

NIMISÕNAFRAASI JA HULGAFRAASI PIIRIMAIL: OSA, ENAMIK JA ENAMUS HULGASÕNADENA

Maarja-Liisa Pilvik, Liina Lindström,
Helen Plado, Carl Eric Simmul

Ülevaade. Artiklis vaatleme hulgasõnast *osa*, *enamik* ja *enamus* ning mitmuslikust komplektisõnast moodustatud fraaside struktuurilist ja semantilist varieerumist tänapäeva eesti keeles. Lähemalt keskendume hulgasõna arvu (nt *osale inimestele*, *osadele inimestele*), komplektisõna käände (*enamikku inimesi*, *enamikku inimestest*) ning verbi vormi valikut (*enamus inimesi läheb ~ lähevad*) mõjutavate tegurite analüüsile. Nii hulgasõna valikut kui ka kõiki kolme varieerumise aspekti on varasemalt seotud käimasoleva keelemuutusega, milles hulgasõnast põhi ja komplektisõnast laiendiga kvantorifraas on asendumas hulgasõnast laiendi ja komplektisõnast põhjaga nimisõnafrasiga. Keelekasutajaid on võinud sealjuures motiveerida vajadus eristada hulgasõna kvantifitseerivat ja määratlevat funktsiooni. Uurimuse tulemused kinnitavad paljusid varasemaid, väiksema materjali põhjal tehtud või intuitsioonil põhinevaid tähelepanekuid, ent toovad lisaks esile fraaside süntaktilise rolli olulisuse: fraasisisese arvuuühildumise võimalikuks lähtekohaks on tõenäoliselt adverbialid, samas kui hulgasõna määratlev funktsioon on kinnistumas pigem subjektina toimivates fraasides. Muu hulgas näeme keelemuutuse levikul ka selget žanri ja tekstilooma spontaansuse mõju.*

Võtmesõnad: keele varieerumine, ühildumine, keelemuutused, korpuslingvistika, eesti keel

1. Sissejuhatus

Artiklis vaadeldakse mitmeliikmelise hulga osahulka väljendavaid fraase, mis koosnevad mõõdunimisõnast *osa*, *enamik* või *enamus* ja kvantitatiivselt piiritletud komplekti väljendavast mitmuslikust nimisõnafrasist, nt *osa mu sõpru*, *enamik kongressi delegaate*.

* Artikli valmimist on toetanud projekt EKKD-TA2 "Eesti keele morfosüntaktiline varieerumine".

Peep Nemvalts (2000) on hulgasõnu *osa*, *enamik*, *muist* nimetanud osalisuskvantoriteks, sest need väljendavad osalisust, vastandudes tervikule või koguhulgale. Tänu sellistele semantilistele omadustele saab neid kasutada lisaks mitmuslikele loendatavatele (*osa asju*) ja ainsuslikele sõnadele (*osa aega*) ka ainsuslike loendatavate nimisõnadega, nt *Osa pead oli kaetud*. Samasse ritta sobib ka *enamus*, mis samuti väljendab osalisust.

Kõnealuste hulgasõnade kasutust on piiranud keelekorraldus. Sõna *enamus* on olnud keeles tähenduses 'arvuline ülekaal' kasutusel juba 19. sajandi teisel poolel (Habicht 1995), sõna *enamik* toodi õigekeelsussõnaraamatutesse tähenduses 'suurem osa' 1925. a (Risberg, Langemets 2021). Ehkki kasutuses on kahe hulgasõna tähendused aja jooksul põimunud, on keelekorraldusallikates alates 1953. aasta "Väikesest õigekeelsuse sõnaraamatust" tähendused lahutatud: tähendus 'suurem osa' kinnistati sõnale *enamik* ja 'arvuline ülekaal' sõnale *enamus*. Seda tähendusvahet on keelekorralduses ja keeleõpikutes tehtud tänini. (Risberg, Langemets 2021)

Kvantori- ehk hulgafrasid on nimisõnu *osa*, *enamik* ja *enamus* hakatud kasutama suhteliselt hiljuti. Triin Lõbu (2011) on vaadanud hulgafraside diakroonilist kujunemist korpusmaterjalide põhjal ning leidnud, et *osa* hakati hulgafrasid kasutama 19. sajandi lõpus, *enamik* 1930-ndatel ning *enamus* 20. sajandi teisel poolel (vaatamata keelekorralduse püüdele seda selles funktsioonis keelata). Need ajalised määratlused on muidugi tinglikud, sest tõenäoliselt kasutati neid hulgafrase enne, kui need ilmusid kirjakeelde ja jõudsid väiksemahulistesse korpusse. Sellegipoolest võib arvata, et hulgafrasi peasõnana on neid sõnu hakatud kasutama võrdlemisi hilja.

Hulgasõnad *osa*, *enamik* ja *enamus* on ühelt poolt tüüpilised hulgafrasi peasõnad: neid võivad laiendada komplektisõnad; teiselt poolt on need nimi-sõnad: neid võivad hulgafrasid laiendada eestäiendid, nt *suur osa inimesi*, *valdav enamik inimestest*. Hulgafrasi teeb keerukaks asjaolu, et laiendi vorm oleneb hulgasõna käändest mitmeti. Laiend on partitiivis (*osa inimesi*), kui hulgasõna ise on nominatiivis või partitiivis (*enamik inimesi*, *enamikku inimesi*). Kui hulgasõna on ainsuse genitiivis (*enamiku*), sõltub laiendi vorm fraasi funktsioonist: kui hulgafras on sihitis, siis on laiend partitiivis (*enamiku mõõtmisi saab teha*), kui fraas on mõni muu lauseliige, siis ühildub laiend käändes peasõnaga (*enamiku inimeste jaoks*). Käändes ühildumine toimub ka siis, kui hulgasõna on semantilises käändes (*enamikule inimestele meeldib mahe muusika*) (Metslang 2017: 470–471). Seega on hulgafrasi peasõna ja laiendi vahel kord rektisioonisuhe, kord ühildumissuhe. Seejuures on hulgafrasile peetud iseloomulikuks seda, et mitmuslik laiend ühildub põhisõnaga käändes, aga mitte arvus (*osale inimestele*, *enamikus riikides*). See nähtus tundub olevat eesti keele eriareng (puududes näiteks soome keelest): nimisõnalise hulgasõnaga fraasid käituvad obliikväänetes mõneti sarnaselt arvsõnalise hulgafrasiga, kus käändeühildumine toimub ainsuslike vormide vahel, vrd *osale inimestele* – *kolmele inimesele* (vt Koptjevskaja-Tamm 2001: 539).

Osalisuskvantorid on loomu poolest ainsuslikud, sest väljendavad üht osahulka tervikust (Nemvalts 2000: 100). Samas on juba mõnda aega täheldatud (nt Erelt, Metslang 1998), et sõnu *osa*, *enamik* ja *enamus* on võimalik kasutada ka mitmuslikult: *osadele inimestele*, *enamikele riikidele*, *osad inimesed*. Selle kasutusviisi puhul asendab nimisõnafras sisuliselt hulgafrasi ning teeb kogu nähtuse analüüsi

keerukamaks, sest hulga fraasi ja nimisõna fraasi vaheldumisega vahetub ka peasõna ja laiendi vahekord.

Artiklis käsitleme hulgasõnu *osa*, *enamik* ja *enamus* sisaldavate fraaside ehituse ja semantilis-struktuuriliste seoste varieerumist, analüüsides nii fraasisisest arvuühildumist (mida võib pidada ka nimisõna fraasi ja hulga fraasi vaheldumiseks), laiendi valikut (elatiiv või partitiiv / muu kääne) kui ka seda, kuidas aluse funktsioonis fraas ühildab finiiitverbi (ainsuses või mitmuses). Kuivõrd sõltuvalt vahelduvast fraasitüübist peaksime erinevalt nimetama fraasi komponente, siis lähtume edaspidi terminite kasutuses semantilistest kategooriatest: räägime **hulgasõnast** ehk **kvantorist**, mis toimib nimisõna fraasis laiendina (*enamikud inimesed*) ja hulga fraasis peasõnana (*enamik inimesi*), ning **komplektisõnast**¹, mis toimib nimisõna fraasis peasõnana (*enamikud inimesed*) ja hulga fraasis laiendina (*enamik inimesi*). Viitamaks mõlemale fraasitüübile korraga, kasutame mõistet **hulgasõnaga fraas**.

Analüüsitava materjal pärineb eesti keele ühendkorpuse 2023 (Koppel jt 2023) perioodika, blogide ja foorumite žanrimärgendiga tekstidest. Ka kõik allpool olevad žanrimärgendiga näited on pärit samast allikast. Kolme varieeruva aspekti (fraasisisese arvuühildumise, komplektisõna käände ning verbiühildumise) ulatuse kaardistamiseks kasutame sagedusanalüüsi, kõnelejate valikuid mõjutavate tegurite väljaselgitamiseks aga varieerumise analüüsi levinud meetodeid: hii-ruut-teste, tingimuslikke otsustuspuid, tingimuslikke juhumetsi ning logistilist regressioonianalüüsi.

2. Taust

2.1. Fraasisisene arvuühildumine

Nagu mainitud, on laieneva keelemuutusena täheldatud sõnade *osa*, *enamik* ja *enamus* tendentsi ühilduda komplektisõnaga ka arvus, kusjuures keelekorraldus on muutusi aegade jooksul tõrjunud (nt Kindlam 1978: 62, ettevaatlikumalt, hulga suurst väljendavas tähenduses soovib arvuühildumist vältida Erelt 2019: 106–107, Erelt jt 2020: 492). Ühildumine toimub mitte ainult käändeühildumist nõudvates käänetes (*osadele inimestele*, *enamikes kohtades*), vaid ka siis, kui hulgasõna on nominatiivis: *osad inimesed*, *enamikud inimesed*. Kui hulka väljendava fraasi moodustajad ühilduvad arvus, ei eristu hulga fraas struktuurilt nimisõna fraasist, mille laiend on kvantor (nt *mõned inimesed*) või määratleja (nt *need inimesed*). See tähendab, et rektsoonisuhte asemel on sõnade vahel ühildumissuhe ning kogu fraasi võib pidada nimisõna fraasiks. Subjekti rollis ühildab arvus ühilduva kvantori ja komplektisõnaga fraas ka öeldise (*osad inimesed pesevad hambaid ainult hommikuti*), samas kui ainsusliku kvantoriga võib öeldise vorm varieeruda (*osa inimesi peseb/pesevad*).

Mati Erelt ja Helle Metslang (1998) on kvantori ja komplektisõna arvuühildumise teket ja laienemist näinud muutusena, mis on vähemalt osaliselt seletatav vajadusega eristada kvantifitseerivat ja määratlevat funktsiooni. Kvantifitseeriv

¹ "Eesti grammatika" (EG 2023: 619–621) nimetab komplektisõnadeks mitmuslikke sõnavorme, mis tähistavad leksikaalse tähenduse poolest piiritletud objektihulki (nt *vanemad*, *kingad*, *malendid*), vastandades neid piiritlemata tähendusega juhuühildumisele (nt *lilled*, *aknad*). Sealjuures võib käändevaheldus võimaldada kategooriate vahetamist (nt *Sul on ilusad kingad* -> *Sul on ilusaid kingi*). Siin nimetame lihtsuse huvides komplektisõnadeks kõiki hulgasõnaga fraaside mitmuslikke asja- ja ainesõnu, mis võivad väljendada nii kvantitatiivselt piiritletud kui ka piiritlemata objektihulki.

funktsioon osutab hulga suurusele (vastab küsimusele *kui palju?*, nt *enamik inimesi oli juba koju läinud*), määratlev funktsioon aga piiritleb referente (sarnaselt määratlejatega *need*, *mitu*: *need inimesed* / *mitmed inimesed* / *osad inimesed*). Käändes ja arvus ühildumine on seotud ennekõike määratleva funktsiooniga; kõige selgemalt ilmneb see kvantoriga, mis on mitmuse nimetavas (*osad inimesed*). Lisaks peavad Erelt ja Metslang (1998) *osad inimesed* -tüüpi konstruktsiooni levi- kut soodustavateks asjaoludeks veel ikoonilisust (sisulist mitmuslikkust tähistab mitmus), ökonoomsust (muutmisel valitseb kogu paradigmas ühildumisseos) ning vormihomonüümia vältimist (nominatiiv, genitiiv ja partitiiv eristuvad sel- gemini). Arvuühildumine on olnud kasvav tendents paljude hulgasõnade puhul (*paljud inimesed*, *mitmed mehed*, Erelt 2000: 184), seega on tegemist laiema nähtusega eesti keeles. Sama meelt on olnud ka Nemvalts, kes arvabki, et *osa* ja *enamiku* mitmuslik kasutus on kujunenud sõnade *palju* ja *mitu* analoogia alusel. Peale selle peab ta struktuure *osale inimestest*, *osas autodest* keelekasutajale keerukaks, mistõttu on nende muutumine ootuspärane. (Nemvalts 1999: 760, 2000: 101)

2.2. Komplektisõna kääne

Teine varieeruv nähtus hulgasõnadega *osa*, *enamik* ja *enamus* on komplektisõna kääne. Eesti hulgafraasi tüüpiline laiend on partitiivis (nt *kaks meest*, *palju/osa mehi*). Lisaks on nimisõnalisel hulgafrasil kasutusel elatiivne laiend: *osa selts- konnast*, *enamik/enamus korteritest* (Metslang 2017: 473–474).

Elatiivset laiendit ei ole peetud partitiivse laiendiga päris võrdväärseks. On leitud, et elatiivne laiend on pigem hulgafrasiga kui terviku, mitte hulgasõna laiend (Erelt jt 2020: 493). Elatiivse laiendiga hulgafrasiga on seetõttu nähtud kui vaeg- fraasi, milles partitiivne või muus käändes komponent on jäänud sõnastamata (nt *osa poistest* täisfraas oleks *osa poisse neist poistest*) (EKG II 1993: 145). Elatiivne laiend väljendab seejuures tervikut, mille *osa* väljendab hulgafras. “Eesti keele grammatika” lisab, et elatiivne laiend võib hulgafrasist lahus esineda (EKG II 1993: 146), aga partitiivne laiend mitte, vrd näiteid (1a) ja (1b).

(1a) [---] nüüd antakse **minu palgast osa** laiskadele

(1b) *nüüd antakse **minu palka osa** laiskadele

“Eesti keele käsiraamatu” põhjal partitiivne laiend vaid nimetab objekte, ent seest- ütlevas käändes laiend märgib kindlat tervikut. Näitena on toodud hulgafras, mille peasõna on arvsõna: *Saatke kaks meest koormat maha võtma* (s.o kaks mistahes meest) vs. *Saatke kaks (neist) meestest koormat maha võtma* (s.o kaks meest mingist kindlast rühmast) (Erelt jt 2020: 492). Partitiiviga markeeritakse sel juhul pigem indefiniitset kvantifitseerimist (väljendab liiki või tüüpi), elatiiviga väljenda- takse aga definiitse hulga alamhulka – nagu on väidetud soome keele kohta (vt ka Koptjevskaja-Tamm 2001: 536–537). Hulga definiitsuse erinevus ei ole nii selge siinses artiklis vaadeldavate nimisõnaliste kvantorite puhul (*osa*, *enamik*, *enamus*), mille enda tähendus eeldab samuti tervikhulka. Lausetes *Saatke osa mehi* / *osa meestest koormat maha võtma*, *Saatke enamik mehi* / *enamik meestest koormat maha võivad* partitiivne ja elatiivne laiend tähistada sama kindlat koguhulka.

Seetõttu ei sõltu nende hulgasõnade puhul elatiivse või partitiivse laiendi valik vähemalt mitte nii selgelt semantikast.

Tiia Laanekask (2000) on näidanud bakalaureusetöös, et elatiivse komplektisõna osakaal on kvantoriga *osa* alates 1890-ndatest suurenenud, olles 1980-ndate ja 1990-ndate korpuses ülekaalus (vastavalt 66% ja 58,5%). 1990-ndate korpuses oli märgatav tendents, et elatiivi kasutati juhtudel, kui sõna *osa* ees oli omakorda omadussõnaline või omastavas käändes laiend (*suur, suurem, valdav, väike* jne); selliseid oli tema andmestikus 69% elatiiviga laienditest. Hulgasõna *enamik* esines elatiivse laiendiga harvem kui *osa* (1990-ndate korpuses 29%), ent sageli oli elatiivseks laiendiks asesõna (*enamik neist, kellest enamik*), mis partitiivse laiendiga pole võimalik. Seega võib oletada, et hulgafraasi laiendi vormi valikut mõjutavad ühtaegu mitmed erinevad tegurid, nagu nimisõnale omase eestäiendi esinemine (2) või asesõnaline laiend (3).

- (2) Sellises nukleiidis moodustavad *osa* prootoneid ja **suurem osa neut-ronitest** sisetuumad (perioodika)
- (3) [---] õnnetusi juhtub, aga **enamikul neist** pole ränka tagajärge ja lapsed saavad neist ainult õppida (foorumid)

Ka elatiivse komplektisõna kasutamist on keelekorraldus varem tõrjunud, pidades “keeleomaseks” fraase *väikesel osal taimedel, enamik andmeid, enamikul juhtudel* ning tõrjudes elatiivse laiendiga fraase *väikesel osal taimedest, enamik andmetest, enamikul juhtudest* (Kindlam 1978: 61–62). “Eesti keele käsiraamat” (Erelt jt 2020: 492) sõnastab soovitusena ettevaatlikumalt “seestütlevaga ei maksa liialdada”.

2.3. Verbi ühildumine

Mitmusliku hulgasõnaga fraas (nagu *enamikud inimesed, osad sõbrad*), mis toimib subjektina, ühildab öeldise mitmuses (4). Ainsusliku hulgasõnaga subjektifraasi korral võib öeldis olla nii ainsuses (5) kui ka mitmuses (6). “Eesti keele grammatikas” (EKG II 1993: 28) on märgitud, et kui aluseks on hulgasõna *osa, enamik* või *muist* ning selle laiend on loendatav (nt *osa inimesi, enamik inimesi*), on öeldisverb ainsuses siis, kui alus on reema (s.t lause lõpus), kui aga alus on teema, s.t paikneb lause algul rõhutus positsioonis, võib öeldis olla ka mitmuses.

- (4) [---] *osad valged šokolaadid* **tahavad** sulada väga madalal temperatuuril. (foorumid)
- (5) Selles fondis **kogub** pensionit *enamik Tuleva liikmeid ja asutajaid*. (blogid)
- (6) *Enamus puhmaid ja põõsaid* **seisavad** kenasti püsti. (blogid)

Ehkki vormiliselt hulgafraasi ainsusliku põhjaga ühilduvaks võib pidada vaid ainsuslikku verbi, võib semantiliselt ühilduvaks pidada mõlemat. Ainsusliku hulgasõnaga fraas, mis ühildab öeldise mitmuses (6), keskendub komplekti mitmele liikmele, st juhindub komplektisõna arvust. Ainsusliku hulgasõnaga fraas, mis ühildab öeldise ainsuses (5), keskendub komplekti ühele alamhulgale, st juhindub hulgasõna arvust. Sellele, et öeldise ühildamisel ja situatsiooni tõlgendamisel võivad mängida koos

rolli nii infostruktuur kui ka valitud semantiline perspektiiv, viitab ka see, et öeldise vorm võib hulgafraasi ees olla ka mitmuses (7) ning hulgafraasi järel ka ainsuses (8).

(7) Siiski **lõppevad** *enamik Jaigi novelle* õnnelikult (perioodika)

(8) *Enamik töid läheb* alltöövõtjatele, firma töölised saavad raha. (foorumid)

Laanekask (2000) on 1980-ndate ja 1990-ndate korpuse suhteliselt piiratud materjali peal näidanud, et hulgasõnaga *osa* ja loendatava mitmusliku komplekti-sõnaga hulgafrasist aluse korral kasutatakse valdavalt ainsuslikku öeldist (89%) ja seda ka lausealgulise teema positsioonis paikneva aluse korral (1980-ndatel 84%, 1990-ndatel 86%), *enamik* puhul olid need suhtarvud enam-vähem samad (1980-ndatel 98%, 1990-ndatel 89%). Küll aga leidis Laanekask, et mitmust kasutati vahel ka siis, kui alus oli reema, s.t lause lõpus. Lisaks on leitud, et *olema*-verbi arvühildumist (minevikus, sest olevikus ainsuse ja mitmuse 3. isikut ei eristata) mõjutab mitmusliku öeldistäite olemasolu. Laanekase 1990-ndate korpuse andmestikus oli mitmusliku öeldistäite korral 73,5% lausetest mitmusliku öeldisega.

Seega võib verbi ühildumist mõjutada nii subjekti tuntus, asukoht, vorm kui ka öeldistäite vorm. Potentsiaalseid ühildumist mõjutavaid tegureid võib olla veelgi, näiteks kogu konstruktsiooni muud osad (kas partitiivne või elatiivne laiend), täiendi olemasolu nimisõna ees jms. Kärt-Kristiin Jaagu (2023) analüüsis magistritöös verbi ühildumist arvsõna sisaldavate hulgafrasidega (nt *kaksteist koera koorisid käbisid*) ning leidis, et lisaks võivad ühildumist mõjutada veel verbi ajavorm (olevik või minevik), lause transitiivsus, verbi lemma, subjektifraasi referendi elusus jm. Kasutame osa neist tunnustest ka selles analüüsis.

3. Andmed

Kuna hulgafraside varieerumist ja muutumist on käsitletud ka keelekorralduslikust vaatepunktist ning esitatud on kasutussoovitusi, eeldame, et rohkem varieerumist ning soovitudele mittevastavat kasutust leiame žanridest, kus teksti planeerimiseks on vähem võimalusi või kus kirjakeele normidel ja ametlikul, formaalsel toonil on väiksem tähtsus.

Uurimuse aluseks olevad andmed pärinevad eesti keele ühendkorpuse 2023. aasta versiooni (Koppel jt 2023) veebikorpuste perioodika, blogide ja foorumite žanrimärgendiga tekstidest. Need kolm žanri on valitud esindama teksti loomise formaalsuse-informaalsuse ja spontaansuse-toimetatuse erinevaid astmeid, ehkki mitte päris puhtal kujul (näiteks sisaldab perioodika ka veebiväljaannete kommentaariumit). Andmestike koostamiseks pärisime Sketch Engine'i (Kilgarriff jt 2004, Kilgarriff jt 2014) konkordantsitööriista abil igast žanrist kõiki kontekste, kus lemmadele *osa*, *osasi*², *osas*, *enamik* või *enamus* järgnes mõni mitmuslik käändsõna, kusjuures hulgasõna ja mitmusliku käändsõna vahele võis jääda veel 0–5 nimi-, omadus- või asesõna (nt *osale inimestele*, *enamikud tähtsad inimesed*, *enamust nendest väga tähtsatest inimestest*, *enamikku neist*, *enamused hullud*). Selles uurimuses vaatleme niisiis ainult sellise sõnajärjega fraase, milles hulgasõna eelneb nimisõnale (nt *enamik õpilastest*) ega pööra eraldi tähelepanu vastupidise sõnajärjega fraasitüüpidele (nt *õpilastest enamik*). Saadud vastete hulgas valisime

² Lemmad *osas* ja *osasi* kaasati, kuna korpuses kasutatud automaatne morfoloogiline märgendus ei suuda analüüsida kirjakeele normile mittevastavat mitmuse partitiivi vormi *osasi* (nt *osasi inimesi*) kui lemma *osa* vormi. Selle asemel määratakse vormile *osasi* eraldi lemma.

iga žanri puhul lähianalüüsiks juhuslikud 750 vaatlust, välja jättsime elliptilised, ainult hulgasõnaga kasutusjuhud (9) ning kasutusjuhud, kus hulgasõna moodustas fraasi tegelikult koos ainsusliku nimisõnaga (10). Samuti jäid lõppandmestikust välja kasutusjuhud, milles otsitud lemmasid kasutati muus kui hulgasõnaga fraasi kontekstis (11, 12). Nende hulka lugesime ka elatiivse kompleksisõnaga koos esi-neva *osa* kasutuse predikatiivina, kus kontekstis on primaarsem nimisõna *osa* leksikaalne tähendus (13).

- (9) mäest alla jäid elumajad, **osad** vanad, **osad** uuemad ehitised (blogid)
- (10) Andsin nädal aega tagasi **osa oma varustust** vanematele (foorumid)
- (11) Poisid **said osa** kogemismängudest (perioodika)
- (12) Autorite seas on **enamuses** toonaste sündmuste vahetud osalised (perioodika)
- (13) Järjepidevus on **tähtis osa headest tulemustest** (blogid)

Kokku jäi lõplikku andmestikku alles 1838 vaatlust (vt tabel 1). Andmestik ja selle analüüsiks kasutatud kood on vabalt kättesaadavad OSF-i andmehoidlas leheküljel <https://osf.io/vm6j2/> (Pilvik jt 2025).

Tabel 1. Hulgasõnaga fraaside jagunemine andmestikus vastavalt hulgasõnale ja žanrile

Hulgasõna	Perioodika		Blogid		Foorumid	
	arv	osakaal	arv	osakaal	arv	osakaal
<i>osa</i>	302	50,5%	268	44,3%	274	43,1%
<i>enamik</i>	213	35,6%	169	27,9%	100	15,8%
<i>enamus</i>	83	13,9%	168	27,8%	261	41,1%
Kokku	598	100%	605	100%	635	100%

Tabelist 1 näeme, et kolmest vaadeldud hulgasõnast enim on kõigis žanrides kasutusel *osa*. *Enamiku* ja *enamuse* kasutuse osakaal korreleerub žanri tekstide toimetatuse astmega: eeldatavasti enim toimetatud tekstitüübis – perioodikas – kasutatakse oluliselt rohkem kvantorit *enamik*, spontaanseimas ja vähim kontrollitud tekstitüübis – foorumites – aga rohkem kvantorit *enamus*. Nende vahele jäävad blogid, milles kaht kvantorit kasutatakse võrdsel määral. Kvantori lemmade kasutuse erinevus žanrides on statistiliselt oluline ($X^2(N = 1838, df = 4) = 133,45, p < 0,001$, Craméri $V = 0,19$). Kui võrrelda ainult kvantoreid *enamik* ja *enamus*, on tegemist ka tugeva seosega ($X^2(N = 994, df = 2) = 128,11, p < 0,001$, Craméri $V = 0,36$).

Kõigi vaatluste puhul kodeerisime eraldi žanri, hulgasõna käände ja arvu, kompleksisõna käände, hulgasõnaga fraasi koosseisu kuuluva kompleksisõnast peasõnaga nimisõna fraasi sõnade arvu (ilma hulgasõna ja selle täienditeta)³, hulgasõna eestäiendite olemasolu, kompleksisõna lemma ja selle referendi elususe ning hulgasõnaga fraasi süntaktilise funktsiooni. Kasutasime ka Sketch Engine'i sõnasageduste leidmise tööriista (*Wordlist*), mille abil lugesime korpuse kolme žanri tekstide juhusliku 10 miljoni sõna hulgas kokku kompleksisõna lemmade esinemissagedused. Kuna väärtused olid korpuse sõnasagedustele omaselt väikeste väärtuste suhtes tugevalt kaldu, logaritmisime need, et tunnuse jaotus

³ Fraasi sõnade arvu võimalikud väärtused on mõjutatud päringu tegemise viisist, kus mitmuslikku nimisõna otsiti hulgasõnadest 1–6 sõna kauguselt. Seetõttu ei ole andmestikus väga pikki fraase ning enam kui kuut sõna sisaldavad fraasid on sinna sattunud vaid seetõttu, et sisaldasid mitmuslikku nimisõna ka kompleksisõna eestäiendite hulgas (nt *enamus nendes uutes kutsekoolides* omandatud äri- ja IT-eesliitega erialadest, perioodika).

oleks sümmeetrilisem. Seletavate tunnuste hulka kaasasime kompleksisõnade andmestikus esinemise ja kogu korpus esinemise sageduste põhjal ka mõõdiku, mis väljendab, kui ebaootuspärane on iga lemma puhul hulgasõnaga *osa*, *enamik* või *enamus* fraasi kasutus. Mõõdiku arvutamiseks võtsime negatiivse logaritmi (alusel 2) tõenäosusest kasutada uuritavate hulgasõnadega fraasi, tingimusel, et kasutatakse mingit konkreetset lemmat (sama mõõdiku kasutamist *want to* ja *wanna* varieerumise seletamisel vt Levshina, Lorenz 2022). Lemmad, mis esinevad muude süntaktiliste konstruktsioonide kõrvalt suhteliselt sagedasti ka hulgasõnaga fraasis, saavad madala ebaootuspärasuse skoori; lemmad, mille puhul hulgasõnaga fraasi kasutus on aga ebaharilik ja harv, saavad kõrge ebaootuspärasuse skoori. Kuna hulgasõnaga fraaside andmestik on väike, kipuvad mõõdiku järgi ootuspäraste sõnade hulgas domineerima korpus esinevad sõnad, mis oma 1–2 esinemiskorra juures ainult hulgasõnaga fraasides esinevadki; ebaootuspäraste sõnade hulgas omakorda domineerivad väga sagedased sõnad. Seetõttu kaalusime ebaootuspärasuse skoorid läbi kompleksisõna lemma korpusageduse pöördarvuga. Et muuta väikeste ootuspärasuse skooride suhtes kaldus jaotus ka enam-vähem sümmeetriliseks, korrutasime saadud skoorid läbi 100 000ga ning logaritmisime tulemuse. Subjekti funktsioonis fraaside puhul märkisime omakorda ära subjektiga seotud verbifraasi finiiitverbi lemma, aja ja arvu ning finiiitverbi paiknemise hulgasõnaga fraasi suhtes. Kodeerimisskeem on esitatud tabelis 2.

Tabel 2. Eesti keele ühendkorpus (2023) kogutud andmestikus kodeeritud tunnused

Tunnus	Väärtused
Žanr	perioodika blogid foorumid
Hulgasõna kääne	11 käänat (terminatiivi, essiivi ja abessiivi ei kasutatud)
Hulgasõna arv	ainsus (nt <i>enamik</i> , <i>enamikul</i>) mitmus (nt <i>enamikud</i> , <i>enamikel</i>)
Komplektisõna kääne	11 käänat (terminatiivi, essiivi ja abessiivi ei kasutatud)
Fraasi sõnade arv (ilma kvantori ja selle täienditeta)	1–10 (keskmine = 1,5, <i>SD</i> = 0,9, mediaan = 1)
Kvantori eestäiendite olemasolu	jah (nt <i>suur osa</i> , <i>valdav osa</i> , <i>teatud osa</i>) ei
Komplektisõna lemma	N = 930 (unikaalsed lemmad, sh liitsõnad)
Komplektisõna lemma logaritmitud sagedus korpus kolmes žanris	0–14,2 (keskmine = 8,8, <i>SD</i> = 3, mediaan = 9,2)
Konstruktsiooni kasutamise kaalutud ebaootuspärasus kompleksisõna lemma puhul	0–15,7 (keskmine = 7,1, <i>SD</i> = 3,7, mediaan = 6,9)
Komplektisõna referendi elusus	inimene (nt <i>inimene</i> , <i>mees</i> , <i>töötaja</i>) organisatsioon (nt <i>riik</i> , <i>pere</i> , <i>ettevõtte</i>) elus_mu (nt <i>kala</i> , <i>isend</i> , <i>bakter</i>) elutu (nt <i>auto</i> , <i>koht</i> , <i>toode</i>) elutu_abstraktne (nt <i>juht(um)</i> , <i>küsimus</i> , <i>teema</i>)

Tunnus		Väärtused
Fraasi süntaktiline funktsioon		subjekt (nt enamik inimesi tahab näha kena välja) objekt (nt seadus tühistaks enamiku abielusid) predikatiiv (nt sõdades on enamus ohvreid tsiviilelanikud) adverbiaal (nt osades valdades tegutseb varivalitsus) genitiivatribuut (nt osade vigastatute seisund on tõsine) kaassõna komplement (nt osa riikide jaoks)
Subjekti funktsioonis fraasidel	finiitverbi lemma	N = 255 (unikaalsed lemmad)
	finiitverbi lemma logaritmitud sagedus korpuse kolmes žanris	4,7–18,5 (keskmine = 15,7, SD = 2,9, mediaan = 15,9)
	finiitverbi aeg	olevik (nt <i>enamus inimesi</i> arvab/arvavad) minevik (nt <i>enamus inimesi</i> arvas/arvasid)
	finiitverbi arv	ainsus (nt <i>enamus inimesi</i> arvab/arvas) mitmus (nt <i>enamus inimesi</i> arvavad/arvasid)
	finiitverbi paiknemine	ees (nt arvab <i>enamus inimesi</i>) taga (nt <i>enamus inimesi</i> arvab)

Jagasime vaatlused kodeeritud tunnuste põhjal analüüsiks kolme põhitüüpi: 1) hulgafrasid (kus hulgasõna ei ühildu komplektisõnaga arvus, nt *enamik sõpru*, *enamikul sõpradest*), 2) nimisõnafrasid (kus hulgasõna ühildub komplektisõnaga arvus, nt *osad sõbrad*, *enamikel sõpradest*⁴) ja 3) muutumatu kvantoriga nimisõnafrasid, kus hulgasõna ei ühildu komplektisõnaga arvus ega käändes (nt *enamus sõpradel*). Ainsuse nimetavas hulgasõna ja mitmuse nimetavas komplektisõna koosinemist (*enamus inimesed*) ei ole arvestatud ühilduva fraasina, vaid muutumatu kvantoriga fraasina.

Kaassõna komplementi kui harva funktsiooni arvasime analüüsis genitiivatribuudi hulka. Predikatiivseid fraase esineb andmestikus vaid neli (kolm kvantoriga *enamus* ja üks kvantoriga *enamik*), mistõttu jätsime ka predikatiivina toimivad fraasid süntaktilise rolli kaasamisel analüüsist välja. Komplektisõna sisaldava fraasi sõnade arvu (ilma kvantori ja selle täienditeta) kaasasime analüüsi selle ebasümmeetrilise jaotuse tõttu binaarse faktortunnusena (fraasis on kas ainult üks sõna või rohkem sõnu).

4. Meetodid

Vaatleme süsteemse statistilise analüüsi abil eraldi kolme nimetatud varieerumise aspekti: 1) fraasisisest arvuühildumist, 2) komplektisõna kääned ning 3) verbi arvuühildumist subjekti rollis oleva hulgasõnaga fraasiga. Esmalt anname kõigi kolme puhul ülevaate varieeruva nähtuse suhtelisest sagedusest andmetes vastavalt žanrile ja kvantori lemmale. Seejärel kasutame tingimuslikke otsustuspuid (Hothorn jt 2006), et visualiseerida mitmetunnuselises analüüsis olulisemaid

⁴ Sellised mitmusliku kvantori ja elatiivse laiendiga fraasid oleme arvanud nimisõnafraasideks eeskätt arvus ühildumise tõttu, ent ka seetõttu, et kui näha fraasi taustal elliptilist vaegfraasi (*enamikel [sõpradel] sõpradest*), toimuks atributiivse hulgasõna ühildumine komplektisõnaga nii arvus kui ka käändes.

tunnusekomplekte, mis andmestiku vaatlusi uuritava tunnuse (näiteks ainsusliku või mitmusliku kvantori valiku) väärtuste põhjal klassidesse jagavad. Kuna üksikud puumudelid võivad olla ebastabiilsed ning sõltuda tugevalt puu hierarhilises mudelis varasemalt toimunud jagunemistest, täiendame analüüsi ka tingimuslike juhumetsadega (Breiman 2001, Strobl jt 2007), et paljude otsustuspuude agregeeritud tulemuste alusel määrata kodeeritud tunnuste hierarhia vastavalt nende seotusele uuritava tunnuse väärtuste valikuga. Juhumetsade väljastatud faktorite hierarhiat kasutame omakorda logistilise regressioonanalüüsi mudelite⁵ astmeliseks ehitamiseks (vt nt Agresti 2012: 99–172), mille abil kirjeldame lähemalt kodeeritud tunnuste üldisi, mittehierarhilisi pea- ja koosmõjusid uuritava tunnuse väärtuste valikule. Selles uurimuses kasutatud statistilised meetodid kuuluvad nn sagedusliku tõenäosusteooria (ingl *frequentist statistics*) meetodite hulka, mis tähendab, et statistiliselt oluliste seoste tuvastamiseks kasutame olulisuse tõenäosusi ehk *p*-väärtusi ning loeme seose oluliseks ja mittejuhuslikuks, kui *p*-väärtus on väiksem kui 0,05. Mudelite juures raporteerime nii uuritava tunnuse sagedama taseme suhtelise sageduse valitud alamandmestikus (*lähteosakaal*) kui ka mudeli täpsuse mõõdikud, mille abil hinnata, kui palju seletavate tunnuste kaasamine aitab pelgalt lähteosakaalu kasutamisest paremini uuritava tunnuse väärtusi ennustada. Headuse näitajatenä kasutame klassifitseerimistäpsust (*täpsus*), mis on otseselt võrreldav lähteosakaaluga, väljendades uuritava tunnuse tegelike väärtuste ning mudeli ennustatud väärtuste kokkulangevuse protsenti, ning nn *AUC*-indeksit (*AUC*), mis on paindlikum mudeli eristusvõime mõõdik. Analüüsiks kasutame programmeerimiskeelt ja rakendustarkvara R (R Core Team 2024). Otsustuspuude tegemiseks on kasutatud selles artiklis R-i paketti *partykit* (Hothorn, Zeileis 2015), visualiseerimiseks paketti *ggparty* (Borkovec, Madin 2019). Juhumetsade tegemiseks on kasutatud R-i paketti *party* (Strobl jt 2007).

5. Tulemused

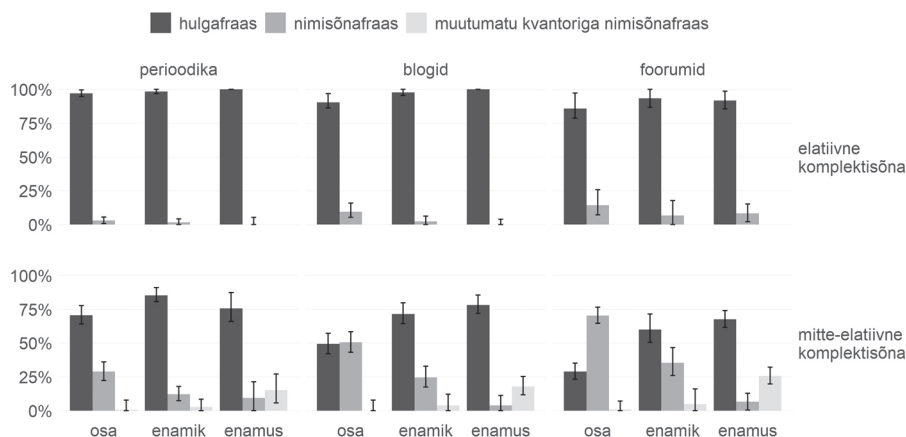
5.1. Fraasisisene arvuühildumine

5.1.1. Arvuühildumise sagedus vastavalt žanrile, hulgasõnale ja kompleksisõna käände

Vaatleme esmalt fraasimallide suhtelisi sagedusi žanride ja hulgasõnade lõikes. Elatiivse kompleksisõna ja mitte-elatiivse hulgasõnaga fraase (*enamikul sõpradest*) vaatame mitte-elatiivse kompleksisõnaga fraasidest eraldi.

Elatiivse kompleksisõnaga fraasid (vt joonis 1 ülemine paneel) on ülivaldavalt ilma arvuühildumiseta hulgafrasid. Teistest hulgasõnadest eristub *osa*, millega esineb nimisõnafrasalist arvus ühilduvat kasutust mõnevõrra rohkem (14), eriti vähem toimetatud tekstides. Muutumatu kvantoriga nimisõnafraase elatiivse kompleksisõnaga ei esine.

⁵ Keeleteaduses ja muudeski valdkondades on üha enam tavaliste regressioonanalüüsi mudelite asemel kasutusel nn segamõjudega mudelid (ingl *mixed-effects models*), mis võimaldavad arvesse võtta vaatluste juhuslikku klasterdumist näiteks kõnelejati või andmestikus esinevate konkreetsete sõnade põhjal. Selliste mudelite kasutamine aitab leevendada andmestikus sagedasti esinevate üksikkõnelejate või sõnade käitumismustrite mõju üldiste, populatsioonitasandil kehtivate tendentside tuvastamisel. Käesolevas uurimuses ei ole segamudelite kasutamine paraku mõistlik, kuna ühendkorpuse veebikorpuste materjalid ei võimalda enamasti tuvastada teksti autorit ning eri korpusdokumente identifitseerivate koodide põhjal pärinevad peaaegu kõik vaatlused erinevatest dokumentidest. Samuti on andmestikus esinevate lemmade jaotus tugevalt vaid 1–2 korda esinevate lemmade suunas kaldu, mistõttu ei ole ka üksiksõnade juhusliku mõjuna segamudelitesse kaasamine kohane. Nõnda lähtume andmeanalüüsil eeldusest, et suurest korpusest võetud juhuvalimi vaatlused on üksteisest sõltumatud.



Joonis 1. Fraasimallide osakaal vastavalt žanrile, hulgasõnale ja kompleksisõna käände

- (14) Vähemalt **osad lugejatest** saavad ilmselt aru, miks ma marru läksin. (foorumid)

Mitte-elatiivse kompleksisõnaga fraaside puhul (vt joonis 1 alumine paneel) näeme oluliselt enam varieerumist. Perioodikas domineerib üldiselt hulgafraas, s.t arvuühildumiseta mall. Võib oletada, et perioodikas on tekste enim toimetatud, mistõttu hulgafraasi valdavus on ootuspärane. Nimisõnafraasi esineb aga enim foorumites, kus teksti vastavus õigekeelsusallikates esitatud normingutele on vähem oluline.

Hulgasõnadest ühildub kompleksisõnaga ehk esineb nimisõnafraasis kõikides žanrides kõige sagedamini *osa*. *Enamus* esineb nimisõnafraasis peamiselt muutumatuna (15), hulgasõnu *osa* ja *enamik* kasutatakse muutumatuna harva (16).

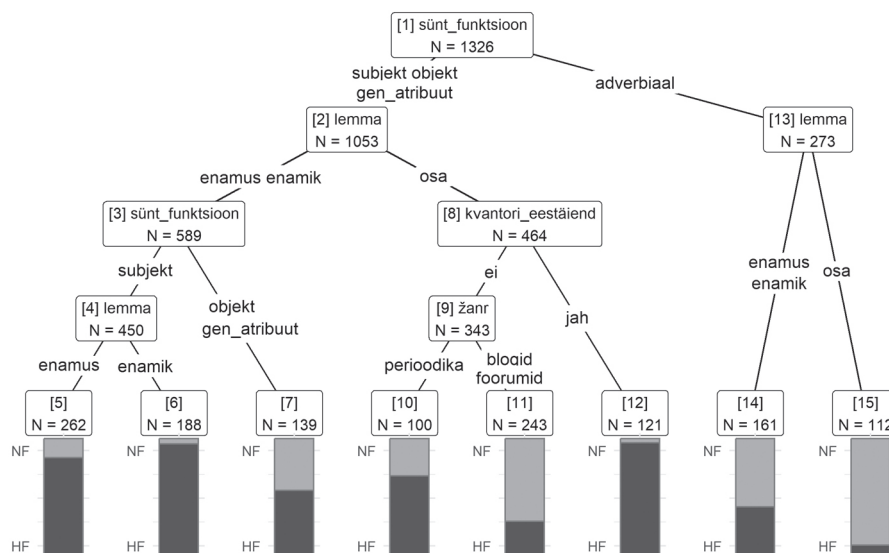
- (15) Eks kuritegevusega on ikka nii, nagu **enamus turismipiirkondades**. (foorumid)
- (16) Mina mängisin ka selle nüüd läbi ja ütleks ära et **osa kohad** olid täiesti utoopiad. (foorumid)

Hulga- ja kompleksisõna arvuühildumist on seostatud fraasi määratleva funktsiooniga (vt Erelt, Metslang 1998). Et *osa* ühildub kompleksisõnaga sagedamini kui *enamik* ja *enamus*, kasutatakse *osa* ka sagedamini määratlemiseks. Selle poolest sarnaneb *osa* näiteks hulgasõnaga *mõni* või inglise keele *some*. Žanripõhiste suhte-liste ühildumissageduste seostamisest määratleva või kvantifitseeriva funktsiooniga aga ei piisa, et arvuühildumist seletada. Järgnevalt vaatlemegi, millised kodeeritud tunnused veel nimisõna- või hulgafraasilist kasutust (s.t arvuühildumist, vt osa 2.1) mõjutavad.

5.1.2. Arvuühildumise mitmetunnuseline analüüs

Kuna elatiivne kompleksisõna tingib peaaegu alati hulgasõna ainsuse, vaatleme mitmetunnuselises analüüsis ainult mitte-elatiivse kompleksisõnaga kasutusjuhte. Muutumatu kvantoriga fraase on materjalis vähem kui teisi fraase ja need seostuvad peamiselt hulgasõnaga *enamus*, mistõttu oleme need analüüsis ühendanud

nimisõnafraasidega, kus hulgasõna ühildub komplektisõnaga mitmuses (*osad, enamikud, enamused*). See tähendab, et lühendiga *HF* viitame klassikalisele hulgafraasile ja lühendiga *NF* viitame mitmusliku või muutumatu hulgasõnaga nimisõnafraasile.



Joonis 2. Mitte-elatiivse laiendiga hulgafraasi (HF) ja nimisõnafraasi (NF) valikut mõjutavate tegurite otsustuspuu (HF lähteosakaal = 0,62, täpsus = 0,77, AUC = 0,84)

Arvuühildumise otsustuspuu mudelist (vt joonis 2) näeme, et esmalt lahkneb andmestik kaheks seoses fraasi süntaktilise funktsiooniga (1. sõlm): paremas harus on adverbiaalid, millele on omane NF-mall, vasakusse harusse jäävad teised lause liikmed. Mõlemad harud jagunevad omakorda kaheks hulgasõna lemma alusel, kuna *osa* esineb sagedamini nimisõnafraasis: adverbiaalina suisa 89% (15. sõlm), samas kui *enamik* ja *enamus* varieeruvad rohkem (43% HF, 57% NF, 14. sõlm). Enamik hulgasõnu, mis esinevad fraasis, mis toimib adverbiaalina, on seega mitmuslikud ja moodustavad nimisõnafraasi (17), eriti hulgasõna *osa* puhul (18).

(17) Rattad on koopas **enamustel BMW-del** mis on stock seades. (foorumid)

(18) Trenniga jäin rahule, sain jällegi raskust juurde lisada **osadele harjutustele** [---] (foorumid)

Asjaolu, et arvuühildumist esineb enim semantiliste käänete vormis ja adverbiaali funktsioonis, on ootuspärane: ka hulgafraasina toimivad fraasid ühilduvad neis käändes, aga mitte arvus, nt *osale harjutustele*. Seega on arvuühildumise teke adverbiaali kontekstis ootuspärane nii käändeühildumise olemasolu kui ka teiste mitmusliku hulgasõnaga fraaside (*mitmetel BMW-del, paljudele harjutustele*) analoogia tõttu, nagu on leidnud ka Nemvalts (1999, 2000). Pole eriti tõenäoline, et arvuühildumise väga sage esinemine adverbiaalides oleks tõlgendatav määratleva funktsiooni ülekaaluga.

Vaadates joonise vasakpoolset haru, näeme, et ka teistes süntaktilistes rollides esineb *osa* sagedamini NF-mallis kui *enamik* ja *enamus*, eriti blogides ja foorumites

(11. sõlm). Eestäiend välistab aga *osa* esinemise nimisõnafraasis peaaegu täielikult: eestäiendiga *osa* moodustab hulgafrasi (*suurem osa inimesi, mingi osa numbrit, arvestatav osa finantsvarasid*, 12. sõlm). Eestäiend laiendabki hulgasõna *osa* andmestikus oluliselt sagedamini (318 juhul 846st) kui hulgasõnu *enamik* (10/483) ja *enamus* (18/512).

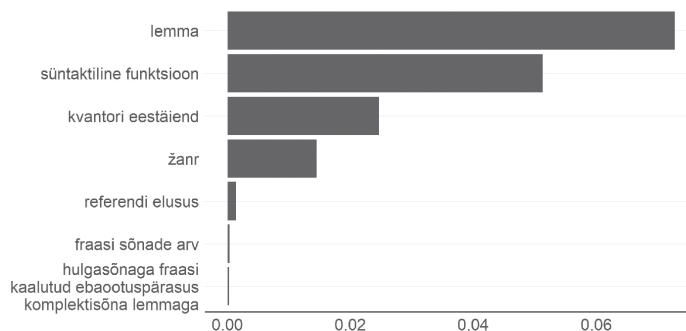
Eestäiendiga hulgasõna moodustab seega hulgafrasi. Ilma eestäiendita kasutatult moodustab hulgasõna ka nimisõnafraase. Eestäiendita hulgasõna on hõlpsamini tõlgendatav määratlejana, eriti sõna *osa* puhul, mis ka kõige sagedamini nimisõnafraasis esineb (vt joonist 1). Määratlejana on selle sõna tähendus semantiliselt pleekinud ja sõna ise grammatiseerunud ning pole enam käsitletav nimisõnana, seetõttu ei saa see võtta enam ka nimisõnale omast täiendit, vaid toimib sarnaselt muude hulgasõnast pärit määratlejatega ise täiendina (vrd *mõni*), vt ka näide (19).

(19) **Osad autod** parkisid poe juurde, edasi enam ei vedanud mäest. (blogid)

Enamik ja *enamus* moodustavad lisaks adverbiaalile nimisõnafraasi sagedamini ka siis, kui fraas toimib genitiivtribuudina (20) või objektina (7. sõlm).

(20) Samas on meil **enamike spetsiifiliste teenuste** koguhind pääliinaga võrreldes mitmekordne, [---] (foorumid)

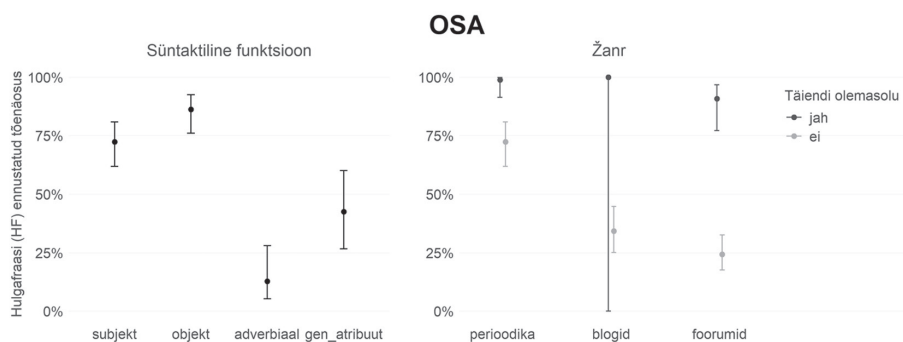
Seletavate tunnuste üldist olulisust on mõõdetud **tingimusliku juhumetsa mudeliga** (vt joonis 3). Juhumetsa mudelis on tunnuste olulisust hinnatud selle põhjal, kui palju konkreetse tunnuse väärtuste juhuslik segiajamine mõjutab mudeli võimet uuritava tunnuse klasse (nt HF vs. NF) eristada, kasutades mõõdikuna nn ROC-kõvera alust pindala (*Area Under the Curve* ehk AUC, vt Janitza jt 2013). Näeme, et metsamudeli põhjal on kõige olulisem tunnus hulgasõna lemma. See kinnitab, et kolm hulgasõna – *osa*, *enamik*, *enamus* – käituvad hulgafrasi või nimisõnafraasi valiku suhtes erinevalt. Peale hulgasõna lemma on olulised fraasi funktsioon, eestäiendi olemasolu ja žanr. Muud tunnused (referendi elusus, hulgafrasi koosseisu kuuluva ja komplektisõna sisaldava nimisõnafraasi sõnade arv ning fraasi ootuspärasus komplektisõna lemmaga) ei paista selles varieerumises rolli mängivat.



Joonis 3. Fraasimalli valikut ennustavate tunnuste AUC-olulisus tingimusliku juhumetsa mudelis (hulgafrasi lähteosakaal = 0,62, täpsus = 0,83, AUC = 0,9)

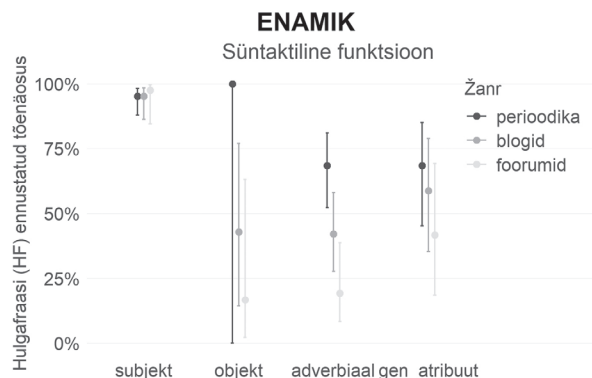
Kuna erinevad hulgasõnad näivad käituvat erinevalt, hindasime **logistilise regressiooni** abil kontekstuaalsete tunnuste mõju fraasisisesele ühildumisele iga

hulgasõna kohta eraldi. Kõik regressioonimudelid ennustavad hulgafraasi esinemise tõenäosust, olenemata sellest, kas konkreetset hulgasõna kasutatakse sagedamini hulgafraasis või nimisõnafraasis.



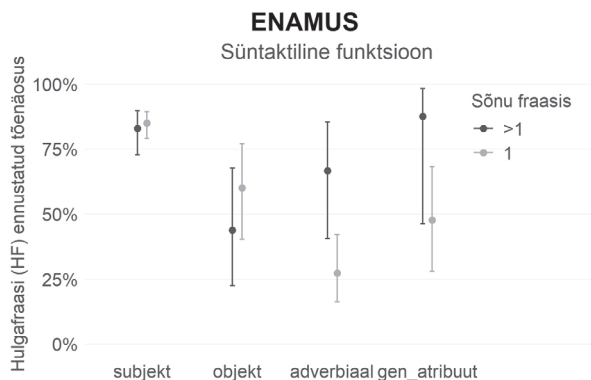
Joonis 4. Fraasimalli valikut mõjutavad tegurid kvantoriga *osa* (nimisõnafraasi lähteosakaal = 0,53, täpsus = 0,81, AUC = 0,9)

Näeme (vt joonise 4 vasakpoolne paneel), et subjekti ja objektina moodustab **osa** tõenäolisemalt hulgafraasi (kus *osa* on ainsuses), genitiivtribuudi ja eriti adverbiaalina moodustab *osa* aga pigem nimisõnafraasi (kus *osa* on mitmuses, vt näide 18). Ka selles mudelis näeme, et eestäiendi olemasolu (parempoolne paneel) välistab hulgasõna mitmuse, mis tähendab, et eestäiendiga *osa* moodustab hulga-
fraasi. Eestäiendita hulgasõnade puhul on aga blogides ja foorumites hulgafraasi kasutamise tõenäosus oluliselt madalam kui perioodikas.



Joonis 5. Fraasimalli valikut mõjutavad tegurid kvantoriga *enamik* (hulgafraasi lähteosakaal = 0,74, täpsus = 0,82, AUC = 0,87)

Ka hulgasõna **enamik** kasutust mõjutavad fraasi süntaktiline roll ja žanr, samuti nende koosmõju (vt joonis 5). Subjektina toimiv fraas on olenemata žanrist tõenäoliselt hulgafraas. Objekti rollis on pilt segasem, sest usaldusvahemikud katuvad ja on väga laiad: andmestikus on liiga vähe vaatlusi objektina toimivatest *enamik*-fraasidest. Adverbiaali rollis moodustab *enamik* aga just foorumites kõige tõenäolisemalt hoopis mitmusliku kvantoriga nimisõnafraasi. Foorumid erinevad teistest žanridest ka genitiivtribuutide puhul (vt näide 20).



Joonis 6. Fraasimalli valikut mõjutavad tegurid kvantoriga *enamus* (hulgafräasi lähteosakaal = 0,72, täpsus = 0,78, AUC = 0,74)

Ka hulgasõna ***enamus*** kasutuses on hulgafräasi mall sagedaim subjekti rollis (vt joonis 6). Oluline, ennekõike adverbiaali rollis, on fräasi pikkus: (mitmusliku või muutumatu kvantoriga) nimisõnafräas on oluliselt tõenäolisem fräasis, kus komplektisõna on ilma täienditeta (nt *enamustel BMW-del*, vt näide 17). Sarnane, kuigi nõrgem tendents valitseb genitiivatribuudi rollis. Erinevalt teistest hulgasõnadest ei tõuse *enamuse* kasutuses esile žanri mõju. Selle põhjus on tõenäoliselt see, et hulgasõna *enamus* on juba žanriliselt määratud: seda kasutatakse peamiselt toimetamata tekstides.

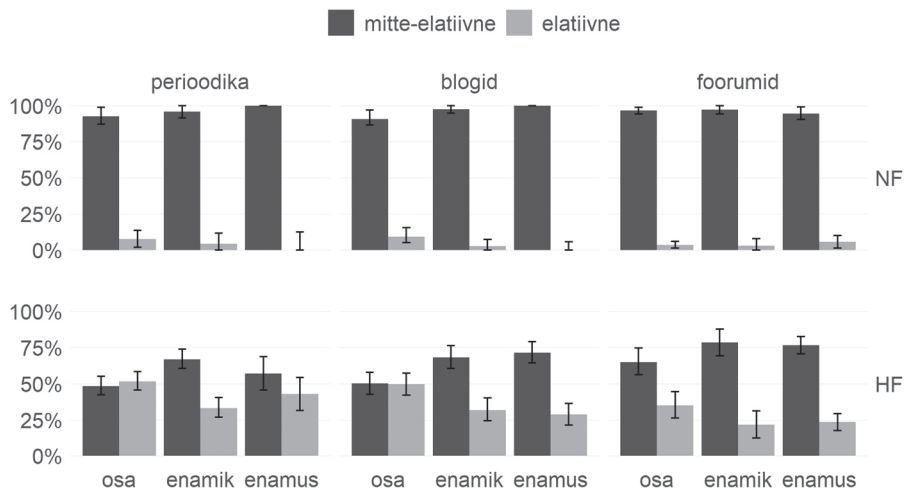
Kokkuvõttes on fräasimalli valikul oluline hulgasõna ise. Samuti on üldiselt oluline fräasi süntaktiline roll: adverbiaali ja genitiivatribuudi rollis esineb tõenäolisemalt mitmusliku kvantoriga NF. Nimisõnafräasi sagedus adverbiaali rollis osutab, et adverbiaali roll on hulga- ja komplektisõna arvühildumise leviku tõenäoline kese. Muutus võib olla tingitud sellest, et adverbiaali rollis valitseb käändeühildumine juba ka hulgafräasis (*osale inimestele*). Nagu juba Nemvalts (1999, 2000) on maininud, on selline ebasümmeetria, käändeühildumine ilma arvühildumiseta, keelekasutajale keeruline. Tõenäoliselt on käändes ühilduvad hulga- ja komplektisõnad hakanud ühilduma ka arvus (*osadele inimestele*). Siit on see võinud kanduda nimetavasse käändesse, kus hulgasõna *osa* on ümber tõlgendunud umbmäärast hulka väljendava määratlejana, mitte enam kvantorifräasi peasõnana, ning toimib nimisõnafräasis hoopis ise komplektisõna täiendina (*osad inimesed*). Sellel protsessil on analoogia ja eeskuju seoses hulgasõnaga *palju*, mis on läbinud samasuguse protsessi juba varem (vt Lõbu 2011). Kui aga kvantorit *osa* laiendab eestäiend, blokeerib see mitmusliku kvantori võimaluse. Sel juhul käitub *osa* hulgafräasi nimisõnalise põhisõnana, mis moodustab tavalise hulgafräasi (*suur osa inimesi, väiksele osale inimestele*).

5.2. Komplektisõna kääne

5.2.1. Komplektisõna käänete sagedus vastavalt žanrile, hulgasõnale ja fraasimallile

Nagu eespool öeldud (vt 2.2), on hulgafraasi laiend tüüpiliselt (meie valimis mitmuse) partitiivis, kui hulgasõna on nominatiivis, partitiivis või objekti rollis genitiivis, nt *enamik inimesi*, *enamikku inimesi*, (*päästis*) *enamiku inimesi*. Muudes käänetes (ja genitiivis, kui fraas ei toimi objektina) ühilduvad hulgafraasi moodustajad käändes, nt *enamikule inimestele*, *enamiku inimeste (jaoks)*. Komplektisõna võib hulgafraasis olla aga ka elatiivis (*enamik inimestest*, *enamikule inimestest*). Järgnevalt vaatame lähemalt, millest sõltub komplektisõna kääne hulgasõna sisaldavas fraasis.

Joonisel 7 on eristatud nimisõnafraasi ja hulgafraasi üldine jaotumine samal viisil nagu eelmises analüüsis (vt 5.1.1).



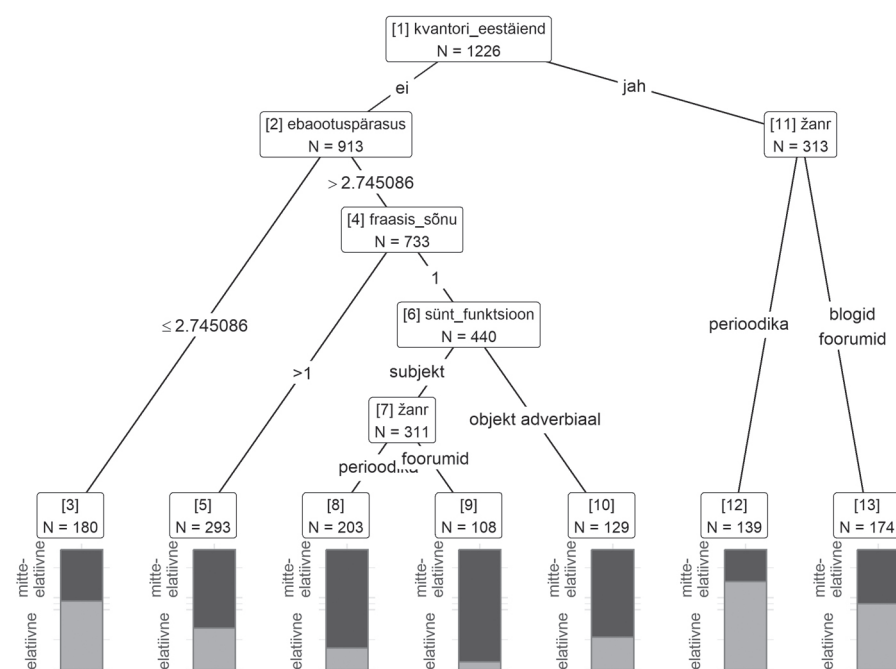
Joonis 7. Komplektisõna käände varieerumine vastavalt žanrile, hulgasõnale ja fraasimallile

Jooniselt 7 näeme joonise 1 omamoodi peegelpildina, et nimisõnafraasis esineb elatiivset komplektisõna marginaalselt: elatiivne laiend ei soosi mitmusliku või muutumatu hulgasõnaga kasutust. Kui lähtuda eeldustest, et komplektisõna elatiivne vorm (nt *inimestest*) väljendab komplekti terviklikkust ja spetsiifilisust ning et hulgasõna, mis ühildub mitmusliku komplektisõnaga arvus (nt *osadel*), määratleb komplekti liikmeid eraldi, on elatiivne hulgafras ja nimisõnafras komplemendaarsed võimalused, kuidas vaadelda mitmeliikmelise komplekti osahulka. Sellel juhul võib elatiivse komplektisõna ja mitmusliku hulgasõna koosinemise puudumist seletada kahe perspektiivi kokkusobimatusega või vähemasti sellise koosinemise kognitiivse kompleksusega. Ülal kirjeldatud eeldustel nõuaks fraas, mille moodustavad mitmuslik hulgasõna ja elatiivne komplektisõna, seda, et kasutaja perspektiiv nihkub keset fraasi osahulga määratletud, ent nimetamata liikmetelt (*osadel [inimestel]*) spetsiifilisele ja definiitsele komplektile-koguhulgale (*[piiritletud hulgast] inimestest*).

Hulgafrasid varieerub kompleksisõna kääne rohkem, ent hulgasõna lemma ja žanriga seoses kompleksisõna käändes selgeid eelistusi ei ilmne: elatiivset kompleksisõna esineb mõnevõrra rohkem perioodikas ning kõigis žanrides seoses kvantoriga *osa*.

5.2.2. Komplektisõna käänete mitmetunnuseline analüüs

Komplektisõna käänte valiku mitmetunnuselises analüüsis uurisime lähemalt edasi vaid hulgafrase, jättes kõrvale nimisõnafraseiks liigitatud mitmusliku ja muutumatu hulgasõnaga kasutusjuhud, kus kompleksisõna kääne ei varieeru. Jätsime välja genitiivtribuudi ja selle alla liigitatud harvaesinevad kaassõna genitiivse komplemendi funktsioonis hulgafrased, mis ei luba arvatavasti nii semantilise kui grammatilise kompleksuse tõttu järelpositsioonis elatiivset kompleksisõna (**enamiku inimestest autod*, **enamiku inimestest juurde*, aga ?*inimestest enamiku autod*, *inimestest enamiku juurde*). Eessõnadega, kus kaassõnafrase eessõnast peasõna ning selle laiendiks oleva hulgafrase kvantoriga peasõna paiknevad lähedikkude, näib kompleksisõna elatiivne järellaiend vastuvõetavam (?*peale enamiku õpetajatest*), ent ükski hulgafrase andmestikus eessõna laiendina ei esinenud.



Joonis 8. Komplektisõna käänet mõjutavate tegurite otsustuspuid (mitte-elatiivse laiendi lähteosakaal = 0,61, täpsus = 0,7, AUC = 0,74)

Jooniselt 8 näeme, et **komplektisõna käänte otsustuspuid** hargneb esmalt selle järgi, kas hulgasõnal on eestäiend (sõlm 1). Hulgasõna eestäiend tõstab oluliselt kompleksisõna elatiivivormi tõenäosust, eriti perioodikas, aga ka teistes žanrides (vt sõlmed 12, 13). On võimalik, et eestäiendiga, st selgelt nimisõnaline kvantor

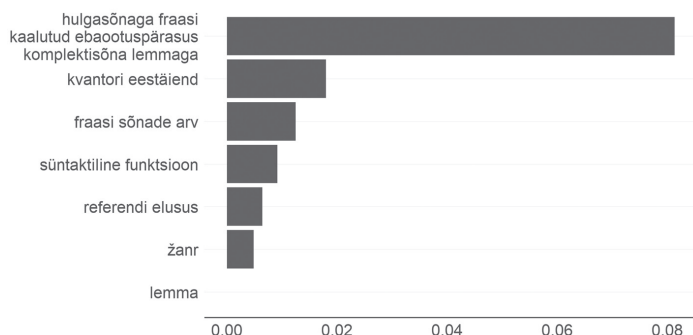
tõstab rohkem esile hulga osalisuse ja seotuse kindla tervikkomplektiga, mille väljendamisel eelistatakse elatiivi (21), (22). Ilma eestäiendita hulgasõna on aga vabam täitma määratlevat funktsiooni, nagu nägime ka eelmises analüüsis. Igal juhul saab kinnitust tendents, mille Laanekask (2000) välja tõi: kvantori omadussõnaline või omastavas käändes laiend suurendab komplektisõna elatiivivormi tõenäosust.

- (21) Seega on **suur osa kroonilistest haigustest** välditavad [---] (perioodika)
 (22) Sest **valdav enamik debati osalistest** ei tulnud ausalt ja avatult ideed kaaluma, [---] (perioodika)

Kui kvantoril eestäiendit ei ole, toimub järgmine hargnemine selle järgi, kui (eba)ootuspärane on hulgasõnaga fraas komplektisõna lemma seisukohast (sõlm 2). Ootuspärane (ehk madala ebaootuspärasuse skooriga) komplektisõna, mille alla kuuluvad 3. sõlmes põhiliselt sõna *inimene* ja asesõnad (*need, meie, nemad*), tingib enam elatiivset kasutust kui mitteootuspärane. Asesõnaliste komplektisõnade elatiivilembust on eelkõige seoses hulgasõnaga *enamik* märkinud ka Laanekask (2000), põhjendades seda partitiivi kasutamise võimatusega (nt **enamik neid*). Vähem ootuspärase komplektisõna korral jagunevad ilma eestäiendita vaatlused edasi selle alusel, kui pikk on hulgasõna laiendav komplektisõna sisaldav nimisõnafraas (sõlm 4). Pikemas kui ühesõnalis fraasis on komplektisõna elatiivis sagedamini (sõlm 5) kui siis, kui komplektisõna paikneb vahetult hulgasõna järel. Näites (23) on kvantori ja komplektisõna vahel 2 muud sõna. Ühelt poolt võib elatiiv markeerida selgemalt pika fraasi moodustajate piiri. Teiselt poolt võib pikem fraas komplekti spetsifitseerida, nii et komplektisõna, mida väljendab pikem fraas, väljendab ühtlasi spetsiifilisemat ja tuntumat komplekti, mida vormistatakse sageli elatiiviga.

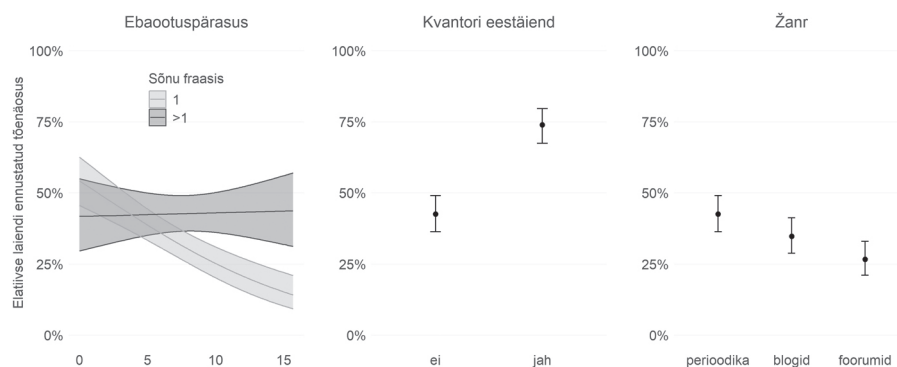
- (23) **Enamik separatistide ridades võitlejatest** on Venemaa sakslased ehk etnilised sakslased, kes on [---] (perioodika)

Ühesõnalis, ainult komplektisõnast koosnevas nimisõnafraasis, mis laiendab ilma eestäiendita hulgasõna, mõjutab komplektisõna käändevormi hulgafraasi süntaktiline roll: objekti ja adverbialina toimivas fraasis (sõlm 10) on komplektisõna elatiiv pisut tõenäolisem (27,7%) kui subjektina toimivas fraasis. Viimaks (sõlm 7) näeme eestäiendita hulgasõnast ja ühest komplektisõnast koosnevate ja subjektina toimivate hulgafraside puhul seda, et perioodikas on komplektisõna elatiiv õige pisut tõenäolisem kui blogides ja foorumites.



Joonis 9. Komplektisõna käände valikut ennustavate tunnuste AUC-olulisus tingimusliku juhumetsa mudelis (mitte-elatiivse komplektisõna lähteosakaal = 0,61, täpsus = 0,79, AUC = 0,87)

Joonisel 9 on **juhumetsa mudel**, mis visualiseerib kõigi tunnuste suhtelist mõju hulgafrasid esineva komplektisõna käände. See kinnitab suures osas eespool mainitud: komplektisõna elatiivi ja teiste käänete (partitiiv jm) varieerumist mõjutavad enim hulgafrasi (eba)ootuspärasus seoses komplektisõnaga, hulgasõna eestäiendi olemasolu ja komplekti väljendava fraasi pikkus.



Joonis 10. Komplektisõna käände valikut mõjutavad tegurid (mitte-elatiivse komplektisõna lähteosakaal = 0,61, täpsus = 0,69, AUC = 0,71)

Joonisel 10, mis kuvab komplektisõna käänet ennustava **logistilise regressiooni-mudeli** kaht olulist peamõju ning üht koosmõju, näeme otsustuspäru juures nähtule sarnaseid tendentse. Kui hulgasõna ei paikne hulgafrasid vahetult komplektisõna ees (s.t komplektisõnal on eestäiendeid), on elatiivi kasutamise tõenäosus suurem, ent see kehtib vaid fraasides, mille komplektisõnaga ei ole hulgasõnaga fraasi kasutamine väga ootuspärane (vt joonise 10 vasakpoolne paneel). Elatiivse komplektisõna kasutamist suurendab oluliselt ka hulgasõna eestäiendi olemasolu (keskmine paneel, vt näide 21). Erinevused on ka žanride vahel, nimelt on perioodikas suurem tõenäosus kasutada elatiivset komplektisõna kui teistes žanrides (parempoolne paneel).

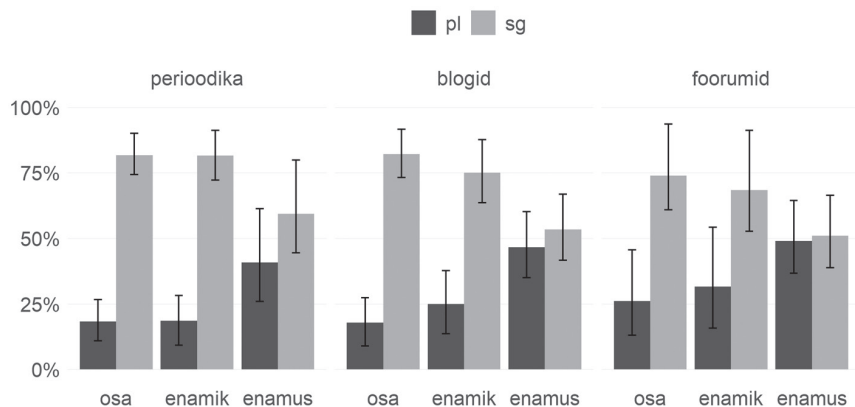
Tuleb aga tähele panna, et elatiivse ja mitte-elatiivse komplektisõna vaheldumise mudelite headuse näitajad ei ole väga kõrged, mis viitab sellele, et käänevaheldust mõjutab muidki tegureid, millega andmestiku kodeerimisel pole arvestatud.

5.3. Verbiühildumine

5.3.1. Verbiühildumise sagedus vastavalt žanrile ja hulgasõnale

Verbiühildumise varieerumist (vt 2.3) vaatlesime ainult kontekstides, milles öeldise vorm varieerus. Seetõttu analüüsisime vaid lauseid, milles subjektiks olid ainsusliku hulgasõnaga hulgafrasid ja predikaadiks verbid kindla kõneviisi jaatavates vormides, kus ainsus ja mitmus selgelt eristuvad. Alamandmestiku kitsendamise vajadus tulenes statistilistest mudelitest, mille kasutamisel tekitavad mittevarieeruvad kontekstid nii tehnilisi kui ka kontseptuaalseid probleeme. Ehkki kogu andmestikus oli valdav enamik fraasidest subjekti rollis (1838-st 1114 ehk 60%), jäi verbiühildumise varieerumise uurimist võimaldavaid vaatlusi nõnda alamandmestikku vaid 414, neist 161 kvantoriga *osa*, 117 kvantoriga *enamik* ja 136 kvantoriga *enamus*. Nendega

kasutatud finiiitverbi vormide ainsuse (*sg*) ja mitmuse (*pl*) jaotust vastavalt žanrile ja kvantori lemmale kuvab joonis 11.

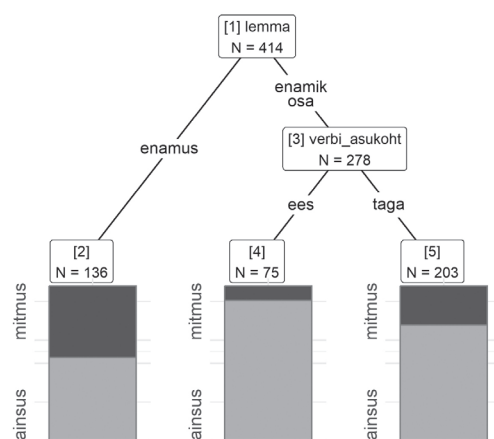


Joonis 11. Finiitverbi vorm subjekti rollis hulgafrasis korral vastavalt žanrile ja hulgasõnale

Ehkki mitmuslikke verbivorme kohtame perioodikast (22%) veidi enam foorumites (40%) ja blogides (31%), on hulgasõna lemmade lõikes ühildumismustrid kolmes žanris sarnased: enim mitmusliku öeldise kasutust kohtame sõna *enamus* puhul, *osa* ja *enamiku* puhul eelistatakse selgemini ainsuslikku öeldist. Seda on keeruline põhjendada, ent vahest on keelekorralduses kvantorina tõrjutud hulgasõna *enamus* ning blogides ja foorumites sagedasema mitmusliku verbivormi seotuse vahendajaks tekstide vähesem toimetatus.

5.3.2. Verbiühildumise mitmetunnuseline analüüs

Mitmetunnuselises analüüsis testisime nii žanri kui lemma mõju statistiliselt. Lisaks kaasasime seletavate tunnustena verbi asukoha hulgasõnaga fraasi suhtes, verbi ajavormi, verbi logaritmitud sageduse korpuses, subjekti referendi elususe, kompleksisõna käände (elatiiv või mitte-elatiiv) ja hulgasõna eestäiendi olemasolu.



Joonis 12. Subjekti rollis hulgafrasis finiiitverbi vormi valikut mõjutavate tegurite otsustuspuu (ainsusliku vormi lähteosakaal = 0,71, täpsus = 0,71, AUC = 0,67)

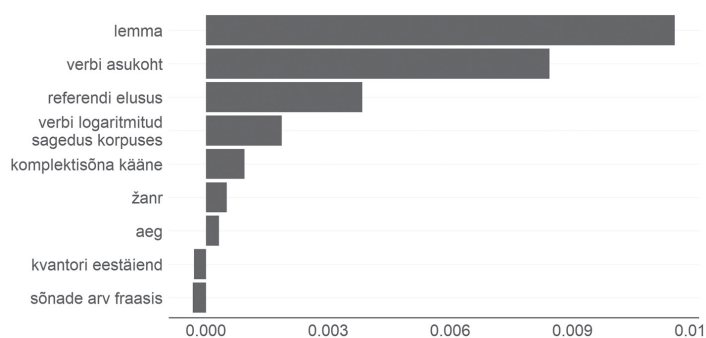
Ka **verbiühildumise tingimusliku otsustuspuu** mudeli põhjal (vt joonis 12) näeme põhilist erinevust hulgasõnaga *enamus* fraaside kasutuses (sõlm 2), mis esinevad subjektina koos mitmuslike finiiitverbi vormidega (*enamus inimesi lähivad*) teiste hulgasõnadega fraasidest oluliselt enam (46%). Hulgasõnade *enamik* ja *osa* puhul on ainsuslikud vormid oluliselt tõenäolisemad, seda eriti juhul, kui finiiitverb paikneb hulgasõnaga fraasi ees (sõlm 4, *läheb enamus inimesi*). Seega tuleb ka meie andmetest välja, et hulgafraasi asukoht mõjutab teatud määral verbi arvuühildumist: kui hulgafraas paikneb verbielses prototüüpses subjekti positsioonis (24) ja verb seega hulgafraasi taga, mõjutab fraasi sisuline mitmuslikkus sagedamini mitmuslikku verbivormi valima kui lause lõpupoole paikneva subjekti puhul (25). Sama tuli välja ka Laanekase tööst (2000) ja Kärt-Kristiin Jaagu (2023) analüüsist arvsõna sisaldava hulgafraasi arvuühildumise kohta.

(24) *Enamik Briti poliitikuid kõigist parteidest tahtsid* jääda EL-i. (perioodika)

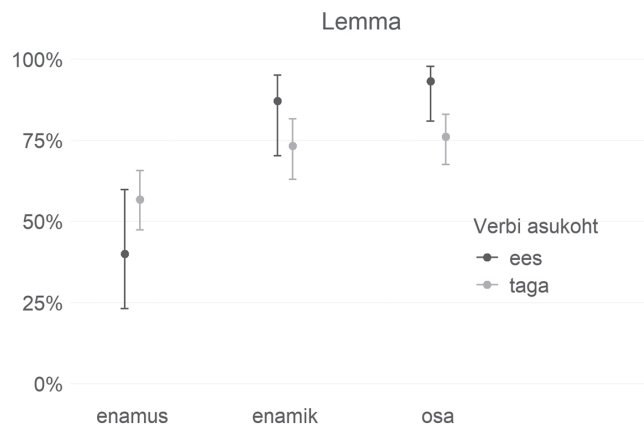
(25) Kindlasti **tuleb** tagasi *osa pensionile jäävaid inimesi*. (perioodika)

Tuleb siiski mainida, et meie andmestikus oli subjekti ees paikneva verbi korral mitmusevorme rohkem kui näiteks Laanekase (2000) andmestikus. Kui verb paiknes subjekti ees, oli meie andmestikus mitmusliku verbiga lauseid 22%; kui vaadelda ainult hulgasõnu *enamik* ja *osa*, siis 9%. Erinevus võib tuleneda sellest, et kui Laanekask käsitles eraldi vaid reema positsioonis (lause lõpus, näide 25) paiknevaid subjekte, mis üksikutele juhtudel tema andmestikus ühildusid verbiga mitmuses, siis meie vaatlesime vaid verbi ja subjekti omavahelist asukohta, sh juhte, kus verb paikneb subjekti ees V2-tendentsist tingitud aluse ja öeldise inversiooni tõttu (26). Sel juhul säilitab subjekt lause teema-reema suhetes oma teema funktsiooni (vt Lindström 2017) ning mitmuse kasutamine on pigem ootuspärane, sarnaselt verbielse subjektiga.

(26)=(7) Siiski **lõppevad** *enamik Jaigi novelle* õnnelikult või vähemalt lootusriikka noodiga.



Joonis 13. Subjekti rollis hulgafraasiga finiiitverbi vormi valikut ennustavate tunnuste AUC-olulisus tingimusliku juhumetsa mudelis (ainsusliku verbivormi lähteosakaal = 0,71, täpsus = 0,76, AUC = 0,83)



Joonis 14. Subjekti rollis hulgafrasiga finitiiverbi vormi valikut mõjutavad tegurid (ainsusliku verbivormi lähteosakaal = 0,71, täpsus = 0,72, AUC = 0,68)

Nii **juhumetsa** kui ka **logistilise regressiooni** mudelid (joonised 13 ja 14) tuvastavad samuti sisuliselt vaid lemma ja verbi asukoha koosmõju, siinjuures ei ole üksik otsustuspuu ega logistilise regressiooni mudel kahe tunnuse põhjal eriti head vaatluste klassifitseerijad ega oska mitmuslikke verbivormide kontekste ainsuslike verbivormide kontekstidest hästi eristada. Tuleb siiski arvestada, et 414 vaatlust on võrdlemisi vähe selleks, et tuvastada andmestikust oluliselt komplekssemaid või tugevamaid seoseid.

6. Kokkuvõtte ja arutelu

Artiklis käsitlesime fraase, mille moodustavad hulgasõnad *osa*, *enamik* ja *enamus* koos mitmuslike komplektisõnadega. Kõigi nende hulgasõnade kasutuses toimub ulatuslik varieerumine, mille siinses uurimuses jagasime kolmeks: esiteks analüüsisime fraasisisest arvuühildumist, teiseks komplektisõna käände valikut ning kolmandaks verbi ühildumist subjekti funktsioonis fraasiga.

Kogu seda varieerumist on varem käsitletud kui keelemuutust, mille käigus tüüpiline hulgafras saab endale nimisõnafraasi omadusi ning sellega seoses muutuvad fraasi põhja ja laiendi suhted. Ka meie uurimus kinnitab, et varieerumine on ulatuslik ning suur osa sellest on seotud nimisõnafraasi ja hulgafraasi suhetega. Kui tüüpilise hulgafraasi puhul (nt *osa inimesi*, *osal inimestest*) on põhjaks hulgasõna (*osa*) ja laiendiks komplektisõna (*inimesi*, *inimestest*), siis nimisõnafraasi puhul (*osad inimesed*, *osadel inimestel*) on põhjaks komplektisõna (*inimesed*) ja laiendiks hulgasõna (*osad*). Kahe fraasitüübi eristumise üheks tagapõhjaks on olnud ilmselt kahe funktsiooni eristumine (Erelt, Metslang 1998): kvantifitseeriv funktsioon toob esile ennekõike kogust ja on hulgafraasi prototüüpne ja algne funktsioon, määratlev funktsioon aga esitab mingisse tüüpi/hulka kuuluvaid entiteete individuaalsemalt ja seda saab siduda pigem nimisõnafraasiga. Samas ei saa fraasitüüpi võrdsustada ühe või teise funktsiooniga, vaid valikud on põimunud mitmeti; samuti ei saa öelda, kas enne toimus hulgafraasi muutumine, mis viis uue funktsiooni tekkimiseni, või vastupidi. Kuivõrd tegemist on käimasoleva keelemuutusega ja sellest tuleneva

varieerumisega, millele on oma jälje jätnud ka keelekorralduslikud soovitusel, proovime tulemusi kokku võtta kolmest vaatenurgast: funktsiooni ja vormi suhted, keelemuutus ja keelekorralduslikud aspektid.

Mitmetunnuselise varieerumise analüüsi meetoditega selgitasime välja kõigi kolme vaadeldud (alam)varieerumise mustrid: varieerumist mõjutavad tegurid ja nende koosmõjud. Püüame siinkohal varieerumismustritest leitud siduda hulgasõnaga fraaside kahe **põhifunktsiooniga**.

Kvantifitseeriv funktsioon, mis rõhutab ennekõike osa-terviku suhteid, tuleb välja hulgafraasile omases fraasistruktuuris, kus peasõnaks on ainsuslik kvantor ja laiendiks mitmuslik komplektisõna. Fraasi tüübi valikut vaatlesime esimeses analüüsis, mis käsitles arvuühildumist hulgasõnaga fraasi sees. Leidsime, et elatiivse komplektisõnaga fraasides on hulgasõna ülivaldavalt ainsuses. Jättes elatiivsed vormid kõrvale, nägime, et hulgafraasi valikut mõjutasid soodustavalt hulgasõna lemma (*enamik* ja *enamus*), fraasi süntaktiline funktsioon (subjekt, objekt) ning hulgasõna eestäiendi olemasolu (*suur osa inimesi*). Viimane toob eriti esile hulgafraasi kvantifitseerivat funktsiooni, sest rõhutab hulga suurust või muud omadust tervikuna, mitte hulga liikmete omadusi.

Hulgafraasis hulgasõna laiendava komplektisõna käände puhul (elatiiv *vs.* muu) ilmestas kvantifitseerivat funktsiooni elatiivi valik, mis rõhutab rohkem koguhulka, mitte selle üksikuid liikmeid. Elatiivi valikut soodustas teises analüüsis hulgasõna eestäiendi olemasolu, mis samuti, nagu juba öeldud, toob esile just nimelt hulga terviklikkuse. Kvantifitseerivat funktsiooni rõhutab hulgafraasist subjekti paiknemine verbi järel, tüüpilises reema positsioonis, nagu selgus kolmandast analüüsist.

Määratlev funktsioon, mis tõstab individuaalsemalt esile mingisse hulka või tüüpi kuuluvaid entiteete, kõlab aga kokku hulga- ja komplektisõna kasutamisega nimisõnafraasitaoliselt, mis avaldub fraasisiseses arvuühildumises. Arvuühildumist soodustab hulgasõna *osa* kasutamine, eriti adverbiaalides (*osadele inimestele*), ja hulgasõna kasutamine ilma eestäiendita (ka nominatiivis: *osad inimesed*). Nominatiivis nimisõnafraas tingib subjektina alati ka arvuühildumise verbiga (*Osad vanemad siis tulid hommikul, ventilaatorid kaelas, kooli*). Hulgafraasi korral soosib verbi semantilist arvuühildumist subjektiga aga hoopis hulgasõna *enamus* (*Enamus norskamisega seotud probleeme vajavad kirurgilist sekkumist*); samuti suurendab verbiühildumise võimalust verbi järelasend hulgafraasi suhtes. Võime muidugi küsida, kas verbiühildumine osutab määratlevale funktsioonile või mitte. Arvestades aga, et arvuühildumine osutab sisulisele mitmuslikkusele ja toob seega esile hulga osi, mitte ei rõhuta hulka ennast, toetab see pigem määratlevat tõlgendust.

Keelemuutuse osas leidsime, et tõenäoliselt on hulga- ja nimisõnafraasi segunemine toimunud esmalt adverbiaali funktsioonis fraasides, kus varasemale käändeühildumisele (*osas autodes*) on lisandunud arvuühildumine (*osades autodes*). Nägime seda esimesest analüüsist, kus vaatasime lähemalt fraasisisest ühildumist ning leidsime, et adverbiaalina esinemine suurendab märgatavalt nimisõnafraasi struktuuri esinemise tõenäosust. Seega võib oletada, et adverbiaalid on olnud keelemuutuse lähtekoht. Sellele osutavad ka varasemad uurijad, kes on toonud keelemuutuse põhjuseks hulgafraasi kasutamise keerukuse ning analoogia teiste hulgasõnadega (eelkõige *palju*), millega on toimunud samalaadsed protsessid (Erelt, Metslang 1998, Nemvalts 1999, 2000). Adverbiaalide eeskujul on arvuühildumine laienenud muudele lauseliikmetele, ehkki oluliselt väiksemal määral.

Subjektina esineva hulgasõnaga fraasi puhul tingib fraasi arvühildumine ühtlasi verbi ühildumise. Võib oletada, et alles selle kasutusviisiga lisandub ka määratlev funktsioon, sest adverbiaali funktsioonis ei ole määratlemine nii selge ega vajalik. Määratleja funktsioonis ei ole võimalik hulgasõnale lisada enam täiendit, sest hulgasõna nimisõnaline tähendus on semantiliselt pleekinud.

Kvantifitseeriva ja määratleva funktsiooni väljakujunemises jäävad kõige ebamäärasemasse rolli lühikesed partitiivse komplektisõnaga ja ainusliku hulgasõnaga fraasid, mis on prototüüpsed hulgafrasid (nt *osa mehi*, *enamik inimesi*). Need hulgafrasid võivad ka verbiühildumises toimida vahel arvus ühildavalt, vahel mitte.

Keelekorralduse mõju ilmneb ennekõike keelemuutuse pidurdamises ja klassikalise partitiivis laiendiga hulgafraasi eelistamises. Nii esimesest kui teisest analüüsist tuli välja žanri mõju, see on aga üsna otseselt seostatav teksti toimetatusega: perioodikas on enim toimetatud tekste, foorumites neid arvatavasti pole üldse, blogid jäävad kahe vahele. Keelekorraldus on tauninud ka sõna *enamus* kasutuse hulgafrasis – seegi ilmneb peamiselt žanridevaheliste erinevustena. Kuigi keelekorraldus on ka elatiivse laiendi kasutamist tauninud, ei ole see päriselt õnnestunud: elatiivi kasutatakse palju ja varasemate uurimuste põhjal (Laanekask 2000) võib öelda, et kasvavas trendis. Samuti ilmnes, et elatiiv on mõnevõrra üllatuslikult tüüpilisem vaadeldud žanridest just enim toimetatud tekstitüübis – perioodikas –, mistõttu ei ole selle kasutamine seletatav teksti loomise informaaalsuse ja spontaansusega.

Kokkuvõttes võib öelda, et vaadeldavad hulgafrasid on muutuses, mis näib seotud olevat kvantifitseeriva ja määratleva funktsiooni kujunemisega. Kvantifitseeriv funktsioon seostub tugevalt elatiivse komplektisõnaga fraasidega ning määratlev funktsioon arvus ühilduvate nimisõnafrasidega. Need on skaala kaks äärmist otsa. Partitiivse laiendiga fraasid jäävad nende kahe vahepeale ning neid tõlgendatakse kord nii, kord naa. Nende kasutamist hoiavad kinni keelekorralduslikud soovitusel.

Viidatud kirjandus

- Agresti, Alan 2012. Categorical Data Analysis. 3rd ed. Wiley Series in Probability and Statistics, 792. Hoboken: John Wiley & Sons.
- Borkovec, Martin; Madin, Niyaz 2019. ggparty: “ggplot” Visualizations for the “partykit” Package. R package version 1.0.0. <https://CRAN.R-project.org/package=ggparty>
- Breiman, Leo 2001. Random forests. – Machine Learning, 45 (1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- EG 2023 = Metslang, Helle; Erelt, Mati; Habicht, Külli; Hennoste, Tiit; Kasik, Reet; Teras, Pire; Viht, Annika; Asu, Eva Liina; Lindström, Liina; Lippus, Pärtel; Pajusalu, Renate; Plado, Helen; Rääbis, Andriela; Veismann, Ann 2023. Eesti grammatika. Tartu: Tartu Ülikooli Kirjastus. <https://doi.org/10.12697/EG>
- EKG II 1993 = Erelt, Mati; Kasik, Reet; Metslang, Helle; Rajandi, Henno; Ross, Kristiina; Saari, Henn; Tael, Kaja; Vare, Silvi 1993. Eesti keele grammatika II. Süntaks. Lisa: Kiri. Tallinn: Eesti Teaduste Akadeemia Keele ja Kirjanduse Instituut.
- Erelt, Mati 2000. Arvühildumisest tänapäeva eesti kirjakeeles. – Keel ja Kirjandus, 3, 180–189.
- Erelt, Mati 2019. Lause õigekeelsus. Juhatused ja harjutused. Neljas, täiendatud trükk. Tallinn: Emakeele Selts.
- Erelt, Mati; Erelt, Tiiu; Ross, Kristiina 2020. Eesti keele käsiraamat. Uuendatud väljaanne. Tallinn: Eesti Keele Instituut, Eesti Keele Sihtasutus.

- Erelt, Mati; Metslang, Helle 1998. Oma või võõras. – Keel ja Kirjandus, 10, 657–668.
- Habicht, Külli 1995. *us*-liitelised abstraktsed nimisõnad eesti vanima isonaraamatuis. – Jaak Peebo (Toim.), Läänemere rahvaste kirjakeelte ajaloo. Tartu Ülikooli eesti keele õppetooli toimetised, 1. Tartu, 15–26.
- Hothorn, Torsten; Hornik, Kurt; Zeileis, Achim 2006. Unbiased recursive partitioning: A conditional inference framework. – Journal of Computational and Graphical Statistics, 15 (3), 651–674. <https://doi.org/10.1198/106186006X133933>
- Hothorn, Torsten; Zeileis, Achim 2015. partykit: A Modular Toolkit for Recursive Partytioning in R. – Journal of Machine Learning Research, 16, 3905–3909.
- Jaagu, Kärt-Kristiin 2023. Arvõnalise kvantorifraasi õhildumine verbiga. MA. Tartu Ülikool. <https://hdl.handle.net/10062/90764>
- Janitza, Silke; Strobl, Carolin; Boulesteix, Anne-Laure 2013. An AUC-based permutation variable importance measure for random forests. – BMC Bioinformatics, 14 (1), A19, 1–11. <https://doi.org/10.1186/1471-2105-14-119>
- Kilgariff, Adam; Baisa, Vít; Bušta, Jan; Jakubiček, Miloš; Kovář, Vojtěch; Michelfeit, Jan; Rychlý, Pavel; Suchomel, Vít 2014. The Sketch Engine: ten years on. – Lexicography, 1, 7–36. <https://doi.org/10.1007/s40607-014-0009-9>
- Kilgariff, Adam; Rychlý, Pavel; Smrž, Pavel; Tugwell, David 2004. The Sketch Engine. – Geoffrey Williams and Sandra Vessier (Eds.), Proceedings of the 11th EURALEX International Congress. Lorient: Université de Bretagne-Sud, 105–116.
- Kindlam, Ester 1978. Sõltumussuhestest. – Keelevoos 76. Tallinn: Valgus, 55–66.
- Koppel, Kristina; Kallas, Jelena; Jürviste, Madis; Kaljumäe, Helen 2023. Eesti keele õhend-korpus 2023. Lexical Computing Ltd. / Eesti Keele Instituut.
- Koptjevskaja-Tamm, Maria 2001. “A piece of the cake” and “a cup of tea”: Partitive and pseudo-partitive nominal constructions in the Circum-Baltic languages. – Østen Dahl, Maria Koptjevskaja-Tamm (Eds.), Circum-Baltic Languages. Volume 2: Grammar and Typology. Studies in Language Companion Series, 55. Amsterdam–Philadelphia: John Benjamins, 523–568. <https://doi.org/10.1075/slcs.55.11kop>
- Laanekask, Tiia 2000. Eesti hulgasõnade sõntaktilisi omadusi. BA. Tartu Ülikool.
- Levshina, Natalia; Lorenz, David 2022. Communicative efficiency and the Principle of No Synonymy: Predictability effects and the variation of want to and wanna. – Language and Cognition, 14 (2), 249–274. <https://doi.org/10.1017/langcog.2022.7>
- Lindström, Liina 2017. Lause infostruktuur ja sõnajärg. – Mati Erelt, Helle Metslang (Toim.), Eesti keele sõntaks. Eesti keele varamu, 3. Tartu: Tartu Ülikooli Kirjastus, 537–565.
- Lõbu, Triin 2011. Hulgasõnade kasutamisest eesti kirjakeeles. MA. Tartu Ülikool.
- Metslang, Helle 2017. Kvantorifraas. – Mati Erelt, Helle Metslang (Toim.), Eesti keele sõntaks. Eesti keele varamu, 3. Tartu: Tartu Ülikooli Kirjastus, 463–478.
- Nemvalts, Peep 1999. Pigem oma kui võõras! (“Muutuv keel”). – Keel ja Kirjandus, 11, 756–763.
- Nemvalts, Peep 2000. Aluse sisu ja vorm. Alusfraasi käändevaheldus tänapäeva eesti kirja-keeles. Tallinn: Eesti Keele Sihtasutus.
- Pilvik, Maarja-Liisa; Lindström, Liina; Plado, Helen; Simmul, Carl Eric 2025. Nimisõnafraasi ja hulgafraasi piirimal: Lisamaterjalid. OSF. <https://doi.org/10.17605/OSF.IO/VM6J2>
- R Core Team 2024. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org>.
- Risberg, Lydia; Langemets, Margit 2021. Paronõumide probleem eesti keeles. – Keel ja Kirjandus, 10, 903–926. <https://doi.org/10.54013/kk766a4>
- Strobl, Carolin; Boulesteix, Anne-Laure; Zeileis, Achim; Hothorn, Torsten 2007. Bias in random forest variable importance measures: Illustrations, sources and a solution. – BMC Bioinformatics, 8 (1), A25, 1–21. <https://doi.org/10.1186/1471-2105-8-25>

ON THE BORDER OF NOUN PHRASES AND QUANTIFIER PHRASES: *OSA* 'SOME', *ENAMIK* 'MOST', AND *ENAMUS* 'MOST; MAJORITY' AS QUANTIFIERS

Maarja-Liisa Pilvik, Liina Lindström,
Helen Plado, Carl Eric Simmul

University of Tartu

In this article, we examine the structural and semantic variation of phrases which are formed with the quantifiers *osa* 'some', *enamik* 'most', or *enamus* 'most; majority', and plural set nouns in contemporary Estonian. We focus specifically on the analysis of factors influencing the number agreement of the quantifier (e.g., *enamikule inimestele*, *enamikele inimestele* 'to some people'), the case form of the set noun (e.g., *enamikku inimesi* 'most people,' *enamikku inimestesti* 'most (of the) people'), and the verbal number agreement (e.g., *enamus inimesi läheb ~ lähevad* 'most people go'). Both the choice of quantifier and all three aspects of variation have previously been linked to an ongoing linguistic change, in which the quantifier as the head and set noun as the modifier in quantifier phrases are being replaced by noun phrases with quantifier as the modifier and the set noun as the head. The motivation behind this change might stem from the need to distinguish the quantifying and specifying functions of the quantifier. This study's results confirm many previous observations, made either on a smaller data set or based on intuition, but additionally highlight the importance of the syntactic role of phrases: the potential source of number agreement within phrases is likely adverbials, while the specifying function of the quantifier is increasingly solidifying in phrases where it serves as the subject. Among other things, we also see a clear effect of genre and the degree of text editing in the expansion of this linguistic change.

Keywords: language variation, agreement, language change, corpus linguistics, Estonian

Maarja-Liisa Pilvik (Tartu Ülikool) tegeleb põhiliselt keele morfosüntaktilise varieerumise ja tuletusmorfoloogia produktiivsuse uurimisega eesti keele eri allkeeltes ja registrites. Jakobi 2, 51005 Tartu, Estonia
maarja-liisa.pilvik@ut.ee

Liina Lindström (Tartu Ülikool) uurib eesti keele varieerumist ja grammatikat, eriti süntaksit ja selle varieerumist nii suulises kui ka kirjalikus keeles, nii murretes kui ka tänapäeva veebikeses või teismeliste keelekasutuses. Jakobi 2, 51005 Tartu, Estonia
liina.lindstrom@ut.ee

Helen Plado (Tartu Ülikool) on eesti ja lõunaeesti keele uurija, kelle peamiseks uurimisvaldkonnaks on morfosüntaks. Eelkõige huvitab teda keele varieerumine ja muutumine. Jakobi 2, 51005 Tartu, Estonia
helen.plado@ut.ee

Carl Eric Simmul (Tartu Ülikool) on doktorant, kes on uurinud *des-*, *mata-* ja *maks-*tarindit ning keelendeid, mis väljendavad kõneleja suhtumist kõnealusesse (nt *vist*, *tegelikult*, *kahjuks*, *kindlasti*). Jakobi 2, 51005 Tartu, Estonia
carl.eric.simmul@ut.ee