

## Murded, varieerumine ja korpusandmed

Eitussõna paiknemine võru ja seto eitusslausetes

MAARJA-LIISA PILVIK, HELEN PLADO, LIINA LINDSTRÖM

Varieerumine on keele olemuslik osa: keel võib varieeruda nii kõnelejate vahel kui ka ühe inimese kasutuses, sõltuvalt registrist (nt teaduskeel vs. ajakirjanduskeel) ja tekstitüübist (nt uudis vs. arutus), suhtluskanalist (nt suuline vs. kirjalik suhtlus) ja suhtlusolukorrast (nt ametlik vs. mitteametlik suhtlus), suhtlejate vanusest, soost, haridusest, tegevusalast, päritolust ja omavahelistest suhetest, aga ka inimese kognitiivsetest võimetest (nagu mälu), vahetust keelelisest ümbrusest jne. Kõige tuntum on keele murdeline varieerumine: sama keelenähtus võib geograafiliselt eri punktides olla väljendatud erinevalt (nt ainsuse 1. isiku *n*-lõpuline variant *ma olin* on levinud laialdaselt põhjaeesti murretes, lõputa variant *ma oli* lõunaeesti murretes ning osaliselt Lääne-Eestis ja saartel, vt kaarti Saareste 1955: 71). Keelenähtus võib varieeruda ka murdeala sees (näiteks võib 1. isiku pöördelõpp vahel esineda, vahel mitte) ning selle kasutust võivad mõjutada erinevad tegurid. Grammatiliste, eriti süntaktiliste joonte kohta ongi leitud, et nende levikupiirid ei ole reeglina aredad, vaid toimub ulatuslik varieerumine, mis on oma olemuselt pigem tõenäosuslik kui kategoriaalne (Kortmann 2010). Selliste varieeruvate nähtuste kasutusmustrite selgitamisel on abiks sagedusandmed, mis võimaldavad mõõta varieerumise ulatust ning varieerumist mõjutavate tegurite rolli kvantitatiivsete meetoditega.

Artiklis vaatleme lähemalt üht varieeruvat nähtust – ees- ja tagaeituse kasutamist (nt *ei olõq ~ olõ-iq*) – lõunaeesti seto ja võru keelevariantides ning rakendame varieerumise analüüsi (ingl *variationist linguistics*) meetodeid, et välja selgitada eituskonstruksioonis esineva varieerumise ulatus ning seda mõjutavad tegurid vaadeldaval keelealal. Uurimus esindab korpuspõhist kvantitatiivset dialektoloogiat (vt ülevaateid Lindström, Pilvik 2018; Szmrecsanyi, Anderwald 2018): uurimuse andmed pärinevad eesti murrete korpuselt (EMK) ja loodavast seto korpuselt (SetKo) ning eitusmustrite analüüsiks kasutame kvantitatiivseid, sagedusandmetel põhinevaid meetodeid. Selle uurimuse kitsam eesmärk on kirjeldada ja seletada keeles esinevat varieerumist, ent oma andmekesksuse ja analüüsivahendite valiku poolest panustab see laiemalt digihumanitaaria valdkonda ning artiklis kirjeldatud meetodeid on võimalik edukalt kasutada ka teistes humanitaaria suundades, nt erinevate klassifitseerimisülesannete lahendamiseks. Seetõttu on meetoditel ja nende kirjeldamisel artiklis ka tavalisest keskem roll.

Artikli juhtumiuuring – eitussõna asendi varieerumine põhisõna suhtes – on huvitav nii läänemeresoome keelte uurimise perspektiivist kui ka laiemalt, sest ühel suhteliselt kitsal alal on kõrvuti kasutusel kaks sõnajärjevarianti, mis kõne kognitiivse töötlemise mõttes on väga erinevad. Tagaeitus on haruldasem nii keeletüpoloogiliselt kui ka Uurali keelkonna sees, eeseitus on aga tavaline läänemeresoome keeltes ja

naaberkeeltes ning domineerib ka keeletüpoloogiliselt (vt Krasnoukhova jt 2021). Esiteks huvitab meid, kuidas tagaeitus võru-seto keelealal on levinud ning kui järsk või sujuv on üleminek tagaeitusele eesitusele. Teiseks huvitab meid, millest sõltub ees- või tagaeituse valik alal, kus on ühtaegu kasutusel mõlemad variandid, ning kas see valik toimub võru ja seto alal sarnaselt või erinevalt.

Artikli esimeses osas anname ülevaate varieerumise analüüsi kujunemisest keeleteaduse meetodiks ning kvantitatiivsete meetodite rakendamisest selle sees. Artikli teises osas tutvustame eituse moodustamise viise seto ja võru keeles, kolmandas ja neljandas osas kirjeldame vastavalt juhtumiuuringu aluseks olevat materjali ja meetodeid. Artikli viiendas osas on esitatud tulemused eitussõna asendi varieerumise kohta võru ja seto keeles.

## 1. Varieerumise analüüsist keeleteaduses

Varieerumise analüüs keskendub ühe keelelise joone (milleks meie näites on eitussõna asukoht põhiverbi suhtes) kahe (või enama) variandi varieerumist mõjutavate piirangute ja tegurite väljaselgitamisele. Varieerumise analüüs sai alguse sotsiolingvistikast 1960-ndatel USA-s. Selle pioneeriks võib pidada William Labovit, kelle huviorbiidis oli ennekõike häälduslik varieerumine ning selle seotus kõnelejate sotsiaalsete omadustega, sh sotsiaalse klassiga. Labovi esimene vastavateemaline artikkel „The social motivation of a sound change” (1963) käsitles häälikumuutusi Martha’s Vineyardi saarel. Ta näitas, et pika aja jooksul toimuvad häälikumuutused võivad sünkrooniliselt avalduda varieerumisena sõnade häälduses ning selle varieerumise taga võivad olla sotsiaalsed tegurid, mis mõjutavad variantide valikut ja selle kaudu ka laiemalt keele muutust (Labov 1963; vt ka Hazen 2011). Labov rõhutas sotsiaalse varieerumise seisukohast olulise uurimisobjektina suulist igapäevast keelt (ingl *vernacular*), mida ta kirjeldas kui „stiili, kus oma kõnele pööratakse minimaalset tähelepanu” (Labov 1972: 208) ning mis just seetõttu annab kõige paremini infot sotsiolingvistilise varieerumise kohta ja võimaldab selle kaudu prognoosida keelemuutuse suunda. Labov on tuntud ennekõike New Yorgi väga erineva sotsiaalse ja etnilise taustaga elanikkonna keelekasutuse uurijana (vt Labov 1972).

Varieerumise analüüs algab varieeruva nähtuse, muutuja (kasutatud on ka terminit *variaabel*, ingl *variable*) tuvastamisest. Sotsiolingvistilise varieerumise analüüsi põhiprintsiip on olnud, et keelelised muutujad varieeruvad mingis keelekogukonnas süstemaatiliselt, mitte juhuslikult, ning keeleuurija ülesanne on üles leida, millest valik sõltub. Varieerumist on võimalik modelleerida – kvantitatiivseid meetodeid kasutades selgitada välja, kuidas varieerumine toimub ning mis keelelisi valikuid mõjutab (Tagliamonte 2012: 2; Walker 2013). Keelelise muutuja mõiste defineerimine oli Labovi varasemates töödes üsna lihtne, ta kirjeldas seda kui sama asja ütlemist kahel (või enamal) erineval viisil (Labov 1972: 188). Eesti keeles võib öelda näiteks *lihtsalt* või *lissalt*, *janu* või *jänu*, *teinud* või *teind*, *julgetud* või *juletud*. Sealjuures on oluline, et iga keelelise muutuja puhul oleksid k õ i k v a r i a n d i d arvesse võetud (ingl *principle of accountability*; Tagliamonte 2012: 9–10) ning et variandid

esineksid võrreldavas kontekstis ja oleksid vastastikku asendatavad. Morfoloogiliste ning süntaktiliste muutujate määratlemine on keerukam, sest vastastikune asendatavus ei ole alati nii selge (vt ülevaadet muutuja mõiste arenemisest Tagliamonte 2012: 15–19). Võib näiteks küsida, kas fraasid *mees, kes jooksis üle tee ning üle tee jooksnud mees* on tekstides alati vastastikku asendatavad (täendus ja funktsioonid ei muutu, kui asendada üks teiseга), ehkki mõlemas laiendatakse sama nimisõna samasisulise propositsiooniga.

Meetodid, kuidas varieerumist sotsiolingvistikas traditsioonis uuriti, võeti üle sotsiaalteadustest: lingvistikasse jõudis statistiline andmeanalüüs, mida rakendati selleks, et kahe samas positsioonis esineva, vastastikku asendatava keelendi valikut statistiliselt modelleerida, võttes arvesse erinevaid valikut mõjutada võivaid tegureid. Sotsiolingvistikas otsitakse neid tegureid kõnelejate sotsiaalsetest omadustest (sugu, vanus, haridus, etniline rühm, sotsiaalne klass vms), ent need võivad olla ka keelelised (nt hääliku hääldamist võib mõjutada sellele eelnev või järgnev häälik). Mida enam on varieerumise analüüs liikunud lingvistika muudesse valdkondadesse, seda enam on meetodit kasutatud pigem selleks, et välja selgitada lingvistilisi, tekstilisi, kognitiivseid vm tegureid, mis funktsionaalselt sarnaste konstruktsioonide valikut mõjutavad (vt nt ülevaadet Szmrecsanyi 2017). Näiteks eesti keele lühikese ja pika sisseütleva käände kasutuse uurimisest on ilmnenu, et pikka sisseütlevat (nt *majasse*) kasutatakse pisut teisiti kui lühikest (*majja*): pikka vormi eelistatakse pärisnimedega (nt *Tartusse* vs. *Tartu*). Samuti mängib pika ja lühikese vormi valikul rolli mitmete morfonoloogiliste ja morfosüntaktiliste tegurite koosmõju (Siiman 2018).

Aastakümneid on varieerumise analüüsi peamiseks meetodiks olnud logistiline regressioonianalüüs, mis võimaldab modelleerida, kuidas iga vaadeldav tegur mõjutab uuritava muutuja variandi valikut (Paolillo 2002: 15). Koos arvutiajastuga loodi selle hõlpsamaks kasutamiseks juba 1970-ndatel, ammu enne digihumanitaaria mõiste sündi ja kvantitatiivset pööret keeleteaduses, spetsiaalne abiprogramm VARBRUL (Cedergren, Sankoff 1974) ja hiljem selle edasiarendatud variandid, nt Goldvarb X (2005) ja Goldvarb Z (2018). Programm oli loodud spetsiaalselt keeleandmete analüüsiks ja võimaldas vajadusel andmeid mugavalt ümber kodeerida, rühmitada jne ning rakendada logistilist regressiooni, mis paljudes muudes statistikaprogrammides polnud sel ajal kättesaadav.

Selle sajandi alguses toimunud kvantitatiivne pööre keeleteaduses (vt nt Kortmann 2021; Veismann jt 2018) on oluliselt muutnud keeleteaduse olemust. Tekkinud on uued kasutuspõhise keeleteaduse (ingl *usage-based linguistics*, vt nt Diessel 2017) suunad, mis seovad keeles toimuvaid protsesse ja muutusi keeleüksuste ja kontekstide esinemissagedusega. Selle taga on arusaam, et keel ei ole oma olemuselt kategoriaalne ning keele kirjeldamiseks sobivad seetõttu pigem statistilised tendentsid (tõenäosused) (vt nt Bod jt 2003). Selle käigus on oluliselt suurenenud ka keele varieerumise uurimine kvantitatiivsete meetodite toel. Laienenud on uurimissuunad: räägitakse näiteks korpuslingvistikas varieerumise uurimisest (Szmrecsanyi 2017), mis on kasutusel kognitiivses ja laiemalt kasutuspõhises lingvistikas (vt nt Klavan, Divjak 2016; Klavan 2018). Tekkinud on uurimissuund, mida nimetatakse tõenäosuslikuks grammatikaks ning mis vaatleb keeleliste joonte varieerumist ning

väidab, et kõikidel kõnelejatel on oma tõenäosuslik keelemudel, mida pidevalt vastavalt keelelisele sisendile kohandatakse (Bresnan 2007; Bresnan, Ford 2010; Grafmiller jt 2018). Kõigi nende suundade ühisosa on keelelise varieerumise uurimine ja statistiline mudeldamine, ent vastuseid küsimusele, miks mingi nähtus varieerub, ei otsita enam niivõrd sotsiaalsetest faktoritest, kuivõrd keelelisest kontekstist, teksti tüübist, inimese kognitiivsetest omadustest (nt mälu seotud piirangutest) jne. Ka meetodite pagas on oluliselt laienenud (vt nt Tagliamonte, Baayen 2012): tavaliseks on saanud otsustuspuud, juhumeetsad, segamudelid jm mitmemõõtmelise statistika meetodid, mille kasutamiseks on mitmeid vabavaralisi ja üldisema funktsionaalsusega tarkvarasid (nt programmeerimiskeelte R ja Python paketid). Koos meetodi laienemisega on toimunud ka nihe terminoloogias. Sotsiolingvistikas traditsioonis kasutatud *muutuja* või *variaabli* asemel kasutatakse üha enam statistikast üle võetud termineid *uuritav* (*sõltuv*) *tunnus* (inglise keeles ka *response*) ning varieerumist seletavate tegurite jaoks terminid *seletav* (*sõltumatu*) *tunnus* (inglise keeles ka *predictor*), millel on omakorda väärtused, tasemed või klassid. Uuritava ja seletava tunnuse termineid kasutame ka siinse artikli metoodika kirjeldamisel ja tulemuste esitamisel, et viidata vastavalt analüüsis operatsionaliseeritud muutujale ja selle varieerumist mõjutavatele teguritele.

Kui võrrelda varieerumise analüüsi sotsiolingvistilist traditsiooni ja hilisemaid korpuslingvistikas levinud praktikaid, siis võib leida neis nii sarnasusi kui ka erinevusi. Benedikt Szmrecsanyi (2017: 687) on välja toonud üldprintsipiibid, mida järgides võib korpuslingvistilisi uurimusi pidada varieerumise analüüsi suuna esindajateks: varieerumisena käsitletakse eri viise sama asja väljendamiseks; defineeritakse muutujad ja nende variandid; tähelepanu keskmes on keeleliste valikute tegemise protsess, mitte niivõrd sagedus tekstis; kasutatakse kvantitatiivseid meetodeid ja statistilist mudeldamist, et varieerumist tingivate tegurite mõju mõõta. Kahe suuna oluliste erinevustena on seevastu välja toodud, et korpuslingvistikas on kasutusel palju mitmekesisem materjalide valik (erinevad piirkonnad, registrid, ajaperioodid jne); sotsiolingvistikas traditsioonis on esikohal häälde uurimine, korpuslingvistikas on aga võimalik varieerumist palju mitmekesisemalt uurida (leksikas, grammatikas, konstruktsioonides); olemasolevaid korpusi kasutades on sotsiolingvistilised tegurid sageli raskesti kättesaadavad ning seetõttu saavad need korpuslingvistikas vähe tähelepanu. Erinevused on ka kvantitatiivsete meetodite rakendamisel – sotsiolingvistikas domineerib logistiline regressioonanalüüs, korpuslingvistikas on meetodeid oluliselt rohkem (Szmrecsanyi 2017: 688–690).

Eestis on varieerumise analüüsi kasutatud meetodina alates 1990-ndatest. Eesti murretega seoses on sotsiolingvistilist varieerumise analüüsi ja VARBRUL-i kasutatud näiteks Karksi sõnalõpulise [e] häälde (Pajusalu 1996), võru inessiivi lõpu (-h ~ -n) vaheldumise (Pajusalu jt 1999), eesti murrete ainsuse 1. isiku pronoomeni väljajätku (Lindström jt 2009) ning võru kõnekeele varieerumiste ja muutuste uurimisel (Mets 2010). Kirjutatud keele põhjal on varieerumise analüüsi meetodiga uuritud näiteks sama või sarnase tähendusega konstruktsioonide valikut (*laua peal* või *laual*, Klavan 2012, 2021; Klavan, Veismann 2017), adessiivis või allatiivis kogeja ja selle väljajätku varieerumist (Lindström, Vihman 2017), demonstratiivpronoomenite

ja -adverbide vaheldust nimisõna täiendina (Reile jt 2020; Hint jt 2021; Taremaa jt 2021), osastava ja sisseütleva käände vormide varieerumist (Siiman 2019) jne.

Murdekorpuse põhjal on võimalik varieerumise analüüsi meetodit edukalt kasutada (nt Klavan jt 2015; Lindström, Uiboed 2017; Lindström jt 2018; Lindström jt 2021), ent enamasti mitte päris sotsiolingvistilises võtmes, sest info kõnelejate sotsiaalsete tunnuste kohta pole sageli piisav. Samuti on murdekorpuse (ja klassikalise murdeuurimise) keelejuhid üldjuhul valitud sarnastel alustel: vanemas eas kohaliku keele kõnelejad, kes on elanud pikalt samas piirkonnas, elu jooksul vähe liikunud ning kelle haridustee on olnud pigem napp (vt ka Lindström 2001). Kombineerituna uuemate andmetega on see teisalt hea võimalus modelleerida muutust või võrrelda eri piirkondade keelt. Uute andmete lisamine võimaldab keelemuutuste lähtekohana tuua enam esile sotsiolingvistilisi tegureid. Käesoleva artikli aluseks ongi nii vanemad, Tartu Ülikooli eesti murrete korpuse andmed (EMK)<sup>1</sup>, kus sotsiolingvistiline info on kohati puudulik, ning uuemad, 2010. aastatel lindistatud seto korpuse andmed (SetKo), milles kõnelejate taust on paremini dokumenteeritud ning mis võimaldavad seega eitussõna asukoha varieerumise seletamisel esile tuua ka kõnelejate sotsiolingvistilisest taustast tingitud erinevusi.

## 2. Eitumusmustrid seto ja võru keeles

Seto ja võru on lõunaeesti keelevariandid. Eesti traditsioonilises murdekäsitluses on keelekuju, mida setod kõnelevad, peetud eesti keele lõunaeesti peamurde Võru murde alla kuuluvaks Setu murrakuks (vt nt Viitso 1998: 115; Keem, Käsi 2002: 21; Pajusalu jt 2002: 52–57; Viikberg 2020) ning on väidetud, et võru ja seto peamine erinevus on mõnes häälaluslikus ja sõnavaralises joones, mis on tingitud peamiselt tugevast vene mõjust (Keem, Käsi 2002: 23; Pajusalu jt 2002: 188–189). Keelelistest erinevustest olulisemateks on peetud religioosid ning kultuurilisi erinevusi: setod on traditsiooniliselt õigeusklikud ning võrukesed luterlased ning neil on erinev identiteet (vt ka Paas 1927: 8–9; Jääts 2000: 651). Jüri Viikberg (2020: 23) on pidanud Setu murde eristamist Võru murdest, nagu seda tehakse Tartu Ülikooli eesti murrete korpuses (vt Lindström 2015), peamiselt sotsiolingvistilistel põhjustel tehtud otsuseks. Siiski on uurimused näidanud, et vähemalt võru ja seto (morfo)süntaktiliste konstruktsioonide kasutussageduses on olulisi erinevusi (vt nt Plado 2015: 199–200; Lindström, Uiboed 2017: 331–333; Lindström jt 2019: 170–174; seto keele süntaktiliste omapärade kohta vt Lindström, ilmumas).

Kui traditsiooniline murdeuurimine käsitleb mõlemat keelevarianti murdena või lausa murrakuna (seto puhul), siis keeleajalooliselt on viimastel aegadel kõneldud siiski lõunaeesti keelest (vt nt Kallio 2007, 2014), millest tänapäeval on kõige elujõulisemad võru ja seto keel. Ka kõnelejad ise eelistavad kõnelda keelest, mitte murdest (Eichenbaum, Pajusalu 2001: 484). Käesolevas käsitluses kasutatakse lihtsuse huvides mõlema keelevariandi kohta terminit *keel*, diskuteerimata murde ja

<sup>1</sup> Kasutajaliides on kättesaadav aadressil <http://www.murre.ut.ee/mkweb> (2. VII 2021).

keele vaheliste suhete üle. Samal põhjusel kasutame artiklis väikest algustähte (*seto*, *võru*).

Käesolevas artiklis rakendatakse sagedusandmeid ning varieerumise analüüsi võimaldavaid statistilisi meetodeid ees- ja tagaeituse leviku ning kasutustingimuste väljaselgitamiseks. Eeseituses paikneb eitussõna põhiverbi ees (*ei olõq*), tagaeituses põhiverbi järel (*olõ-õi(q)*). Tagaeitus on võru-seto keelealale iseloomulik ja välja-paistev morfosüntaktiline joon (Iva 2002: 582–587, 2007: 102) ning eesti kirjakeelest selgelt eristuva joonena on see võetud kasutusele ka võru kirjakeeles (näide 1).

(1) *Ma tiiä-ei külh üttegi umaiälist inemist, kiä olõ-õi koolin üten klassiga tiatrin käänü.* (Uma Leht 25. III 2021)

Keeleajalooliselt on alguurali keelde rekonstrueeritud eitusverb, millele lisan-dusid aja-, kõneviisi- ja isikutunnused (Janhunen 1982). Eitusverb on siiani olemas paljudes tänapäeval Uurali keelkonda kuuluvates keeltes (vt Miestamo jt 2015: 17). Läänemeresoome keeltes on eituse moodustamisel toimunud hulgaliselt muutusi, ent enamikus on eitussõnal säilinud mingi verbi omadus (Miestamo jt 2015: 17), kõige tüüpilisemalt isikutunnus (nt sm *en lue* 'ma ei loe', *et lue* 'sa ei loe' jne). Põhja-eestil põhinevas eesti kirjakeeles on eitussõnal kadunud kõik verbi omadused ning vana eituskonstruktsiooni meenutab vaid spetsiaalne konnegatiivne verbivorm ole-vikus (*ma/sa/ta ei loe*), mineviku moodustamiseks kasutatakse mineviku kesksõna (*ma ei lugenud*). Lõunaeestis (sh võrus ja setos) on aga eitussõnal säilinud aja mar-keerimine: olevikus toimib eitussõnana *ei* ja mineviku eitussõnana *es*, põhiverb on spetsiaalses konnegatiivses vormis nii olevikus kui ka minevikus. Samuti võib mine-vikuvormides esineda nii ees- kui ka tagaeitus (vastavalt 2a ja 2b).

(2a) *ma/sa/tä/mi/ti/nä es süüq*

(2b) *ma/sa/tä/mi/ti/nä süü es ~ süü-üs ~ süü-s*

Eesti (ja kogu uurali) keelealale on üldiselt iseloomulik eeseitus. Tagaeitus on ka maailma keeltes oluliselt haruldasem nähtus kui eeseitus. Eeseitust esineb u 52%-l maailma keeltest, tagaeitust aga u 28%-l (Dryer 2013). Eitussõna tendents paikneda verbi ees on keeletüpoloogias tuntud printsiip, millel arvatakse olevat kognitiivne/psühholoogiline põhjus: kuna eitussõna muudab radikaalselt sõnumi sisu, annab selle paiknemine enne verbi kuulajale selgema aluse järgneva sõnumi tõlgendamiseks, ning aitab vältida segadust sõnumi mõistmisel (Krasnoukhova jt 2021).

Seto-võru keelealal on tagaeitust seostatud enam idapoolse võru keelega, mille hulka on traditsiooniliselt loetud ka seto keelt (Lindström 1997: 144; Keem, Käsi 2002: 53; Iva 2007: 103). Karl Pajusalu jt (2002: 189) ning Paul Hagu (Hagu, Pajusalu 2020: 19) peavad seto keelele omaseks just tagaeitust. Sama on näidanud ka Liina Lindströmi jt (2021) statistiline uurimus idaseto keelepruugi kohta. Läänevõru keeles esineb tagaeitus aga pigem harva.

Võru ja seto keeles moodustatakse standardeitus eitussõna *ei* (olevikus) või *es* (minevikus) ja nn konnegatiivse verbivormi abil. Eeseituse konnegatiivne verbi-

vorm koosneb verbitüvest ja sellele liituvast larüngaalklusiilist (3), mis võib fonoloogilistel põhjustel jääda välja hääldamata või mis võib kokku sulada järgneva sõna alguskonsonandiga (larüngaalklusiili hääldamise kohta võro keeles vt nt Iva 2013: 106–107). Tagaeituse konnegatiivne verbivorm on verbitüvi (4). Võru ja seto keeles on eitusvorm kõikides isikutes sama. Tagaeituse puhul võib eitussõna olla kas verbist eraldiseisev (4a) või sulada eelneva verbiga kokku (4b–4c).

- (3) *ma/sa/tä/mi/ti/nä ei süüq*  
 (4a) *ma/sa/tä/mi/ti/nä süü eiq*  
 (4b) *ma/sa/tä/mi/ti/nä süü-üiq*  
 (4c) *ma/sa/tä/mi/ti/nä süü-iq*

Peale ees- ja tagaeituse võib nii võru kui ka seto keeles kasutada topelteitust, kus esimene eitussõna paikneb vahetult kas konnegatiivse verbivormi ees (5a, 5b) või järel (5c) ja teine eitussõna eitusfookuse lõpus (Lindström 1997: 150–153). Seto materjalist leidsime ka mõne topelteituse kasutusjuhu, kus mõlemad eitussõnad paiknevad enne verbi (5d), mille puhul pole selge, kas tegemist on suulise keele töötlemise eripärast tingitud erandlike kasutusjuhtudega või süsteemse kasutusviisiga.

- (5a) *eij = olõ = ma säänest tennüq ei* (Vas)  
 (5b) *ma=s soovigi ess avaldaq* (ISe)  
 (5c) *Vinnemaal üldse olõ-i säänest eiq* (ISe)  
 (5d) *`näime hobõst = õss `kätte = ss saaq* (PSe)

Ka imperatiivis võib kasutada nii ees-, taga- kui ka topelteitust, kuid ees- ja tagaeituses kasutatakse erinevat eitussõna: eeseituses eitussõna *är(q) ~ ar(q)* (*är võtkuq*) ja tagaeituses eitussõna *ei(q)*, mis võib eelneva verbiga kokku sulada (*võtku eiq ~ võtku-üiq ~ võtku-iq*). Nii ees- kui ka tagaeituses on verbil sufiks *-gu(q)/-ku(q)*. Topelteituses kasutatakse mõlemat eitussõna (*är võtku-iq*). Imperatiivis on topelteituse kasutamine harv, vähemalt idasetos (Lindström jt 2021).

Varasemas artiklis, mille fookuses oli ees- ja tagaeituse varieerumine idaseto alal (Lindström jt 2021), leidsime mõningate kihelkondade andmetele tuginedes, et tagaeituse kasutussagedus langeb järsult võru-seto keeleala piiril. Siinses artiklis täpsustame ees- ja tagaeituse vahekorda seto ja võru alal ning modelleerime võimalikku varieerumist. Samuti laiendame vaadeldavat geograafilist piirkonda kogu seto-võru keelealale ning võrdleme eituskonstruktsiooni valikut tingivaid tegureid eri piirkondades.

### 3. Materjal

Uurimuses võrdleme nii vanemaid korpusandmeid, milles on info kõnelejate sotsiolingvistilise tausta kohta lünklik, kui ka uuemaid andmeid, mis võimaldavad analüüsi tõlgendamisel arvestada ka sotsiaalsete faktoritega.

Vanem seto materjal (põhjaseto, lõunaseto, idaseto) on pärit eesti murrete korpusest (EMK). Seto alalt pärit intervjuud on lindistatud aastatel 1964–1996. Eitusvorme otsisime morfoloogiliselt märgendatud tekstidest. Põhjaseto materjalist kaasasime analüüsi 487, lõunasetost 707 ning idasetost 345 eituse kasutusjuhtu.

Uuem seto materjal on pärit ainult idaseto alalt, mis jääb praeguse Vene Föderatsiooni Petseri rajooni territooriumile. Materjalid on kogutud välitööde käigus aastatel 2010–2013 ning neid hoitakse Tartu Ülikooli eesti murrete ja sugulaskeelte arhiivis (Murdearhiiv). Need on ka loodava seto korpuse (SetKo) osaks. Intervjuud viidi läbi vabas vormis ning need toimusid seto keeles. Analüüsis oleme kasutanud materjali kaheksalt kõnelejalt. Uurimuse andmestiku korjasime intervjuude transkriptsioonidest käsitsi, kokku kaasasime analüüsi 1083 eituse kasutusjuhtu.

Võru materjal on samuti pärit EMK-st ning on salvestatud ajavahemikus 1960–1993. Eitusvorme otsisime morfoloogiliselt märgendatud tekstidest. Uurimuse andmestikus on kaheksa kihelkonna andmed: Rõuge (280 eituse kasutusjuhtu), Vastseliina (438), Urvaste (41), Räpina (193), Põlva (589), Hargla (382), Karula (574) ja Kanepi (255). Eitusjuhtude arv murrakuti erineb sõltuvalt sellest, kui palju teksti on igast piirkonnast EMK-sse kaasatud.

## 4. Meetod

Nagu eelnevalt selgitatud, on käesoleval juhtumiuringul kaks põhilist eesmärki. Esiteks huvitab meid see, kuidas on tüpoloogiliselt haruldane tagaeitus võru-seto keelealal levinud. Selleks võrdleme ees- ja tagaeituse ning topelteiluse suhtelisi kasutusagedusi korpuste materjalides ning visualiseerime tagaeituse levikut sageduskaardil, et saada parem ülevaade sellest, kas üleminekud eeseituse kasutamisel tagaeituse eelistamisele on pigem järsud või sujuvad.

Teiseks keskendume nende piirkondade keelele, kus esineb arvestataval määral ees- ja tagaeituse varieerumist, ning selgitame välja, millised tegurid sel juhul eitusõna paiknemist enim mõjutavad. Kuivõrd võru ja seto andmed erinevad oluliselt ees- ja tagaeituse proportsioonide poolest, huvitab meid, kas varieerumist mõjutavad tegurid neis kahes keelevariantis toimivad sarnaselt või erinevalt. Analüüsi kaasatud tegurite ehk seletavate tunnuste nimekiri on esitatud alapeatükis 5.2.1.

Kasutame logistilise regressiooni segamudeleid (vt nt Pinheiro, Bates 2000; Baayen 2008; Baayen jt 2008; Gries 2015, 2021; Winter 2020), et kontrollida, milline roll on eri teguritel võru ja seto keelevariantides. Regressioonimudel püüab ennustada uuritava nähtuse ehk ees- ja tagaeituse kasutamise tõenäosusi matemaatilise lineaarfunktsiooni kaudu, mis võtab korruga arvesse kõikide seletavate tunnuste kombineeritud mõju, ent lubab samal ajal hinnata iga tunnuse muutumise mõju eraldi, hoides teiste seletavate tunnuste mõju samal ajal fikseerituna. See võimaldab pea katselaadsetes tingimustes testida iga üksiku tunnuse olulisust ja mõju suunda muutuja variantide ehk uuritava tunnuse väärtuste ennustamisel, kui tegelike katseandmete asemel on kasutada hoopis korpuste vaatlusandmed (Hinrichs, Szmrecsanyi 2007: 459). Näiteks saame regressioonimudeli abil testida, kas idasetos ees-



eituse kasutamise tõenäosus kasvab või kahaneb, kui vaatleme oleviku eitussõna *ei* asemel hoopis mineviku eitussõna *es*, ent hoiame ülejäänud konteksti (nt grammatilise isiku ja eelmise eituse vormi) samana. Regressioonimudeli eesmärk on niisiis tuvastada suure hulga tegeliku keelekasutuse näidete põhjal kontekste, milles ees- või tagaeituse valiku tõenäosused oluliselt erinevad.

Segamudel on regressioonimudeli edasiarendus, mis võimaldab ees- ja tagaeituse valikut tingivate tunnuste mõju (ehk fikseeritud mõjude) hindamisel arvestada ka nn juhuslike mõjudega. Juhuslikeks mõjudeks nimetatakse selliseid tunnuseid, mille üksikud väärtused valimis esindavad juhuslikku valikut suuremast populatsioonist. Näiteks meie korpusvalimites esindatud kõnelejad on juhuslik hulk kõikidest seto ja võru keele kõnelejatest ning eitatud tegusõnad on (võrdlemisi) juhuslik hulk kõikidest võimalikest tegusõnadest, mida lauses võiks eitada. Seevastu näiteks eituskonstruktsiooni eitussõna leksikaalse kuju väärtused valimist ei sõltu: ei ole tõenäoline, et leiaksime mõnest teisest juhuslikust valimist peale *ei* ja *es* eitussõna ka mingeid teisi lekseme. Seetõttu saame selle mõju kontrollida erinevates valimites ühtmoodi. Juhuslike mõjude kaasamine võimaldab varieerumise seletamisel kaudselt arvestada aga teguritega, mille täpset olemust on raske määrata või operatsionaliseerida, ning samal ajal leevendada statistiliste mudelite olulist eeldust vaatluste ehk siinses uurimuses üksikute korpustes esinevate eitusjuhtude sõltumatuse kohta.

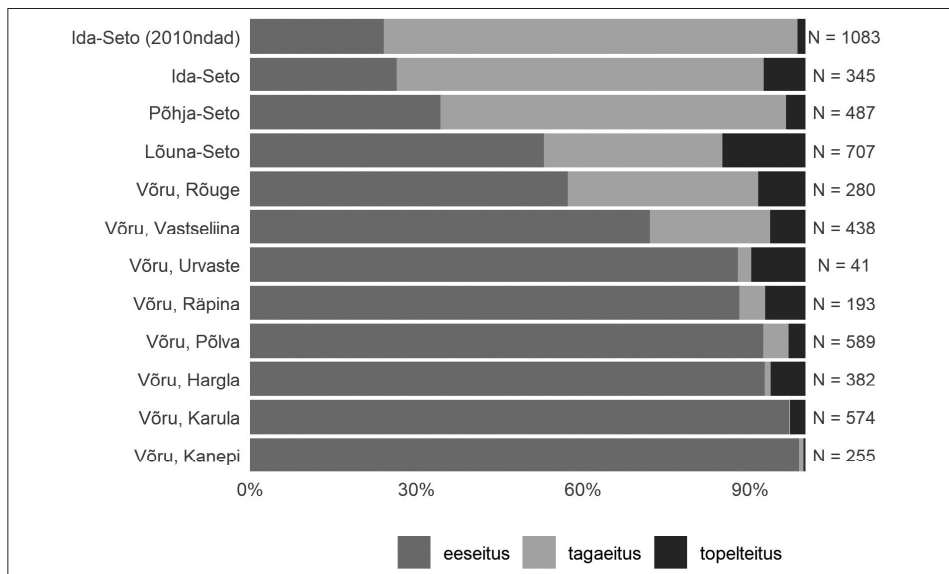
Materjali analüüsisime rakendustarkvaraga R (R Core Team 2020). Mudelite tegemiseks kasutasime R-i paketti *lme4* (vt Bates jt 2015), tulemuste visualiseerimiseks pakette *sjPlot*, *ggeffects* (vt Lüdecke 2018) ja *ggplot2* (vt Wickham 2016), sageduskaardi jaoks ka ruumiandmete analüüsi paketti *sf* (vt Pebesma 2018). Kasutatud andmestikud ja analüüsi skriptid on avalikult kättesaadavad Open Science Frameworki (OSF) platvormil (Pilvik jt 2021).

## 5. Tulemused

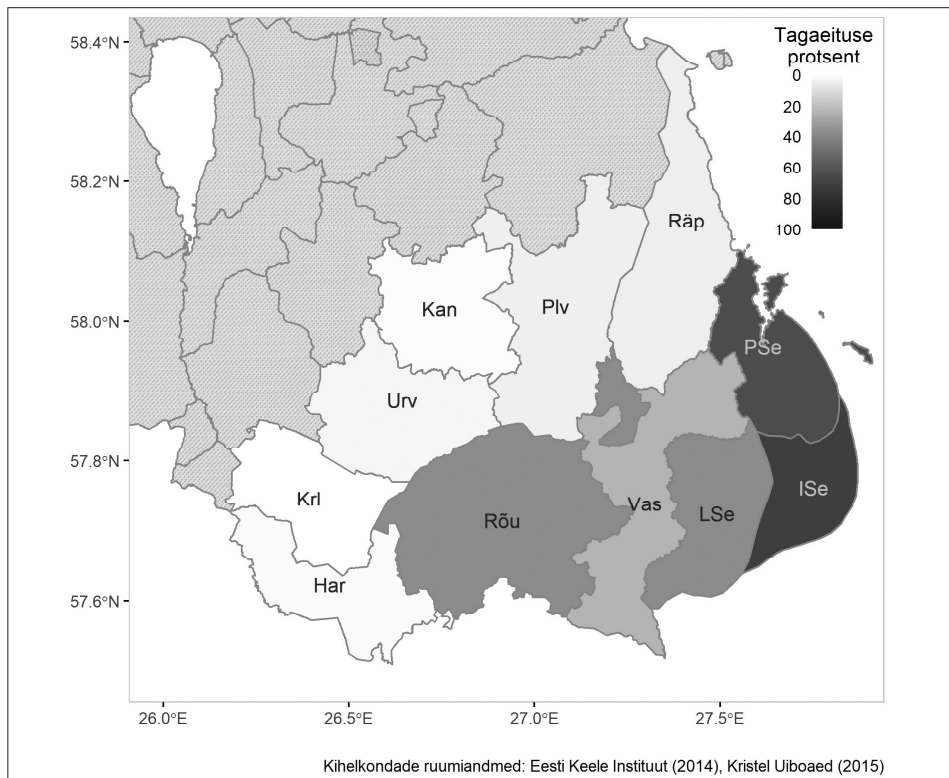
### 5.1. Tagaeituse kasutamine võru-seto keelealal: sagedusandmed

Esmalt vaatleme suhteliste kasutussageduste põhjal, mis on ees-, taga- ja topelt-eituse vahekorrad võru-seto keelealal (joonis 1). Tagaeituse osakaalu illustreerib ka sageduskaart (joonis 2), mis sisaldab vaid EMK andmeid, st välja on jäetud idaseto 2010-ndate andmestik.

Nii joonis 1 kui ka joonis 2 osutavad selgelt, et tagaeitus on ennekõike seto keelele iseloomulik joon, mida esineb vanemas materjalis, st EMK põhjal enim idasetos (66%) ning põhjasetos (62%), kusjuures idasetos on tagaeituse osakaal aja jooksul koguni mõnevõrra kasvanud (SetKo 2010-ndate andmestikus 74%). Tagaeituse kasutamine väheneb idast läände liikudes hüppeliselt, olles marginaalne Karula (0,2%), Kanepi (0,8%), Hargla murrakutes (1%) ning väga väike ka Urvas-tes (2,4%), Põlvas (4,6%) ja Räpinas (4,7%). Üleminekuala seto ja võru vahel moodustavad lõunaseto ala (32%), Vastseliina (22%) ning Rõuge (34%). Seega võib kinnitada, et tagaeitus on iseloomulik ennekõike seto keelele ja teatud määral levi-



**Joonis 1.** Ees-, taga- ja topeliteituse osakaalud võru ja seto keelealadel (eitusjuhtude koguarv N on esitatud tulpadest paremal). Andmed EMK-st, idaseto 2010-ndate andmed SetKo-st.



**Joonis 2.** Tagaeituse leviku osakaal protsentides seto-võru keelealale jäävates kihelkondades. Andmed EMK-st.

nud ka Võrumaa kaguosas. Üllatav on tagaeituse üsna kõrge kasutussagedus Rõuges, mis ei piirne (vähemalt tänapäeval) otseselt seto alaga. Sarnasel põhjusel üllatab ka tagaeituse madal kasutussagedus Räpinas, mis külgneb põhjaseto kõnelemise alaga.

Joonise 1 põhjal on võimalik vaadelda ka topeltheituse kasutussagedust kogu keelealal. Selle kõrgeim kasutussagedus on lõunasetos (15%), millele järgnevad Urvaste (9,8%), Rõuge (8,6%), idaseto (vanemad andmed, 7,5%), Räpina (7,3%), Vastseliina (6,4%), Hargla (6,3%). Idasetos on topeltheituse juhtumeid tänapäeva andmetes (SetKo) oluliselt vähem (2%) ning nende kasutus on vähem süstemaatiline kui EMK andmetes (vt täpsemalt Lindström jt 2021).

## 5.2. Ees- ja tagaeituse varieerumise analüüs

Sagedusandmete põhjal valisime edasiseks analüüsiks välja kaks ala: seto keelt esindavad idaseto 2010-ndate andmed (mida analüüsisime ka artiklis Lindström jt 2021) ning võru keelt Vastseliina ja Rõuge EMK andmed, mille aluseks on põhiliselt 1960.–1970. aastatel tehtud lindistused. Idaseto uuema materjali kohta on olemas ka põhjalikud sotsiolingvistilised andmed, mis toetavad tulemuste tõlgendamist. Valisime Rõuge ja Vastseliina, kuna neis domineerib eeseitus, ent varieerumist esineb piisaval määral (vt joonist 1), et selle analüüs oleks mõttekas ning samade analüüsivahenditega ka võimalik.

### 5.2.1. Tunnused

Analüüsis seletame eitussõna asendit (ees- või tagaeitus) viie tunnuse kaudu, mis kõik iseloomustavad üht aspekti kontekstist, kus ees- või tagaeitus esineb. Seletama oleme valinud põhiliselt grammatilised tunnused: EITUSSÕNA leksikaalne kuju väljendab eituse ajavormi (olevikus *ei*, minevikus *es*); TOPELTEITUS väljendab seda, kas tegemist on topeltheitusega või mitte; SUBJEKT näitab seda, millises vormis on lause alus ehk see, kelle tegevust või kogemist eitatakse; ISIK väljendab seda, milline on kõnealuse olendi või asja osalus kõnesituatsioonis (kas *mina/meie*, *sina/teie*, *tema/nemad* või kasutatakse hoopis umbisikulist tegumoodi). Peale grammatiliste tunnuste hindame ka ühe mälu seotud nähtuse, nn tunnetusliku või struktuuralse praimimise mõju eitussõna paiknemisele (vt Szmrecsanyi 2005; Torres Cacoullós, Travis 2019: 674; eesti keeles nt Tulving 2002). Praimimine tähendab tendentsi korraldada suhtluses aktiveeritud struktuure ning siinses uurimuses operatsionaliseerime seda kõnehetkele viimati eelnenud eituskonstruksiooni sõnajärje kaudu (EELMINE). Nagu eespool öeldud, arvestame mudelites ühtlasi sellega, et kõnelejad võivad oma individuaalse keeletunnetuse ja -kogemuse tõttu kontekstist sõltumata kalduda kasutama pigem ees- või tagaeitust (KÕNELEJA) ning et mingite konkreetsete tegusõnade korral võidakse eituskonstruksioonis eelistada pigem üht või teist sõnajärge (LEMMA). Analüüsi kaasatud seletavad tunnused on esitatud tabelis 1.

Esialguses seletavate tunnustega kodeeritud uuemas idaseto andmestikus oli 1083 vaatlust ning võru andmestikus 718 vaatlust. Jätsime andmestike puhastamise käigus välja juhud, kus aluse ehk subjekti vormi või isikut ei olnud võimalik määrata,

**Tabel 1.** Logistilise regressiooni segamudelisse kaasatud eitussõna asukoha varieerumist seletavad tunnused.

Tunnus	Väärtused	Näited
EITUSSÕNA	<i>ei</i> <i>es</i>	<i>ma ütles=i midägi</i> <i>tä ütles=ess midägi mullõ</i>
EELMINE	<i>ees</i> <i>taga</i>	<i>ja härrä iss tii es's' mitte midägi [es arvõsta] üttegi midägi</i> <i>kuq `miildü = üss siss (.) [anda =es] tap- `viinagiq</i>
TOPELTEITUS	<i>ei</i> <i>jah</i>	<i>millegiberäst ei olõ näid</i> <i>ei olõ vist säänest eiq</i>
SUBJEKT	<i>ei</i> <i>muu</i> <i>nom</i> <i>prt</i>	<i>tiijä=ei tuud joht</i> <i>miil=ei olõq (...) muijalõ tettä=ku `taahha</i> <i>pühä+bäävä es=tii pere+miiss ka es's' tüüd</i> <i>puu+ `luidsaga `süüd'i (.) es=olek= `kahvlit</i>
ISIK	1 2 3 <i>imps</i>	<i>muud ma=ei mõistaq tuust pulma asäst selettäq</i> <i>egä tii=ei taha vast nüüd tsuidsutta</i> <i>ron'giq es käüq viil</i> <i>ja nüüd=ei pandak'k'i noid `her'n'it inämp</i>
KÕNELEJA	N <sub>võru</sub> = 12 N <sub>idaseto</sub> = 8	
LEMMA	N <sub>võru</sub> = 17 N <sub>idaseto</sub> = 22	

ning laused, mis asusid lindistuste algul või pärast katkestusi ning seega polnud võimalik tuvastada vahetult eelneva eituse vormi. Muudel juhtudel märgendasime viimati kasutatud eituskonstruksiooni sõnajärje, olenemata sellest, kui kaugel see kõnehetke suhtes paiknes. Samuti otsustasime jätta välja käskiva kõneviisi juhud (nt *ära `jäl'g'i `jätku `perrä, küsügü=üi*), mida esines väga harva.

Artiklis Lindström jt (2021) mudeldasime sama idaseto andmestikku, kaasates eraldi seletava semantilise tunnuseka verbitüübi (nt kas tegusõna on olemisverb, kognitsiooniverb, liikumisverb vm), ent ei kaasanud tegusõna ennast juhuliku mõjuna. Leidsime, et eitussõna (*ei, es*) ning verbitüübi vaheline koosmõju aitas eitussõna asendit küllalt hästi ennustada: kognitsiooniverbidega oli eesituse esinemine oluliselt tõenäolisem, kui eitussõna oli *ei* ja ajavorm seega olevik (nt *ei tiijäq, ei mälehtä*). Selle põhjustena nägime kognitsiooniverbide suurt rolli parajasti käimasoleva diskursuse vahendajatena ning neid sisaldavate eituskonstruksioonide sagedamat kinnistumist fikseeritud sõnajärjega kõneüksustena. Viimastel on kontaktsituatsioonis suurem ülevõtmise tõenäosus. Teisisõnu võivad tihedamad kontaktid eesti või vene keelega tingida ka eesti või vene keele pärase sõnajärjega eituskonstruksioonide sagedama kasutamise. Selles artiklis toimime vastupidi, kaasates juhuliku mõjuna tegusõna lemma, ent jättes fikseeritud mõjuna välja verbitüübi, mis klassifitseeriks semantiliselt sama tüüpi, ent erineva esinemissagedusega tegusõnad samasse kategooriasse ning uuriks semantilise klassi kui terviku mõju eitussõna paiknemisele. Semantilise klassi asemel verbilemmade kasutamiseks pidime aga andmestikust välja jätma tegusõnad, mida esines väga vähe, vastasel juhul oleks mudelil konkreetse tegusõna ees- või tagaeituse lembust raske hinnata (vt Gries 2021). Seega eemaldasime mõlema keele vaatluste hulgast laused, milles eitatud tegusõna esines andmestikus vähem kui 5 korda. Idaseto andmestikku jäi

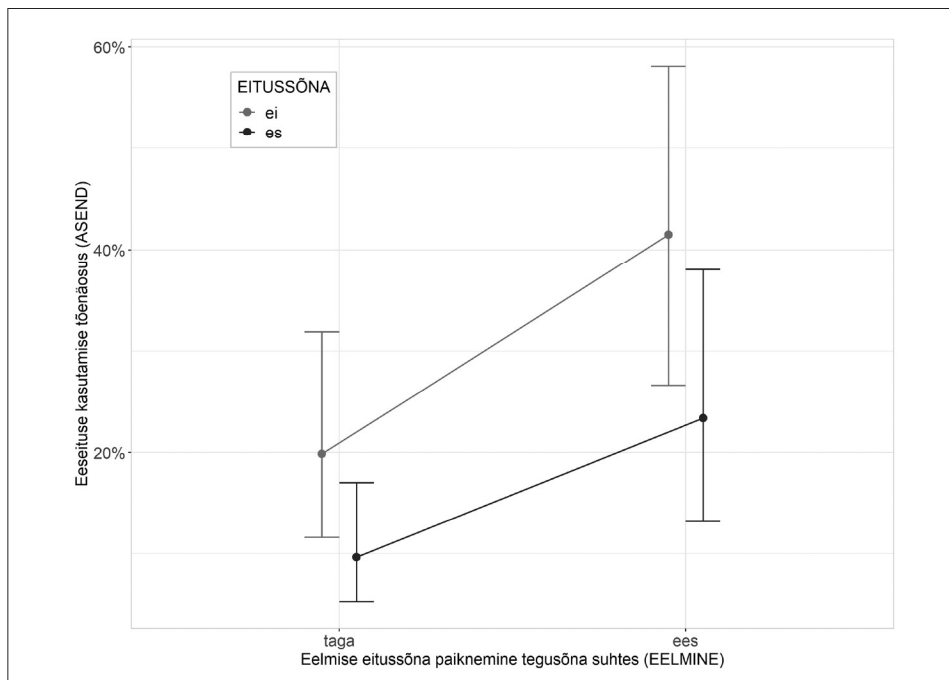
nõnda 922 vaatlust (nendest 25% ees- ja 75% tagaeitusega) ning võru andmestikku 569 vaatlust (nendest 70% ees- ja 30% tagaeitusega). Mõlemas andmestikus lasime mudelil ennustada vähem levinud eituskonstruksiooni esinemise tõenäosust, st idaseto keeles eeseitust ja võru keeles tagaeitust.

### 5.2.2. Eitussõna paiknemist seletavad tunnused idaseto keeles

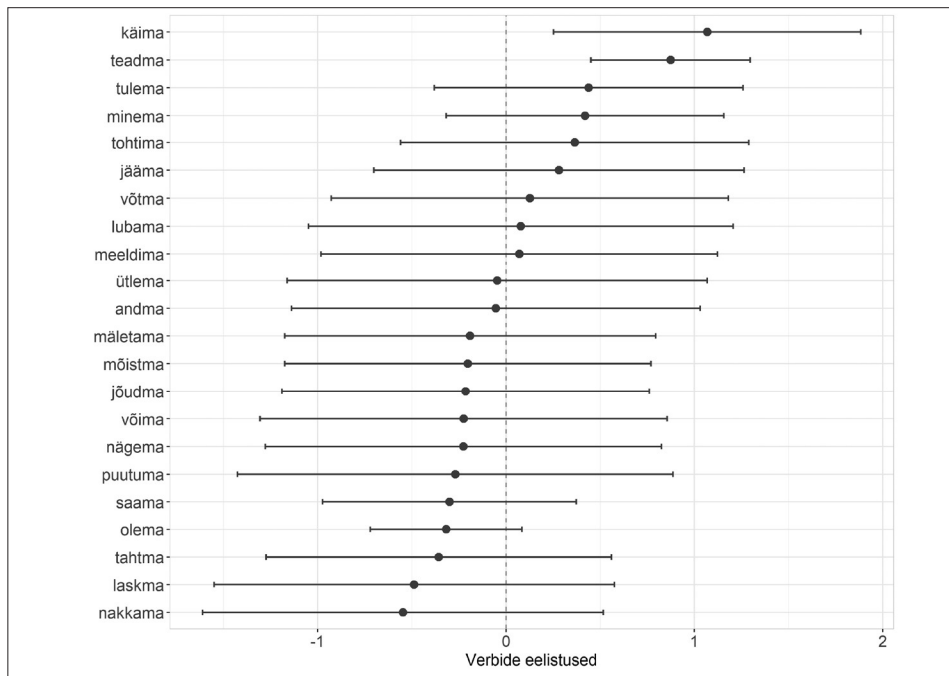
Selleks et selgitada välja, millised eeltoodud tunnustest eitussõna paiknemist tingivad, alustasime kõige lihtsamatest, üht seletavat tunnust sisaldavatest regressioonimudelitest ning liikusime järjest keerukamate kombinatsioonideni, lisades mudelisse ükshaaval seletavaid tunnuseid ning testides ka nendevahelisi koosmõjusid. Erinevate mudelite võrdlemiseks kasutasime tõepärasuhte testi (ingl *likelihood ratio test*, vt Pinheiro, Bates 2000: 83; Gries 2021). Idaseto andmestikus jõudsime optimaalse mudelini, kus eitussõna asendit seletab kõige paremini see, milline oli diskursuses viimati kasutatud eitussõna asenditega suhtes, ning see, kas räägitakse olevikust või minevikust (oleviku eitussõnaga *ei* või mineviku eitussõnaga *es*). Lisaks arvestab mudel ka juhusliku kõneleja- ning verbispetsiifilise varieerumisega.<sup>2</sup> Muud, grammatilised tunnused oluliseks ei osutunud. Varieerumist seletavate tunnuste fikseeritud mõju kujutab joonis 3.

Eitussõna paiknemist mõjutavad tegurid idasetos on tihedalt seotud mälust, suhtlusolukorrast ja suhtlemise sagedusest tulenevate tendentsidega ning keelesisesed, grammatilised tegurid jäävad siinjuures tagaplaanile. Jooniselt 3 näeme, et üldiselt domineerib idasetos tagaeitus ning eeseituse kasutamise tõenäosus jääb igas vaadeldud kontekstis alla 50%. Kui kõnehetkele eelnenud eituskonstruksioon (EELMINE), ükskõik kas öeldud kõneleja enda või intervjuueerija poolt, on olnud samuti tagaeitusega, siis on tagaeituse uuesti kasutamise tõenäosus veelgi suurem ning eeseituse kasutamise tõenäosus seega väike (olenevalt eitussõnast kas 10% või 20%). Kui aga eelnev eituvorm on olnud eeseitus, siis kasvab ka eeseituse uuesti kasutamise tõenäosus (olenevalt eitussõnast on see kas 23% või 42%). Oluline roll eitussõna asukoha ennustamisel on mudelis niisiis ka eitussõnal endal (*ei* või *es*). Eelneva eituskonstruksiooni sõnajärjest olenemata on setole vähem omase eeseituse kasutamise tõenäosus suurem, kui eitussõnaks on *ei*. Seda võib seletada ühelt poolt sellega, et oleviku eitussõna moodustab sagedamini teatud verbidega suhtluses kinnistunud ühendeid (nt *ma=i tiijä*), mida kontaktsituatsioonis võidakse lähtekeele sõnajärjega üle võtta. Teisalt kasutataksegi teatud verbe suhtluses palju sagedamini olevikus kui teisi ning sel juhul peegelduvad eitussõna mõjus ka verbispetsiifilised eelistused. Verbide individuaalseid eelistusi (eitussõnast sõltumata) näitab joonis 4.

<sup>2</sup> Lõpliku, optimaalse mudeli kuju oli  $ASEND \sim EELMINE + EITUSSÕNA + (1|KÕNELEJA) + (1|LEMMA)$ . Mudeli konkordantsiindeks  $C$  (vt Hosmer jt 2013: 173–182) oli 0,82, mis viitab mudeli heale võimele ees- ja tagaeituse kasutamise kontekste eristada (ideaalset eristusvõimet näitaks  $C$ -indeks väärtusega 1).



**Joonis 3.** Eeseituse kasutamise tõenäosus idaseto andmestikus vastavalt eelmise eitussõna paiknemisele ja eitussõna leksikaalsele kujule.



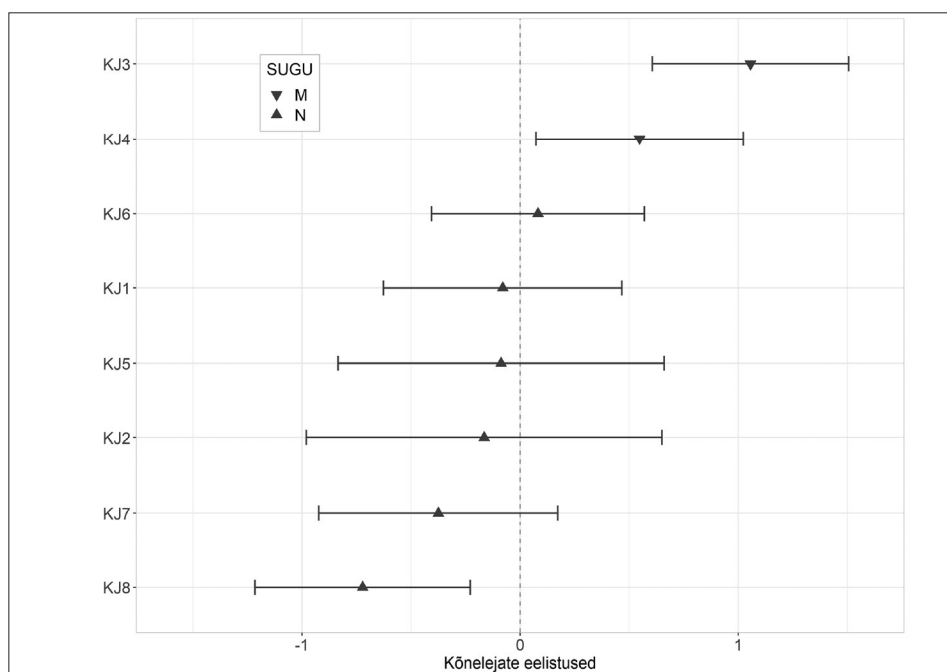
**Joonis 4.** Idaseto verbilemmade individuaalsed eelistused ees- ja tagaeituse kasutamisel.

Positiivsed väärtused joonisel 4 näitavad, et tegusõnal on tendents esineda eesitusega keskmisest rohkem, negatiivsed väärtused jällegi, et tegusõna esineb keskmisest enam tagaeitusega.<sup>3</sup> Mida suurem on punkthinnangute 95% usaldusvahemik (joonisel 4 märgitud n-ö vurrudega), seda vähem kindlust mingi tegusõna eelistuse puhul on (üldjuhul seetõttu, et tegusõna esineb harva). Kui usaldusvahemik ei kata nulli, võib olla üsna kindel, et tegusõna käitub teistest erinevalt. Sellised tegusõnad on idasetos *käima* ja *teadma*, mis esinevad teistest tegusõnadest sagedamini eesitusega (*ei/es käü, ei/es tiijäq*). Tegusõnad *nakkama*, *laskma*, *tahtma* ja *olema* aga kalduvad esinema keskmisest enam tagaeitusega (nt *olõ-õiq, taha-ass*), ehkki tegusõnade punkthinnangute usaldusvahemik on lai ja katab ka nulli ning seetõttu ei saa hinnangutes populatsiooni kohta väga kindel olla. Üksikuid tegusõnu eraldi vaadeldes näeme ka, et tegusõnad ei käitu eitussõna asukoha määramisel semantilise klassi poolest sugugi ühtselt. Näiteks kognitsiooniverbidel *teadma*, *meeldima*, *mäletama*, *mõistma*, *nägema* ja *tahtma* on väga erinevad eelistused ning kognitsiooniverbide klass tervikuna näib eelistavat eesitust (vt Lindström jt 2021) pelgalt seetõttu, et eesituselembene *teadma* on neist ülekaalukalt kõige sagedasem ja esineb eitussõnaga tõenäoliselt suhtluses võrdlemisi kinnistunud järjendis. Ühelt poolt võidakse suhtluses niisiis võtta kontaktkeeltest üle fikseeritud sõnajärjega ühendeid, teisalt aga võivad mingid sagedamad järjendid kinnistuda ka keelesiseselt (nt *es käü, olõ-õiq*), kusjuures siin võib olulist rolli mängida verbitüve morfofonoloogiline vorm.

Segamudel arvestab asjaoluga, et andmestikus esinevad eituse juhud võivad ka kõnelejati koonduda: kuigi kogukonna ühist, jagatud keelt võivad suunata mingid keelelised ja keelevälised tegurid, räägib iga kõneleja ikka pisut isemoodi. Nõnda võib mõni kõneleja kasutada rohkem eesitust, teine tagaeitust. Neid eelistusi võib omakorda seostada näiteks kodukeele, meediatarbimise, hariduse, suhtlusringkonna jm sotsiaalsete teguritega. Seda eriti seto puhul, kus tihedad kontaktid vene keele kõnelejate, vene kultuuri ja religiooniga on võru keelega võrreldes põhiliseks fonoloogiliste, leksikaalsete ja ka süntaktiliste erinevuste allikaks (vt ptk 2). Samuti mõjutavad eriti Venemaa alal elavaid idaseto kõnelejaid üha kahanevad võimalused seto keelt igapäevaselt kasutada.

Kui vaatame idaseto kõnelejate eelistusi joonisel 5, siis näeme, et valimi kaks ainsat meessoost kõnelejat (KJ3 ja KJ4) kalduvad kasutama oma kõnes eesitust teistest oluliselt enam. Siiski ei ole põhjust arvata, et meessoost kõnelejad üldiselt eelistavad eesitust, vaid tõenäoliselt on tegemist sotsiaalsete tegurite mõjuga: neil on olnud elu jooksul tihedad kontaktid Eestiga (on töötanud Eestis või külastanud sageli Eestit) ning nad on abielus venelannadega (vt Lindström jt 2021). Võib seega oletada, et nende meeskõnelejate keeles on teiste keelte (vene, eesti) mõju.

<sup>3</sup> Kuna logistilise regressiooni mudelid ennustavad vaikumisi uuritava tunnuse ühe väärtuse esinemise (antud juhul eesituse kasutamise) šansse logaritmitud skaalal, näitab  $x$ -telje skaala joonisel muutust šansi logaritmis. Kui väärtus on positiivne, siis suureneb konkreetse tegusõnaga eesituse kasutamise logaritmitud šanss ning seega ka šanss ja tõenäosus. Kui väärtus on negatiivne, siis tuleb jällegi eesituse kasutamise (logaritmitud) šansse ja tõenäosust kohandada tagaeituse kasuks.



Joonis 5. Idaseto kõnelejate ees- ja tagaeituse kasutamise eelistused.

### 5.2.3. Eitussõna paiknemist seletavad tunnused võru keeles

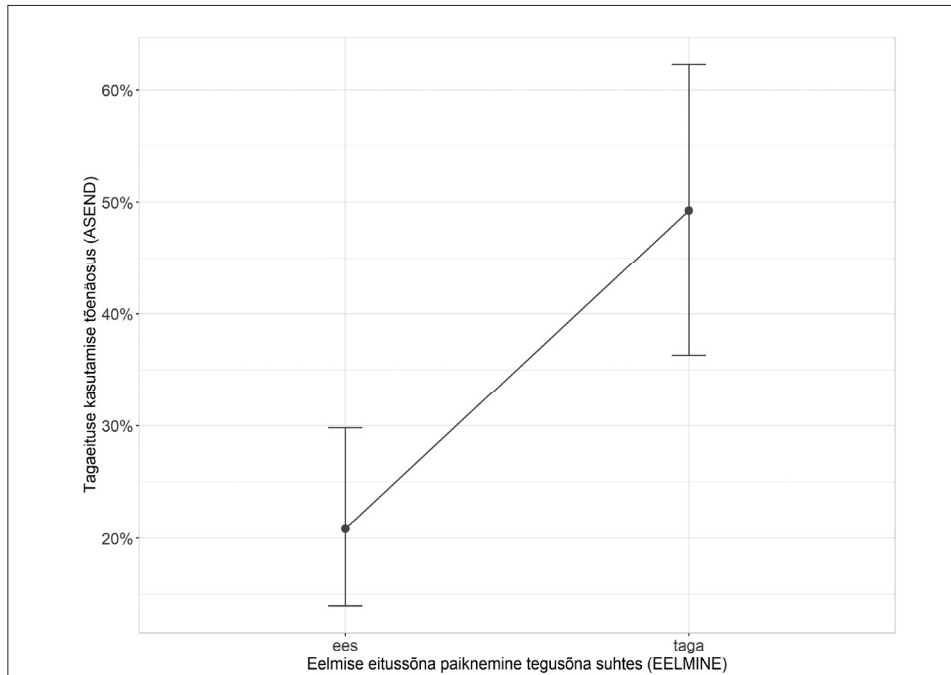
Võru keel ja selle variandid näitasid üldjoontes väga tugevat eelistust just eesituse kasutamisele (vt jooniseid 1 ja 2), mis ühtib standardeituse sõnajärgjega ka eesti, vene ja läti keeles. Ainult kahes kihelkonnas – Rõuges ja Vastseliinas – esines arvestataval hulgal tagaeituse kasutamist. Võrreldes (ida)setoga on neis olukord vastupidine: üle kahe kolmandiku eituse kasutusjuhtudest esineb eesitusega ja vaid alla ühe kolmandiku juhtudest tagaeitusega. Rõuge ja Vastseliina andmeid koondavas regressioonimudelis, mis ennustas harvema eituskonstruksiooni ehk sedapuhku tagaeituse kasutamise tõenäosust, osutus aga ainsaks tugevaks varieerumist seletavaks tunnuseks eelmise eitussõna asend tegusõna suhtes (joonis 6), seega sama tunnus, mis oli oluline ka seto materjalis.

Jooniselt 6 näeme, et primimise efekt võru keeles on idasetoga samapidine: kui kõnehetkele eelnenud eituskonstruksioon esines eesitusega, on palju tõenäolisem eesituse uuesti kasutamine (tagaeituse kasutamise tõenäosus on sel juhul ainult 21%); kui aga eelmine eituskonstruksioon on olnud tagaeitus, tõuseb märgatavalt ka tagaeituse uuesti kasutamise tõenäosus, jäädes siiski napilt alla 50%.

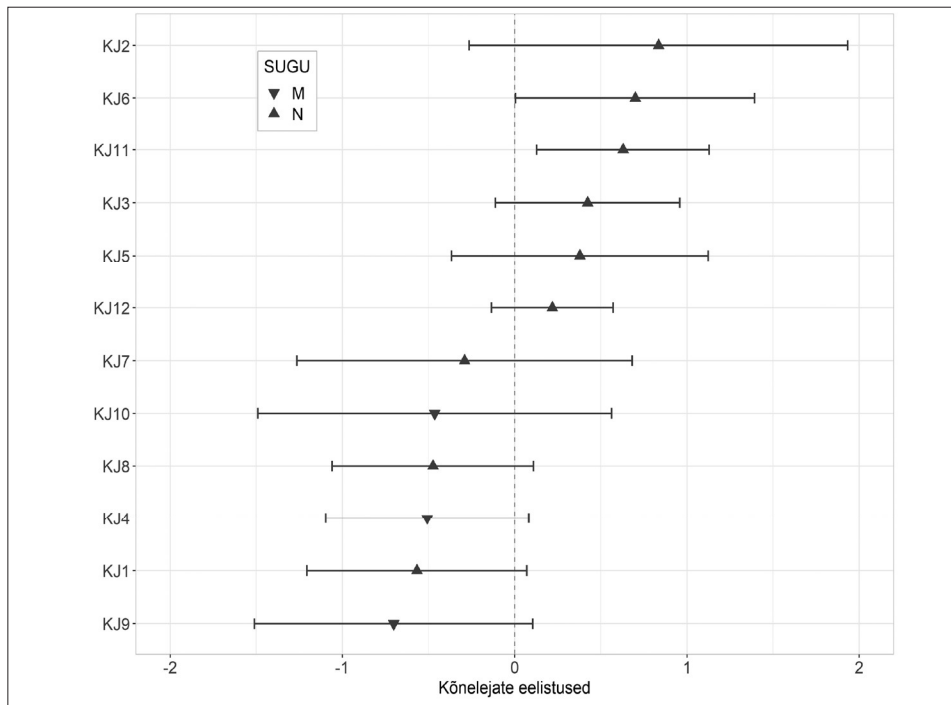
Ka kõnelejate juhuslik mõju aitab eitussõna positsiooni valiku varieerumist seletada,<sup>4</sup> tegusõna lemma võru keeles aga mitte. See, et tegusõna osutub oluliseks

<sup>4</sup> Lõpliku mudeli kuju oli  $ASEND \sim EELMINE + (1|KÕNELEJA)$ . Mudeli C-indeks oli 0,74, mis näitab mudeli keskmist eristusvõimet.





**Joonis 6.** Tagaeituse kasutamise tõenäosus võru andmetes vastavalt eelmise eitussõna paiknemisele.



**Joonis 7.** Võru keele kõnelejate ees- ja tagaeituse kasutamise eelistused.

varieerumise seletajaks idasetos, aga mitte võrus, osutab vähemalt kaudselt piirkonna keelekontaktide mõju üldisele suunale: idaseto on saanud mõjutusi ümberkaudsetest keeltest, kus on valdav eeseitus (eesti, vene, võru) ning mis on mõjutanud eeseituse kasutust mõningate sagedaste verbidega (nt *ei tiijääq*). Vastupidi see aga ei näi toimivat: seto tagaeitusega keelena ei näi võru keelt samal viisil mõjutavat. See on ka tõenäoline, sest andmed on pärit 1960.–1970. aastatest, mil seto keele prestiiž oli võrukete silmis madal ning laenamine ebatõenäoline. Samuti ei olnud sel ajal kasutuses võru kirjakeelt, kus tänapäeval on normiks kujundatud tagaeitus. Võru kõnelejate individuaalseid eelistusi kujutab joonis 7.

Jooniselt 7 näeme, et ka võru keele kõnelejad erinevad mõnevõrra individuaalsete eelistuste poolest ees- ja tagaeituse kasutamisel, kusjuures kõik meessoost kõnelejad näivad taas eelistavat keskmisest enam eeseitust. Paraku ei ole EMK-s võru murrukute kõnelejate kohta sama põhjalikku taustinfot kui idaseto kõnelejate kohta ning seetõttu on raske neid eelistusi sotsiolingvistiliste tegurite kaudu seletada. Küll aga näitab ka võru keele mudel, et kõnelejate idiosünkraatilise käitumisega arvestamine keele varieerumise uurimisel on oluline ning võimaldab arvesse võtta ka seda varjatud osa varieerumisest, mis fikseeritud lingvistiliste või ekstralingvistiliste kategooriatega ei pruugi ilmnedada.

Statistilise mudeldamise strateegiaid on erinevaid (vt nt Hosmer jt 2013; Winter 2020; Gries 2021) ning nii keelekasutuse formaalsete mudelite koostamist kui ka tõlgendamist peab alati saatma dialoog keeleteooria ja varasemate uurimustega, eriti kui eesmärk on keelekasutuse võimalikult adekvaatne kirjeldamine, mitte mudelite parim ennustatäpsus. Mudelid, milleni siinse analüüsi käigus jõudsime, on meie hinnangul kasutada olevate andmete põhjal optimaalsed, kirjeldamaks olulisemaid tegureid, mis valimitesse sattunud kõnelejate ees- ja tagaeituse valikut suunavad. Tuleb aga rõhutada, et ehkki raporteerisime ruumipuudusel  $p$ -väärtuste alusel vaid lõplikud mudelid statistiliselt oluliste tunnustega, oli meie mudeldamisstrateegia pigem uuriv kui kinnitav (vt nt Winter 2020: 274–280). See tähendab, et meil puudusid tugevad eeldused eituskonstruktsiooni sõnajärge tingivate tegurite kohta, ning testisime kõiki võimalikke seoseid ja koosmõjusid, mida oleksime suutnud põhjendada ning mida andmestik lubas analüüsi kaasata. Selle uurimuse andmestikust koorunud eelistusmuustrite paikapidavust tuleks seega võimalusel testida uute andmestikega. Hoolimata sellest, et mudelid ei ole kunagi ideaalsed esitused keeles esinevast varieerumisest, on mitmemõõtmelise statistika meetoditest abi komplekssete seoste tuvastamisel ja hindamisel suurtes andmehulkades, millest palja silmaga mustreid leida on keerukas.

## 6. Kokkuvõte

Artiklis näitlikustasime, kuidas korpusandmete kvantitatiivne analüüs võimaldab kirjeldada ja seletada keeles esinevat varieerumist, analüüsides juhtumiuuringus eitussõna paiknemist võru ja seto suulises keelekasutuses. Analüüsitud andmed pärinevad loodavast seto interdistsiplinaarsest korpusest, mis sisaldab muu hulgas

aastatel 2010–2013 Pihkva oblastis läbi viidud välitöödel tehtud intervjuusid, ning eesti murrete korpusest, kus on samalaadsed tekstid 1960.–1980. aastatest.

Võru ja seto on nii häälduse, vormimoodustuse kui ka sõnavara poolest sarnased, mistõttu on neid dialektoloogias käsitletud ühe Võru murdena. Ometi avaldub lause- tasandil ka erinevusi, mida põhjendatakse sageli eesti ja vene üldkeele mõju erineva ulatusega keeleala eri piirkondades. Üks silmapaistvamaid võru ja seto erinevusi on eitussõna asukoht eitusslauses. Ehkki mõlemas on võimalik paigutada eitussõna nii tegusõna ette kui ka taha, mis tüpoloogiliselt on küllalt haruldane, eelistavad võru keele kõnelejad kasutada ees- ja seto kõnelejad tagaeitust.

Artiklis vaatlesime esmalt eitustumustrite levikut võru-seto keelelal, et kaardistada ees- ja tagaeituse eelistustumustreid ning tuvastada üleminekuala ühe eelistamiselt teisele. Selgus, et setos on valdav tõepoolest tagaeitus, moodustades kolmveerandi kõikidest eituskonstruksiooni kasutamise juhtudest. Võrus seevastu prevaleerib selgelt eeseitus ning ainult kahes kihelkonnas – Rõuges ja Vastseliinas – võis täheldada märgatavat ees- ja tagaeituse varieerumist.

Seejärel uurisime mitmemõõtmelise statistilise analüüsi abil, kas kahe keele- variandi sees esinevat varieerumist suunavad ühesugused või erinevad tegurid. Selleks rakendasime segamõjudega logistilist regressiooni, mis võimaldab hinnata grammatiliste, semantiliste ja suhtluslike tegurite üldist rolli eituskonstruksiooni valikul, võttes samal ajal arvesse ka kõneleja- ja tegusõnaspetsiifilisi eelistustumustreid. Tulemused näitasid nii võrus kui ka setos selget tendentsi korrata juba kõneleja või vestluspartneri aktiveeritud vorme: ehkki setos domineerib taga- ja võrus eeseitus, kasvab harvemini esineva variandi kasutamise tõenäosus mõlemas keeles märkimis- väärselt siis, kui vastav eituskonstruksioon on hiljuti juba suhtluses aktiveeritud.

Idasetos osutus oluliseks ka kasutatud eitussõna, mis vastab kas oleviku või mine- viku ajavormile: oleviku eitussõnaga *ei* on eesti ja vene keele pärase eeseituse kasu- tamise tõenäosus oluliselt suurem kui mineviku eitussõnaga *es*. See on tõenäoliselt seotud kontaktkeeltest üle võetud suhtluses sagedaste kinnistunud sõnajärgjega eitus- konstruksioonidega (nt *ei tiijäq*).

Sellistes võrdlemisi väikestes valimites ei saa mööda vaadata kõnelejate indi- viduaalsest mõjust varieerumisele, ehkki selle mõju sisuline tõlgendamine võib kõneleja tausta teadmata osutada keeruliseks. Kõneleja osutus oluliseks nii seto kui ka võru eitustumustrite eelistuste seletamisel. Setos saime siduda selle kõnelejate eri- neva kokkupuutega kontaktkeeltega (eesti, vene), ent võru kohta vastav info paraku puudus. Ka tegusõnade idiosünkraatiline käitumine võib heita valgust nii keele- kontaktide rollile suhtluses sagedaste väljendite kasutamisel kui ka keelesisestele protsessidele, milles järjendid kasutussageduse ja/või morfofonoloogiliste tegurite mõjul võivad eituskonstruksioonis kindla sõnajärgjega kinnistuda ja automatisee- ruda.

Artikli valmimist on toetanud Euroopa Liit Euroopa Regionaalarengu Fondi kaudu (Eesti-uuringute Tippkeskus) ning EKKD projekt „Interdistsiplinaarne seto korpus“.

KIHELKONDADE LÜHENDID:

**Hargla**; **ISe** = Ida-Seto; **Kanepi**; **Krl** = Karula; **LSe** = Lõuna-Seto; **Plv** = Põlva; **PSe** = Põhja-Seto; **Rõuge**; **Urvaste**; **Vastseliina**.

VEEBIVARAD

**EMK** = Eesti murrete korpus. <https://www.keel.ut.ee/et/keelekogud/murdekorpus>

**ggplot2**. Create Elegant Data Visualisations Using the Grammar of Graphics. <https://cran.r-project.org/package=ggplot2>

**Goldvarb Z**. A multivariate analysis application for Macintosh. Toronto: Department of Linguistics, University of Toronto, 2018. <http://individual.utoronto.ca/tagliamonte/goldvarb.html>

**Goldvarb X**. A variable rule application for Macintosh and Windows. Toronto: Department of Linguistics, University of Toronto, 2005. <http://individual.utoronto.ca/tagliamonte/goldvarb.html>

**lme4**. Linear Mixed-Effects Models using 'Eigen' and S4. <https://cran.r-project.org/package=lme4>

**Murdearhiiv** = Tartu Ülikooli eesti murrete ja sugulaskeelte arhiiv. <https://murdearhiiv.ut.ee>

**R Core Team 2020**. R: A Language and Environment for Statistical Computing. Vienna: R Foundation for Statistical Computing. <https://www.R-project.org>

**SetKo** = Interdistsiplinaarne seto korpus. <https://setko.ut.ee>

**sf**. Simple Features for R. <https://cran.r-project.org/package=sf>

**sjPlot**. Data Visualization for Statistics in Social Science. R package version 2.8.7. <https://CRAN.R-project.org/package=sjPlot>

KIRJANDUS

**Baayen, R. Harald 2008**. Analyzing Linguistic Data: A Practical Introduction to Statistics Using R. Cambridge, UK: Cambridge University Press. <https://doi.org/10.1017/CBO9780511801686>

**Baayen, R. Harald; Davidson, Douglas J.; Bates, Douglas M. 2008**. Mixed-effects modeling with crossed random effects for subjects and items. – Journal of Memory and Language, kd 59, nr 4, lk 390–412. <https://doi.org/10.1016/j.jml.2007.12.005>

**Bates, Douglas; Maechler, Martin; Bolker, Ben; Walker, Steve 2015**. Fitting linear mixed-effects models using lme4. – Journal of Statistical Software, kd 67, nr 1, lk 1–48. <https://doi.org/10.18637/jss.v067.i01>

**Bod, Rens; Hay, Jennifer; Jannedy, Stefanie (toim) 2003**. Probabilistic Linguistics. Cambridge, MA: MIT Press. <https://doi.org/10.7551/mitpress/5582.001.0001>

**Bresnan, Joan 2007**. Is syntactic knowledge probabilistic? Experiments with the English dative alternation. – Roots: Linguistics in Search of its Evidential Base. Toim Sam Featherston, Wolfgang Sternefeld. Berlin: Mouton de Gruyter, lk 75–96.

- Bresnan, Joan; Ford, Marilyn 2010.** Predicting syntax: Processing dative constructions in American and Australian varieties of English. – *Language*, kd 86, nr 1, lk 168–213. <https://doi.org/10.1353/lan.0.0189>
- Cedergren, Henrietta J.; Sankoff, David 1974.** Variable rules: Performance as a statistical reflection of competence. – *Language*, kd 50, nr 2, lk 333–355. <https://doi.org/10.2307/412441>
- Diessel, Holger 2017.** Usage-based linguistics. – *Oxford Research Encyclopedia of Linguistics*. Toim Mark Aronoff. Oxford: Oxford University Press. <https://doi.org/10.1093/acrefore/9780199384655.013.363>
- Dryer, Matthew S. 2013.** Order of negative morpheme and verb. – *The World Atlas of Language Structures Online*. Toim M. S. Dryer, Martin Haspelmath. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://wals.info/chapter/143> (11. VII 2019).
- Eichenbaum, Külli; Pajusalu, Karl 2001.** Setode ja võrokeste keelehoiakutest ja identiteedist. – *Keel ja Kirjandus*, nr 7, lk 483–489.
- Grafmiller, Jason; Szmrecsanyi, Benedikt; Röthlisberger, Melanie; Heller, Benedikt 2018.** General introduction: A comparative perspective on probabilistic variation in grammar. – *Glossa: A Journal of General Linguistics*, kd 3, nr 1, lk 94. <https://doi.org/10.5334/gjgl.690>
- Gries, Stefan Th. 2015.** The most under-used statistical method in corpus linguistics: Multi-level (and mixed-effects) models. – *Corpora*, kd 10, nr 1, lk 95–125. <https://doi.org/10.3366/cor.2015.0068>
- Gries, Stefan Th. 2021.** (Generalized linear) mixed-effects modeling: A learner corpus example. – *Language Learning*, kd 71, nr 3, lk 757–798. <https://doi.org/10.1111/lang.12448>
- Hagu, Paul; Pajusalu, Karl 2020.** Lühikene seto keeleoppus. Lühike seto keeleõpetus. Seto Instituut. <http://www.setoinstituut.ee/download/luhikene-seto-keeleopetus> (2. VII 2021).
- Hazen, Kirk 2011.** Labov: Language variation and change. – *The SAGE Handbook of Sociolinguistics*. Toim Ruth Wodak, Barbara Johnstone, Paul Kerswill. SAGE Publications Ltd, lk 24–39. <https://doi.org/10.4135/9781446200957.n3>
- Hinrichs, Lars; Szmrecsanyi, Benedikt 2007.** Recent changes in the function and frequency of Standard English genitive constructions: A multivariate analysis of tagged corpora. – *English Language and Linguistics*, kd 11, nr 3, lk 437–474. <https://doi.org/10.1017/S1360674307002341>
- Hint, Helen; Taremaa, Piia; Reile, Maria; Pajusalu, Renate 2021.** Demonstratiivpronoome-nid ja -adverbid määratlejatena. Miks me oleme *siin ilmas, selles olukorras*? – *Eesti ja soome-ugri keeleteaduse ajakiri / Journal of Estonian and Finno-Ugric Linguistics*, kd 12, nr 1, lk 79–111.
- Hosmer, David W. Jr.; Lemeshow, Stanley; Sturdivant, Rodney X. 2013.** *Applied Logistic Regression*. Hoboken, NJ: John Wiley and Sons. <https://doi.org/10.1002/9781118548387>
- Iva, Sulev 2002.** *Võro-eesti synaraamat*. (Võro Instituudi toimõndusõq 12.) Võro: Võro Instituut.
- Iva, Sulev 2007.** *Võru kirjakeele sõnamuutmissüsteem*. (Dissertationes philologiae estonicae Universitatis Tartuensis 20.) Tartu: Tartu Ülikooli Kirjastus.

- Iva, Sulev 2013.** Võru ja seto kõrihäälikud *h* ja *q*. – Tartu Ülikooli Lõuna-Eesti keele- ja kultuuriuuringute keskuse aastaraamat XI–XII. Tartu: Tartu Ülikooli Lõuna-Eesti keele- ja kultuuriuuringute keskus, lk 102–116.
- Janhunen, Juha 1982.** On the structure of Proto-Uralic. – *Finnisch-Ugrische Forschungen*, nr 44, lk 23–42.
- Jääts, Indrek 2000.** Ethnic identity of the Setus and the Estonian-Russian border dispute. – *Nationalities Papers*, kd 28, nr 4, lk 651–670. <https://doi.org/10.1080/00905990020009665>
- Kallio, Petri 2007.** Kantasuomen konsonanttihistoriaa. – *Sámit, sánit, sátnehámit: Riepmočála Pekka Sammallahtii miessemánu 21. beaivve 2007.* (Mémoires de la Société Finno-Ougrienne 253.) Toim Jussi Ylikoski, Ante Aikio. Helsinki: Société Finno-Ougrienne, lk 229–249.
- Kallio, Petri 2014.** The diversification of Proto-Finnic. – *Fibula, Fabula, Fact: The Viking Age in Finland.* (Studia Fennica Historica 18.) Toim Joonas Ahola, Frog, Clive Tolley. Helsinki: Suomalaisen Kirjallisuuden Seura, lk 155–168.
- Keem, Hella; Käsi, Inge 2002.** Võru murde tekstid. (Eesti murded VI.) Tallinn: Eesti Keele Instituut.
- Klavan, Jane 2012.** Evidence in Linguistics: Corpus-Linguistic and Experimental Methods for Studying Grammatical Synonymy. (Dissertationes linguisticae Universitatis Tartuensis 15.) Tartu: Tartu Ülikooli Kirjastus.
- Klavan, Jane 2018.** Kognitiivne keeleteadus arvude rägastikus. – *Keel ja Kirjandus*, nr 8–9, lk 697–712.
- Klavan, Jane 2021 (ilmumas).** The alternation between exterior locative cases and postpositions in Estonian web texts. – *Eesti ja soome-ugri keeleteaduse ajakiri / Journal of Estonian and Finno-Ugric Linguistics*, kd 12, nr 1.
- Klavan, Jane; Divjak, Dagmar 2016.** The cognitive plausibility of statistical classification models: Comparing textual and behavioral evidence. – *Folia Linguistica*, kd 50, nr 2, lk 355–384. <https://doi.org/10.1515/flin-2016-0014>
- Klavan, Jane; Pilvik, Maarja-Liisa; Uiboaed, Kristel 2015.** The use of multivariate statistical classification models for predicting constructional choice in spoken, non-standard varieties of Estonian. – *SKY Journal of Linguistics*, nr 28, lk 187–224.
- Klavan, Jane; Veismann, Ann 2017.** Are corpus-based predictions mirrored in the preferential choices and ratings of native speakers? Predicting the alternation between the Estonian adessive case and the adposition *peal* 'on'. – *Eesti ja soome-ugri keeleteaduse ajakiri / Journal of Estonian and Finno-Ugric Linguistics*, kd 8, nr 2, lk 59–91. <https://doi.org/10.12697/jeful.2017.8.2.03>
- Kortmann, Bernd 2010.** Areal variation in syntax. – *Language and Space: An International Handbook of Linguistic Variation. Theories and Methods.* Toim Peter Auer, Jürgen E. Schmidt. Berlin–New York: De Gruyter, lk 837–864. <https://doi.org/10.1515/9783110220278.837>
- Kortmann, Bernd 2021.** Reflecting on the quantitative turn in linguistics. – *Linguistics*. <https://doi.org/10.1515/ling-2019-0046>
- Krasnoukhova, Olga; Auwera, Johan van der; Crevels, Mily 2021 (ilmumas).** Introduction: Postverbal negation: What, where, why. – *Studies in Language*.

- Labov, William 1963.** The social motivation of a sound change. – *Word*, kd 19, nr 3, lk 273–309. <https://doi.org/10.1080/00437956.1963.11659799>
- Labov, William 1972.** *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.
- Lindström, Liina 1997.** Eitus Võru murde suulises kõnes. – *Õdagumeresoomõ lõunapiir*. Toim Karl Pajusalu, Jüvä Sullöv. (Võro Instituudi Toimõtiseq 1.) Võro: Võro Instituut, lk 143–154.
- Lindström, Liina 2001.** Eesti murrete korpuse iseloomustus argivestlustega võrrelduna. – *Keele kannul. Pühendusteos Mati Erelti 60. sünnipäevaks 12. märtsil 2001.* (Tartu Ülikooli eesti keele õppetooli toimetised 17.) Toim Reet Kasik. Tartu: Tartu Ülikooli Kirjastus, lk 212–221.
- Lindström, Liina 2015.** Ülevaade eesti murrete korpusest seisuga 17.11.2015. [https://www.keel.ut.ee/sites/default/files/www\\_ut/emk\\_teejuht2015.pdf](https://www.keel.ut.ee/sites/default/files/www_ut/emk_teejuht2015.pdf) (2. VII 2021).
- Lindström, Liina 2021 (ilmumas).** Seto lause põhijooned. – *Setomaa 3. Keel, rahvaluule ja tänapäeva kultuur*. Toim Andreas Kalkun, Karl Pajusalu, Ergo-Hart Västrik. Tartu: Eesti Rahva Muuseum.
- Lindström, Liina; Kalmus, Mervi; Klaus, Anneliis; Bakhoff, Liisi; Pajusalu, Karl 2009.** Ainsuse 1. isikule viitamine eesti murretes. – *Emakeele Seltsi aastaraamat 54 (2008)*. Tallinn: Teaduste Akadeemia Kirjastus, lk 159–185.
- Lindström, Liina; Pilvik, Maarja-Liisa 2018.** Korpuspõhine kvantitatiivne dialektoloogia. – *Keel ja Kirjandus*, nr 8–9, lk 643–662.
- Lindström, Liina; Pilvik, Maarja-Liisa; Plado, Helen 2018.** Nimetamiskonstruktsioonid eesti murretes: murdeerinevused või suuline süntaks? – *Mäetagused*, nr 70, lk 91–126. [https://doi.org/10.7592/MT2018.70.lindstrom\\_pilvik\\_plado](https://doi.org/10.7592/MT2018.70.lindstrom_pilvik_plado)
- Lindström, Liina; Pilvik, Maarja-Liisa; Plado, Helen 2021 (ilmumas).** Variation in negation in Seto. – *Studies in Language*. <https://doi.org/10.1075/sl.19063.lin>
- Lindström, Liina; Pilvik, Maarja-Liisa; Ruutma, Mirjam; Uihoaed, Kristel 2019.** On the use of perfect and pluperfect in Estonian dialects: Frequency and language contacts. – *Multilingual Finnic – Language Contact and Change*. Toim Sofia Björklöf, Santra Jantunen. Helsinki: Finno-Ugrian Society, lk 155–193. <https://doi.org/10.33341/uh.85035>
- Lindström, Liina; Uihoaed, Kristel 2017.** Syntactic variation in 'need'-constructions in Estonian dialects. – *Nordic Journal of Linguistics*, kd 40, nr 3, lk 313–349. <https://doi.org/10.1017/S0332586517000191>
- Lindström, Liina; Vihman, Virve-Anneli 2017.** Who needs it? Variation in experiencer marking in Estonian 'need'-constructions. – *Journal of Linguistics*, kd 53, nr 4, lk 789–822. <https://doi.org/10.1017/S0022226716000402>
- Lüdecke, Daniel 2018.** ggeffects: Tidy data frames of marginal effects from regression models. – *Journal of Open Source Software*, nr 3 (26), lk 772. <https://doi.org/10.21105/joss.00772>
- Mets, Mari 2010.** Suhtlusvõrgustikud reaajas: võru kõnekeele varieerumine kahes Võrumaa külas. (Dissertationes philologiae estonicae Universitatis Tartuensis 25.) Tartu: Tartu Ülikooli Kirjastus.
- Miestamo, Matti; Tamm, Anne; Wagner-Nagy, Beáta 2015.** Negation in Uralic languages. Introduction. – *Negation in Uralic Languages*. Toim M. Miestamo, A. Tamm, B. Wagner-Nagy. Amsterdam–Philadelphia: John Benjamins, lk 1–41. <https://doi.org/10.1075/tsl.108.01int>

- Paas, Friedrich-Eugen 1927.** Sega-abelud ja nende mõju rahvusele piiriäärsetes maakondades Eestis. Tartu: Eesti Vabariigi Tartu Ülikool.
- Pajusalu, Karl 1996.** Multiple Linguistic Contacts in South Estonian: Variation of Verb Inflection in Karksi. (Turun yliopiston suomalaisen ja yleisen kielitieteen laitoksen julkaisuja. Publications of the Department of Finnish and General Linguistics of the University of Turku 54.) Turku: Turku University.
- Pajusalu, Karl; Hennoste, Tiit; Niit, Ellen; Päll, Peeter; Viikberg, Jüri 2002.** Eesti murded ja kohanimed. Tallinn: Eesti Keele Sihtasutus.
- Pajusalu, Karl; Velsker, Eva; Org, Ervin 1999.** On recent changes in South Estonian: Dynamics in the formation of the inessive. – *International Journal of the Sociology of Language*, nr 139, lk 87–103. <https://doi.org/10.1515/ijsl.1999.139.87>
- Paolillo, John C. 2002.** Analyzing Linguistic Variation. (CSLI Lecture Notes 114.) Stanford: CSLI Publications.
- Pebesma, Edzer 2018.** Simple features for R: Standardized support for spatial vector data. – *The R Journal*, kd 10, nr 1, lk 439–446. <https://doi.org/10.32614/RJ-2018-009>
- Pilvik, Maarja-Liisa; Lindström, Liina; Plado, Helen 2021.** Murded, varieerumine ja korpusandmed: eitussõna paiknemine võru ja seto eituslausetes. Lisamaterjalid. <https://doi.org/10.17605/OSF.IO/GCFT7>
- Pinheiro, José C.; Bates, Douglas M. 2000.** Mixed-effects Models in S and S-PLUS. New York: Springer. <https://doi.org/10.1007/978-1-4419-0318-1>
- Plado, Helen 2015.** *des-* ja *mata-*konverbi kasutusest eesti murretes. – *Emakeele Seltsi aastaraamat 60* (2014). Tallinn: Teaduste Akadeemia Kirjastus, lk 195–218. <https://doi.org/10.3176/esa60.10>
- Reile, Maria; Plado, Helen; Gudde, Harmen B.; Coventry, Kenny R. 2020.** Demonstratives as spatial deictics or something more? Evidence from Common Estonian and Võro. – *Folia Linguistica*, kd 54, nr 1, lk 167–195. <https://doi.org/10.1515/flin-2020-2030>
- Saareste, Andrus 1955.** Petit atlas des parlers estoniens. Väike eesti murdeatlas. Uppsala: Almqvist & Wiksell.
- Siiman, Ann 2018.** Variation of the Estonian singular long and short illative form: A multivariate analysis. – *SKY Journal of Linguistics*, nr 31, lk 139–167.
- Siiman, Ann 2019.** Vormikasutuse varieerumine ning selle põhjused osastava ja sisseütleva käände näitel. (Dissertationes philologiae estonicae Universitatis Tartuensis 45.) Tartu: Tartu Ülikooli Kirjastus.
- Szmrecsanyi, Benedikt 2005.** Language users as creatures of habit: A corpus-based analysis of persistence in spoken English. – *Corpus Linguistics and Linguistic Theory*, kd 1, nr 1, lk 113–149. <https://doi.org/10.1515/clt.2005.1.1.113>
- Szmrecsanyi, Benedikt 2017.** Variationist sociolinguistics and corpus-based variationist linguistics: Overlap and cross-pollination potential. – *Canadian Journal of Linguistics / Revue canadienne de linguistique*, kd 62, nr 4, lk 685–701. <https://doi.org/10.1017/cnj.2017.34>
- Szmrecsanyi, Benedikt; Anderwald, Lieselotte 2018.** Corpus-based approaches to dialect study. – *The Handbook of Dialectology*. Toim Charles Boberg, John Nerbonne, Dominic Watt. Malden, MA: Wiley-Blackwell, lk 300–313. <https://doi.org/10.1002/9781118827628.ch17>



- Tagliamonte, Sali A. 2012.** Variationist Sociolinguistics: Change, Observation, Interpretation. Chichester: Wiley-Blackwell.
- Tagliamonte, Sali A.; Baayen, R. Harald 2012.** Models, forests, and trees of York English: *Was/were* variation as a case study for statistical practice. – Language Variation and Change, kd 24, nr 2, lk 135–178. <https://doi.org/10.1017/S0954394512000129>
- Taremaa, Piia; Hint, Helen; Reile, Maria; Pajusalu, Renate 2021.** Constructional variation in Estonian: Demonstrative pronouns and adverbs as determiners in noun phrases. – Lingua, kd 254, nr 103030. <https://doi.org/10.1016/j.lingua.2021.103030>
- Torres Cacoullous, Rena; Travis, Catherine E. 2019.** Variationist typology: Shared probabilistic constraints across (non-)null subject languages. – Linguistics, kd 57, nr 3, lk 653–692. <https://doi.org/10.1515/ling-2019-0011>
- Tulving, Endel 2002.** Mälu. Tartu: Tartu Ülikooli Kirjastus.
- Veismann, Ann; Klavan, Jane; Öim, Haldur 2018.** Teoreetiline keeleteadus ja kvantitatiivsed meetodid. – Keel ja Kirjandus, nr 8–9, lk 609–621.
- Viikberg, Jüri 2020.** Eesti murrete grammatika. (Eesti keele varamu VIII.) Tartu: Tartu Ülikooli Kirjastus.
- Viitso, Tiit-Rein 1998.** Estonian. – The Uralic Languages. Toim Daniel Abondolo. London–New York: Routledge, lk 115–148.
- Walker, James A. 2013.** Variation analysis. – Research Methods in Linguistics. Toim Robert J. Podesva, Devyani Sharma. Cambridge: Cambridge University Press, lk 440–459.
- Wickham, Hadley 2016.** ggplot2: Elegant Graphics for Data Analysis. New York: Springer-Verlag. [https://doi.org/10.1007/978-3-319-24277-4\\_9](https://doi.org/10.1007/978-3-319-24277-4_9)
- Winter, Bodo 2020.** Statistics for Linguists: An Introduction Using R. New York: Routledge. <https://doi.org/10.4324/9781315165547>

**Maarja-Liisa Pilvik** (snd 1989), MA, Tartu Ülikooli doktorant ja rakendusliku dialektoloogia nooremteadur (Jakobi 2–430, 51005 Tartu), [maarja-liisa.pilvik@ut.ee](mailto:maarja-liisa.pilvik@ut.ee)

**Helen Plado** (snd 1981), PhD, Tartu Ülikooli eesti keele teadur ja lektor ning Võru Instituudi teadur (Jakobi 2–426, 51005 Tartu), [helen.plado@ut.ee](mailto:helen.plado@ut.ee)

**Liina Lindström** (snd 1973), PhD, Tartu Ülikooli tänapäeva eesti keele professor ning digihumanitaaria ja infoühiskonna keskuse juhataja (Jakobi 2–443, 51005 Tartu), [liina.lindstrom@ut.ee](mailto:liina.lindstrom@ut.ee)