasvama*, *järgi tulema*, *pealt minema*, *sees käima*, *valmis kuduma*, *ära ehitama*, *ära mõskma*, *ära tuulama*, *ära värvima*, *ülal käima* jt.

Lääne- ja saarte murdes suhteliselt sage *ära teadma* vaste võiks kirjakeeles olla lihtsalt *teadma* või harvem ka *ära arvama*. Ühendverb *ära teadma* on hea näide sellest, kuidas afiksaaladverb *ära* märgib murretes perfektiiivsust väga erinevates situatsioonides ja võib kombineeruda paljude verbidega nii, nagu kirjakeeles seda ei toimu.

Paljud kirjakeele püsiühendite andmebaasist puuduvad ja murretes esinevad ühendverbid on sellised, mille üks komponent on murdeline või kõnekeelsem, nt *manu panema*, *valla pääsema*, *tagakätt viskama*, *ära koolema*.



Joonis 1. Murrete ja kirjakeele 15 sagedasemat ühendverbe moodustavat afiksaaladverbi.

Erinevusi tuleneb ka murrete ja kirjakeele tekstide sisust. Murretes on palju selliseid ühendverbe, mis kirjakeeles oleksid täiesti aktsepteeritavad. Sellised on näiteks käsitöö ja majapidamistöodega seotud ühendverbid, mida mõjutab murdekorpuse tekstide temaatika.

Murrete ja kirjakeele ühendverbe moodustavad adverbid

Järgnevalt vaatlen täpsemalt ühendverbide koosseisu kuuluvaid sagedasemaid murrete ja kirjakeele afiksaaladverbe. 15 kõige sagedasemat ühendverbe moodustuvat afiksaaladverbi on täpselt samad nii murretes kui ka kirjakeeles. Erinev on ainult nende sagedusjärjestus (vt joonist 1).

Suurima erinevusena torkab silma afiksaaladverbi *ära* väga kõrge produktiivsus murretes. Murretes moodustab *ära* ühendverbi peaaegu 500 erineva verbiga. Kirjakeeles on *ära* samuti väga produktiivne ühendverbide moodustaja (sageduselt teine), kuid tema produktiivsus on peaaegu neli korda väiksem kui murretes. Seda võib pidada kõnekeelele omaseks nähtuseks, sest murretes väljendatakse perfektiiivsust sageli lauses topelt, see tähendab, et ka koos verbidega, mis oma olemuselt on juba perfektiiivsed (*ära kasvatama, ära lahkuma, ära poegima*).

Teiste ühendverbe moodustavate afiksaaladverbide sagedused on enamjaolt sarnased. Kirjakeeles sagedasim *välja* ja sageduselt teine *ära* on peaaegu võrdselt produktiivsed. Kirjakeeles on pisut produktiivsemad *tagasi, üle* ja *ette*, kõik teised adverbid on murretes produktiivsemad.

Murretes mitteesinevaid ühendverbe moodustavad afiksaaladverbid on ka kirjakeeles suhteliselt madala sagedusega (näiteks *esile*, *ammuli*, *sirgu*). Sama kehtib murrete afiksaaladverbide kohta, kuid siin on siiski märgata suuremaid erinevusi. Näiteks *ühes* (*ühes võtma*, *ühes tulema*), *takka* (*takka kaema*) on murretes täiesti tavalised ühendverbi komponendid, samal ajal kui kirjakeele 300 000-sõnalises uurimismaterjalis ei esinenud neid kordagi ning kogu ühendverbide andmebaasis olid need suhteliselt madala produktiivsusega.

Samuti torkab silma mõne afiksaaladverbi (näiteks *valmis*, *katki*, *täis*) keskmine produktiivsus murretes, samal ajal kui kirjakeeles on nad pigem madala produktiivsusega.

Ühendverbe moodustavad afiksaaladverbid erinevad ka murrete lõikes. Alutaguse murde sagedasemate ühendverbide erinevaima koosseisu üks põhjusi on kindlasti afiksaaladverbide erinev produktiivsus. Nimelt kasutatakse Alutaguse murdes afiksaaladverbi *ära* tunduvalt vähem kui teistes murretes. Samuti pole sagedasemate afiksaaladverbide esinemissagedus nii suure erinevusega kui muudes murretes, jaotudes enam-vähem ühtlaselt.

Kui Mulgi murre erines ühendverbide koosseisu poolest märgatavalt teistest lõunaeeesti murretest, siis sagedasemate ühendverbe moodustavate afiksaaladverbide puhul on erinevused peaaegu olematud. Mulgi murde sagedasemate afiksaaladverbide seas on suhteliselt sage *järgi*, samal ajal kui üheski teises murdes see sagedasimate hulgas pole. Selle adverbil põhjal otsustades sarnaneks Mulgi murre rohkem oma murderühma murretega, kus vastav adverb on ühendverbi moodustajana igas murdes esindatud.

Ühe suurema omapärana torkab veel silma Setu murde afiksaaladverbide koosseis. Nimelt on Setu murdes ühendverbe moodustavaid afiksaaladverbe tunduvalt vähem kui teistes murretes. Kui muudes murretes (ka teistes lõunaeeesti murretes) on sagedusloendi üleminekud suhteliselt sujuvad, siis Setu murdes on need järsud ning põhilised ühendverbide adverbilised komponendid piirduvad umbes viieteistkümne afiksaaladverbiga, mille hulgas on sagedasemad *ära*, *välja*, *peale*, *üles*, *sisse*, *maha*, *tagasi*. Eriti märgatav on sagedasima ja sageduselt teise ühendverbi suur esinemissageduse erinevus: *ära* on Setu murdes üle 7,5 korra produktiivsem kui sageduselt teine afiksaaladverb *välja*. Nii suurt erinevust ei esine üheski teises murdes (sama erinevuse põhjal järgmises, Võru murdes, erinevad kahe sagedasima afiksaaladverbi sagedused alla viie korra).

Murderühmiti on sagedasemate afiksaaladverbide koosseis suhteliselt sarnane. Erinevustena võiks välja tuua adverbid *tagasi* ja *lahti*. Kui põhjaeeesti ja kirderanniku murderühmas on *lahti* suhteliselt produktiivne ühendverbide moodustaja, siis lõunaeeesti murderühmas on see pigem madala produktiivsusega. Isegi kui arvestada *lahti* lõunaeeestilisi esinemiskujusid *valla* ja *valla*, jääks selle adverbil kasutussagedus pigem kirderanniku ja põhjaeeesti murderühma üheks jooneks ning eristaks mõlemat murderühma lõunaeeesti omast.

Afiksaaladverbi *tagasi* produktiivsus on sarnasem põhja- ja lõunaeeesti murderühmas ning kirderanniku murderühma omapäraks on selle adverbil madalam produktiivsus.

Kokkuvõte

Sagedasemate ühendverbide puhul tuli välja erinevusi kõigis murretes. Sagedasim ühendverb oli sama ainult Setu ja Võru murdes. Enim erines teistest Alutaguse murre, mille ühendverbide sagedusloendi tipus oli suurem osa selliseid ühendeid, mis kõigi murrete koosarvestuses sagedasimad polnud. Leidsin 1065 ühendverbi, mida kirjakeele püsiühendite andmebaas ei sisaldanud. Ühendverbide kasutussageduse suurim erinevus ei ilmnunud murderühmade vaid piirkondade vahel. Idapoolsed alad kasutavad analüütilist väljendusviisi tunduvalt vähem kui läänepoolsed, kusjuures analüütilisema piirkonna moodustavad lääne-, saarte, kesk- ja Mulgi murre ning selgelt vähem analüütilise ala kirderanniku murderühm ning ida-, Tartu, Võru ning Setu murre. Kõige vähem kasutatakse ühendverbe Alutaguse murdes ning kõige rohkem keskmurdes. Murrete analüütilisust on plaanis uurida ka kaassõnade osas, et näha, kas selline eristus tõepoolest eksisteerib teistegi keelenähtuste puhul.

Afiksaaladverbide kasutamise suurim erinevus on afiksaaladverbi *ära* kõrge produktiivsus murretes, ületades selle kirjakeelega võrreldes peaaegu neljakordselt. 15 kõige produktiivsemat afiksaaladverbi on samad nii murretes kui ka kirjakeeles, mõnevõrra erineb vaid nende sagedusjärjestus ning produktiivsus. Setu murre torkab silma väheste produktiivsete afiksaaladverbidega.

Töös kasutatud materjal on ühendverbide leidmiseks piisav, kuid põhjalikumate järelduste tegemiseks oleks vaja korpuse mahtu kindlasti suurendada. Lisaks korpuse mahu suurendamisele katsetan tulevikus ka sõnadevahelise seose tugevuse mõõtmist. See võimaldaks tulemusi võrrelda, mis omakorda aitaks välja selgitada, kas ehk leidub murdematerjali jaoks mõni statistiline meetod, mis töötab sel puhul paremini kui sageduste arvestamine.

Kirjandus

- Eesti keele süntaks I. Prooviartikleid lihtlause süntaksi alalt. Eesti NSV Teaduste Akadeemia Keele ja Kirjanduse Instituut. Tallinn, 1964.
- EKG II = Mati Ereht, Reet Kasik, Helle Metslang, Henno Rajandi, Kristiina Ross, Henn Saari, Kaja Tael, Silvi Vare, Eesti keele grammatika II. Süntaks. Tallinn: Eesti TA Keele ja Kirjanduse Instituut, 1993
- Kaalep, Heiki-Jaan, Muischnek, Kadri 2009. Püsiühendid arvutilingvistikas: miks ja kuidas. – Eesti rakenduslingvistika ühingu aastaraamat 5. Tallinn: Eesti Keele Sihtasutus, lk 157–172.
- Krenn, Brigitte, Evert, Stefan 2001. Can We Do Better than Frequency? A Case Study on Extracting PP-Verb Collocations. Proceedings of the ACL Workshop on Collocations. Toulouse, lk 39–46. – <http://www.cogsci.uni-osnabrueck.de/~severt/PUB/KrennEvert2001.pdf>.
- Lindström, Liina 2009. Ülevaade eesti murrete korpusest. – <http://www.murre.ut.ee>.
- Muischnek, Kadri, Kaalep, Heiki-Jaan 2006. Multi-Word Verbs in a Flective Language – the Case of Estonian. Proceedings of the EACL 2006 Workshop on Multi-word-expressions in a Multilingual Context. Trento, lk 57–64.
- Muuk, Elmar 1938. Verbide ja verbaalnoomenite kokkukirjutamisest. – Eesti Keel, nr 1, lk 4–17.

- Pajusalu, Karl 2002. Eesti murrete uurimine. – Emakeele Seltsi aastaraamat 48. Tallinn, lk 197–216.
- Rätsep, Huno 1969. Ühendverbide rektsioonstruktuuride iseärasustest eesti keeles. – Emakeele Seltsi aastaraamat 14–15. Tallinn, lk 59–77.
- Rätsep, Huno 1978. Eesti keele lihtlause tüübid. (Eesti NSV Teaduste Akadeemia Emakeele Seltsi toimetised nr 12.) Tallinn: Valgus.
- Valgma, Johannes, Remmel, Nikolai 1968. Eesti keele grammatika. Käsiraamat. Tallinn: Valgus.

Murrakud: **Avinurme**; **Hanila**; **Hargla**; **Hlj** = Haljala; **Hls** = Halliste; **HMd** = Harju-Madise; **Juuru**; **Jöelähtme**; **Jõhvi**; **Kambja**; **Keila**; **Krk** = Karksi; **Kuusalu**; **Käina**; **LSe** = Lääne-Setu; **Lüganuse**; **Mihkli**; **Muhu**; **Otepää**; **Palamuse**; **Phl** = Pühalepa; **Pilistvere**; **Plv** = Põlva; **PSe** = Põhja-Setu.

Phrasal Verbs in the Corpus of Estonian Dialects

Keywords: Estonian language, computational dialectology, dialectology, computational linguistics, multiword expressions.

The development of electronic language resources enables us to conduct more extensive studies on different linguistic phenomena, but it also requires development of tools for retrieving data of interest. Since 1998, the Corpus of Estonian Dialects (CED) has been compiled to make (so far only) recorded data available to researchers. This article describes phrasal verbs in Estonian dialects and the possibility of automatic retrieval of phrasal verbs in the CED. The purpose of the work was to develop an automatic method for extracting phrasal verbs from the CED, which would enable us to discover the differences and similarities in the usage of phrasal verbs in all dialects and to compare them with standard Estonian usage. The resulting method for automatic phrasal verb extraction achieved a 98% recall and 78% precision. The results indicate bigger differences of phrasal verb usage between the west and the east of the country. The eastern part, including the Coastal, Alutaguse, Eastern, Tartu, Setu and Võru dialects, uses a less analytical way of expression, whereas the western part, comprising the Central, Mulgi, Western and Island dialects, uses a significantly larger number of phrasal verbs. Surprisingly, the Mulgi dialect appeared more similar in this respect to the North-Estonian dialect group than to its traditional group. About 1065 phrasal verbs used in the dialects were not found in the standard Estonian database of multiword expressions. Although the current scope of the CED can be considered sufficient for the study described, enlargement of the corpus will be essential for drawing more thorough conclusions.

Kristel Uiboaed (b. 1984), doctoral student in Estonian, University of Tartu, Institute of Estonian and General Linguistics, kristel.uiboaed@ut.ee