

TARTU ÜLIKOOL
HUMANITAARTEADUSTE JA KUNSTIDE VALDKOND
EESTI JA ÜLDKEELETEADUSE INSTITUUT

Kärt-Kristiin Jaagu

ARVSÕNALISE KVANTORIFRAASI ÜHILDUMINE VERBIGA

Magistritöö

Juhendajad Liina Lindström ja Maarja-Liisa Pilvik

Tartu 2023

Sisukord

Sissejuhatus.....	5
1. Eesti keele kvantorifraas	7
1.1. Arvsõnaline kvantorifraas	8
1.2. Muud kvantorifraaside liigid ja kvantorsõnad.....	9
2. Ühildumine ja eesti keele kvantorifraasi ühildumisest verbiga.....	10
2.1. Ühildumine	10
2.2. Aluse ja öeldise ühildumine	11
2.3. Eesti keele kvantorifraas ja selle ühildumine verbiga	13
3. Materjal ja meetod.....	16
4. Käsitletavad tunnused	20
4.1. Verbivormi arvukategooria	20
4.2. Verbi ajavorm.....	21
4.3. Verbi paiknemine kvantorifraasi suhtes	22
4.4. Verbi transitiivsus.....	23
4.5. Kvantorifraasi arvsõna	26
4.6. Ebamäärane hulk	26
4.7. Subjekti referendi elusus	27
4.8. Lause tekstitüüp.....	31
5. Andmete ühetunnuseline analüüs	33
5.1. Verbi ajavorm.....	35
5.2. Verbi paiknemine kvantorifraasi suhtes	35

5.3.	Verbi transitiivsus.....	36
5.4.	Kvantorifraasi arvsõna	39
5.5.	Ebamäärane hulk	41
5.6.	Subjekti referendi elusus	42
5.7.	Lause tekstitüüp.....	44
5.8.	Verbid.....	45
6.	Tunnuste koosmõju analüüs	47
6.1.	Otsustuspuu	47
6.2.	Juhumets.....	50
7.	Järeldused.....	52
	Kokkuvõte.....	57
	Kirjandus	59
	Summary. Subject-Verb Agreement with Numerical Quantifier Phrase.....	64

Autorsuse kinnitus

Kinnitan, et olen käesoleva lõputöö ise kirjutanud ning toonud korrekselt välja teiste autorite panuse. Töö on kirjutatud lähtudes Tartu Ülikooli eesti ja üldkeeleteaduse instituudi lõputöö nõuetest ning on kooskõlas heade akadeemiliste tavadega.

Sissejuhatus

Keele varieerumine on tavapärane nähtus, loomuliku keele lahutamatu osa. Varieerumine võib hõlmata kõiki keele tasandeid, sõnavarast süntaksini, ning see annab märku keele elujõulisusest. Võistlevate keelendite ja nende kasutust mõjutavate tegurite väljaselgitamiseks on oluline tegeliku keelekasutuse andmete uurimine, et jõuda võimalikult objektiivse ja usaldusväärse tulemuseni ning näha, mis keeles päriselt toimub (Krug jt 2013: 1–3).

Keele varieerumist ei mõjuta ainult keelekõneleja omadused, sotsiaalne taust ja isiklikud eelistused ehk sotsiolingvistikas tuntud nähtused nagu vanus ja sugu, amet ja haridus, suhtlusolukord, murdetaust, keelekontaktid ja inimese päritolu. Märkimisväärset mõju keele varieerumisele avaldavad ka keelesisesed muutujad, näiteks sõnade järjekord lauses või verbi transitiivsus. Siinses magistritöös uurin tegelikul keelekasutusel põhineva andmestiku, 2021. aasta eesti keele ühendkorpusest (EKÜ 2021) pärinevate lausete abil, millised keelesisesed muutujad mõjutavad üht eesti keeles varieeruvat nähtust – lauses aluseks oleva kvantorifraasiga ühilduva verbi vormi valikut.

Minu magistritöö teemat, arvsõnalise kvantorifraasi ühildumist verbiga ei ole eesti keeles varem tegeliku keelekasutuse andmete põhjal käsitletud, seega aitab uurimus täita tühimikku Eesti keeleteaduses ning selgitada, millistest keelelistest teguritest võiks sõltuda verbi vormi varieerumine kvantorifraasiga ühildumisel. Magistritöö eesmärk on vaadelda lauses aluseks oleva kvantorifraasiga ühilduva verbi vormi varieerumist ning selgitada välja, millistest keelesisestest muutujatest sõltub verbivormi arvukategooria ehk ainsuse või mitmuse vormi valik (nt *kaks sõbralikku meest tuleb/tulevad*). Täpsemad uurimisküsimused on järgnevad:

- millised keelesisesed muutujad aitavad seletada verbi vormi valikut?
- millise keelesisese muutuja mõju verbi vormi valikule on kõige suurem?

Uurimus keskendub sellistele arvsõnalistele kvantorifraasidele, kus kvantor paikneb vahetult teda laiendava nimisõna ees ning kus kvantorifraas on lauses alusena. Töös analüüsin lauseid, mis vastavad järgmistele konstruktsioonidele:

- kolm (pikka) meest tuleb/tuli, kolm (pikka) meest tulevad/tulid;
- tuleb/tuli kolm (pikka) meest, tulevad/tulid kolm (pikka) meest.

Keelesisesed muutujaid ehk seletavaid tunnuseid, mis võiksid aidata verbi vormi valikut seletada, on siinses töös seitse. Nendeks on verbi ajavorm, verbi paiknemine kvantorifraasi suhtes, verbi transitiivsus lähtuvalt sellest, kas lauses on olemas eksplitsiitne ehk selgesti väljendatud objekt, kvantorifraasi arvsõna, kvantorifraasiga väljendatava hulga ebamäärasus, kvantorifraasi subjekti referendi elusus ja tekstitüüp.

Magistritöö koosneb seitsmest peatükist. Esimeses peatükis annan teoreetilise ülevaate eesti keele kvantorifraasist, teises teooriapeatükis kirjeldan ühildumist kui nähtust ning kirjutan eesti keele kvantorifraasi ja verbi ühildumisest. Kolmas peatükk tutvustab magistritöö andmestikku ja uurimisel kasutatud meetodeid. Neljandas peatükis kirjeldan tunnuseid, mille kaudu verbivormi arvukategooria varieerumist uurin. Viies peatükk sisaldab andmete ühetunnuselise analüüsi, kuendas peatükis analüüsin tunnuste koosmõju. Seitsmendas peatükis esitan tulemuste kokkuvõtte ja olulisemad järeldused.

1. Eesti keele kvantorifraas

Selles peatükis selgitan, mis on kvantorifraas ja millisena see lauses esineda võib. Samuti toon välja, millised on kvantorifraasi liigid.

Kvantori funktsioon on määratleda nimisõnafraasiga tähistatu kogust, hulka (Metslang 2017a: 463). Eesti keeles on kvantoreid kahte liiki: atributiivsed kvantorid (*üks inimene, mõned inimesed*), mis laiendavad nimisõna ning ühilduvad sellega käändes ja arvus; ja põhjana toimivad kvantorid (*hulgale inimestele, kaks inimest*), mis määravad nimisõna käände ja on ise fraasi tuumaks (samas: 463). Sellist fraasi, mis koosneb põhisõnaks olevast kvantorsõnast ja laiendavast substantiivifraasist, nimetataksegi **kvantorifraasiks** või hulgafraasiks. Kvantorifraasid väljendavad hulka, mõõtu ja määra ning nende põhisõna võib olla:

- 1) põhiarvsõna või asesõna (*kolm poissi, paar päeva*) – sellised fraasid on **arvsõnalised kvantorifraasid**;
- 2) mõõtu, hulka, kogu, rühma või osa märkiv kvantiteedinimisõna (*meeter, kari, enamik*) – sellised fraasid on **nimisõnalised kvantorifraasid**;
- 3) kvantiteedimäärsõna (*palju, veidi, rohkesti, vähe*) – sellised fraasid on **määrsõnalised kvantorifraasid**. (Metslang 2017a: 463, 466, 470, 474; EKK 2020: 491)

Kvantori ja tema laiendi, st nimisõna vahel on kvantorifraasis reksiooniseos. Nimetavas käändes kvantor tingib kvantorifraasis nimisõna partitiivse vormi, nt *kolm meest* (Metslang 2017a: 464), muude käänete puhul on kvantor ja tema laiend samas käändes ehk kvantor ühildub käändes fraasi põhjaga, nt *kolmele mehele* (EKK 2020: 490).

Kvantor paikneb tavaliselt oma laiendi ees, ent võib teatud tingimustel paikneda ka mujal.

Näiteks:

- 1) laiendi järel, juhul kui märgitakse ligikaudset arvu (*kraadi kolm*) või on laiendiks elatiivis nimisõna (*vigastada saanud meestest kolm*),
- 2) laiendist lahus, juhul kui kvantor on fookustatud (*mune on ostetud tosin, seltsi liikmeid on paarkümmend*). (Metslang 2017a: 464–465)

Lauses võib kvantorifraas esineda nii aluse (*kolm meest tuleb*), sihitise (*kuusk kasvatas neli kabi*), öeldistäite (*maja kõrgus on üheksa meetrit*), määruse (*seitsmele koerale kümnest on omanik leitud*) kui ka täiendina (*tema kümne aasta tegemisi on kajastatud*) (Metslang 2017a: 465). Selles magistritöös keskendun nendele arvsõnalistele kvantorifraasidele, kus kvantor paikneb vahetult teda laiendava nimisõna ees ning kus kvantorifraas on lauses alusena.

1.1. Arvsõnaline kvantorifraas

Arvsõnalised kvantorifraasid koosnevad põhjaks olevast põhiarvsõnast (*kaks*), asearvsõnast (*paar, mitu*) või kvaasinumeraalist (oskuskeeles nt *n, x*) ja seda laiendavast loendatavast nimisõnast (*koer, pealtnägija, kord*). Kvantorifraasi laiendavas substantiivifraasis võivad olla lisaks substantiivile ka seda laiendavad sõnad, näiteks *seitse viimast päeva*. Samuti võib kvantorifraasi substantiivi asemel olla teine kvantorifraas, mis tähendab, et üht kvantorifraasi laiendab teine kvantorifraas, nt *70 protsenti vastajaid, kolm paari pükse*. (Metslang 2017a: 466–468)

Arvsõnalise kvantorifraasiga saab näidata osa tervikust või kogu tervikut. Selleks, et näidata osa tervikust, kasutatakse kvantorifraasi laiendamisel elatiivset adverbiaaltribuuti – *umbes kolmveerand osalistest, kaheksa tema sõpradest*. (Metslang 2017a: 469) Elatiivset laiendit saab kasutada ka siis, kui kvantorifraas ei ole arvsõnaline, nt *osa loomadest*, ja kvantori elatiivne laiend on tõenäoliselt arenenudki kvantorifraasi elatiivsest laiendist (*kaheksa sõpra tema sõpradest* asemel *kaheksa tema sõpradest*). Samas kajastub terviku ja osa suhe sageli nt hulgasõna osatähenduses ja sel

juhul saab seestütleva käände asendada osastavaga, nt *osa loomi*. (EKK 2020: 492) Kui arvsõnalise laiendiga näidata kogu tervikut, esitatakse laiendav nimisõna fraasis ühel korral: *üheksa meest kümnest (mehest)*. (Metslang 2017a: 468–469)

1.2. Muud kvantorifraaside liigid ja kvantorsõnad

Nimisõnaliste kvantorifraaside põhisõna on nimisõna. Nimisõnad võivad märkida mõõtu (nt *liiter, tükk, klaas, sort*), komplekti (nt *hulk, trobikond*) või osa tervikust (*osa, lõik*). Sellised fraasid esinevad lauses tüüpiliselt aluse ja sihitisena. (Metslang 2017a: 470)

Määrsõnaliste kvantorifraaside põhisõna on määrsõna (nt *palju, veidi*). Määrsõna laiendavad partitiivis nimisõnad, mille tähendus on kvantitatiivselt piiritlemata, nt *vähe vett*. Määrsõnalised kvantorifraasid toimivad tavaliselt lauses aluse, sihitise või määrusena. (Metslang 2017a: 474–475)

Lisaks arv-, nimi- ja määrsõnalistele kvantorifraasidele on olemas hulgasõnad, mis kvantorifraasi ei moodusta: näiteks *üks, mõlemad, kumbki*. Need sõnad toimivad omadussõnalise täiendina kogu käände paradigmas, nt *üks mees tuli, mõlemad mehed tulid, kumbki mees ei tulnud*. (EKK 2020: 493)

2. Ühildumine ja eesti keele kvantorifraasi ühildumisest verbiga

Selles peatükis selgitan, mis on ühildumine. Samuti toon erinevate allikate põhjal välja, millised on eesti keele kvantorifraasi ja sellega seotud verbi ühildumise üldised seaduspärasused, mida varem on täheldatud.

2.1. Ühildumine

Ühildumine ehk kongruents on lisaks käändemärgistusele ja sõnajärjele üks vahenditest, mis aitab eristada lause tuumargumente – alust ja öeldist. Seega on ühildumine üks tähtsamaid nähtusi, mida keeled grammatiliste seoste väljendamiseks kasutavad. (Kroeger 2005: 118) Ühildumine on maailma keeltes väga levinud ning aitab keekekasutajal nii tekstis kui ka kõnes paremini järge pidada ja mõista, kellele või millele parasjagu viidatakse. Niisiis ei ole tegemist vaid liiasusega, nagu varasemalt on arvatud. (Corbett 1999: 12, 17)

Ühildumise näol on tegemist alistuse märkimise viisiga, mis näitab põhja ja laiendi vahelist suhet – ühe moodustaja grammatiline vorm määrab teise moodustaja vormi, see võib toimuda nii fraasi kui ka lause tasandil (Erelt 2017b: 67). Näiteks määravad nimisõna või nimisõnafraasi grammatilised omadused nimisõna või selle fraasiga süntaktiliselt seotud sõna omadused (Kroeger 2005: 111). Ühildumise ülesanne, seega, ongi näidata, kuidas lauseliikmed omavahel seostuvad ja kuidas kaks sõnaliiki omavahel sobitada (Karlsson 2002: 195, Corbett 1999: 12).

Selleks, et ühildumine toimuks, on vajalik kontrolliva sõna või fraasi (ingl *controller*) olemasolu, sest kontrolliva sõna/fraasi semantilised või vormilised omadused määravad sihtsõna (ingl *target*) vormi. (Corbett 2006: 4). Nii võib näiteks tegusõna pöördelõpu järgi teada saada, kes on lauses tegija. (Karlsson 2002: 195) Magistritöös vaatlen, kuidas kvantorifraasist aluse sisuline mitmuslikkus mõjutab öeldise arvukategooriat. Kontrollija

on kvantorifraas, mis toimib alusena, ning sihtsõna on sama lause öeldisverb. Uurin, millised on need kvantorifraasi omadused, mis mõjutavad sihtsõnaks oleva verbi vormi.

Ühildumist on üritatud ühildumise tüüpide ja nende funktsioonide järgi süstematiseerida. Kõige üldisemad ühildumiskategooriad maailma keeltes on soo-, isiku-, arvu-, käände- ja possessiivsuskategooria. Sooühildumine avaldub lauses tavaliselt kontrolliva sõna täiendina või öeldistäitena ning see esineb näiteks vene keeles, kus on olemas mees-, nais- ja kesksugu. (Karlsson 2002: 192, Corbett 1999: 12) Isikuühildumise puhul on kontrollivaks sõnaks grammatiline alus, mis tingib verbi vormi (Karlsson 2002: 194). Näiteks eesti ja soome keeles toimub isikuühildumine nii ainsuse kui ka mitmuse kolmes pöördes ja ainsuse esimeses isikus olev *mina* tingib selle, et ka alusega ühilduv verb oleks ainsuse esimeses isikus, nt *mina tee-n*, mitte *mina tee-b* või *mina tee-me*. Arvuühildumisel määrab kontrolliv sõna sihtsõna arvuvormi, nt võib kontrolliv nimisõna määrata omadussõna (eesti keeles nt *valge-le koera-le* ja *valge-te-le koer-te-le*) või verbi (*koer tee-b* ja *koera-d tee-vad*) arvuvormi. Käändeühildumisel tingib kontrolliv sõna sihtsõna, nt omadussõnalise täiendi (eesti keeles nt *valge koer* ja *valgest koerast*) käände, ent sel juhul pole neljas viimases käändes (rajav, olev, ilmaütlev ja kaasaütlev) ühildumine täielik (*valge koeraga*, mitte *valgega koeraga*) (EKK 2020: 477). Possessiivsuseühildumine näitab omamist, kellelegi kuulumist, ja kontrolliv sõna ehk omaja tingib omatava objekti vormi. Selline ühildumine on tavaline näiteks soome keeles, kus omajast tingituna lisatakse omatavale objektile possessiivsufiks (nt *hänen koiransa* ehk *tema koer*), aga omane ka nt uurali keeltele. (Karlsson 2002: 193–194) Eesti keeles esineb ühildumine arvu-, käände ja isikukategoorias. Possessiivsuseühildumine on eesti keelest kadunud ning sooühildumist ei ole.

2.2. Aluse ja öeldise ühildumine

Aluse ja öeldise ühildumisel on tavapärane, et alus ühildab endaga öeldise ehk öeldise vorm sõltub alusest (Erelt 2017a: 202–203, Erelt 2017b: 67). Öeldisverbid ühilduvadki enamasti vaid ühe argumendiga, milleks on sageli subjekt (Kroeger 2005: 112), ja öeldise arvuvormi valiku määrab subjekti vorm või sisu. (EKK 2020: 399; Erelt 2017a: 203) Eesti

keele grammatika eristab öeldise vormis vaid ainsust ja mitmust, mitte nt kahele objektile viitavat duaali nagu mõned muud maailma keeled. Seega saab verb alusega arvus ühilduda ainsuslikus või mitmuslikus vormis.

Küll aga ei pruugi aluse ja öeldise ühildumine alati ühene olla, sest eesti keeles puudub selgus selle kohta, kuidas toimida siis, kui lauses olev alus on sisult mitmuslik, kuid vormilt ainsuslik, nagu põhiarvsõnaline kvantorifraas (Erelt 1999: 9). Kontrolliva sõna või fraasiga on võimalik kasutada nii ainsuses kui ka mitmuses olevat sihtsõna, milleks on verb. D. A. Cruse on osutanud, et ka inglise keeles eksisteerib sarnane probleem – kui nimisõna on sisult mitmuslik, ent vormilt ainsuslik, siis võidakse verbi kasutada nii mitmuses kui ainsuses sõltuvalt sellest, kas lähtutakse nimisõna sisust või vormist (nt *the government has/have decided to act* ehk *valitsus on otsustanud tegutseda*) (Cruse 1999: 270).

Mati Erelt on märkinud, et ühildumisest rääkides on tüpoloogiline tendents toetuda Greville G. Corbetti 1979. aastal loodud ühildumishierarhiale (Erelt 2000: 185):
atribuut – predikaat – relatiivpronoomen – personaalpronoomen (Corbett 1979: 204).

Corbetti järgi on vasakul pool oleva elemendi puhul suurem tõenäosus, et ühildumisel lähtutakse vormist. Paremal pool olev element lähtub ühildumisel pigem sisust. (Corbett 1979: 204) Seega, kui võtta näiteks kvantorifraas *kolm meest*, siis saab öeldis ehk predikaat lähtuda nii vormist kui sisust, ent kuna predikaat asub hierarhias pigem vasakul pool, siis on tõenäolisem, et lähtutakse vormist ja valitakse ainsuslik verbivorm *tuleb*. Tegelikuses varieerub verbi arvuvorm siiski omajagu, näiteks lausetes *Viis kohtunikku leidis üksmeelselt, et Behring Breivik on süüdiv* ja *Kaks konna kelkisid kõrge kuuse kõrval*. Minu magistritöö eesmärk ongi välja selgitada, millistest keelesisestest muutujatest sõltub verbivormi arvukategooria ehk ainsuse või mitmuse vormi valik.

2.3. Eesti keele kvantorifraas ja selle ühildumine verbiga

Kvantor nõuab tavaliselt partitiivis olevat laiendit, mis põhiarvsõnade puhul on ainsuses, nt *kolm meest* (Metslang 2017a: 464). Kvantorifraasi ja verbi ühildumisel on võimalik nii ainsuslik kui ka mitmuslik verbivorm. Verbivormi varieerumine arvsõnalise kvantorifraasiga ühildumisel ei ole uudne nähtus, sest see esineb ka lähedases sugulaskeeles – soome keeles – ja on võimalik, et varieerumist mõjutavad vähemalt osaliselt samad tegurid. Järgnevalt toon välja verbiga ühildumise seaduspärasused ja tendentsid eesti keeles.

Helle Metslang ja Mati Erelt on ajakirjas Keel ja Kirjandus ilmunud artiklis „Oma või võõras“ (1998) maininud, et mitmuse kasutamine muutub kvantorsõnade puhul järjest tavalisemaks. Peamiselt keskenduvad nad küll kvantorifraasi muutumisele ja kirjeldavad näiteks seda, kuidas fraasi *osa inimesi* asemel on hakatud kasutama *osad inimesed*, mistõttu muudetakse ka lauses olev verb ainsuslikust mitmuslikuks. (Samas: 659–660)

Mati Erelt on artiklis „Arvuühildumisest tänapäeva eesti kirjakeeles“ (2000) kirjutanud, et aja jooksul on arvuühildumine suurenenud ja seda mitte ainult põhiarvsõnaliste kvantorifraaside ja verbide puhul, vaid ka kvantiteedinimisõnaliste kvantorite nagu nt *enamik* ja *osa* puhul – *enamikule* vs. *enamikele meestele* ja *osa mehi* vs. *osad mehed*; ja predikatiivadverbiaalide puhul, nagu nt lauses *poisid istusid nukra(te)na nurgas*. Erelt täheldab, et kvantorifraasi puhul (nt *kolm meest*) on tegemist sisult mitmusliku ja vormilt ainsusliku tarindiga, mistõttu võib näiteks verbi vormi valides lähtuda ühest või teisest. Tegemist on mitteprototüüpse ühildumisega ja ainsuse kasutamine võib märkida hoopis ühildumatust. Ereli sõnul on kvantorifraaside korral mitmuse kasutamine levinum kui näiteks adjektiivatribuudiga. (Samas: 185, 187)

Paralleelina võib ka soome keele aluseks oleva kvantorifraasi ja verbi ühildumine sõltuda sellest, kas lause subjekt on semantiliselt või vormilt mitmuses või ainsuses (Hakulinen jt 2004: § 1285). Lausetes, kus kvantor asub vahetult oma laiendi ees, kasutatakse „Iso suomen kielioppi“ andmetel ainsuses olevat verbi, nt *kolme miestä ryösti*

pankin (eesti keeles *kolm meest röövis pankka*). Kui verb panna mitmusesse, siis tähendab see, et kvantorifraasis olev nimisõna on selgemalt määratletud ja on kindlamalt teada, kellest räägitakse, nt *kolme miestä jäivät kiinni* (eesti keeles *kolm meest jäid vahele*). (Hakulinen jt 2004: § 790) Seda kinnitab ka Kielitoimiston Ohjepankki kodulehekül, mille kohaselt kirjutatakse kvantorifraasiga ühilduv verb enamasti ainsuses, kuid kui mõeldakse konkreetset rühma või kõiki rühmaliikmeid, siis pannakse verb mitmusesse (Kielitoimiston ohjepankki 2022).

See on sarnane eesti keelega. „Eesti keele käsiraamatu“ järgi saab arvsõnalise kvantorifraasi puhul kasutada nii ainsuslikku kui mitmuslikku verbi. Kui fraas on lause alguses või tähistab täpset hulka, siis kasutatakse verbi pigem mitmuses. Kui fraas on lõpus või ebamäärase hulga väljendaja, eelistatakse ainsust. (EKK 2020: 400)

Preskriptiivsest aspektist on Mati Erelt osutanud, et hulgamäärsõnade ja selliste hulganimisõnade puhul, mis väljendavad tervikut, tuleks eelistada ainsuslikku verbi. (Erelt 2019: 16–17) Ainsuslik öeldis käib enamasti kaasas ka kogust väljendava hulgasõnaga – erandina võib öeldis olla nii ainsuses kui mitmuses, kui selline hulgasõna on lause alguses ja tal on loendatav nimisõnaline laiend. Erelt ütleb, et põhiarvsõnaline kvantorifraas ühildub nii ainsusliku kui ka mitmusliku verbivormiga, ent soovitab kasutada:

- mitmust, kui hulgafraas väljendab täpset hulka või asub lause alguses (nt *kaks koera jooksevad*);
- ainsust, kui hulgafraasiga väljendatav hulk on ebamäärane või kui hulgafraas asub lause lõpus (nt *jookseb umbes kümme koera*). (Samas:17)

„Eesti keele süntaksi“ järgi sõltub verbivormi arvukategooria lause subjekti vormist või sisust, kuid erinevusi võib kohata kvantorifraaside puhul ehk näiteks nendel juhtudel, mil subjekt on vormilt ainsuslik, kuid sisult mitmuslik. Kui lause alus on kvantorifraas, mida laiendab loendamatu nimisõna, on öeldisverb ainsuses (nt *liiter vett valgub laiali*); kui kvantorifraasi laiendab loendatav nimisõna, sõltub verbi vorm hulgasõnast, lausetüübist ja kvantorifraasi definiitsusest. (Erelt 2017a: 203–204)

Erelti tähelepanekud on järgmised:

- pronumeraalide (*paar*) ja põhiarvsõnadega saab kasutada ainsust ja mitmust, aga põhiarvsõnaga kaasneb normaallauses enamasti mitmus;
- eksistentsiaallause ebamäärast hulka väljendava kvantorifraasiga kasutatakse pigem ainsust (*umbes kaheksa poissi*);
- definiitse määratleja või mitmusliku täiendiga arvsõnafraas esineb mitmusliku verbiga (*need kõik kaheksa poissi*);
- omajalause kaasneb lause lõpus oleva arvsõnafraasiga ainuslikus vormis öeldis (*mul oli kolm õuna*);
- kvantitatiivset ebamäärasust väljendavad määrsõnad (*natuke, vähe, palju*) nõuavad ainsuslikku verbi;
- nimisõnad, mis väljendavad osa (*osa, enamik*), peavad eksistentsiaallauses ja omajalause käima koos ainsusliku verbiga (*neil oli kodus mitu koera*), normaallauses sobib kasutada ka mitmuslikku verbi (*enamik poistest sai/said kommi*). (Erelt 2017a: 204–206)

Magistritöö analüüsisosas kajastan ainsuse ja mitmuse kasutamist kvantorifraasi arvsõnadega, samuti vaatlen ebamäärase hulgaga kasutatavaid verbivorme.

3. Materjal ja meetod

Siinses magistritöös keskendun põhi- ja asearvsõnalistele kvantorifraasidele, sest nende puhul oli verbi varieerumine kõige ulatuslikum. Asearvsõnaliste kvantorifraaside puhul pean silmas selliseid kvantorifraase, mille puhul on võimalik oletada, millist arvu mõeldakse: nt *kolmteist, paar* või *kümmekond*. Uurimusest jäid välja asearvsõnalised kvantorifraasid, milles kvantor oli nt *mitu*, sest sel juhul ei olnud mul võimalik arvata, mis arvu mõeldakse ja kvantorifraasis oleva arvu kui seletava tunnuse mõju hindamine olnuks keerulisem. Töö materjalina kasutasin SketchEngine'i keskkonnas kättesaadavat korpust nimega eesti keele ühendkorpus 2021 (EKÜ 2021).

Magistritöö materjali piirasin selliselt, et valimisse jäid vaid laused, milles kvantor asub vahetult oma laiendi ees, nt *kolm sõbralikku meest tuli/tulid*. Lauseid, milles kvantor ei paikne vahetult oma laiendi ees (nt *mehi tuli kohale kolm, õues on kraade kümme*) ma selles uurimuses ei vaata. Viimati nimetatud juhtumeid on SketchEngine'i eesti keele ühendkorpusest avaldiste abil keeruline kätte saada, sest korpuses ei ole fraase märgendatud. Esialgsete katsete käigus korpuse materjali analüüsida ei täheldanud ma selliste konstruktsioonide puhul ka märkimisväärset verbivormi varieerumist. Seega keskendun järgmistele konstruktsioonidele:

- kolm (pikka) meest tuleb/tuli, kolm (pikka) meest tulevad/tulid (nt *900 sõdurit kaitseb maja; Kaks kanget kivi jahvatavad aeglaselt*);
- tuleb/tuli kolm (pikka) meest, tulevad/tulid kolm (pikka) meest (nt *Tähtajaks vastas viis erakonda; 12. juulil sooritasid kaks lendobjekti kolmesaja kilomeetrise raadiusega ringikujulise lennu Lõuna Soome kohal*).

Töö jaoks materjali kogumisega alustades ja korpusest sobivate lausete leidmiseks avaldasi katsetades olid peamiseks murekohaks laused, kus nimisõna laiendas lisaks arvsõnale veel mõni, näiteks omadus- või asesõna (*kolm head sõpra tulevad, kolm minu sõpra tulevad*). Selleks, et mitte uurida ainult arv-, nimi- ja tegusõnast koosnevaid lauseid,

otsustasin avaldisele lisada tingimuse, et arv- ja nimisõna vahel võib olla 0 kuni 1 sõna. See tähendas, et pidin analüüsiks sobivad laused korpusest kogutud andmestikust käsitsi tuvastama. Minu uurimusest jäid sedapuhku siiski välja ka sellised laused nagu *kolm väga pikka meest tulevad ja kolm minu väga head sõpra tulevad*, sest kui lubada avaldisel leida ka selliseid lauseid, kus arv- ja nimisõna vahel on rohkem kui üks sõna, saan väga palju materjali, mis minu uurimuseks ei sobi, ning ajakulu selle läbitöötamiseks ei oleks mõistlik. Samuti ei analüüsi ma selles töös eraldi mittearvsõnalise laiendi asukoha mõju ehk juhtumeid, kus nimisõna laiend esineb arvsõna ees (nt *viimased kolm meest*), sest sellised laused leiab kasutatud avaldis üles vaid siis, kui lause sõnajärjes on esimesel kohal arvsõna, sellele järgneb nimisõna ja viimaks verb.

Avaldis, millega korpusest lauseid otsisin, on järgmine:

```
( [tag="N"] [tag!="J"] {0,1} [tag="S"&features="sg_p"] [tag="V"&features="s|sid|b|vad"&word!="on"] within<s/> ) | ( [tag="V"&features="s|sid|b|vad"&word!="on"] [tag="N"] [tag!="J"] {0,1} [tag="S"&features="sg_p"] within<s/> )
```

Päring leiab lauseid, mis sisaldavad

- 1) arvsõna, millele järgneb 0 sõna või 1 sõna, mis ei ole sidesõna. Neile omakorda järgneb ainsuse osastavas käändes nimisõna ja seejärel verb, mis on ainsuse või mitmuse olevikus või minevikus. Verbi puhul on välistatud sõna *on*, sest olema-verbil olevikuvormi puhul ei ole võimalik mõista, kas kasutatud on ainsust või mitmust. Laused, mida päringu esimene pool leiab, on nt *kaks sõbralikku meest tuleb ja kaks sõbralikku meest tulid*;
- 2) või verbi, mis on ainsuse või mitmuse olevikus või minevikus, millele järgneb esmalt arvsõna ning seejärel 0 sõna või 1 sõna, mis ei ole sidesõna. Konstruktsiooni viimane liige on taas ainsuse osastavas käändes nimisõna. Laused, mida päringu teine pool leiab, on nt *tuli kaks sõbralikku meest ja tulevad kaks sõbralikku meest*.

Pärast päringu sisestamist kasutasin SketchEngine'i *shuffle*-funktsiooni ja sain laused juhuslikus järjestuses. *Shuffle*-funktsiooni kasutamine oli vajalik, sest kui SketchEngine'i otsingusse sisestada avaldis, mis sisaldab mitut otsitavat, siis on esimesed just need tulemused, mis vastavad päringu esimesele otsitavale. Kuna analüüsin üldist keelekasutust, pidin võtma juhusliku arvu juhuslikele otsitavatele konstruktsioonidele vastavaid lauseid. Seega laadisin Sketch Engine'ist XLSX-formaadis alla esimesed 10 000 päringuga leitud tulemust ja alustasin magistritöös analüüsimiseks sobivate lausete otsimist. Selleks, et iga seletava tunnuse mõju hindamiseks oleks piisavalt andmepunkte, pidin leidma vähemalt 2000 lauset, mis otsitavat konstruktsiooni esindaksid. Lõpptulemusena tuli läbi vaadata 7000 lauset selleks, et leida magistritöös kasutatud 2264 lauset.

Järgmisena märgendasin iga sobiva lause puhul 8 tunnust: uuritava tunnuse verbivormi arvukategooria ehk verbi vormi ainsuslikkus või mitmuslikkus ning seletavate tunnustena verbi transitivsus lause kontekstis (ehk objekti olemasolu lauses), verbi ajavorm, kvantorifraasi arvsõna, kvantorifraasis oleva nimisõna referendi elusus, verbi paiknemine kvantorifraasi suhtes, kvantorifraasiga väljendatava hulga ebamäärasus ja tekstitüüp (ühendkorpuse alamkorpuse). Selle jaoks koostas Exceli tabeli, millesse märkisin tunnuste kategooriad. Pärast kodeerimist kasutasin andmete kvantitatiivseks analüüsiks statistikaprogrammi R (R Core Team 2022).

R-is viisin läbi ühetunnuselise analüüsi, et hinnata iga tunnuse olulisust ja selle mõju uuritavale tunnusele ehk verbivormi ainsuse või mitmuse valikule. Selleks kasutasin hiirruut-testi, mille tulemuste põhjal sain teada, kas seletav tunnus aitab selgitada uuritava tunnuse valikut. Craméri V seosekordaja abil mõõtsin uuritava ja seletava tunnuse vahelise seose tugevust. Kvantorifraasi arvsõna tunnuse puhul kasutasin Wilcoxon'i astaksummatesti ehk Manni-Whitney U-testi, et näha, kas kvantorifraasi aluseks olevad põhiarvsõnadega väljendatud arvude keskmised on ainsuses ja mitmuses kasutatud verbivormidega statistiliselt oluliselt erinevad. Samuti vaatlesin andmestikus vähemalt 10 korral esinenud verbide jaotumist ainsuslike ja mitmuslike vormide vahel, et näha, kas ka üksikute verbidega joonistuvad välja eelistusmusterid. Mitmetunnuselise analüüsi jaoks

kasutasin otsustuspuud, mis kuvab olulisemaid seoseid ja seletavate tunnuste mõju hierarhilisel graafikul. Otsustuspuu usaldusväarsuse kontrollimiseks kasutasin juhumetsa (*random forest*) meetodit.

Valitud andmeanalüüsi meetodeid on varem Eesti keele- ja süntaksiuurimustes edukalt kasutatud (hii-ruut-testi kohta vt nt Metslang (2015), Siiman (2016), Teras (2018); otsustuspuude ja juhumetsade kohta nt Klavan jt (2015), Lindström ja Uiboed (2017), Lindström jt (2018), Teras (2019)). Seega leian, et arvestades minu andmeid ja uurimuse eesmärki, on need siingi sobilikud.

Analüüsi käigus kogutud tulemuste põhjal leidsin vastused oma uurimisküsimustele ning tegin järeldused selle kohta, millised keelesisesed muutujad mõjutavad verbivormi arvukategooria valikut.

4. Käsitletavad tunnused

Magistritöös uuritav nähtus on öeldisena esineva verbivormi arvukategooria ehk ainsuse ja mitmuse vorm lauses, kus aluseks on arvsõnalise põhisõnaga kvantorifraas. Valikut seletada võivateks tunnusteks olen valinud verbi ajavormi, verbi paiknemise kvantorifraasi suhtes, verbi transitiivsuse lähtuvalt sellest, kas lauses on olemas eksplitsiitne ehk selgesti väljendatud objekt, kvantorifraasi arvsõna, kvantorifraasiga väljendatava hulga ebamäärasuse, kvantorifraasi subjekti referendi elususe ja tekstitüübi. Järgnevalt kirjeldan kõiki tunnuseid.

4.1. Verbivormi arvukategooria

Verbivormi arvukategooria ehk verbi vormi ainsuslikkus ja mitmuslikkus on magistritöö uuritav tunnus, mis varieerub lausetes, milles aluseks on kvantorifraas. Ainsus ja mitmus väljenduvad verbi pöörde kategoorias ja erinevad teineteisest pöörde lõppude poolest. Eesti keeles on kolm ainsuslikku ja kolm mitmuslikku pöörde vormi, millest magistritöös vaatlen ainsuse ja mitmuse 3. pööret ehk nt *tuleb* või *tulevad*. (Erelt 2017a: 196–199)

Ajalooliselt ei ole eesti keele 3. pöörde ainsusel ja mitmusel olnud oma tunnuseid, kuid tänapäeva eesti keeles loetakse tunnusteks järgnevat:

- ainsuse 3. pöörde tunnused on *-b*, mis esineb kindla kõneviisi olevikus (nt *ta tuleb*), ja *0*, mis esineb kõigil muudel juhtudel, nt kindla kõneviisi lihtminevikus (*tuli*), tingivas kõneviisis (*tuleks*), käskivas kõneviisis (*tulgu*);
- mitmuse 3. pöörde tunnus saab olla *-vad*, mida kasutatakse kindla kõneviisi oleviku puhul (nt *nad tulevad*); *d*, kui verb on kindla kõneviisi lihtminevikus või mõnes tingiva kõneviisi vormis (*tulid*); *0*, kui verb on käskivas kõneviisis (*tulgu*). (EKK 2020: 239–240)

Märgendasin andmestikus ainsuse ja mitmuse tunnusena ainsuse, kui verb oli kindla kõneviisi ainsuse 3. pöördes, ning mitmuse, kui verb oli kindla kõneviisi mitmuse 3. pöördes. Ajaliselt võivad verbid olla nii olevikus kui ka liht- või enneminevikus: *tuleb*, *tuli* ja *oli tulnud* või *tulevad*, *tulid* ja *olid tulnud*. Täisminevikus olnud verbid jäid uurimusest välja, sest nende puhul polnud võimalik eristada, kas need on ainsuses või mitmuses (nt *on märganud* on sama nii ainsuse kui ka mitmuse vormis).

Seletavate tunnuste abil püüan välja selgitada, mis mõjutab verbivormi arvukategooria varieerumist.

4.2. Verbi ajavorm

Verbi ajavormi olen valinud üheks magistritöös uuritava tunnuse valikut seletada võivaks tunnuseks. On võimalik, et kui verbivormil on vaja markeerida üht grammatilist kategooriat (minevikku), siis markeeritakse tõenäolisemalt ka teisi (mitmust) või püütakse, hoopis vastupidi, vorme võimalikult ökonoomseks teha ja eelistatakse minevikuajajas verbidega ainsuslikke vorme.

Aeg ehk tempus on verbi kategooria, mille abil saab väljendada ajalisi suhteid verbiga märgitava tegevuse ja kõnehetke või mõne muu tegevuse toimumise hetke vahel. (EKK 2020: 242) Eesti keele ajavormid on olevik, lihtminevik, täisminevik, enneminevik, kaudses ja tingivas kõneviisis ka üldminevik (Erelt 2017a: 129, 136). Eesti keeles saab ajamääruste ja tulevikutähendusega verbidega väljendada tulevikku, kuid tuleviku väljendamiseks ei ole eraldi morfoloogilist vormi (EKK 2020: 407–408).

Magistritöös tegelen nii oleviku- kui ka minevikuvormis verbidega, mis on jaatavas kõneliigis, isikulises tegumoes ja kindlas kõneviisis. Mineviku puhul ei saa analüüsida täisminevikus olevaid verbe, sest verbi *olema* olevikuvorm on nii ainsuse kui ka mitmuse kolmandas pöördes *on* ning seetõttu pole võimalik vahet teha, kas tekstis on mõeldud ainsust või mitmust. Seega on verbi *olema* puhul arvesse võetud vaid minevikuvorme. Enneminevikus olnud laused liigitusid kodeerides mineviku kategooria alla, sest

enneminevikulisi verbe esines andmestikus liiga vähe, et neid eraldi analüüsida. Niisiis on töös märgendatud kategooriatena olevik ja minevik.

Järgnevalt kaks näidet lausetest:

- 1) Võimas video: kaks küürvaala *ähvardavad* kalamehe kajaki ümber lükata.
- 2) Septembris *lendas* 51 960 reisijat.

Esimeses lauses (1) on verb märgendatud oleviku kategooriasse kuuluvana, teises lauses (2) mineviku kategooriasse kuuluvana.

4.3. Verbi paiknemine kvantorifraasi suhtes

Verbi paiknemine, edaspidi töös viidatud ka kui sõnajärg, on eesti keeles üsna vaba. Seda saab muuta vastavalt infostruktuurile ehk tuntud ja uute referentide esitamise vajadusele ning fookusele. Samuti võib sõnajärg muutuda sõltuvalt lausetüübist ja kommunikatiivsest eesmärgist. (Lindström 2017: 547) Tavaliselt paikneb lause alguses see, mille kohta lauses midagi öeldakse ehk lause teema, lause lõpuossa aga see, mida öeldakse ehk lause reema (EKK 2020: 388).

Põhisõnajärg eesti keeles on SVX: subjekt ehk alus ehk S, verb ehk öeldis ehk V, öeldise laiend ehk X, kusjuures X võib olla objekt (O), öeldistäide (P) või määrus (A). (Lindström 2017: 547) Öeldisverbi positsioon lauses varieerub teatud määral, kuid üldiselt paigutatakse eesti keeles öeldisverb lauses teisele kohale, olenemata, millega lause algab (Lindström 2017: 551). Nõnda on nt lausetes *kolm meest tulid eile* ja *eile tulid kolm meest* aluseks kvantorifraas ja verb teisel kohal, ehkki fookus on mõlemal lausel erinev. Sõnajärg on mõningal juhul oluline ka lause tuumargumentide eristamisel, kui ühildumisest ja käändemärgistusest ei piisa, et teha kindlaks, milline on subjekt ja milline objekt. Neil juhtudel – nominaalfraasi vormihomonüümia transitiivses lauses (*muna õpetab kana*); asesõnade vormihomonüümia, kui nii subjekti kui ka objekti asendis on asesõna *ta* (*ta tõigi ta ära*); subjekt ja objekt on mitmuse nominatiivis (*politseinikud pidasid kinni kurjategijad*); subjekti, objekti või mõlema argumenti peasõna on

(semantiliselt ühilduv) kvantorifraas (*päästjad toimetasid kaldale kolm kalameest*); öeldisverb lauses on möönvas kõneviisis (hoolekogu valigu direktor) – aitavad subjekti ja objekti eristada sõnajärg ning semantika ja üldised teadmised. Kõnealuste juhtumite puhul paikneb subjekt lauses enamasti enne objekti. (Lindström 2004: 40–49)

„Eesti keele käsiraamat“ märgib, et kui kvantorifraas on lause alguses ning sellele järgneb verb, siis on soovituslik kasutada verbi mitmuses (nt *kaks koera jooksevad*). Kui kvantorifraas on lause lõpus ja verb on kvantorifraasi ees, siis tuleks eelistada ainsust (nt *jookseb umbes kümme koera*). (EKK 2020: 400) Ka Mati Erelt ütleb, et põhiarvsõnaline kvantorifraas ühildub ainsuse ja mitmusega, ent soovitab olenevalt verbi asukohast ja hulga täpsusest kasutada üht või teist (Ereilt 2019: 17).

Eelnevalt väljatoodu põhjal on sõnajärg verbivormi ainsuse või mitmuse valikut mõjutada võiv tunnus. Järgnevalt esitan näited verbi paiknemise kohta kvantorifraasi suhtes:

- 3) Eestis *kasvab* kakskümmend pajuliiki.
- 4) Neli meest *olid* nüüd kohal.

Näites 3 paikneb verb *kasvab* kvantorifraasi *kakskümmend pajuliiki* ees. On näha, et kvantorifraas paikneb lause lõpus ning verb on ainsuses. Näites 4 paikneb verb *olema* kvantorifraasi *neli meest* järel. Kvantorifraas paikneb lause alguses ja verbi on kasutatud mitmuslikus vormis.

4.4. Verbi transitiivsus

Transitiivsus on verbi omadus, millest sõltub, kas verb seob endaga objekti ning kas verbi tegevus on suunatud teisele entiteedile ja mõjutab selle seisundit (Hopper, Thompson 1980). Lihtsustatult öeldes on need laused, milles on olemas sihitis ehk objekt, transitiivsed laused; laused, milles sihitist ei ole, on intransitiivsed laused (Metslang 2017b: 261). Transitiivsus ei ole siiski vaid grammatiline kontseptsioon märkimaks sihitise olemasolu lauses, vaid see aitab keelekasutajal mõista ja luua seoseid tegude ja sündmuste vahel. See tähendab, et lause transitiivsus ei ole alati kindlalt määratletav, vaid

sõltub sageli keelest, kontekstist ja on nähtusena skalaarne. (Hopper, Thompson 1980) Eesti keele verbide transitiivsuse määramist on käsitlenud nt Natalia Vaiss (2020), kes on samuti välja toonud, et transitiivsus ei ole selgelt piiritletav ja üheselt mõistetav, vaid mitmete häägusate äärealade ja alamkategoriatega. Transitiivsuse tugevuse määramiseks võib kasutada prototüüpset transitiivsust ehk olukorda, kus transitiivsus on kõige klassikalisemal moel määratletud – verb kirjeldab tegevust, mis põhjustab muutuse patsiendi seisundis – ning on selge, et verb nõuab sihitist (Tsunoda 1999: 384). Sedavõrd kompleksne nähtus, mis põhineb suuresti kontekstil, mõjutab tõenäoliselt ka verbi arvuvormi valikut.

Hopper ja Thompson on transitiivsust skalaarse nähtusena käsitlenud ja välja toonud, et lause on seda transitiivsem, mida paremini on täidetud järgmised tingimused: lause on jaatav; lauses on suure agenttiivsusega agent ning täielikult kaasatud ja selgelt piiritletav objekt; kindlas kõneviisis verbiga väljendatakse aktiivset tegevust, mis on teeline, tahtlik, määratud lõpuga (Hopper, Thompson 1980). Transitiivsuse tugevuse ehk minu töö puhul sihitise ehk objekti olemasolu määramiseks olen otsustanud lähtuda kontekstist. Seega vaatlen lauset kui tervikut, mitte ei keskendu vaid verbile, sest on verbe, mis võivad kontekstist mõjutatuna olla kas transitiivsed või intransitiivsed ning vaid verbi analüüsimine ei annaks täpseid tulemusi. Seejuures ei erista ma nt täis- või osasihitist, vaid ainult seda, kas sihitis on lauses olemas või mitte.

Sihitist nõudvad verbid väljendavad abstraktset suhet või põhjustamist ja sellised verbi on nt *põhjustama* ja *eeldama*. Samas on mitmeid transitiivseid verbe, mida saab kasutada ilma sihitiseta, nt *sööma* ja *harjutama*. Selliste verbidega moodustatud laused saavad niisiis olla transitiivsed või intransitiivsed olenevalt sellest, kas lauses on sihitis. (Metslang 2017b: 261) Intransitiivsed ehk sihitist mitte nõudvad verbid, nagu nt *magama* või *elama* esinevad lauses ilma sihitiseta. Vahel aga võivad intransitiivsed verbid saada täpsustava sihitise, nt *käima* võib olla transitiivne, kui lauses on *käima pikka teed*. (Metslang 2017b: 262) Kontekstipõhiselt transitiivsust muutvaid verbe ja nende sihitisi on lähemalt vaadanud nt Anne Tamm (vt Tamm 2012).

Eksplitsiitne sihitis ehk grammatiline objekt on vormilt enamasti osastavas, omastavas või nimetavas käändes noomeni- ehk käändsõnafraas, näiteks *kolm meest nägid koera*. Lisaks saab sihitis esineda *da-* või *vat-*infinitiivina, nt *kolm meest kavatsevad loota*, või kõrvallausena, nt *kolm meest teadsid, et nii läheb*. (EKK 2020: 420)

Magistritöös olen lause transitiivsust määranud lausest lähtuvalt, et saaksin võimalikult täpse ülevaate selle kohta, kas transitiivsus võib mõjutada verbivormi ainsuse või mitmuse valikut. Vaatasin igat andmestiku lauset eraldi ning kontrollisin esmalt, kas lauses olev verb on lause kontekstis transitiivne või intransitiivne. Seejärel vaatasin, milline on objekti vorm. Kui lauses oli intransitiivne ja täpsustava sihitiseta verb, märkisin lause intransitiivseks. Kui lause oli transitiivne, vaatasin, kas lauses olev objekt on noomenifraas, *da-* või *vat-*infinitiiv või kõrvallause ning grupeerisin laused objekti vormi põhjal. Järgnevalt toon iga liigituse kohta näite oma andmestikust:

- intransitiivne lause – *Septembris lendas 51 960 reisijat*. Lauses on verb *lendama* intransitiivne ja sihitist ei ole;
- noomenifraasilise sihitisega lause – *Kaks noort keeleteadlast tutvustasid oma liiviteemalisi doktoritöid*. Verb *tutvustama* on transitiivne, objekt lauses on noomenifraas *oma liiviteemalisi doktoritöid*;
- *da-* või *vat-*infinitiivist sihitisega lause – *Kaheksa inimest otsustas tulla ristimisele*. Verb *otsustama* on transitiivne, sihitis lauses on *da-*infinitiiv *tulla*.
- kõrvallauselise sihitisega lause: *Kuus kohtunikku ütlesid, et selles olukorras oleks olnud vajalik luua erand, mis lubab valitsusel peatada põhiõigused*. Verb *ütleva* on transitiivne, sihitis lauses on kõrvallause *et selles olukorras oleks olnud vajalik luua erand, mis lubab valitsusel peatada põhiõigused*.

Minu andmestikus polnud lauseid, kus oli verb, mille seotud laiend on sihitis, kuid kus sihitist vormiliselt ei olnud, (nt *kolm meest olid eeldanud [mida?]* või *kolm meest olid põhjustanud [mille?]*), mistõttu sain märkida vaid intransitiivsust ja transitiivse lause objekti tüüpi.

4.5. Kvantorifraasi arvsõna

Kvantorifraasi arvsõna ehk kvantorifraasis olevat arvu märgin arvsõna põhjal ja kvantorifraasi arvsõna on samuti arvu verbivormi ainsuslikkust või mitmuslikkust seletav tunnus. Arvsõnad ehk numeraalid on sõnad, millega saab väljendada hulka või järjestust ning mis jagunevad põhiarvsõnadeks (nt *kaks*, *viidendik*) ja järgarvsõnadeks (nt *kolmekümnes*). (Erelt 2017b: 59)

Kuna eesti keele kvantorifraasi ühildumist verbi ainsuse ja mitmuse vormiga ei ole varem uuritud, põhineb tunnuse valik suuresti oletusel, et arvu suurusjärk võib mõjutada verbivormi valikut. Oletust toetab õigekirjasoovitus, millele on osutanud Mati Erelt: kui kvantorifraasi kaudu väljendatava hulga suurus on ebamäärane, siis tuleks verbivormi puhul eelistada ainsust (Erelt 2019: 17). Mida suuremaks muutub kvantorifraasis olev arv, seda ebamäärasem ja vähem hoomatav see inimese jaoks on. See omakorda tähendab, et suuremate arvudega võib keelekasutaja eelistada pigem ainsuslikku verbivormi. Töös tegelen põhi- ja asearvsõnaliste kvantorifraasidega, nt *kaks*, *kaheksasada* ja *paar*.

4.6. Ebamäärane hulk

Täpset hulka väljendava kvantorifraasi puhul on Mati Erelt soovitanud kasutada mitmuslikku verbivormi, ebamäärase kvantorifraasi puhul aga ainsuslikku (Erelt 2019: 17). Ehkki magistritöös ei mõista ma mingist kindlast arvust suuremat arvu ebamäärasena, sest selle piiri märkimine oleks keeruline, valisin ebamäärasuse üheks võimalikuks seletavaks tunnuseks ning leidsin, milliste vahenditega ebamäärasust kajastatakse.

Ebamäärasuse tunnuseks mõistan andmestikus seda, kui kvantorifraasi arvuna on esitatud arvude vahemik (nt *3–4*, *100–1000*) või on arvu väljendatud umbmääraselt, nt sõnaga *paar*, *mõni* või *kümmekond*. Samuti pean ebamääraseks neid kvantorifraase, millele eelneb vahetult ligikaudsust näitav sõna. Minu andmestikus esinesid järgmised ligikaudsust näitavad sõnad: *ligi*, *ligemale*, *pea*, *umbes*, *mingi*, *kuskil*, *vähemalt*.

Kodeerides märgendasin ebamäärasena kõik laused, milles esines arvuvahemik, arvu umbmäärane väljendamine või ligikaudsust näitav sõna. Kategooriad tekkisid magistritöö andmestiku põhjal ning järgnevalt toon näited iga kategooria kohta:

- arvuvahemik – *72 000– 89 000 inimest saab aru ja veidi räägib*. Selles lauses on kvantorifraasi arv esitatud vahemikuna 72 000 – 89 000. Ei ole täpselt teada, mitut inimest on silmas peetud ja see muudab lause ebamääraseks;
- arvu umbmäärane väljendamine – *Kümmekond inimest mõtlevad kiiruga välja reformi mis tuleb 100% praak*. Lauses on kvantorifraasi täpne arv asendatud sõnaga *kümmekond*, mis tähendab, et inimesi oli kümne ringis. Ei ole selge, kas neid oli näiteks 9 või 11, niisiis on hulk ebamäärane;
- ligikaudsust näitav sõna – *Umbes 100 politseinikku sai kokkupõrgetes kannatada*. Kvantorifraasile eelneb ebamäärastav sõna *umbes*, mis näitab, et politseinikke võis olla 100, aga võis olla ka veidi vähem või rohkem. Seega on hulk lauses ebamäärane, sest ei saa öelda, et täpselt sada politseinikku sai kokkupõrgetes kannatada.

Ebamääraseid hulki sisaldavate lausete vähesuse tõttu ei erista ma analüüsis ebamäärasuse kategooriate mõju. Töös uurin, kas väidetav seos ebamäärasuse ja verbivormi valiku vahel ning sellekohaste keelekorralduslike soovitude mõju kajastub ka minu andmetes.

4.7. Subjekti referendi elusus

Elusust saab määrata subjekti nimisõna põhjal ja siinses töös lähtun peamiselt referendi bioloogilisest elususest. Bioloogiliselt elus on nt inimesed ja loomad, elutud aga nt kivid ja vesi. (Kittilä jt 2011: 9) Olen töös lugenud elutuks ka taimed.

Elususe valisin verbivormi valikut seletada võivaks tunnuseks, sest on täheldatud, et subjekti referendi elusus võib mõjutada lause sõnajärge (vt nt Myhill 1992, Helasvuo 2001) ja mida rohkem elus on see, kellele lauses viidatakse, seda rohkem kaldub see referent esinema pigem lause alguses (Sornicola 1999: 375, 377). Liina

Lindström on leidnud, et subjekti referendi elusus on sõnajärje puhul oluline ka eesti keeles ning kirjutanud, et kui suulise keele lausungi subjekt on inimene, asub subjekt tõenäolisemalt verbi ees. Kui subjekt on mitteinimene, valitakse pigem vastupidine sõnajärg, kus subjekt asub verbi järel. (Lindström 2001: 94) Kuna sõnajärjest ehk verbi vormi asukohast lauses on eesti keele kvantorifraasi ja sellega ühilduva verbi vormi valimiseks antud õigekirjasoovitusi (Erelt 2019: 17), võib eeldada, et subjekti referendi elusus võib oluline olla ka verbi vormi valimisel.

Ehkki elusust skalaarse nähtusena on käsitletud mitmete hierarhiate kaudu (vt nt Silverstein 1976, Dixon 1979), on üks universaalsemaid elusushierarhiaid Crofti hierarhia:

1. ja 2. isik > 3. isik > pärisnimi/sugulusnimi > inim-NP > elus NP > elutu NP (Croft 2003: 310).

Selle põhjal on kõige rohkem elus see, mis käitub lauses kõige tõenäolisemalt agendina. Kõige rohkem elus referent paikneb skaala vasakul otsal, ning kõige vähem elus on see, mis paikneb paremal otsal. Seega, 1. ja 2. isikut (*mina, sina, meie, teie*) käsitletakse keeles rohkem elus olevatena kui 3. isikut (*tema, nemad*), 3. isikule järgneb pärisnimi või sugulasnimi (*Mari, õde*), sellele omakorda inimesele viitav nimisõnafraas (*inimene, tüdruk*), seejärel muudele elusatele entiteetidele viitav nimisõnafraas (*koer, hiir*) ja viimaks elututele objektidele viitav nimisõnafraas (*laud, kivi*). (Kittilä jt 2011: 9–10) Selle skaala puhul tuleb meeles pidada, et tegelikult on elusa nimisõnafraasiga viidatud referent sama elus kui 1., 2. või 3. isikuna märgitud referent, seega mõjutab siinkohal elususe määramist ka kõneleja positsioon ja see, kui suurt empaatiat kõneleja referendi vastu tunneb. Lisaks on oluline referendile viitava fraasi definiitsus, nt asesõnad ja nimed on alati definiitsed, seevastu elutule viitav nimisõnafraas on vähem definiitne. (Whaley 1997: 172–173)

Eesti keele kohta on täheldatud, et enamasti kehtib tendents viidata elusale referendile sõnaga *kes*, elutule referendile sõnaga *mis*. Küll aga pole see alati nii, näiteks on asesõna *kes* kasutust elutule referendile viitamisel eesti keeles uurinud

Hanna Pook (vt Pook 2019). Erinevusi võib olla ka loomade puhul, mil võidakse kasutada ka sõna *mis*, või siis, kui referente on mitu. Kui referentide hulka on võimalik vaadelda eraldi, kasutatakse pigem relatiivpronoomenit *kes*, kui aga koos, siis *mis*. Mida aktiivsema ja elusamana tundub referent lauses olevat, seda tõenäolisemalt saab seda asendada sõnaga *kes*. (Erelt 2017c: 743)

Erinevad hierarhiad kajastavad kõneleja hinnangut erinevate entiteetide elususe kohta. Näiteks võib üks elutule referendile viitav nimisõna asetseda hierarhia kõrgemal astmel kui teine. Selles olukorras on üks referentidest kõneleja jaoks rohkem elus, sest inimesele omased prototüüpsed tunnused väljenduvad hierarhias kõrgemal asuva referendi puhul tajutavamalt. (Ji, Liang 2018) Eesti keeles on elususele lähenetud näiteks elusus- ja individuaalsushierarhiate kaudu.

Elusushierarhia on järgnev:

inimene > kõrgem loom > madalam loom > mitteolend.

Elusushierarhia kohaselt on inimene kõige rohkem elus, talle järgneb kõrgem loom ehk näiteks põder, talle omakorda madalam loom ehk näiteks hiir. Eluta on mitteolend ehk nt diivan või valgus.

Individuaalsushierarhia on järgmine:

indiviid > distributiivne hulk > kollektiiv.

Individuaalsushierarhia järgi on kõige rohkem elus indiviid ehk inimene või koer – miski, mida on üks ja millel on elus olemise tunnused. Indiviidile järgneb distributiivne hulk ehk selline hulk, milles saab hulga elemente eraldi vaadelda: nt iga laps hulgast. Kollektiiv on individuaalsushierarhias kõige vähem elus, nt valitsus või brigaad. (Erelt 2017c: 743)

Minu andmestiku põhjal tundus, et lisaks elususele oleks vajalik eristada ka seda, kas subjekti referent on konkreetne või abstraktne. „Eesti keele käsiraamatus“ on märgitud, et tähenduse põhjal jagunevad nimisõnad konkreetseteks ja abstraktseteks selle järgi, kui

tajutav on sõnaga tähistatav: meeleliselt tajutav on konkreetne (nt *puu*), meeltega haaramatu aga abstraktne (nt *idee*) (EKK 2020: 158).

Magistritöös lähtun referendi elususe määramisel kvantorifraasi laiendava nimisõnafraasi bioloogilisest elususest, sellest, kui konkreetne või abstraktne nimisõna on ning sellest, kuhu paigutub subjekti referent individuaalsushierarhias. Samuti vaatan konteksti, milles referent esineb, sest nt *asi* või *teema* võib olla nii konkreetne ja elutu kui ka abstraktne ja elutu. Kategooriate loomisel lähtusin oma andmestikust ja sellest, et kategooriad ei oleks liialt üldistavad (nt elus ja elutu), samas mitte liialt kitsad (nt inimene, naine, mees). Ma ei saanud täielikult tugineda Crofti elusushierarhiale (Croft 2003: 310) või elus- ja individuaalsushierarhiatele (Erelt 2017c: 743), sest minu andmestikus oli mitmeid abstraktseid mõisteid, mida neis ei eristata ning seetõttu jäänuks need hierarhiad minu töö jaoks liiga üldiseks. Sellest tulenevalt otsustasin luua kategooriad, mis põhinevad minu andmestikul ja kajastavad seda kõige paremini.

Magistritöö andmestikus kodeeritud elususe kategooriad on:

- inimene, nt lauses *Vanemas vanuserühmas võistles kümme lauljat*;
- elus ja konkreetne, nt lauses *kaks jõehobu lendasid, kolmas pööras paremale*;
- elutu ja konkreetne, nt lauses *Teid kasutas 6344 sõidukit*;
- elutu ja abstraktne, nt lauses *2 jääraja kogemust näitavad, et UrS4 seisab paremini rajal, kui S2*;
- aeg, nt lauses *Möödusid kaks minutit*;
- protsent, nt lauses *44 protsenti ütles, et loeb vähem paberil ajalehti ja ajakirju*.

Kategooriad tekkisid kvantorifraasi laiendite põhjal: kui kvantorifraas oli nt *kolm meest*, liigitus subjekti referent elususest inimese kategooria alla; kui kvantorifraas oli nt *kolm ideed*, liigitasin referendi elususe järgi elutusse ja abstraktsesse kategooriasse. Inimese kategooria alla liigitasin nii sõna *inimene* kui ka inimest tähistavad nimed või muud inimesele viitavad sõnad, nt *tüdruk* või *õpetaja*. Elus ja konkreetsete mõistete alla liigitusid loomad, nt *lammas*, linnud, nt *kukk*, ja kollektiivid, nt *meeskond*. Elutute sõnade kategoriseerimisel pöörasin tähelepanu ka sellele, kas sõnaga tähistatav on füüsiliselt

olemas või ettekujutatav. Elutusse ja konkreetseesse kategooriasse kuuluvad mh asjad, nt *lamp*, sõidukid, nt *buss*, asutused, nt *kool* jms. Elutu ja abstraktse kategooria alla liigitusid mõisted, mis tähistavad midagi, millel pole bioloogilisi elususe tunnuseid ning mida ei saa piiritleda, silmaga näha või käega katsuda, nt *kogemus*, *idee* ja *asjaolu*. Aja kategoorias olevad sõnad märgivad ajaliselt piiritletut, nt *päev*, *hetk*, *aasta*. Protsendi kategooria vajalikkuse tingis see, et andmestikus oli lauseid, milles kvantorifraasis oli subjektina protsent, kuid millest polnud aru saada, kas mõeldakse nt protsenti küsitlusele vastajatest või kampaaniat toetanud asutustest. Seepärast otsustasin kõik laused, milles aluseks oleva kvantorifraasi subjekt on sõna *protsent*, märkida elususele protsendi kategooria alla.

4.8. Lause tekstitüüp

Magistritöös uurin eesti üldkeelt. Kuna töös uuritava nähtuse – verbivormi varieerumine aluseks oleva arvsõnalise kvantorifraasiga ühildumise – osas on Mati Erelt jaganud keelelisi soovitusi (Erelt 2019: 17), võib nähtuse tegelik kasutus sellest mõjutatud olla. Korpuste põhjal toimetamise mõju siiski otseselt hinnata ei saa, sest pole teada, millised tekstid on toimetatud, millised toimetamata. Võib vaid oletada, et teadustekstid ja ajakirjandustekstid on sagedamini toimetatud, foorumipostitused on reeglina toimetamata. Kuna eri tekstitüüpidel on erinevad varieeruvad keelekasutuse tavad ja stiil, on võimalik, et tekstitüüp mõjutab ka kvantorifraasiga seotud verbi vormi valikut.

Seetõttu oli andmestiku kogumiseks oluline jälgida, et andmestik sisaldaks nii toimetatud kui toimetamata tekste ning erinevaid tekstitüüpe ehk žanre. Kuna eri tekstitüüpidel on erinevad varieeruvad keelekasutuse tavad ja stiil, on võimalik, et tekstitüüp mõjutab ka kvantorifraasiga seotud verbi vormi valikut.

Eesti keele 2021. aasta ühendkorpuses on 11 allkorpust, mis omakorda sisaldavad erinevaid keeleregistreid ja tekstitüüpe (Koppel, Kallas 2022: 209). Oma andmestiku piiritlemiseks otsustasin uuritavateks allkorpusteks valida veebikorpused (Web 2013, Web 2017, Web 2019, Web 2021) ning avatud lähtekoodiga teadusartiklite

corpuse (DOAJ). Neis sisalduda võivateks tekstitüüpideks valisin foorumipostitused, ajakirjanduslikud väljaanded ja teaduskirjanduse. Tuginedes eelnevalt välja toodud soovitusel eelistada mitmuslikku verbivormi täpset hulka väljendava ja lause alguses asuva kvantorifraasiga ning ainsuslikku verbivormi ebamäärast hulka väljendava ja lause lõpus asuva kvantorifraasi puhul (Erelt 2019: 17), võib oletada, et ka tekstide toimetatus on verbivormi varieerumisel oluline tegur. Eri registrites võib toimetatud tekstide hulk siiski varieeruda ja ei ole täpselt teada, kui palju eri allkorpustes toimetatud tekste on. Kuna andmestiku valim on juhuslik, pärineb igast tekstitüübist eri arv lauseid.

Eesti keele ühendkorpuses olid laused automaatselt märgendatud vastavalt sellele, millisest tekstitüübist need pärinevad. Järgnevalt kirjeldan, kuidas olid laused tekstitüübist lähtuvalt tähistatud ning toon näitelause igast tekstitüübist:

- foorumipostitused olid märgitud sõnaga *forums* ja foorumipostituste näitelause on *Kolm tegijat hakivad kartulid väikesteks kuubikuteks*;
- ajakirjanduse laused olid märgitud sõnaga *periodicals* ja ajakirjanduslikust väljaandest pärinev lause on *Võistlusel osales üheksa võistlejat Eestist ning 2 Lätist*;
- teaduskirjanduse laused olid märgitud sõnaga *academic* ning näitelause on *126 ametnikku sooritasid ELi põhiteadmiste testi*.

5. Andmete ühetunnuseline analüüs

Selles peatükis kirjeldan andmete ühetunnuselise analüüsi tulemusi. Kokku oli minu andmestikus 2264 lauset, analüüsisin neis kõiki märgendatud tunnuseid ja vaatasin eraldi ainsuse ja mitmuse esinemist verbides, mis kordusid andmestikus vähemalt 10 korda. Esmalt esitan koondtabeli (Tabel 1), millesse on märgitud tunnused, nende väärtused ja sagedused. Sulgudes on näidatud iga väärtuse esinemissagedus protsendiliselt.

Tabel 1. Tunnused, nende väärtused ja sagedused

Tunnus	Väärtused	Sagedus
Verbi arvuvorm	ainsus mitmus	1345 (59%) 919 (41%)
Verbi ajavorm	olevik minevik	845 (37%) 1419 (63%)
Verbi paiknemine kvantorifraasi suhtes	verb fraasi ees verb fraasi järel	1317 (58%) 947 (42%)
Subjekti referendi elus	inimene elus ja konkreetne eluta ja konkreetne eluta ja abstraktne aeg protsent	1140 (50%) 150 (7%) 601 (27%) 206 (9%) 126 (5%) 41 (2%)
Lause žanr	teadustekst foorumitekst ajakirjandustekst	20 (1%) 483 (21%) 1761 (78%)
Ebamäärane hulk	ebamäärane mitte ebamäärane	100 (4%) 2164 (96%)
Verbi transitiivsus	intransitiivne transitiivne: noomenifraas transitiivne: <i>da-</i> või <i>vat-</i> infiniitiv transitiivne: kõrvallause	1689 (75%) 448 (20%) 78 (3%) 49 (2%)
Kvantorifraasi arvsõna	0–850000000	keskmine = 1352493 standardhälve = 30080810 mediaan = 5

Tabelis 2 on toodud näiteread minu andmestiku kodeeringust.

Tabel 2. Näide andmestikust

Tekstitüüp	Lause	Verbi arvuvorm	Transitiivsus	Aeg	Sõnajärg	Arv	Elusus	Ebamäärasus
aja- kirjandus	Eestis kasvab kaks- kümme- pajuliiki	ainsus	intr	olevik	verb ees	20	elutu, konkreetne	ei

Tabelis 2 on lause ja selle kodeering. Lause *Eestis kasvab kakskümme pajuliiki* on pärit ajakirjandustekstide tekstitüübist. Lause verbivorm *kasvab* on ainsuslik ja selle ajavorm on olevik. Kuna *kasvab* on intransitiivne ehk sihitiseta, on transitiivsuse lahtrisse märgitud *intr*, mis tähistab intransitiivsust. Verb paikneb lauses kvantorifraasi ees, seega on sõnajärje lahtrisse märgitud *verb ees*. Kvantorifraasis arv on 20 ning fraasi subjekt *pajuliik* on elutu ja konkreetne. Kuna lauses ei kasutata ebamääraastavaid vahendeid nagu arvuvahemik, arvu umbmäärane väljendamine või ligikaudsust näitav sõna, on hulga ebamäärasuse tunnuseks märgitud mitte-ebamäärasust tähistav *ei*.

Järgnevalt esitan kogutud andmestiku põhjal risttabelid seletavate tunnuste ja väärtuste esinemissagedusest ainsuslike ja mitmuslike vormidega. Hii-ruut-testi (Plackett 1983) abil kontrollin, kas tulemus on statistiliselt oluline ehk kas seletava tunnuse väärtuse teadmine aitab ennustada verbi ainsuse või mitmuse valikut. Statistiliselt oluliseks pidasin seoseid siis, kui testi p-väärtus oli väiksem kui 0,05 ehk kui tõenäosus saada selliseid andmeid, juhul kui kahe tunnuse vaheline seos on tegelikult juhuslik, on väiksem kui 5%. Kategoriaalsete tunnuste puhul mõõtsin Craméri V (Cramér 1946) seosekordaja kaudu uuritava ja seletava tunnuse vahelise seose tugevust. Craméri V tulemus jääb alati vahemikku 0–1, mis võimaldab eri seletavate tunnuste seoste tugevust verbi arvukategooriaga omavahel võrrelda. Seletavate tunnuste seoste tugevuse hindamisel arvestan ka vabadusastmeid (Cohen 1988).

5.1. Verbi ajavorm

Lauseid, milles verbi ajavorm on olevikus (*jookseb, jooksevad*), on andmestikus 845 ehk 37% kõikidest lausetest. Lauseid, milles verbi ajavorm on minevikus (*jooksis, jooksid, oli jooksnud, olid jooksnud*), on andmestikus 1419 ehk 63% kõikidest lausetest, seega on rohkem neid lauseid, milles verb on olevikus. Tabelis 3 on esitatud ainsuse ja mitmuse jagunemine oleviku ja mineviku verbivormidega.

Tabel 3. Ainsuse ja mitmuse jagunemine oleviku ja mineviku verbivormidega

	Olevik	Minevik
Ainsus	447 / 53%	898 / 63%
Mitmus	398 / 47%	521 / 37%
Kokku	845 / 100%	1419 / 100%

Tabelist 3 nähtub, et nii olevikulise kui ka minevikulise verbivormi põhjal on tõenäolisem, et verb on ainsuses. 845 oleviku verbivormist 447 ehk 53% olid ainsuses (nt *kolm meest tuleb* või *tuleb kolm meest*), 398 ehk 47% mitmuses (*kolm meest tulevad* või *tulevad kolm meest*). 1419 mineviku verbivormist 898 ehk 63% olid ainsuses (*kolm meest tuli, tuli kolm meest*), 521 ehk 37% mitmuses (*kolm meest tulid, tulid kolm meest*).

Hii-ruut-test näitas, et seos on statistiliselt oluline ning ainsuse ja mitmuse verbivormide proportsioonid olevikus ja minevikus on piisavalt erinevad: $X^2(1, N = 2264) = 23,255$, p-väärtus $< 0,001$. Craméri V näitas siiski, et seos uuritava ja seletava tunnuse vahel on sellise vabadusastmete arvu (1) korral nõrk: 0,102.

5.2. Verbi paiknemine kvantorifraasi suhtes

Lauseid, milles verb on kvantorifraasi ees (nt *tuli kolm meest, tulevad kolm meest*), on andmestikus 1317 ehk 58% kõikidest lausetest. Lauseid, milles verb on fraasi järel (nt *kolm meest tulid, kolm meest tuleb*), on andmestikus 947 ehk 42% kõikidest lausetest.

Tabelis 4 on esitatud verbivormi ainsuse ja mitmuse jagunemine sõltuvalt verbi paiknemisest.

Tabel 4. Ainsuse ja mitmuse jagunemine sõltuvalt verbi paiknemisest

	Verb fraasi ees	Verb fraasi järel
Ainsus	1054 / 80%	291 / 31%
Mitmus	263 / 20%	656 / 69%
Lauseid kokku	1317 / 100%	947 / 100%

Tabelist 4 nähtub, et 1317 lausest, kus verb asus kvantorifraasi ees, on 1054 ehk 80% ainsusliku verbivormiga (nt *tuli kolm meest*), 263 ehk 20% mitmuslikuga (nt *tulevad kolm meest*). 947 lausest, kus verb asus kvantorifraasi järel, on 291 ehk 31% ainsusliku verbivormiga (*kolm meest tuleb*) ja 656 ehk 69% mitmuslikuga (*kolm meest tulid*). Järeldub, et juhtudel, kus verb on kvantorifraasi ees, on eelistatud pigem ainsuslikku verbivormi. Juhtudel, kus verb on kvantorifraasi järel, on valitud pigem mitmuslik verbivorm. See toetab „Eesti keele käsiraamatu“ tähelepanekut, et lause lõpus asuva kvantorifraasiga kasutatakse pigem ainsuslikku öeldist, lause alguses oleva kvantorifraasiga mitmuslikku (EKK 2020: 400).

Hii-ruut-test näitas, et seos on statistiliselt oluline: $X^2(1, N = 2264) = 553,22$, p-väärtus $< 0,001$. Craméri V näitas, et seos uuritava ja seletava tunnuse vahel on sellise vabadusastmete arvu (1) korral tugev: 0,5.

5.3. Verbi transitiivsus

Andmestikus oli 2264 lausest 1689 lauses 75% kõikidest lausetest verb intransitiivne (nt *kolm meest jooksid*). Ülejäänud 575 lauses ehk 25% kõikidest lausetest oli verb transitiivne (*kolm meest vaatasid Kuud*).

Tabelis 5 on esitatud ainsuse ja mitmuse jagunemine intransitiivsete ja transitiivsete verbidega.

Tabel 5. Ainsuse ja mitmuse jagunemine intransitiivsete ja transitiivsete verbidega

	Intransitiivne	Transitiivne
Ainsus	1142 / 68%	203 / 35%
Mitmus	547 / 32%	372 / 65%
Kokku	1689 / 100%	575 / 100%

Tabelist 5 nähtub, et lausetest, kus verb esines intransitiivsena, on 1142 lause ehk 68% puhul kasutatud ainsuslikku verbivormi (nt *seal jooksis kolm meest*), 547 lause ehk 32% puhul mitmuslikku verbivormi (*seal jooksid kolm meest*). Lausetest, kus verb oli transitiivsena, on suhe vastupidine: 203 lause ehk 35% puhul kasutatud ainsuslikku verbivormi (*kolm meest vaatas Kuud*), 372 lause ehk 65% puhul mitmuslikku verbivormi (*kolm meest vaatasid Kuud*).

Hii-ruut-test näitas, et seos on statistiliselt oluline: $X^2(1, N = 2264) = 184,36$, p-väärtus $< 0,001$. Craméri V näitas, et seos uuritava ja seletava tunnuse vahel on sellise vabadusastmete arvu (1) korral nõrga ja keskmise piirimail: 0,286.

Selleks, et mõista, kas ainsusliku ja mitmusliku verbivormi valik sõltub ka objekti tüübist, märgendasin andmestikus, kas objekt lauses on noomenifraas, *da-* või *vat-*infinitiiv või kõrvallause. Tabelis 6 on esitatud, kuidas andmestik selle suhtes jaotus.

Tabel 6. Objekti tüüpide jagunemine transitiivsetes lausetes

	Noomenifraas	Infinitiiv	Kõrvallause
Esinemine	448 / 78%	78 / 14%	49 / 8%

Tabelist 6 selgub, et 448 lauses ehk 78% kõikidest transitiivsetest lausetest oli lauses objekti tüübi ehk sihitisena noomenifraas (*kolm meest vaatasid Kuud*), 78 lauses ehk 14%

lausetest oli sihitis *da*-infinitiiv või *vat*-infinitiiv (*kolm meest tahtsid joosta*) ning 49 lauses ehk 8% lausetest oli sihitis kõrvallause (*kolm meest tahtis, et Kuu hakkaks jooksuma*).

Tabelis 7 on esitatud ainsuse ja mitmuse jagunemine objekti tüübi põhjal.

Tabel 7. Ainsuse ja mitmuse jagunemine objekti tüübi põhjal

	Noomenifraas	Infinitiiv	Kõrvallause
Ainsus	145 / 32%	38 / 49%	20 / 41%
Mitmus	303 / 68%	40 / 51%	29 / 59%
Kokku	448 / 100%	78 / 100%	49 / 100%

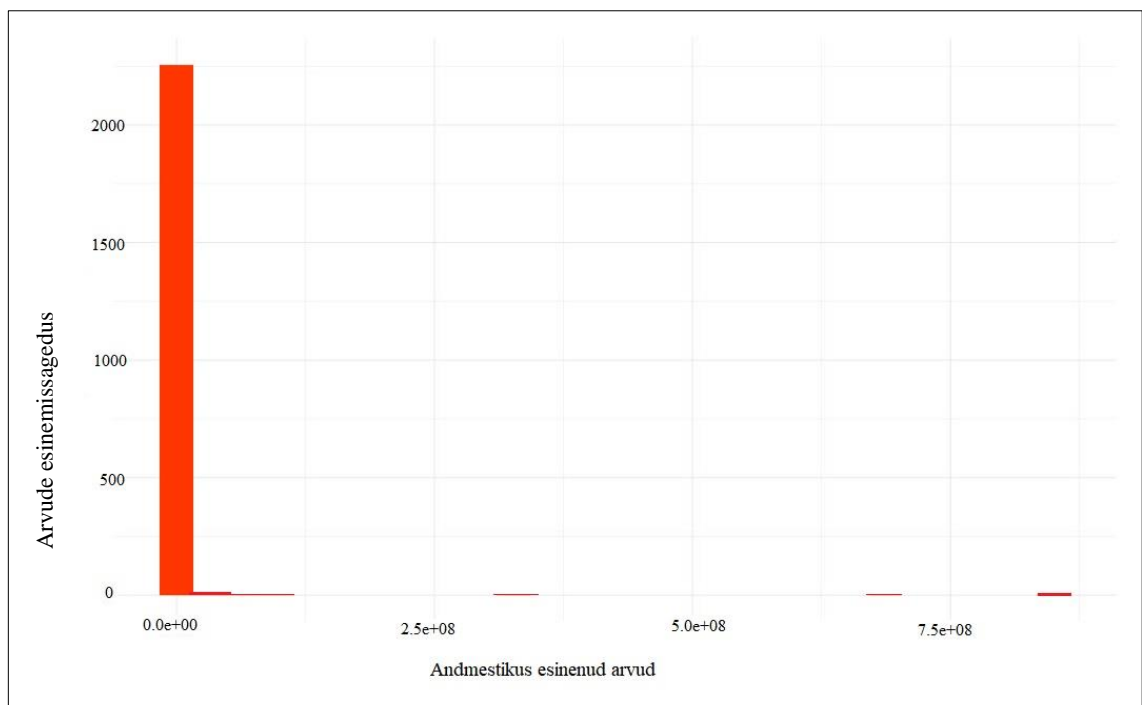
Tabelist 7 nähtub, et lausetest, kus verb oli transitiivisena ja objekti tüüp ehk sihitis oli noomenifraas, oli 145 lauses ehk 32% kõikidest lausetest kasutatud ainsuslikku verbivormi (*kolm meest vaatas Kuud*), 303 lauses ehk 68% mitmuslikku (*kolm meest vaatasid Kuud*). Lausetest, kus transitiivisena esinenud verbi sihitis oli *da*-infinitiiv või *vat*-infinitiiv, oli 38 lauses ehk 49% lausetest kasutatud ainsuslikku verbivormi (*kolm meest tahtis joosta*), 40 lauses ehk 51% lausetest mitmuslikku (*kolm meest tahtsid joosta*). Lausetest, kus transitiivisena esinenud verbi sihitis oli kõrvallause, oli 20 lauses ehk 41% lausetest kasutatud ainsuslikku (*kolm meest tahtis, et Kuu hakkaks jooksuma*), 29 lauses ehk 59% lausetest mitmuslikku verbivormi (*kolm meest tahtsid, et Kuu hakkaks jooksuma*).

Selgub, et kui verb on intransitiivne, kasutatakse verbivormi tõenäolisemalt ainsuses. Kui verb on transitiivne, valitakse pigem mitmuslik verbivorm. Seda enim noomenifraasilise, seejärel kõrvallauselise, väga väikese ülekaaluga ka *da*- või *vat*-infinitiivi sihitise puhul. Hii-ruut-test näitas, et seos on statistiliselt oluline: $X^2(2, N = 575) = 8,4896$, p -väärtus = 0,01434. Craméri V näitas, et seos uuritava ja seletava tunnuse vahel on sellise vabadusastmete arvu (2) korral nõrk: 0,122.

5.4. Kvantorifraasi arvsõna

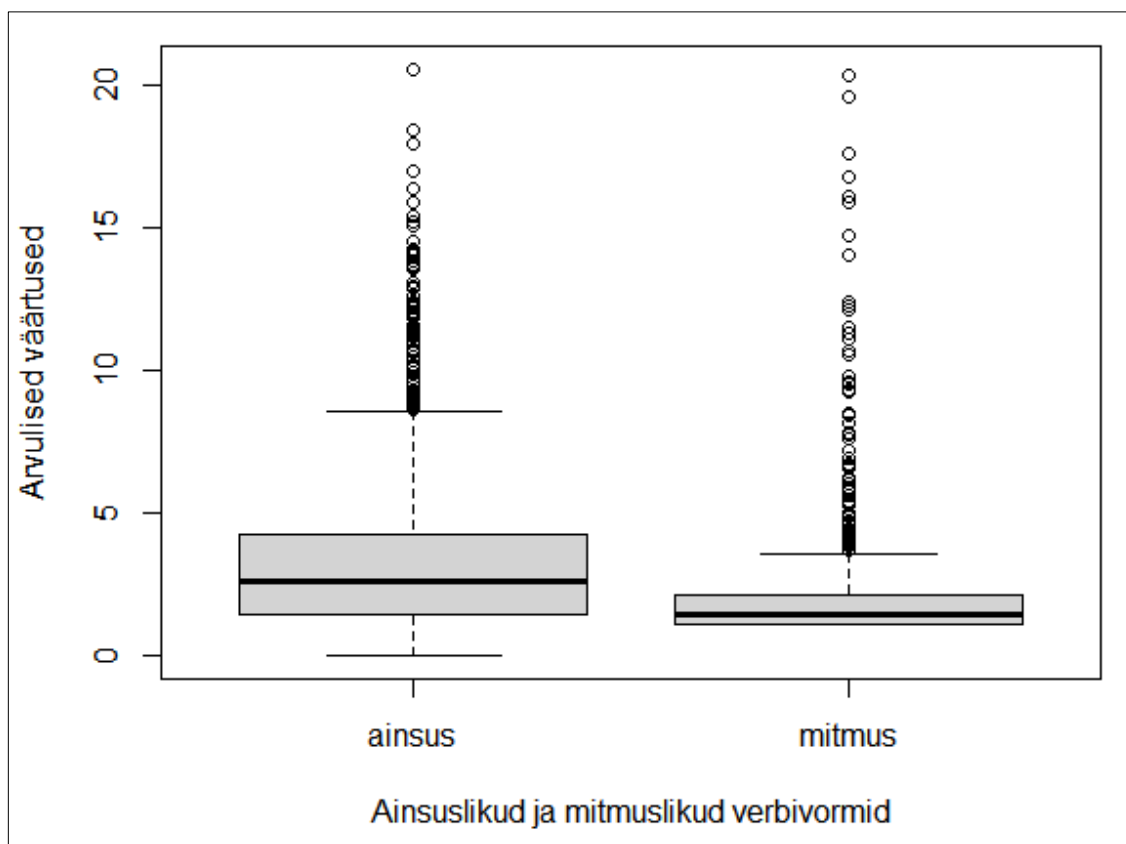
Kvantorifraasi arvsõna ehk kvantorifraasis oleva arvu ühetunnuseliseks analüüsiks esitan esmalt histogrammi, millelt on näha, kuidas arvude sagedus andmestikus jagunes. Seejärel esitan karpdiagrammid, mis näitavad väärtuste jagunemist, sh mediaanväärtusi ainsuslike ja mitmuslike verbivormide puhul. Viimaks kirjeldan, kas arvu tunnus on verbivormi ainsuse või mitmuse valikut mõjutav tunnus. Selleks kasutan mitteparameetrilist Wilcoxon'i astaksummatesti testi (Wilcoxon 1945) ning Cliffi delta seosekordajat (Cliff 1993).

Joonisel 1 kujutatud histogramm annab ülevaate selle kohta, milline oli andmestikus esinenud arvude esinemissagedus.



Joonis 1. Arvude esinemissagedus andmestikus

Joonisel 1 on punase värviga märgitult näha, et andmestikus esines palju väikseid väärtusi, suuri väärtusi oli seevastu vähe. Selleks, et rõhutada erinevusi väiksemate väärtuste vahel, logaritmisin naturaallogaritmiga kasutades fraasi peasõnana esinevaid arve. Selleks, et logaritmid ka arvu 0 (mille logaritm oleks miinus lõpmatus), liitsin esmalt kõikidele arvudele 1. Logaritmitud kvantorifraasi arvude esinemine ainsuslike ja mitmuslike verbivormidega on kujutatud karpdiagrammidena joonisel 2.



Joonis 2. **Kvantorifraasi arvude esinemine ainsuslike ja mitmuslike verbivormidega**

Joonisel 2 on kujutatud kaks karpdiagrammi, millelt on nähtav, kuidas paikneb arvude mediaan ainsuslike (joonisel vasakpoolne diagramm) ja mitmuslike verbivormide (joonisel parempoolne diagramm) puhul. Joonise vasakus servas olevad arvud on minu andmestikus olnud arvude logaritmid.

Jooniselt ilmneb, et suurem osa (50%) logaritmitud väärtusi jääb ainsuslike verbivormide puhul 1,386 ja 4,263, mitmuslike verbivormide puhul 1,099 ja 2,079 vahele, mis logaritmitamata vastavad arvudele 3, 70, 2 ja 7. Ainsuslike verbivormidega lausetes kasutatud kvantorifraasi arvude mediaanväärtus (2,565 ehk logaritmitamata 12) on suurem kui mitmuslike verbivormidega (1,386 ehk logaritmitamata 3). See tähendab, et ainsuslikke verbivorme kasutati rohkem siis, kui kvantorifraasis olev arv oli suurem (nt *kolmteist meest tuli*), mitmuslikke verbivorme eelistati, kui arv oli väiksem (nt *kolm meest tulid*).

Selleks, et teada kvantorifraasi arvsõna kui seletava tunnuse mõju ainsusliku või mitmusliku verbivormi valikul, kasutan ühetunnuselises analüüsis mitteparameetrilist Wilcoxonit astaksummatesti ehk Manni-Whitney U-testi ning Cliffi delta seosekordajat. Wilcoxonit testiga võrdlesin kvantorifraasis olevate arvude mediaane ainsuslike ja mitmuslike verbivormidega. Testi põhjal on kvantorifraaside põhiarvsõnade mediaankeskmine ainsuslike ja mitmuslike verbivormidega statistiliselt oluliselt erinev: $W = 859116$, $p < 0,001$. Seos kvantorifraasis oleva arvu ning verbivormi ainsuse või mitmuse valiku vahel on mõõdukas: ainsuslike verbivormidega koos esinevates kvantorifraasides on keskmiselt enam suuri arve tähistavaid arvsõnu kui mitmuslike verbivormidega fraasides (Cliffi delta = 0,39).

5.5. Ebamäärane hulk

Lauseid, milles oli kasutatud kvantorifraasiga tähistatud hulga ebamäärastamise vahendeid (ligikaudsust näitav sõna, arvuvahemik), oli andmestikus 100 ehk 4% (nt *3–4 inimest*, *umbes viis teemat*). Lauseid, milles hulga ebamäärastamise vahendeid kasutatud ei olnud, oli 2164 ehk 96%.

Tabelis 8 on esitatud ainsuse ja mitmuse jagunemine ebamääraste ja mitte-ebamääraste kvantorifraasidega.

Tabel 8. Ainsuse ja mitmuse jagunemine ebamääraste ja mitte-ebamääraste kvantorifraasidega

	Ebamäärane	Mitte-ebamäärane
Ainsus	67 / 67%	1278 / 59%
Mitmus	33 / 33%	886 / 41%
Kokku	100 / 100%	2164 / 100%

Tabelist 8 nähtub, et nii ebamääraste kui ka mitte-ebamääraste kvantorifraaside puhul eelistati kasutada ainsuslikku verbivormi. Lausetest, kus esines hulga ebamäärasust, on 67 lauset ehk 67% ainsusliku verbivormiga (nt *umbes kümme meest tuli*), 33 lauset ehk 33% mitmuslikuga (*umbes kümme meest tulevad*). Lausetest, kus hulga ebamäärasust ei esinenud, on 1278 ehk 59% ainsusliku verbivormiga (*kümme meest tuli*) ja 886 ehk 41% mitmuslikuga (*kümme meest tulevad*).

Hii-ruut-test näitas, et seos ei ole statistiliselt oluline: $X^2(1, N = 2264) = 2,182$, p-väärtus = 0,1396.

5.6. Subjekti referendi elusus

Lauseid, milles kvantorifraasis olev nimisõna viitab inimesele (nt *inimene, tüdruk*), on 1140 ehk 50% kõikidest lausetest. Elusa ja konkreetse referendiga (nt *koer*) lauseid on 150 ehk 7%; eluta ja konkreetse referendiga (nt *puu*) lauseid on 601 ehk 27%; eluta ja abstraktse referendiga (nt *idee*) lauseid on 206 ehk 9%; ajaga seotud referendiga (nt *nädal*) lauseid on 126 ehk 5%; protsenti väljendavaid (nt *15 protsenti*) lauseid on 41 ehk 2%. Sageduse järgi on enim lauseid, kus subjekti referent on inimene, seejärel lauseid, kus subjekti referent on eluta ja konkreetne, eluta ja abstraktne, elus ja konkreetne, aeg ja viimaks protsent.

Tabelis 9 on esitatud ainsuse ja mitmuse jagunemine vastavalt subjekti referendi elususele.

Tabel 9. Ainsuse ja mitmuse jagunemine vastavalt subjekti referendi elususele

	Inimene	Elus, konkreetne	Eluta, konkreetne	Eluta, abstraktne	Aeg	Protsent
Ainsus	661 / 58%	77 / 51%	356 / 59%	123 / 60%	95 / 75%	33 / 80%
Mitmus	479 / 42%	73 / 49%	245 / 41%	83 / 40%	21 / 25%	8 / 20%
Kokku	1140 / 100%	150 / 100%	601 / 100%	206 / 100%	126 / 100%	41 / 100%

Tabelist 9 nähtub, et 1140 lausest, kus subjekt oli inimene, oli 661 korda ehk 58% verbivorm ainsuslik (nt *kolm inimest tuli*), 479 korda ehk 42 % mitmuslik (*tulevad kolm inimest*). Kui subjekt oli: elus ja konkreetne (150 lauset), oli verb 77 ehk 51% ainsuslik (*kolm koera tuleb*), 73 korda ehk 49% mitmuslik (*kolm koera tulevad*). Kui subjekt oli eluta ja konkreetne (601 lauset), oli verb 356 korda ehk 59% ainsuslik (*kolm puud seisab*), 245 korda ehk 41% mitmuslik (*seisavad kolm puud*). Kui subjekt oli eluta ja abstraktne (206 lauset), oli verb 123 korda ehk 60% ainsuslik (*tärkab kolm ideed*), 83 korda ehk 40% mitmuslik (*tärkavad kolm ideed*). Kui subjekt oli aeg (126 lauset), oli verb 95 korda ehk 75% ainsuslik (*möödus kolm nädalat*), 21 korda ehk 25% mitmuslik (*kolm nädalat mööduvad*). Kui subjekt oli protsent (41 lauset), oli verb 33 korda ehk 80% ainsuslik (*vastas 3 protsenti*), 8 korda ehk 20% mitmuslik (*3 protsenti vastasid*). Kõikide subjekti referendi elususe kategooriatega on eelistatud niisiis ainsuslikku verbivormi, ehkki selle eelistuse ülekaal varieerub ning näiteks elusa ja konkreetse referendi puhul on nii ainsusliku kui ka mitmusliku verbivormiga lauseid peaaegu võrdselt.

Hii-ruut-test näitas, et seos on statistiliselt oluline: $X^2(5, N = 2264) = 25,944$, p-väärtus $< 0,001$. Craméri V näitas, et seos uuritava ja seletava tunnuse vahel on sellise vabadusastmete arvu (5) korral nõrk: 0,107.

5.7. Lause tekstitüüp

Andmestiku lausetest pärinesid ajakirjandustekstidest 1761 lauset ehk 78% lausetest, foorumitest 483 lauset ehk 21% ja teadustekstidest ainult 20 lauset ehk 1%. Tabelis 10 on esitatud ainsuslike ja mitmuslike verbivormide jagunemine tekstitüübiti.

Tabel 10. Ainsuse ja mitmuse jagunemine vastavalt tekstitüübile

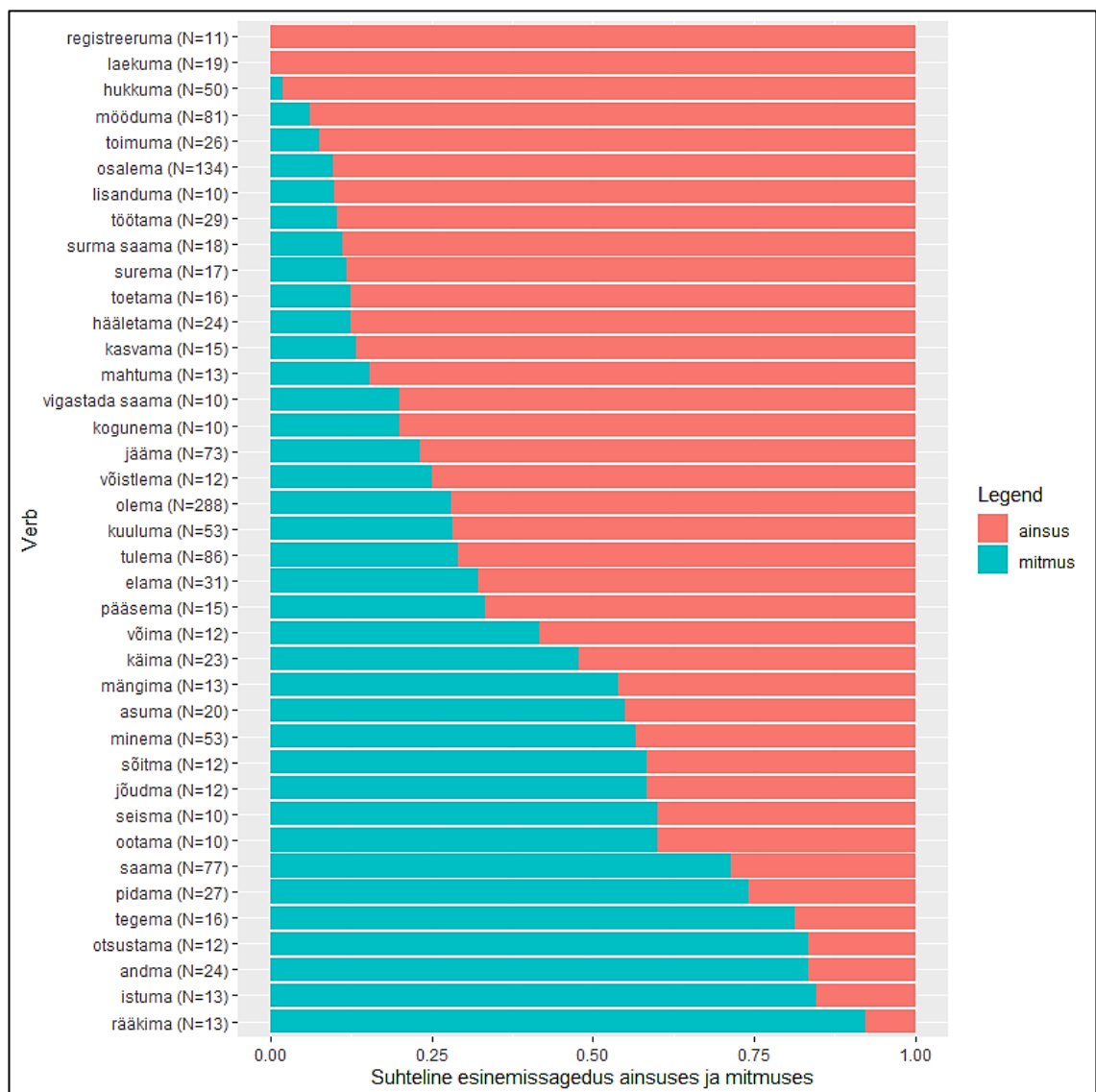
	Teadustekstid	Foorumid	Ajakirjandus
Ainsus	10 / 50%	254 / 53%	1081 / 61%
Mitmus	10 / 50%	229 / 47%	680 / 39%
Kokku	20 / 100%	483 / 100%	1761 / 100%

Tabelist 10 on näha, et ainsus ja mitmus jagunevad tekstitüübi suhtes järgnevalt: teadustekstides oli 10 lauses ehk 50% selles tekstitüübis olevatest lausetest verbivorm ainsuses (nt *osales 39 eelkooliealist last*), 10 lauses ehk 50% verbivorm mitmuses (*kaheksa keelejuhti hääldasid*). Foorumitekstidest 254 lauses ehk 53% selle tekstitüübi lausetest oli verbivorm ainsuses (*hindas 6 liiget*) ja 229 lauses ehk 47% mitmuses (*3 venda kasutavad*). Ajakirjandustekstides oli 1081 lauses ehk 61% selle tekstitüübi lausetest verbivorm ainsuses (*viis kohtunikku leidis*) ja 680 lauses ehk 39% mitmuses (*kaks naist otsustasid*). Seega on minu andmestikus olevate teadustekstide lausetes kasutatud võrdselt nii ainsust kui ka mitmust, kuid foorumi- ja eriti ajakirjandustekstide puhul on eelistatud ainsuslikku verbivormi.

Hii-ruut-test näitas, et seos on statistiliselt oluline: $X^2(2, N = 2264) = 12,906$, p-väärtus = 0,001576. Craméri V näitas siiski, et seos uuritava ja seletava tunnuse vahel on sellise vabadusastmete arvu (2) korral nõrk: 0,076, kuid tulemust võib mõjutada ka tekstitüüpide väga ebavõrdne esindatus andmestikus.

5.8. Verbid

Vaatlesin ka ainsuse ja mitmuse vormide osakaalu nende verbide puhul, mis esinesid andmestiku lausetes vähemalt 10 korda. Joonisel 3 on kujutatud nende verbide esinemine ainsuse ja mitmuse vormides.



Joonis 3. Vähemalt 10 korral esinenud verbide suhteline esinemissagedus ainsuses ja mitmuses

Joonise 3 vasakul serval on ainsuses esinemise osakaalu vähenemise järjekorras esitatud verbide järjestus. Verbi järel on sulgudesse märgitud, mitu korda verb minu andmestikus esines (nt N=11 näitab, et verb esines 11 lauses). Joonise alumises servas on kujutatud suhteline esinemissagedus. Ainsuse osakaalu on kujutatud roosa värviga, mitmuse osakaalu sinisega.

Selgub, et andmestikus vähemalt 10 korral esinenud verbidest oli kaks verbi, mis esinesid vaid ainsusliku verbivormiga: *registreeruma*, mis esines kokku 11 korral, ning *laekuma*, mis esines 19 korral. Ükski vähemalt kümnel korral esinenud verbist ei kasutanud ainult mitmuslikku vormi. Kõige mitmuselembesem oli verb *rääkima*, mis esines 13 lauses, neist 12 lauses mitmuses ja ühes lauses ainsuses.

Tüüpiliselt lauses intransitiivsena esinevate verbidega (nt *hukkuma*) eelistatakse kasutada pigem ainsuslikku verbivormi, tüüpiliselt transitiivsete verbidega (nt *otsustama*) seevastu mitmuslikku. Nagu transitiivsuse ühetunnuselisest analüüsistki, saab ka verbilekseemide analüüsi põhjal järeldada, et transitiivsus on mitmuse kasutamise poole suunav tegur.

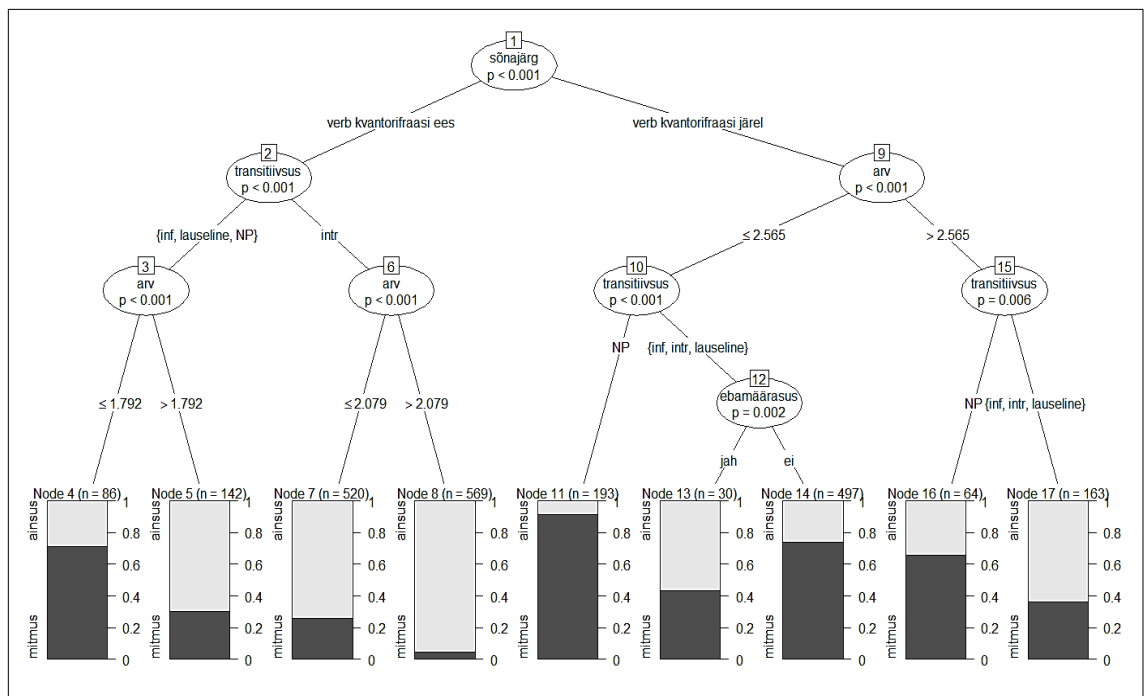
6. Tunnuste koosmõju analüüs

Selles peatükis analüüsin kõikide seletavate tunnuste koosmõju uuritavale tunnusele ning leian, millised seletavad tunnused uuritava tunnuse varieerumisega kõige tugevamalt seotud on, kui arvestada samal ajal ka teiste seletavate tunnuste mõju. Olen analüüsis kasutanud otsustuspuud ja juhumetsa.

6.1. Otsustuspuu

Selleks, et teada saada, millise seletava tunnuse mõju uuritavale tunnusele on kõige tugevam, koostasin R-is otsustuspuu (Breiman 1984). Otsustuspuu visualiseerib seletavate tunnuste seoseid ja koosmõjusid ning aitab visualiseerida, kuidas andmestiku uuritava tunnuse väärtused jagunevad, sõltuvalt mitme seletava tunnuse koosmõjust. Puus valitakse välja see seletav tunnus, mis on verbivormi ainsuse või mitmuse valikul kõige suurema mõjuga, ning jagatakse andmed selle tunnuse järgi kaheks, mistõttu tekib kaks puuharu. Tekkinud harude sõlmedes valitakse järgmine tunnus, mille väärtuste põhjal andmestik uuesti kaheks jagada. Seda mustrit korratakse nii kaua, kuni ükski seletav tunnus ei määra uuritava tunnuse vormi. Lehtedes ehk otsustuspuu joonise allservas asuvates tulpades on kujutatud, millist uuritava tunnuse väärtust ja millise tõenäosusega puu seletavate tunnuste mõju põhjal ennustab. (Strobl jt 2009: 325–327) Selles töös kasutan tingimuslikku otsustuspuud, mis kasutab hargnemisi tekitavate tunnuste valikuks p-väärtusi (Hothorn jt 2006).

Minu otsustuspuu klassifitseerimistäpsus on 0,7928445. See tähendab, et otsustuspuu klassifitseerib umbes 79% andmestikus olevatest vaatlustest õigesti ja puud võib pidada üsna tugevaks mudeliks. Joonisel 4 on otsustuspuu, millel on kujutatud ainsusliku või mitmusliku verbivormi valikut kõige paremini seletavad tunnused.



Joonis 4. Verbivormi ainsuse või mitmuse valikut mõjutavate tunnuste otsustuspuu

Joonise 4 ehk otsustuspuu põhjal on verbivormi ainsuse või mitmuse valikul lauses, kus aluseks on kvantorifraas, kõige olulisem seletav tunnus verbi paiknemine kvantorifraasi suhtes ehk sõnajärg (joonisel sõlm 1). Sõltuvalt sellest, kas lauses on verb kvantorifraas ees või järel, erineb see, kas verbivorm on pigem ainsuses või mitmuses: kui verb on kvantorifraasi ees (VS-sõnajärg), kasutatakse pigem ainsuslikku verbivormi; kui verb on kvantorifraasi järel (SV-sõnajärg), kasutatakse pigem mitmuslikku verbivormi.

Esmalt kirjeldan otsustuspuu sõlmest 1 hargnevat vasakut haru ehk seda, mis mõjutab verbivormi valikut juhul, kui verb on kvantorifraasi ees (nt *tuli kolm meest*).

Kui verb asub kvantorifraasi ees, mõjutab verbivormi arvukategooria valikut see, kas lauses olev verb on lause kontekstis transitiivne või mitte (joonisel sõlm (Node) 2). Kui verb on transitiivne (sõlmest 2 hargnev vasakpoolne haru) ehk lauses on sihitiseks noomenifraas, *da-* või *vat-*infinitiiv või kõrvallause, sõltub verbivormi valik omakorda kvantorifraasi arvsõnast (joonisel sõlm 3). Kui kvantorifraasi arvsõna logaritmi on 1,792 või väiksem (mis töös kasutatud teisenduse puhul vastab arvule 5), valib keelekasutaja

tõenäolisemalt mitmusliku verbivormi (joonisel sõlm 4) (*Kuud vaatasid viis meest*). Kui kvantorifraasi arvsõna on suurem kui 5, valib keelekasutaja tõenäolisemalt ainsusliku verbivormi (*Kuud vaatas viissada meest*).

Kui verb on intransitiivne (sõlmest 2 hargnev parempoolne haru), mõjutab verbivormi valikut taas arvsõna. Kui kvantorifraasi arvsõna logaritm on 2,079 või väiksem (mis töös kasutatud teisenduse puhul vastab arvule 7), on mitmuse kasutamise tõenäosus veidi suurem (joonisel sõlm 7) kui siis, kui arv on suurem kui 7 (joonisel sõlm 8). Siiski eelistatakse sõnajärje puhul, milles verb asub kvantorifraasi ees, intransitiivsete verbidega pigem kasutada ainsuslikku verbivormi (*jookseb seitse meest*).

Otsustuspuu sõlmest 1 hargnev parempoolne haru näitlikustab, millised tegurid mõjutavad verbivormi valikut siis, kui verb asub kvantorifraasi järel (nt *kolm meest tulevad*).

Kui verb asub kvantorifraasi järel, mõjutab verbivormi arvukategooria valikut kvantorifraasi arvsõna (joonisel sõlm 9). Kui kvantorifraasi arvsõna logaritm on 2,565 või väiksem (mis töös kasutatud teisenduse puhul vastab arvule 12), mõjutab verbivormi valikut verbi transitiivsus (joonisel sõlm 10). Kui verb on lauses transitiivne ja sihitis on noomenifraas, valib keelekasutaja tõenäolisemalt mitmusliku verbivormi (joonisel sõlm 11) (*10 meest vaatavad Kuud*). Kui verb on transitiivne ja sihitiseks on *da-* või *vat-* infinitiiv või kõrvallause või on verb intransitiivne, mõjutab verbivormi valikut see, kas kvantorifraasi on ebamäärastatud (joonisel sõlm 12). Kui ebamäärastavaid vahendeid on kasutatud, valib keelekasutaja pigem ainsusliku verbivormi (joonisel sõlm 13) (*umbes 10 meest tahab joosta; umbes 10 meest tahab, et Kuu hakkaks jooksuma; umbes 10 meest tuleb*). Kui ebamäärastavaid vahendeid kasutatud ei ole, valitakse pigem mitmuslik verbivorm (joonisel sõlm 14) (*10 meest tulevad*).

Kui kvantorifraasis olev arv on suurem kui 12 (sõlmest 9 hargnev parempoolne haru), mõjutab verbivormi valikut transitiivsus (joonisel sõlm 15). Kui verb on transitiivne ja sihitiseks on noomenifraas, valib keelekasutaja pigem mitmusliku verbivormi

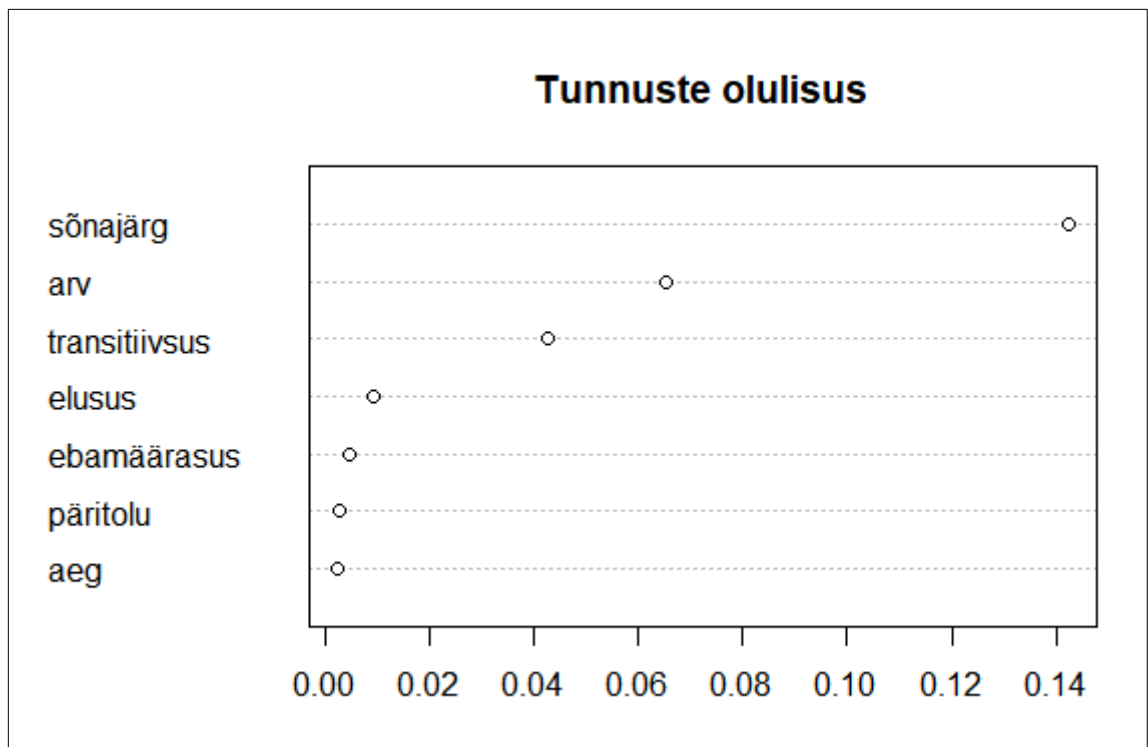
(joonisel sõlm 16) (*kakssada meest vaatavad Kuud*). Kui verb on transitiivne ja sihitiseks on *da-* või *vat-*infinitiiv või kõrvallause või on verb intransitiivne, valib keeleteadaja pigem ainsusliku verbivormi (joonisel sõlm 17) (*kakssada meest tahab joosta; kakssada meest tahab, et Kuu hakkaks jooksuma; kakssada meest tuleb*).

Otsustuspuusse ei ole sattunud selliseid seletavaid tunnuseid nagu verbi ajavorm, lause tekstitüüp ja kvantorifraasi subjekti referendi elusus. Seega saab otsustuspuu põhjal väita, et need tunnused ei ole verbivormi ainsuse või mitmuse valikul teisi seletavaid tunnuseid arvesse võttes sama olulised.

6.2. Juhumets

Otsustuspuu ennustuste täiendamiseks ja usaldusväärsemaks muutmiseks kasutan juhumetsa (Breiman 2001). Juhumetsas kasvatatakse mitu otsustuspuud ja iga puu jaoks kasutatakse juhuslikku valimit olemasolevast andmestikust, nii et osa tegelikke vaatlusi jääb puu kasvatamisest kõrvale. Iga juhumetsa puu harude tekitamiseks kasutatakse juhuslikku hulka kõigist seletavatest tunnustest, et ka nõrgema mõjuga tunnused saaksid kajastatud (Cutler jt 2007). Kuna juhumetsa ei saa kuvada joonisena, määratakse klassifitseerimistäpsuse ja C-indeksi, mida kasutatakse ka teiste klassifitseerimisülesannete puhul (Hosmer jt 2013) ja mis võimaldavad määrata, kui hästi juhumets seletavate tunnuste alusel suudab uuritava tunnuse klasse (ainsuslikke ja mitmuslikke verbivorme) eristada. Juhumetsa loomiseks kasvasin R-is 500 puud ja lasin iga haru tekkimiseks vaadata kolme juhuslikult valitud tunnust.

Minu metsamudeli klassifitseerimistäpsus on 0,79 ehk 79%. Metsamudeli C-indeks on 0,86. See tähendab, et mudel eristab ainsuslikke ja mitmuslikke verbivorme seletavate tunnuste kombinatsioonide põhjal küllaltki hästi. Selleks, et näha, kuidas seletavate tunnuste suhteline olulisus juhumetsas jaotub, kuvasin R-is tunnuste olulisuse graafiku, mis on nähtav joonisel 5.



Joonis 5. Tunnuste olulisuse graafik

Jooniselt 5 selgub, et kõige olulisem verbivormi ainsuse või mitmuse valikut selgitav tunnus on verbi paiknemine ehk sõnajärg. Sellele järgnevad kvantorifraasis olev arv, verbi transitiivsus, kvantorifraasi subjekti referendi elusus, kvantorifraasiga väljendatava hulga ebamäärasus, lause tekstitüüp ja verbi ajavorm. Võrreldes otsustuspuud ja juhumetsa tunnuste olulisust, on viimase järgi kvantorifraasiga väljendatava hulga ebamäärasusest olulisem kvantorifraasi subjekti referendi elusus, mida otsustuspuus ei kajastunud. Samas võib nii sõnajärge kui sihitise olemasolu pidada elususega vähemalt osaliselt korreleerivateks tunnusteks ning kuna otsustuspuus tulid tunnustena tugevalt esile nii sõnajärg kui ka sihitise olemasolu, on subjekti referendi elususe tunnuse puudumine otsustuspuust mõneti ootuspärane.

7. Järeldused

Ilmnes, et 2021. aasta eesti keele ühendkorpusest juhuvalimi võtmise teel kogutud 2264 lause suuruses andmestikus olevate lausete hulgas oli rohkem ainsuslike kui mitmuslike verbivormidega lauseid. Andmestiku põhjal võiks öelda, et suurema tõenäosusega valib keelekasutaja ainsusliku verbivormi, kuid seejuures on keelesisesid muutujaid, mis siiski verbi vormi valikut mõjutavad.

Kõige olulisem mõjutaja verbivormi arvukategooria valikul on otsustuspuu ja juhumetsa põhjal verbi paiknemine kvantorifraasi suhtes. Ka Craméri V tulemus näitas, et seletava tunnuse seos uuritava tunnusega on tugev. Andmestikus oli rohkem neid lauseid, kus verb asus kvantorifraasi ees (nt *tuli kolm meest*), vähem neid lauseid, kus verb oli paigutatud kvantorifraasi järele (nt *kolm meest tulid*). Kui verb on kvantorifraasi ees, on eelistatud pigem ainsuslikku verbivormi, ning juhtudel, kus verb on kvantorifraasi järel, on valitud pigem mitmuslik verbivorm. Protsentuaalsed erinevused kinnitavad, et keelekasutaja lähtub kvantorifraasiga ühilduva verbi ainsusliku või mitmusliku vormi valikul suuresti just sõnajärjest. See teguviis järgib ka Ereleti õigekirjasoovitust (Erelt 2019: 17): kasutada mitmust, kui kvantorifraas asub lause alguses ja verb selle järel, ning ainsust, kui kvantorifraas asub lause lõpus ja verb selle ees. Edaspidi tasub põhjalikumalt uurida, kas selliste õigekirjasoovituste järgimisel avalduks tekstitüüpide suhtes tasakaalustatud valimis enam ka tekstide toimetatuse mõju.

Kvantorifraasis arvsõna analüüsist ilmnes, et lausetes olevad arvud olid pigem väiksemad ja seda kinnitas ka karpdiagrammi jooniselt nähtu. Samas selgus, et ainsuslikke verbivorme eelistati sagedamini siis, kui kvantorifraasis oli suurem arv. Mitmuslikke vorme esines rohkem siis, kui arv oli väiksem. Wilcoxon testi tulemuse põhjal on kvantorifraasis oleva arvu kui tunnuse mõju verbivormi valikule mõõdukas. Otsustuspuu põhjal kujunesid olulisteks arvudeks ehk mingi lausetüübi puhul verbi valikut mõjutavateks piirarvudeks arvud 5, 7 ja 12. Selline tendents näitab, et suuremate ehk

ebamäärasemate ja hulgaliselt vähem hoomatavate arvudega eelistab keeleteadlane ainsuslikku verbivormi, mitte-ebamääraste ja hoomatavamate arvudega valitakse pigem mitmuslik verbivorm.

Verbi transitiivsuse määramisel lähtusin objekti ehk sihitise olemasolust lauses. Selle põhjal oli andmestikus enim intransitiivseid lauseid, sageduselt järgmised olid laused, mille sihtis oli noomenifraas, seejärel *da-* või *vat-*infinitiiv ning viimaks kõrvallause. Intransitiivsete lausetega oli rohkem kasutatud ainsuslikku verbivormi, sihitisega lausetes domineeris mitmuslik verb. Hii-ruut-test näitas, et sihitise olemasolu on statistiliselt oluline tunnus, mis võib mõjutada verbivormi valikut. Craméri V näitas, et seos uuritava ja seletava tunnuse vahel on nõrk. Seda, et transitiivsus on verbivormi valikul oluline mõjutaja, kinnitasid otsustuspuu ja juhumets. Tõenäoliselt on transitiivsuse mõju olulisus seotud argumentide – subjekti ja objekti – eristamise vajadusega, mis tingib sagedasema mitmusliku verbivormi kasutuse transitiivsetes lausetes. Otsustuspuu põhjal erineb transitiivsuse mõju sõltuvalt verbi paiknemisest kvantorifraasi suhtes ehk lause sõnajärjest. Kui verb on kvantorifraasi ees, eristuvad selgelt transitiivselt ja intransitiivselt kasutatud verbid: esimesel juhul on mitmuse kasutamine oluliselt tõenäolisem kui teisel. Kui verb on kvantorifraasi järel, eristuvad noomenifraasid teistest transitiivsetest objektidest ja intransitiivsetest verbide kasutusjuhtudest: noomenifraasidega on mitmuse kasutamine oluliselt tõenäolisem kui ülejäänud juhtudel. Noomenifraasiliste objektidega toimub ühildumine lähtuvalt kvantorifraasi sisust, mitte vormist ilmselt seetõttu, et lauses objekti ja subjekti paremini eristada.

Kvantorifraasi subjekti referendi elususe kategooriate sageduse järgi oli andmestikus enim lauseid, kus subjekti referent on inimene. Sageduselt järgnesid laused, kus subjekti referent on eluta ja konkreetne, eluta ja abstraktne, elus ja konkreetne, aeg ja viimaks protsent, ent kõikide elususe kategooriate puhul oli rohkem kasutatud ainsuslikku kui mitmuslikku verbivormi. Hii-ruut-testi tulemus on statistiliselt oluline, seega erinesid vähemalt kaks seletava tunnuse kategooriat teineteisest selle poolest, kuidas ainsuslikud ja mitmuslikud verbivormid nendega koos esinesid jaotusid. Craméri V tulemus näitas, et seos uuritava tunnusega on nõrk. Ehkki tunnuste olulisuse skaalal paigutus subjekti

referendi elusus olulisuselt neljandale kohale, ei näidanud otsustuspuu ja juhumets siiski, et tunnus oleks kvantorifraasiga ühilduva verbi vormi valikul määrav. Kuna subjekti referendi elususe ja lause sõnajärje vahel on leitud seos (Lindström 2001: 94), võib arvata, et subjekti referendi elusus ja verbi asukoht lauses on osaliselt korreleeruvad tunnused ja seepärast jääb elususe tunnus otsustuspuus kajastamata. Samas jääb Lindstörmi tähelepaneku (kui suulise keele lausungi subjekt on inimene, asub verb tõenäolisemalt subjekti järel) valguses siinkohal vastuseta küsimus, miks on andmestikus, mille lausetest pooltes on subjekti referent inimene, domineeriv sõnajärg, kus verb asub fraasi ees. On võimalik, et see on tingitud asjaolust, et minu andmestikus ei olnud mitte suulise keele lausungid, vaid kirjalikud laused, kuid see vajaks edasist uurimist.

Ebamäärastest hulkadest kui tunnusest lähtudes olid suurem osa andmestiku lausetes kajastatud hulkadest mitte-ebamäärased, vaid 100 lauses oli hulk ebamäärane. Mõlemal juhul, nii ebamäärase kui ka mitte-ebamäärase hulga puhul oli kasutatud rohkem ainuslikku verbivormi. Hii-ruut-testi põhjal ei olnud tulemus statistiliselt oluline. Hulga ebamäärasus paigutub tunnuste olulisuse graafikul viiendale kohale – ent see võib tuleneda ka ebamääraste hulkade vähesest esindatusest andmestikus – ja mõjutab verbi vormi valikut lauses, kus on täidetud järgmised tingimused: verb asub kvantorifraasi järel, kvantorifraasis olev arv on väiksem kui 12 või 12 ja lause on inifiniitvise või kõrvallauselise sihitisega või intransitiivne. Sel juhul on tõenäoline, et kui lauses on kasutatud kvantorifraasi ebamäärastavat vahendit, valib keelekasutaja pigem ainsusliku verbivormi. Kui ebamäärastamist ei ole, valitakse tõenäolisemalt mitmuslik verbivorm. Tulemus on kooskõlas õigekeelsussoovitustega, mille kohaselt tuleks ebamäärase hulga puhul eelistada ainsust, täpse hulga puhul aga mitmust. See, miks hulga ebamäärasus mõjutab verbi vormi valikut vaid lauses, kus verb asub kvantorifraasi järel ja hulga fraasis arvsõnaga viidatud arv on pigem väike (*nt kolm meest tulid*), on ilmselt mõjutatud sellest, et sellises kontekstis muid ebamäärasuse ja sellest tulenevalt ainsuslike verbivormide eelistamisega tugevamalt seotud elemente (nt suured arvud) ei esine ning hulga ebamäärasust väljendav arvuvahemik või umbmäärastav sõna saab ise seda funktsiooni kanda.

Lause tekstitüüpidest – teadustekstidest, foorumipostitustest ja ajakirjandusest – pärineb kõige rohkem andmestiku lauseid ajakirjandusest, sageduselt järgmine tekstitüüp on foorumipostitused ning viimaks teaduskirjandus. Kuna juhusliku valimiga koostatud andmestikus pärines enim lauseid ajakirjandusest, on tõenäoline, et suurem osa andmestikust oli vähemalt osaliselt toimetatud ja sellest tulenevalt keelesoovitusi järgiv. Kui eeldada, et toimetatud tekstides on preskriptiivsetel soovitud suurem tähtsus, võiks neis sagedamini kohata ka esitatud keelesoovituse järgimist – kasutada ainsuslikku verbivormi, kui kvantorifraas asub lause lõpus ja väljendab ebamäärast hulka, ning mitmuslikku, kui kvantorifraas asub lause alguses ja väljendab täpset hulka (Erelt 2019: 17). Tulemustest selgus, et teaduskirjanduses eelistati ainsust ja mitmust samal määral. Foorumipostituses ja ajakirjanduses eelistati rohkem kasutada ainsuslikku verbivormi. Hii-ruut-testi põhjal on tulemus statistiliselt oluline ja saab öelda, et vähemalt kahe kategooria puhul eelistati ainsuslikku verbivormi kasutada rohkem kui juhuse läbi võiks oodata. Samas näitas Craméri V, et seos uuritava ja seletava tunnuse vahel on nõrk. Ka ei tulnud otsustuspuid ja juhumetsa põhjal siiski esile, et lause tekstitüüp oleks määrav verbivormi ainsuse või mitmuse valikul, tunnuste olulisuse graafiku järgi on tekstitüüp tunnuseks eelviimasel kohal ja selle mõju, seega, väike.

Analüüsides verbi ajavorme, selgus, et nii olevikulise kui ka minevikulise verbivormiga eelistati pigem kasutada ainsuslikku verbi. Kuna hii-ruut-testi tulemus oli statistiliselt oluline, siis tähendab see, et mineviku ajavormiga eelistati siiski kasutada rohkem ainsuslikku verbivormi kui juhuse läbi võiks oodata. Nii tunnuste olulisuse graafik kui ka Craméri V tulemus indikeerivad, et seos uuritava tunnusega on nõrk. Otsustuspuid ja juhumets ei näidanud, et ajavorm oleks kvantorifraasiga ühilduva verbi ainsusliku või mitmusliku vormi valikul määrav. Tunnuste olulisuse graafikul paigutus aja tunnus viimasele kohale, mis tähendab, et selle mõju on kõige väiksem.

Uurides andmestikus vähemalt 10 korral esinenud verbide käitumist, tuli samuti välja, et tüüpiliselt intransitiivsete verbidega (nt *hukkuma*) eelistatakse pigem ainsuslikku verbivormi, tüüpiliselt transitiivsete verbidega (nt *otsustama*) aga mitmuslikku. Seega võib järeldada, et mida transitiivsem on verb, seda suurem tõenäosus on, et lauses, milles

on aluseks kvantorifraas, eelistab keelekasutaja valida mitmusliku verbivormi. See on loomulik, sest transitiivsuse olemasolu aitab subjekti ja objekti lauses paremini eristada.

Eelnevast tulenevalt võib öelda, et suurima mõjuga verbivormi arvukategooria valikule on lause sõnajärg ehk verbi asukoht lauses. Samuti on olulise mõjuga verbi transitiivsus ja kvantorifraasi arvsõna, mõningatel juhtudel ka see, kas kvantorifraasiga tähistatav hulk on ebamäärane või mitte. Keelekasutajat mõjutavad ainsuslikku verbivormi kasutama ja kvantorifraasi (nt *kolm meest*) vormist lähtuma verbi kvantorifraasieelne asend, kvantorifraasis olev suurem arv ja verbi intransitiivsus – *tuleb kolmkümmend meest, septembris lendas 51 960 reisijat*. Need on tunnused, mis sobituvad preskriptiivse soovitusena kasutada ainsuslikku verbi, kui kvantorifraasiga väljendatakse ebamäärast hulka või kui kvantorifraas asub lause lõpus. Mitmuslikku verbivormi kasutama mõjutavad tegurid on verbi kvantorifraasijärgne asend, kvantorifraasis olev väike arv ja noomenifraasilise sihitise olemasolu – *kolm meest vaatavad Kuud, kaks noort keeleteadlast tutvustasid oma liiviteemalisi doktoritöid*. Need tunnused, lisaks subjekti referendi elusolemine (mis korreleerub osaliselt sõnajärje ja transitiivsusega), on omased prototüüpsele transitiivlausele. Kuna ühildumise peamine ülesanne on eristada subjekti ja objekti, on tulemus ootuspärane. Verbivormi arvukategooria varieerumise korral on arvuühildumine sagedasem lausetes, mis sarnanevad rohkem prototüüpse transitiivlausega ning milles on subjekti ja objekti eristamine vajalik.

Kokkuvõte

Magistritöös „Arvsõnalise kvantorifraasi ühildumine verbiga“ uurisin, millistest keelesisestest muutujatest ehk tunnustest sõltub verbivormi arvukategooria ehk verbi ainsusliku või mitmusliku vormi valik, kui lauses on aluseks kvantorifraas.

Töö seitsmest peatükist esimeses andsin teoreetilise ülevaate eesti keele kvantorifraasist, teises teooriapeatükis kirjeldasin ühildumist kui nähtust ning kirjutasin eesti keele kvantorifraasi ja verbi ühildumisest. Kolmas peatükk tutvustas magistritöö andmestikku, mis pärines Eesti keele ühendkorpusest (EKÜ 2021), ja uurimisel kasutatud meetodeid. Neljandas peatükis tõin välja, milliste tunnuste kaudu verbivormi arvukategooria varieerumist uurin – nendeks olid verbi ajavorm, verbi paiknemine kvantorifraasi suhtes ehk sõnajärg, verbi transitiivsus, kvantorifraasi arvsõna, kvantorifraasiga väljendatava hulga ebamäärasus, subjekti referendi elusus ja lause tekstitüüp. Viies peatükk sisaldas andmete ühetunnuselise analüüsi, kuuendas peatükis analüüsisin tunnuste koosmõju. Seitsmendas peatükis esitasin tulemuste kokkuvõtte ja olulisemad järeldused.

Magistritöö eesmärk, mis minu hinnangul sai täidetud, oli vaadelda üht eesti keeles varieeruvat nähtust – kvantorifraasiga seotud verbi vormi valikut – ning selgitada välja, millest sõltub verbivormi valik. Töös keskendusin arvsõnalise kvantorifraase sisaldavatele lausetele, mis vastavad järgmistele konstruktsioonidele:

- kolm (pikka) meest tuleb/tuli, kolm (pikka) meest tulevad/tulid;
- tuleb/tuli kolm (pikka) meest, tulevad/tulid kolm (pikka) meest.

Minu uurimisküsimused olid: millised keelesisesed muutujad aitavad seletada verbi vormi valikut ja millise keelesise muutuja mõju verbi vormi valikule on kõige suurem. Need said vastused: verbivormi valikut aitavad kõige paremini seletada sõnajärg – millel on verbivormi valikule ka kõige suurem mõju –, verbi transitiivsus, kvantorifraasi arvsõna ja mõningatel juhtudel ka kvantorifraasiga väljendatava hulga ebamäärasus.

Ainsuslikku verbivormi eelistab keelekasutaja siis, kui verb paikneb kvantorifraasi ees ja on intransitiivne ning kui kvantorifraasiga väljendatav hulk arvult suur või ebamäärane. See on loomulik, sest intransitiivses lauses ei ole subjekti ja objekti eristamine ning seetõttu ka verbivormi arvukategooria ühildumine vajalik. Samuti on arvult suur või ebamäärane hulk inimese jaoks vähem hoomatav kui väike või mitte-ebamäärane hulk, mis tähendab, et keelekasutaja valib sellise hulga puhul pigem ainsusliku verbivormi.

Mitmuslikku verbivormi kasutatakse pigem siis, kui verb paikneb kvantorifraasi järel ja on transitiivne ning kvantorifraasis olev arv on väike. Transitiivsuse olemasolu puhul aitab verbivormi arvukategooria ühildumine lause argumente paremini markeerida, st lause subjekti ja objekti paremini eristada. Samuti on loogiline, et väikesearvulise ja mitte-ebamäärase hulga puhul on inimesel kergem hulga elemente hoomata, millest tulenevalt on mitmusliku verbivormi kasutamine loomulik.

Töö andis esimese põhjalikuma ülevaate teemasse, mida ei ole eesti keeles tegeliku keelekasutuse andmete põhinedes käsitletud.

Kirjandus

- Breiman, Leo 1984.** Introduction to Tree Classification. *Classification and Regression Trees*. London: Chapman and Hall/CRC.
- Breiman, Leo 2001.** Random forests. *Machine Learning* 45 (1), 5–32. <https://doi.org/10.1023/A:1010933404324>.
- Cliff, Norman 1993.** Dominance statistics: Ordinal analyses to answer ordinal questions. *Psychological Bulletin* 114, 494–509.
- Cohen, Jacob 1988.** *Statistical power analysis for the behavioral sciences*. Second edition. Hillsdale, N.J: L. Erlbaum Associates.
- Corbett, Greville G. 1979.** Agreement hierarchy. *Journal of Linguistics* 15 (02), 203–224. DOI: 10.1017/S0022226700016352.
- Corbett, Greville G. 1999.** Agreement. – *Concise Encyclopedia of Grammatical Categories*. Ed. by Keith Brown, Jim Miller, R. E. Asher. Amsterdam, Lausanne, New York, Oxford, Shannon, Singapore, Tokyo: Elsevier, 12–18.
- Corbett, Greville G. 2006.** Agreement. *Cambridge textbooks in linguistics*. Cambridge: Cambridge University Press.
- Cramér, Harald. 1946.** *Mathematical Methods of Statistics*. Princeton: Princeton University Press.
- Croft, William. 2003.** *Typology and universals*. Second edition. Cambridge: Cambridge University Press.
- Cruse, David Alan 1999.** Number and Number Systems. – *Concise Encyclopedia of Grammatical Categories*. Ed. by Keith Brown, Jim Miller, R. E. Asher. Amsterdam, Lausanne, New York, Oxford, Shannon, Singapore, Tokyo: Elsevier, 267–271.
- Cutler jt 2007 = Cutler, David Richard, Thomas C. Edwards, Karen H Beard, Adele Cutler, Kyle T. Hess, Jacob Gibson, Joshua J. Lawler 2007.** Random forests for classification in ecology. *Ecology* 88 (11), 2783–2792. DOI: 10.1890/07-0539.1.
- Dixon, Robert M. W. 1979.** Ergativity. – *Language* 55 (1), 59–138.

- EKK 2020 = Erelt, Mati, Tiiu Erelt, Kristiina Ross 2020.** Eesti keele käsiraamat. Uuendatud väljaanne. Tallinn: Eesti Keele Sihtasutus.
- EKÜ 2021 = Koppel, Kristina, Jelena Kallas 2022.** Eesti keele ühendkorpus 2021. Estonian National Corpus 2021 (Estonian NC 2021). <https://doi.org/10.15155/3-00-0000-0000-0000-08D17L>.
- Erelt, Mati 1999.** Agreement in Estonian. – Estonian: Typological studies III. Toim. Mati Erelt. Publications of the department of Estonian of the University of Tartu II. Tartu: Tartu Ülikool, 7–46.
- Erelt, Mati 2000.** Arvuühildumisest tänapäeva eesti kirjakeeles. Keel ja Kirjandus 3, 180–189.
- Erelt, Mati 2017a.** Öeldis. – Eesti keele süntaks (= Eesti keele varamu III). Toim. Mati Erelt, Helle Metslang. Tartu: Tartu Ülikooli Kirjastus, 94–239.
- Erelt, Mati 2017b.** Sissejuhatus süntaksisse. – Eesti keele süntaks (= Eesti keele varamu III). Toim. Mati Erelt, Helle Metslang. Tartu: Tartu Ülikooli Kirjastus, 53–89.
- Erelt, Mati 2017c.** Liitlause. – Eesti keele süntaks (= Eesti keele varamu III). Toim. Mati Erelt, Helle Metslang. Tartu: Tartu Ülikooli Kirjastus, 647–755.
- Erelt, Mati 2019.** Lause õigekeelsus. Juhatused ja harjutused. Neljas, täiendatud trükk. Tallinn: Emakeele Selts.
- Erelt, Mati, Helle Metslang 1998.** Oma või võõras? Keel ja Kirjandus 10, 657–668.
- Hakulinen jt 2004 = Hakulinen, Auli, Maria Vilkuna, Riitta Korhonen, Vesa Koivisto, Tarja-Riitta Heinonen, Irja Alho 2004.** Iso Suomen Kielioppi. Helsingi: Suomalaisen Kirjallisuuden Seura.
- Helasvuo, Marja-Liisa 2001.** Syntax in the Making: The Emergence of Syntactinits in Finnish Conversation. Studies in Discourse and Grammar 9. Amsterdam/Philadelphia: Benjamins.
- Hopper, Paul J., Sandra A. Thompson 1980.** Transitivity in Grammar and Discourse. Language 56 (2), 251-299. <https://doi.org/10.2307/413757>.
- Hosmer jt 2013 = Hosmer, David W. Jr., Stanley Lemeshow, Rodney X. Sturdivant 2013.** Applied Logistic Regression. Third Edition. Hoboken: John Wiley & Sons (Wiley).

- Hothorn, Torsten, Kurt Hornik, Achim Zeileis 2006.** Unbiased recursive partitioning: A conditional inference framework. *Journal of Computational and Graphical Statistics* 15 (3), 651-674.
- Ji, Jie, Maocheng Liang 2018.** An animacy hierarchy within inanimate nouns: English corpus evidence from a prototypical perspective. *Lingua* 205, 71–89.
- Karlsson, Fred 2002.** Üldkeeleteadus. Tallinn: Eesti Keele Sihtasutus.
- Kielitoimiston ohjepankki 2022.** Verbi yksikössä tai monikossa: viisi hiihtäjää keskeytti vai keskeyttivät? <http://www.kielitoimistonohjepankki.fi/ohje/318>. Vaadatud 20.10.2022.
- Kittilä jt 2011 = Kittilä, Seppo, Katja Västi, Jussi Ylikoski 2011.** Introduction to case, animacy and semantic roles. – Case, Animacy and Semantic Roles. Toim. Seppo Kittilä, Katja Västi, Jussi Ylikoski. Amsterdam: John Benjamins.
- Klavan jt 2015 = Klavan, Jane, Maarja-Liisa Pilvik, Kristel Uihoaed 2015.** The Use of Multivariate Statistical Classification Models for Predicting Constructional Choice in Spoken, Non-Standard Varieties of Estonian. *SKY Journal of Linguistics* 28, 187–224.
- Koppel, Kristina; Kallas, Jelena 2022.** Eesti keele ühendkorpuste sari 2013–2021: mahukaim eestikeelsete digitekstide kogu. *Eesti Rakenduslingvistika Ühingu aastaraamat = Estonian papers in applied linguistics* 18, 207–228. DOI: 10.5128/ERYa18.12.
- Kroeger, Paul R. 2005.** Analyzing Grammar. An Introduction. – Case and agreement. Cambridge: Cambridge University Press.
- Krug jt 2013 = Krug, Manfred, Julia Schlüter, Anette Rosenbach 2013.** Introduction: Investigating language variation and change. – Research Methods in Language Variation and Change. Ed. By Manfred Krug. Cambridge: Cambridge University Press, 1–13.
- Lindström jt 2018 = Lindström, Liina, Maarja-Liisa Pilvik, Helen Plado 2018.** Nimetamiskonstruktsioonid eesti murretes: murdeerinevused või suuline süntaks? *Mäetagused* 70, 91–126.
- Lindström, Liina 2001.** Veel kord subjekti ja predikaadi vastastikusest asendist laiendi järel. *Emakeele Seltsi aastaraamat* 47, 87–106.

- Lindström, Liina 2004.** Sõnajärg lause tuumargumentide eristajana eesti keeles. – Lauseliikmeid eesti keeles. Tartu Ülikooli eesti keele õppetooli preprintid 1. Toim. L. Lindström. Tartu, 40–49.
- Lindström, Liina 2017.** Lause infostruktuur ja sõnajärg. – Eesti keele süntaks (= Eesti keele varamu III). Toim. Mati Ereht, Helle Metslang. Tartu: Tartu Ülikooli Kirjastus, 537–565.
- Lindström, Liina, Kristel Uiboed 2017.** Syntactic variation in ‘need’-constructions in Estonian dialects. *Nordic Journal of Linguistics* 40 (3). 313–349. <https://doi.org/10.1017/S0332586517000191>.
- Metslang, Ann 2015.** Ainsuse pika ja lühikese sisseütleva valiku olenemine morfofonoloogilistest tunnustest – korpusanalüüs. Emakeele Seltsi aastaraamat. Tallinn: Teaduste Akadeemia Kirjastus, 127–147.
- Metslang, Helle 2017a.** Kvantorifraas. – Eesti keele süntaks (= Eesti keele varamu III). Toim. Mati Ereht, Helle Metslang. Tartu: Tartu Ülikooli Kirjastus, 463–478.
- Metslang, Helle 2017b.** Sihitis. – Eesti keele süntaks (= Eesti keele varamu III). Toim. Mati Ereht, Helle Metslang. Tartu: Tartu Ülikooli Kirjastus, 258–277.
- Myhill, John 1992.** *Typological discourse analysis: quantitative approaches to the study of linguistic function.* Oxford: Blackwell.
- Plackett, Robin Lewis 1983.** Karl Pearson and the Chi-Squared Test. *International Statistical Review / Revue Internationale de Statistique* 51 (1), 59–72.
- Pook, Hanna 2019.** The pronoun kes ‘who’ and its referent’s animacy in Estonian dialects. *SKY Journal of Linguistics* 32, 105–144.
- R Core Team. 2022.** *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org>.
- Siiman, Ann 2016.** Ainsuse sisseütleva vormi valiku seos morfosüntaktiliste ja semantiliste tunnustega – materjali ning meetodi sobivus korpusanalüüsiks. Emakeele Seltsi aastaraamat. Tallinn: Teaduste Akadeemia Kirjastus, 207–232.
- Silverstein, Michael 1976.** Hierarchy of features and ergativity. – *Grammatical Categories in Australian Languages.* Ed. by Robert M. W. Dixon. Canberra: Australian Institute of Aboriginal Studies, 112–171.

- Sornicola, Rosanna 1999.** Topic, Focus and Word Order. – Concise Encyclopedia of Grammatical Categories. Ed. by Keith Brown, Jim Miller, R. E. Asher. Amsterdam, Lausanne, New York, Oxford, Shannon, Singapore, Tokyo: Elsevier, 375–383.
- Strobl jt 2009 = Strobl, Carolin, James Malley, Gerhard Tutz 2009.** An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychological Methods* 14 (4), 323–348. DOI: 10.1037/a0016973.
- Tamm, Anne 2012.** Scalar Verb Classes: Scalarity, Thematic Roles, and Arguments in the Estonian Aspectual Lexicon. Firenze: Firenze University Press. <https://doi.org/10.36253/978-88-6655-055-6>.
- Teras, Pire 2018.** The phonetic variation of short intervocalic /h/ in Estonian. *Proc. Speech Prosody* 2018, 883-887. doi: 10.21437/SpeechProsody.2018-178.
- Teras, Pire 2019.** Sõnaalgulise /h/ häälde samadel kõnelejel avalikus ja argisuhtluses. – Eesti ja soome-ugri keeleteaduse ajakiri = *Journal of Estonian and Finno-Ugric Linguistics* 10 (1), 211–231. DOI: 10.12697/jeful.2019.10.1.11.
- Tsunoda, Tasaku 1999.** Transitivity. – Concise Encyclopedia of Grammatical Categories. Ed. by Keith Brown, Jim Miller, R. E. Asher. Amsterdam, Lausanne, New York, Oxford, Shannon, Singapore, Tokyo: Elsevier, 383–391.
- Vaiss, Natalia 2020.** Verbide transitiivsuse kontinumist eesti keeles. *Emakeele Seltsi aastaraamat* 66, 344–386.
- Whaley, Lindsay J. 1997.** Introduction to Typology: The Unity and Diversity of Language. Thousand Oaks: Sage Publications.
- Wilcoxon, Frank 1945.** Individual Comparisons by Ranking Methods. *Biometrics Bulletin* 1 (6), 80–83.

Summary. Subject-Verb Agreement with Numerical Quantifier Phrase

The Master's thesis „Subject-Verb Agreement with Numerical Quantifier Phrase“ belongs to the field of language variation. Language variation is influenced not only by speaker characteristics, social background, and personal preferences, which are well-known phenomena in sociolinguistic, but also by linguistic variables, such as word order in a sentence or verb transitivity. Since linguistic variables have a significant impact on language variation, this master's thesis focuses on the influence of linguistic variables on variation in Estonian –, specifically the choice of verb form in sentences containing a numerical quantifier phrase that acts as the subject.

The aim of the master's thesis was to examine the variation of verb forms agreeing with numerical quantifier phrases and determine the linguistic variables that affect the choice of the verb form's number category (e.g., *kaks sõbralikku meest tuleb* (singular) / *tulevad* (plural)). The specific research questions were:

- which linguistic variables help to explain the choice of verb form?
- which linguistic variable has the greatest influence on the choice of verb form?

The first chapter of the thesis provided a theoretical overview of the quantifier phrase in the Estonian language, while the second theoretical chapter described agreement as a phenomenon and discussed the agreement between the quantifier phrase and the verb in Estonian. The third chapter introduced the dataset used in the study, which was from the Estonian National Corpus (EKÜ 2021), and outlined the research methods I used. In the fourth chapter, I identified the features or linguistic variables through which the variation in the number category of the verb form was investigated. The features included verb tense, word order, transitivity of the verb, quantifier numeral, vagueness of the quantifier phrase, animacy of the subject referent and genre of text. The fifth chapter presented an analysis of the data, while in the sixth chapter, I analyzed the interaction of the features. The seventh chapter provided a summary of the results and the main conclusions.

The aim of my master's thesis was achieved and both research questions were answered. The answers to these questions were as follows: word order is the best predictor of verb form choice and has the greatest influence, followed by verb transitivity, quantifier numeral, and in some cases, the vagueness of the quantifier phrase.

The topic of my master's thesis has not been previously addressed in Estonian linguistics using natural language data. Therefore, this study helps to fill a gap in Estonian language research and explains what kind of linguistic factors may influence the variation of verb forms in agreement with quantifier phrases.

Lihtlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks

Mina, Kärt-Kristiin Jaagu,

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) minu loodud teose „Arvsõnalise kvantorifraasi ühildumine verbiga“, mille juhendajad on Liina Lindström ja Maarja-Liisa Pilvik, reprodutseerimiseks eesmärgiga seda säilitada, sealhulgas lisada digitaalarhiivi DSpace kuni autoriõiguse kehtivuse lõppemiseni.
2. Annan Tartu Ülikoolile loa teha punktis 1 nimetatud teos üldsusele kättesaadavaks Tartu Ülikooli veebikeskkonna, sealhulgas digitaalarhiivi DSpace kaudu Creative Commons'i litsentsiga CC BY NC ND 4.0, mis lubab autorile viidates teost reprodutseerida, levitada ja üldsusele suunata ning keelab luua tuletatud teost ja kasutada teost ärieesmärgil, kuni autoriõiguse kehtivuse lõppemiseni.
3. Olen teadlik, et punktides 1 ja 2 nimetatud õigused jäävad alles ka autorile.
4. Kinnitan, et lihtlitsentsi andmisega ei riku ma teiste isikute intellektuaalomandi ega isikuandmete kaitse õigusaktidest tulenevaid õigusi.

Kärt-Kristiin Jaagu

25.05.2023