

INDEFINIITPRONOOMENITE KEEGI JA MISKI KÄÄNDEVORMIDE VARIEERUMINE SUULISES KEELES

Annika Kängsepp

Ülevaade. Artikkel käsitleb *gi/-ki*-liiteliste indefiniitpronoomenite *keegi* ja *miski* käändevormide varieerumise ulatust ja varieerumist mõjutavaid tegureid kahe eesti keele suulise kõne korpuse, Eesti Rahvusringhäälingu raadiosaadete korpuse ja Eesti taskuhäälingu-korpuse põhjal. Kokku analüüsisin 975 käändevormi, millest 487 moodustasid pronoomeni *keegi* ja 488 *miski* vormid. Tulemustest selgus, et *keegi* puhul esinesid vormid, kus *-gi/-ki* paiknes käändelõpu järel, ning vormid, kus see paiknes käändelõpu ees või kahe ühesuguse käändelõpu vahel üsna võrdselt, vastavalt 54,2% ja 45,8%. *Miski* puhul oli nimetatud vormide osakaal 85,3% ja 14,7%. *-gi/-ki* asukoha varieerumist käändevormides mõjutas *keegi* puhul statistiliselt kõige tugevamalt kõnetempo ning *miski* puhul kõneliik. Binomiaalse logistilise regressiooni segamudeli järgi on 65% käändevormide varieerumisest seletatav kõnelejate individuaalsete erinevustega.*

Võtmesõnad: keelekasutus, varieerumine, korpuslingvistika, morfo-süntaks, morfeemika, statistilised meetodid, eesti keel

1. Sissejuhatus

Indefiniitpronoomenite *keegi*, *miski*, *kumbki* ja *ükski* semantiliste käänete vormides esineb varieeruvust, kus *-gi/-ki* võib paikneda käändelõpu järel (nt *kellelegi*), käändelõpu ees (nt *kellegile*), kahe ühesuguse käändelõpu vahel (nt *kellelegile*) või nii käändelõpu ees kui ka järel (nt *kellegilegi*) (vt nt Rull 1917, Saareste 1923, Saareste 1936). Käändevormide varieerumisel indefiniitpronoomenites on tugev murdeline taust: vormid, kus *-gi/-ki* paikneb käändelõpu järel, on olnud levinud vaid Lõuna- ja Kirde-Eestis (Saareste 1955: 16). *-gi/-ki* paiknemist käändelõpu ees on põhjendatud kinnitava tähenduse 'ka, isegi' kadumisega (Saareste 1936: 554), mistõttu on liide leksikaliseerunud tüve osaks ja sõnad tervikuna omandanud indefiniitse tähenduse (Nevis 1984: 214–215). Siiski on

* Artikli valmimist on toetanud teadusprojekt "Eesti keele morfosüntaktiline varieerumine" (EKKD-TA2).

kirjakeele norminguna jäänud püsima *-gi/-ki* positsioon käändelõpu järel (vt Pant 2020: 14–16).

Kuigi nimetatud varieerumist on täheldatud üle saja aasta, on nähtust siiani süstemaatiliselt vähe uuritud ja uurimustes piiratud vaid kirjaliku keelega (vt Pant 2018, Pant 2020, Kängsepp 2024). 2021. aasta eesti keele ühendkorpuse põhjal tehtud uuringust selgus, et ootuspäraselt esines kõigi nelja pronoomeni kohta enim käändevorme, kus *-gi/-ki* paiknes käändelõpu järel – kokku 85,3%. Pronoomeni *keegi* puhul oli selliseid vorme aga vaid 78,6%. Varieerumist mõjutava tegurina esildus sama pronoomeni puhul statistiliselt kõige tugevamalt žanr: enamasti toimetatud perioodikatekstides ja osaliselt toimetatud blogitekstides oli tõenäolisem praegusele normingule vastav positsioon ja toimetamata foorumitekstides norminguvastane positsioon. (Kängsepp 2024: 1034) Et kirjaliku keele põhjal on võimalik vaadata vaid piiratud hulka tegureid, on sotsiolingvistiliste tegurite ja prosoodia mõju välja selgitamiseks ning varieerumise ulatusest ja selle põhjustest terviklikuma ülevaate saamiseks vaja uurida ka suulist keelt.

gi-/ki-liiteliste indefiniitpronoomenite käändevormide varieerumisest tänapäeva suulises keeles seni ülevaade puudub, mistõttu on siinse artikli eesmärk selgitada kahe suurima eesti keele suulise kõne korpuse, Eesti Rahvusringhäälingu raadiosaadete korpuse (Lippus jt 2023a) ja Eesti taskuhäälingukorpuse (Lippus jt 2023b) põhjal välja pronoomenite *keegi* ja *miski* käändevormide varieerumise ulatus ning uurida varieerumise analüüsi meetodite abil nende pronoomenite käändevormide varieerumist mõjutavaid tegureid.

2. Suulise keele eripära

Üks peamisi suulise ja kirjaliku keele erinevusi on suulise keele lineaarsus, st rääkimine põhineb kõneleja ja kuulaja omavahelisel reaajas toimuval dialoogil. Suulisele keelele on iseloomulikud ka kaduvus, tagasikerimatus ja sünkroonsus. Erinevalt kirjalikust keelest talletavad kõneleja ja kuulaja suulist keelt ainult enda mälus, mis tähendab, et öeldu säilib nii kaua, kui vestluspartnerid seda mäletavad. (Auer 2009: 1–3) Kuna täiskasvanute lühimälu maht on keskmiselt neli sõna (Cowan 2001: 94–97), seab mälu kõne tootmisele ja töötlemisele piirangud. Seetõttu kasutatakse kõnes lühemaid üksusi ja lihtsamaid konstruktsioone. Suulises keeles öeldut on võimalik täpsustada, ümber sõnastada ja parandada, kuid seda ei ole võimalik tagasi võtta. Tagasikerimatus väljendub eelkõige kõhkluste, eneseparanduste ja uuesti alustamise kaudu. Samad protsessid võivad toimuda ka kirjalikus keeles, kuid need pole teksti vastuvõtjale nähtavad, sest toimuvad enne, kui tekst on talle kättesaadavaks tehtud. Suulises keeles töötleb kuulaja kõneleja öeldut vaid väikese ajalise viitega. Seevastu kirjalikus keeles on kõne tootmise ja vastuvõtmise vahel märkimisväärne ajaline vahe. (Auer 2009: 2–3) Et kõneleja toodab kuulajale kõnet väljaütlemisega samal ajal sõnahaaval, on kõne tootmine seotud ka kõnetempoga (EG 2023: 987). Kõnetempo on inimesiti varieeruv. See sõltub näiteks kõnelejate individuaalsetest erinevustest, kõnesituatsioonist, kõneleja soost, vanusest ja haridusest (Jacewicz jt 2009: 233–234). Pärtel Lippus jt (2024: 155) leidsid oma korpuspõhises uurimuses, et eesti keelt kõnelevate täiskasvanute kõnetempo on spontaanses vestluses keskmiselt 4,7 silpi sekundis, varieerudes 2,5 silbist kuni 6,8 silbini sekundis.

Et suuline kõne toimub reaajas, on kaduv, kuulamisega sünkroonne ja seda ei ole võimalik tagasi kerida, on see ka spontaansem ja loomulikum kui kirjalik keel. See tähendab, et kõnelejal ei ole aega öeldut nii palju planeerida ega kontrollida, mis võimaldab omakorda loomulikult varieerumisel rohkem esile tulla. Erinevalt kirjaliku keele uurimisest saab suulise keele uurimisel arvestada ka sotsiolingvistikulisi tegureid (nt sugu) ja prosoodiat (nt kõnetempo).

3. Materjal ja meetod

Suulise keele materjali käändevormide varieerumise ulatusest ülevaate saamiseks ja varieerumist mõjutavate tegurite uurimiseks kogusin Eesti Rahvusringhäälingu raadiosaadete korpusest ja Eesti taskuhäälingukorpusest. Eesti Rahvusringhäälingu raadiosaadete korpus (edaspidi raadiosaadete korpus) sisaldab 53 000 raadiosaadet kogukestusega 16 000 tundi, mis on salvestatud aastatel 1930–2022. Kokku on korpus 109 miljonit sõna. Eesti taskuhäälingukorpus (edaspidi taskuhäälingukorpus) sisaldab 10 633 episoodi 184 eri taskuhäälingust kogukestusega 10 918 tundi. Saated on salvestatud aastatel 2018–2022. Kokku on korpus 85 miljonit sõna. (Lippus jt 2023a, 2023b)

Indefiniitpronoomenite käändevormide varieerumise ulatusest ülevaate saamiseks otsisin korpustest regulaaravaldisi kasutades välja sõnad, mis sisaldasid järjendit *kelle* või *mille*, millele järgnes *-gi/-ki*. Seejuures võis sõnades esineda muid tähemärke nii enne liidet kui ka pärast seda. Et otsingutulemuste hulka sattus ka muid sõnu, filtreerisin poolautomaatselt välja kõik pronoomenite semantiliste käänete vormid. Järgmiseks otsisin vormide abil välja kõik lausungid, kus käändevorm esines. Kuna ühes lausungis võis olla mitu käändevormi, lähtusin põhimõttest, et üks vorm on üks esinemisjuht. Lausungid võtsin välja koos sama kõneleja ühe tervikkõnevooriga. Kõnevooruks peetakse ühe kõneleja jätkuvat rääkimist (EG 2023: 991). Kõnevoorude määratlemisel lähtusin automaatse kõnetuvastuse määratud voorupiiridest. Lisaks võtsin korpusest metainfona välja saate nime, aasta, failinime, kõnevooru algus- ja lõpuaja sekundites, saate pikkuse, kõneleja ID ning kõneleja nime ja soo, kui see oli olemas.

Käändevormide varieerumist mõjutavate tegurite väljaselgitamiseks moodustasin pronoomenite *keegi* ja *miski* käändevormide põhjal juhuvalimi. Tegureid uurisin vaid nende kahe pronoomeni alusel, sest kirjaliku keele põhjal tehtud uurimuses esines neis varieerumist nii protsentuaalselt kui ka käändeti kõige rohkem (vt Kängsepp 2024: 1020–1021). Valimi moodustamiseks võtsin ajalise kõrvutatavuse huvides mõlemast korpusest ainult 2018.–2022. aastal salvestatud saated. Mõlemas korpus on saated transkribeeritud Tallinna Tehnikaülikooli automaatse kõnetuvastusega (Alumäe jt 2018), mille kvaliteet sõltub palju sellest, missuguste andmete peal on kõnetuvastus treenitud. Et kõnetuvastus on treenitud pigem ametliku kui spontaanse kõne materjalil, kaldub see ka ise transkriptsiooni normingupäraseks muutma. Seetõttu on usaldusväärsete andmete saamiseks vormivariantide ülekuulamine hädavajalik.

Jahuvalimi moodustamisel oli eesmärk kaasata mõlema pronoomeni kohta kummastki korpusest 50 episoodi, kus 1) igas episoodis oleks vähemalt kaks uuritavat käändevormi, 2) iga episoodi pikkus jääks kõikide episoodide pikkuste

1. ja 3. kvartiili vahele, 3) sellise pikkusega episoodidest saaks võtta võimalusel juhuvalimisse igast saatest ainult ühe episoodi, 4) kokku saaks ühe pronoomeni kohta ühest korpusest vähemalt 250 käändevormi. Seejuures lähtusin juhuvalimi moodustamisel kriteeriumide eeltoodud järjekorrast. Kui nende kriteeriumide alusel ei olnud võimalik 50 episoodist 250 käändevormi leida, võtsin järjest ühe juhusliku episoodi. Sel viisil sain raadiosaadete korpusest kokku 251 pronoomeni *keegi* käändevormi 98 episoodist (17 saatest mitu episoodi) ja 251 pronoomeni *miski* käändevormi 102 episoodist (16 saatest mitu episoodi). Taskuhäälingukorpusest sai samal viisil valimisse 253 pronoomeni *keegi* käändevormi 71 episoodist (kõik episoodid eri saadetest) ja 250 pronoomeni *miski* käändevormi 100 episoodist (8 saatest mitu episoodi). Nii kuulus valimisse esialgu 1005 käändevormi.

Juhuvalimiga saadud vormide ülekuulamisel selgus, et mõned vormid tuleb eemaldada, sest vormi kääne oli ebasobiv (nt genitiivvormid *kellegi*, *millegi*), tuvas-tatud vormi asemel oli mõni muu sõna või vorm, kõneleja ei olnud eesti emakeelega, rääkis mõnes muus keeles (nt võru keel, inglise keel) või täpselt sama lausung esines nii tutvustuslõigus kui ka saate sees. Pärast selliste lausungite eemaldamist jäi valimisse 975 käändevormi, neist 487 pronoomeni *keegi* vormi (244 raadio-saadete korpusest ja 243 taskuhäälingukorpusest) ning 488 pronoomeni *miski* vormi (246 raadiosaadete korpusest ja 242 taskuhäälingukorpusest).

Uuritava tunnuseksena vaatlesin *-gi/-ki* asukohta käändevormides ning varieerumist seletavate tunnustena korpust, kõneliiki, pronoomeni asukohta osalauses¹, rõhumäärsõna olemasolu osalauses, pronoomeni esinemist fraasisisese laiendina, eitussõna *mitte* esinemist käändevormi ees, lisarõhutatust, kõneleja sugu, käännet ja kõnetempot. Osa neist teguritest kattub võrreldavuse huvides ka kirjaliku keele uurimuses vaadeldud teguritega (vt Kängsepp 2024). Uuritaval tunnusel oli kaks väärtust: TLgi² ja TgiL/TLgiL. Teine väärtus ühendab kahte vormimalli, sest vorme, kus *-gi/-ki* paiknes kahe ühesuguse käändelõpu vahel, oli vähe (kokku 14 juhtu), mistõttu lugesin need analüüsis kokku vormidega, kus *-gi/-ki* paiknes käändelõpu ees. Varieerumist mõjutavaid tegureid ja nende kategooriaid kirjeldan lähemalt artikli viiendas osas.

Et saada teada, millest sõltub *-gi/-ki* positsioon käändevormides, tegin esmalt ühetunnuselise analüüsi, kus rakendasin hii-ruut-testi (Agresti 2007: 34–41), vajadusel Fisheri täpset testi (Agresti 2007: 45–46) ja t-testi (R Core Team 2022). Nende testide tulemuse lugesin statistiliselt oluliseks, kui testi p-väärtus oli väiksem kui 0,05. Tunnustevahelise seose tugevuse hindamiseks kasutasin Craméri V-d (Cramér 1946) ja Coheni d-d (Cohen 1988). Craméri V väärtus väiksem kui 0,1 näitab väga nõrka seost, 0,1–0,3 nõrka seost, 0,3–0,5 keskmist seost ja suurem kui 0,5 tugevat seost. Coheni d väärtus väiksem kui 0,2 viitab olematule mõjule, 0,2–0,5 väikesele mõjule, 0,5–0,8 keskmisele mõjule ja suurem kui 0,8 suurele mõjule. Tunnustevaheliste seoste mõjude visuaalseks hindamiseks ja nende koosmõju hindamiseks *-gi/-ki* asukoha varieerumisele tegin lisaks mitmetunnuselise analüüsi, kus rakendasin tingimuslikku otsustuspuu mudelit (Hothorn jt 2006) ja binomiaalset logistilise regressiooni segamudelit (Bates jt 2015).

Sobivate vormide leidmiseks ja andmete analüüsimiseks kasutasin raketustarkvara R (R Core Team 2022). Tulemuste visualiseerimiseks olid kasutusel paketid ggplot2 (Wickham 2016), partykit (Hothorn, Zeileis 2015), see (Lüdecke

¹ Osalausetes alla on loetud nii liitlausesse kuuluvad osalaused, iseseisvad lihtlaused kui ka nende kahe piirijuhtumid. Suulises tekstis on pea igas lauses markeeritud seos mõne teise lausega ning piir osalause ja iseseisva lause vahel on hajus. (Vt EG 2023: 1016–1021)

² TLgi ja TgiL/TLgiL on vormimallid, kus T tähistab tüve, L käändelõppu ja gi liidet *-gi/-ki*.

jt 2021) ja sjPlot (Lüdecke 2023). Tunnustevahelise seose headuse hindamiseks kasutasin paketti lsr (Navarro 2015) ja puumudeli headuse hindamiseks paketti pROC (Robin jt 2011).

4. Indefiniitpronoomenite keegi ja miski käänevormide varieerumise ulatus

Järgnevalt annan pronoomenite kaupa ülevaate sellest, kui palju ja missuguseid keegi ja miski käänevorme raadiosaadete korpuses ja taskuhäälingukorpuses leidis. Täpsemalt kirjeldan vaid ülekuulatud vorme. Vormide sagedustest ja osakaaludest annan ülevaate tabelite abil juba varasemates töödes kasutatud nelja vormimalli põhjal: TLgi (nt kelle/le/gi), TgiL (nt kelle/gi/le), TLgiL (nt kelle/le/gi/le) ja TgiLgi (nt kelle/gi/le/gi).

Raadiosaadete korpuses ja taskuhäälingukorpuses oli kokku 30 767 pronoomeni keegi käänevormi, mis liigitusid kolme malli alla: TLgi, TgiL ja TLgiL. Vormimalli TgiLgi järgi moodustatud vorme korpustes ei esinenud. Automaatse kõnetuvastuse järgi paiknes 29 956 vormis (97,4%) -gi/-ki käändelõpu järel, 806 vormis (2,6%) käändelõpu ees ja 5 (0%) vormis kahe käändelõpu vahel. Kokku kuulasin üle 504 keegi käänevormi, mille sai samuti liigitada samade vormimallide alla. Kuulamise tulemusel selgus, et automaatne kõnetuvastus oli tuvastanud õigesti 269 käänevormi (53,4%) ja valesti 235 käänevormi (46,6%), st peaaegu iga teine vorm oli valesti tuvastatud. Tabeli 1 põhjal annan ülevaate kuulamisega kontrollitud vormide esinemise sagedustest ja osakaaludest vormimallide järgi. Tabelis olen esitanud 487 vormi, sest pidin juhuvalimist eemaldama 17 käänevormi (vt selgitusi osast 3).

Tabel 1. Pronoomeni keegi käänevormide esinemise sagedused ja osakaalud juhuvalimis vormimallide järgi raadiosaadete korpuse ja taskuhäälingukorpuse põhjal

Kääne	TLgi nt kellelegi	TgiL nt kellegile	TLgiL nt kellelegile	Kokku
Illatiiv	–	1 (100%)	–	1 (0,2%)
Inessiiv	1 (100%)	–	–	1 (0,2%)
Elatiiv	12 (100%)	–	–	12 (2,5%)
Allatiiv	125 (67,6%)	60 (32,4%)	–	185 (38%)
Adessiiv	58 (36%)	91 (56,5%)	12 (7,5%)	161 (33,1%)
Ablatiiv	21 (75%)	7 (25%)	–	28 (5,7%)
Translatiiv	5 (100%)	–	–	5 (1%)
Terminatiiv	–	–	–	–
Essiiv	1 (50%)	1 (50%)	–	2 (0,4%)
Abessiiv	–	–	–	–
Komitatiiv	41 (44,6%)	51 (55,4%)	–	92 (18,9%)
Kokku	264 (54,2%)	211 (43,3%)	12 (2,5%)	487 (100%)

Tabelist 1 selgub, et sarnaselt kirjaliku keele uurimusega (Kängsepp 2024: 1020) esines suulises keeles enim käänevorme, kus -gi/-ki paiknes käändelõpu järel (54,2%). Käänevormid, kus -gi/-ki paiknes käändelõpu ees, olid ülekaalus illatiivis

(100%) (siiski vaid 1 juht), adessiivis (56,5%) ja komitatiivis (55,4%), kirjaliku keele uurimuses vaid abessiivis (65,1%) (Kängsepp 2024: 1020). Mõlemas uurimuses oli vorme, kus *-gi/-ki* paiknes kahe käändelõpu vahel, enim adessiivis: kirjalikus keeles 1,4% (2918 juhtu) (Kängsepp 2024: 1020) ja suulises 7,5% (12 juhtu). Kui kirjalikus keeles moodustasid vormid, kus liide paiknes mujal kui käändelõpu järel, kokku 21,3% (Kängsepp 2024: 1020), siis suulises keeles moodustasid need 45,8% kõigist juhuvalimis olevatest vormidest. Teisisõnu oli kirjalikus keeles *-gi/-ki* asukoht peaaegu igas viiendas vormis mujal kui käändelõpu järel, samas kui suulises keeles oli see nii peaaegu igas teises vormis.

Pronoomeni *miski* käändevorme oli mõlemas korpuses kokku 5775. Erinevalt pronoomenist *keegi* liigitusid pronoomeni *miski* vormid nelja vormimalli alla: TLgi, TgiL, TLgiL ja TgiLgi. Automaatse kõnetuvastuse järgi paiknes *-gi/-ki* käändelõpu järel 4896 (84,8%), käändelõpu ees 862 (14,9%), kahe käändelõpu vahel ühes (0%) ning käändelõpu ees ja järel 16 vormis (0,3%). Kokku kuulasin üle 501 pronoomeni *miski* vormi, mis samamoodi nagu pronoomeni *keegi* vormid liigitusid kolme vormimalli alla: TLgi, TgiL ja TLgiL. Ühtegi malli TgiLgi järgi moodustatud vormi juhuvalimisse ei sattunud. Kuulamise tulemusel selgus, et kõnetuvastus oli tuvastanud õigesti 419 vormi (83,6%) ja valesti 82 vormi (16,4%), st valesti oli tuvastatud peaaegu iga kuues vorm. Tabelis 2 annan ülevaate pronoomeni *miski* käändevormide esinemise sagedustest ja osakaaludest vormimallide järgi ülekuulatud vormide põhjal. Täpsemalt annan ülevaate 488 vormi esinemise sagedustest ja osakaaludest, sest pidin 13 vormi juhuvalimist eemaldama (vt täpsemalt osas 3).

Tabel 2. Pronoomeni *miski* käändevormide esinemise sagedused ja osakaalud juhuvalimis vormimallide järgi raadiosaadete korpuse ja taskuhäälingukorpuse põhjal

Kääne	TLgi nt millelegi	TgiL nt millegile	TLgiL nt millelegile	Kokku
Illatiiv	9 (100%)	–	–	9 (1,8%)
Inessiiv	57 (100%)	–	–	57 (11,7%)
Elatiiv	190 (97,4%)	4 (2,1%)	1 (0,5%)	195 (40%)
Allatiiv	37 (78,7%)	10 (21,3)	–	47 (9,6%)
Adessiiv	1 (50%)	1 (50%)	–	2 (0,4%)
Ablatiiv	–	–	–	–
Translatiiv	39 (92,9%)	2 (4,8%)	1 (2,4%)	41 (8,6%)
Terminatiiv	1 (100%)	–	–	1 (0,2%)
Essiiv	–	1 (100%)	–	1 (0,2%)
Abessiiv	1 (50%)	1 (50%)	–	2 (0,4%)
Komitatiiv	81 (61,4%)	51 (38,6%)	–	132 (27%)
Kokku	416 (85,3%)	70 (14,3%)	2 (0,4%)	488 (100%)

Tabelist 2 on näha, et ka pronoomeni *miski* puhul esines sarnaselt kirjaliku keele uurimusega (Kängsepp 2024: 1021) kõige enam vorme, kus *-gi/-ki* paiknes käändelõpu järel (85,3%). Essiivis (100%) (vaid 1 juht), adessiivis (50%) (vaid 1 juht) ja abessiivis (50%) (vaid 1 juht) oli kõige rohkem või sama palju selliseid vorme, kus liide paiknes käändelõpu ees. Vorme, kus *-gi/-ki* paiknes kahe käändelõpu vahel, oli kirjalikus keeles nii inessiivis, elatiivis, adessiivis, translatiivis kui ka komitatiivis

(Kängsepp 2024: 1021). Siinses uurimuses esines neid vaid kaks: üks elatiivis ja üks translatiivis. Kirjaliku keele uurimuses oli vorme, kus liide paiknes mujal kui käändelõpu järel, kokku 7,5% (Kängsepp 2024: 1021), siinses uurimuses aga 14,7% kõigist vormidest. See tähendab, et sarnaselt pronoomeniga *keegi* varieerub ka *miski* puhul *-gi/-ki* asukoht suulises keeles rohkem kui kirjalikus. Pronoomeni *miski* vormides paikneb *-gi/-ki* käändelõpu ees või kahe käändelõpu vahel peaaegu igas seitsmendas vormis.

Kuna suuline keel on spontaansem ning vähem planeeritud ja kontrollitud kui kirjalik, võib selline varieeruvuse ulatus (eriti pronoomeni *keegi* käändevormides) näidata, et *-gi/-ki* asukoht käändelõpu ees on keelekasutajatele peaaegu niisama loomulik kui käändelõpu järel.

5. Pronoomenite *keegi* ja *miski* käändevormide varieerumist mõjutavate tegurite ühetunnuseline analüüs

Selles osas annan allosade kaupa ülevaate pronoomenite *keegi* ja *miski* käändevormide varieerumist mõjutavatest teguritest ja nende väärtustest. Et testitud tegureid oli kokku kümme, kirjeldan täpsemalt ainult neid, mis *-gi/-ki* asukoha varieerumist käändevormides statistiliselt oluliselt mõjutasid. Kõikidest teguritest ja nende mõju olulisusest käändevormide varieerumisele annan ülevaate tabelis 3.

Tabel 3. Pronoomenite *keegi* ja *miski* käändevormide varieerumist mõjutavad tegurid

Tegur (väärtused)	<i>keegi</i>	<i>miski</i>
	Statistiliselt oluline	Statistiliselt oluline
Korpus (raadiosaated, taskuhäälingud)	x	
Kõneliik (eitav, jaatav)		x
Pronoomeni asukoht osalauses (alguses, keskel, lõpus)	x	x
Rõhumäärsõna olemasolu osalauses (ei, jah)		
Pronoomeni esinemine fraasisises laiendina (ei, jah)		
Eitussõna mitte esinemine käändevormi ees (ei, jah)		
Lisarõhutatus (ei, jah)		
Kõneleja sugu (mees, naine)	x	
Kääne (elatiiv, allatiiv, adessiiv, ablatiiv, translatiiv, komitatiiv)	x	x
Kõnetempo (silpi sekundis)	x	

-gi/-ki asukoht pronoomeni *keegi* käändevormide varieerumisel sõltus statistiliselt oluliselt korpusest, pronoomeni asukohast osalauses, kõneleja soost, käändest ja kõnetempost (vt tabel 3). Võrreldes tulemusi kirjaliku keelega (vt Kängsepp 2024: 1025–1032) nähtub, et kirjalikus keeles sõltus see žanrist, pronoomeni esinemisest laiendina ja pronoomeni funktsioonist osalauses. Seega on mõlemas uurimuses vähemalt pronoomeni *keegi* puhul -gi/-ki asukoha varieerumisel oluline osa sellel, kas tegu on ametliku (kirjalikus keeles toimetatud) või mitteametliku (kirjalikus keeles toimetamata) keelega. Pronoomeni *miski* puhul mõjutasid käändevormide varieerumist statistiliselt oluliselt kõneliik, pronoomeni asukoht osalauses ja kääne. Niisiis sõltub mõlema pronoomeni käändevormide varieerumine suulises keeles pronoomeni asukohast osalauses ja käändest.

5.1. Korpus

Korpus on uurimuses oluline tegur, et näha, kas käändevormide varieerumine sõltub situatsiooni ametlikkusest, nii nagu kirjalikus keeles sõltus see žanrist (vt Kängsepp 2024: 1025–1027). Mõlemad korpused esindavad avalikku suulist registrit, kuid erinevad ametlikkuse poolest. Et raadiosaated on Eesti Rahvusringhäälingu saated, avaldub neis ametlik, sotsiaalse kontrolliga arvestav keelekasutus. Taskuhäälingu materjal seevastu esindab vabamat, mitteametlikku keelekasutust. Raadiosaadete alla kuuluvad näiteks “Käbi ei kuku”, “Stuudios on peaminister”, “Samost ja Sildam” ning taskuhäälingute alla näiteks “Mehed ei nuta”, “Be1St” ja “Kinnisvararadio”. Tulemustest selgus, et korpus osutus statistiliselt oluliseks vaid pronoomeni *keegi* käändevormide varieerumise seletamisel. Tabelis 4 olen esitanud nende käändevormide esinemise sagedused ja osakaalud vormimallide järgi raadiosaadete korpuse ja taskuhäälingu korpuse alusel.

Tabel 4. Pronoomeni *keegi* käändevormide esinemise sagedused ja osakaalud vormimallide järgi korpuste alusel

Käändevorm	Raadiosaated	Taskuhäälingud	Kokku
TLgi	150 (61,5%)	114 (46,9%)	264 (100%)
TgiL, TLgiL	94 (38,5%)	129 (53,1%)	223 (100%)
Kokku	244 (100%)	243 (100%)	487 (100%)

Tabelist 4 on näha, et raadiosaadete korpuses esines enim käändevorme, kus -gi/-ki paiknes käändelõpu järel (61,5%). Sellele vastupidiselt esines taskuhäälingukorpuses rohkem käändevorme, mis olid moodustatud mallide TgiL ja TLgiL järgi (53,1%). Ka hii-ruut-testi tulemus kinnitab, et -gi/-ki asukoht käändevormides sõltub korpusest. Testi tulemus on statistiliselt oluline: $\chi^2(1, N = 487) = 9,822$, $p = 0,002$, Craméri $V = 0,14$.

5.2. Kõneliik

Kõneliik väljendab lausega kirjeldatava sündmuse jaatust või eitust (Erelt 2017: 181). Eitus (näide 1) on eesti keeles erinevalt jaatusest (2) grammatiliselt markeeritud. Kuna indefiniitpronoomenite esmane tõlgendus eitavas lauses on negatiivne (Sang 1983: 121–122), võib oletada, et eitavas lauses paikneb *-gi/-ki* indefiniitsuse rõhutamiseks pigem käändelõpu ees. Uurimistulemustest selgus, et kõneliik oli statistiliselt oluline vaid pronoomeni *miski* käändevormide varieerumise seletamisel. Tabelis 5 olen esitanud pronoomeni *miski* käändevormide esinemise sagedused ja osakaalud vormimallide järgi kõneliigi alusel.

- (1) Seetõttu otseselt me hetkel **millegiga** ei tegele [---] (Raadiosaated)
- (2) [---] et ma tahaks rääkida **millestki** muust ka kui spordist [---] (Taskuhäälingud)

Tabel 5. Pronoomeni *miski* käändevormide esinemise sagedused ja osakaalud vormimallide järgi kõneliigi alusel

Käändevorm	Eitav	Jaata	Kokku
TLgi	147 (91,3%)	269 (82,3%)	416 (100%)
TgiL, TLgiL	14 (8,7%)	58 (17,7%)	72 (100%)
Kokku	161 (100%)	327 (100%)	488 (100%)

Nii eitava kui ka jaatava kõneliigiga lausungites esineb rohkem selliseid käändevorme, kus *-gi/-ki* paikneb käändelõpu järel, vastavalt 91,3% ja 82,3% (vt tabel 5). Siiski on jaatava kõneliigiga lausungites oluliselt rohkem selliseid käändevorme, kus *-gi/-ki* paikneb käändelõpu ees või kahe käändelõpu vahel (17,7%). Ka hii-ruut-testi tulemusel sõltub *-gi/-ki* paiknemine käändevormis kõneliigist: χ^2 (1, N = 488) = 6,3114, p = 0,01, Craméri V = 0,11.

5.3. Pronoomeni asukoht osalauses

Käändevormi paiknemine osalauses on seotud fookusega. Kui pronoomen paikneb lause alguses (3), võib see väljendada kontrastiivset fookust. Lause keskel (4) paiknev pronoomen ei ole tavaliselt fookuses. Tüüpiliselt on fookuses pronoomen, mis paikneb lause lõpus (5), kus see väljendab uut või teate seisukohalt olulisimat infot. (vt Lindström 2017: 545–546) Seega võiks eeldada, et kui indefiniitpronoomen on fookuses ja seeläbi rõhutatud, paikneb *-gi/-ki* pigem käändelõpu ees kui järel. Pronoomeni asukoht osalauses osutus statistiliselt oluliseks nii pronoomeni *keegi* kui ka *miski* käändevormide varieerumise seletamisel. Tabelis 6 olen esitanud *keegi* ja tabelis 7 *miski* käändevormide esinemise sagedused ja osakaalud vormimallide järgi vormi asukoha alusel osalauses.

- (3) [---] et **millestki** veel sellisest põnevast ja spetsiifilisest räägiksime [---] (taskuhäälingud)
- (4) [---] neil ei teki **kellegiga** mingit sügavat klikki. (Taskuhäälingud)
- (5) [---] või see nagu ei kohusta meid **millekski** [---] (Taskuhäälingud)

Tabel 6. Pronoomeni *keegi* käändevormide esinemise sagedused ja osakaalud vormimallide järgi vormi asukoha alusel osalauses

Käändevorm	Alguses	Keskel	Lõpus	Kokku
TLgi	54 (43,2%)	159 (58,5%)	51 (56,7%)	264 (100%)
TgiL, TLgiL	71 (56,8%)	113 (41,5%)	39 (43,3%)	223 (100%)
Kokku	125 (100%)	272 (100%)	90 (100%)	487 (100%)

Tabelist 6 nähtub, et kui pronoomeni käändevorm esineb osalauses keskel või lõpus, paikneb *-gi/-ki* selles pigem käändelõpu järel, vastavalt 58,5% ja 56,7%. Kui käändevorm on aga osalauses alguses, paikneb *-gi/-ki* pigem käändelõpu ees või kahe käändelõpu vahel (56,8%). Selline kasutus ei ole siinses uurimuses seotud siiski kontrastiivse fookusega, vaid omajätähendusega (6). Hii-ruut testi tulemusel mõjutab pronoomeni *keegi* asukoht osalauses seda, kus *-gi/-ki* käändevormis paikneb. Testi tulemus on statistiliselt oluline: $\chi^2 (2, N = 487) = 8,2985$, $p = 0,02$, Craméri $V = 0,13$.

(6) [---] **kellelgil** ikka on oma selline silmaterra. (Taskuhäälind)

Tabel 7. Pronoomeni *miski* käändevormide esinemise sagedused ja osakaalud vormimallide järgi vormi asukoha alusel osalauses

Käändevorm	Alguses	Keskel	Lõpus	Kokku
TLgi	23 (100%)	271 (86%)	122 (81,3%)	416 (100%)
TgiL, TLgiL	–	44 (14%)	28 (18,7%)	72 (100%)
Kokku	23 (100%)	315 (100%)	150 (100%)	488 (100%)

Erinevalt pronoomenist *keegi* ei ole pronoomeni *miski* puhul oluline, kas käändevorm asub osalauses alguses, keskel või lõpus, *-gi/-ki* paikneb alati rohkem käändelõpu järel, vastavalt 100%, 86% ja 81,3% (vt tabel 7). Siiski on oluline märkida, et *miski* puhul ei paikne *-gi/-ki* käändelõpu ees ega kahe käändelõpu vahel juhul, kui käändevorm asetseb osalauses alguses. Ka Fisheri täpse testi tulemusel sõltub *-gi/-ki* paiknemine käändevormis pronoomeni *miski* asukohast osalauses. Testi tulemus on statistiliselt oluline: $p = 0,04$, Craméri $V = 0,11$.

5.4. Kõneleja sugu

Naiste keelekasutust on peetud standardkeelele lähemaks (vt Trudgill 1972, Labov 1990), mistõttu võib oletada, et naised paigutavad *-gi/-ki* käändelõpu järele sagedamini kui mehed. Uurimistulemustest selgus, et sugu mõjutab statistiliselt oluliselt vaid pronoomeni *keegi* käändevormide varieerumist. Tabelis 8 olen esitanud pronoomeni *keegi* käändevormide esinemise sagedused ja osakaalud vormimallide järgi kõneleja soo alusel.

Tabel 8. Pronoomeni *keegi* käändevormide esinemise sagedused ja osakaalud vormimallide järgi kõneleja soo alusel

Käändevorm	Mees	Naine	Kokku
TLgi	171 (50,6%)	93 (62,4%)	264 (100%)
TgiL, TLgiL	167 (49,4%)	56 (37,6%)	223 (100%)
Kokku	338 (100%)	149 (100%)	487 (100%)

Tabelist 8 selgub, et kui kõneleja on naine, paikneb *-gi/-ki* tema öeldud käändevormis pigem käändelõpu järel (62,4%). Meeste puhul on vormid, mis on moodustatud malli TLgi ning mallide TgiL ja TLgiL järgi, kasutusel peaaegu võrdselt, vastavalt 50,6% ja 49,4%. Hii-ruut-testi tulemusel sõltub *-gi/-ki* asukoht käändevormis kõneleja soost: χ^2 (1, N = 487) = 5,3582, p = 0,02, Craméri V = 0,1.

5.5. Kääne

Kääne on uurimuses oluline tegur, et näha, kas *-gi/-ki* asukoha varieerumine sõltub kindlastest semantilistest käänetest. Kirjaliku keele uurimusest selgus, et *-gi/-ki* asukoht varieerus käändeti eri määral (vt Kängsepp 2024: 1020–1021). Seetõttu võib eeldada, et ka suulises keeles sõltub vormi varieerumine konkreetsest käändest. Et uurimuses on tegemist juhuvalimisega, oli valimisse võimalik sattuda kõigil käändevormidel alates illatiivist kuni komitatiivini. Kui ükskõik kumma malli järgi moodustatud vormide arv oli null või mõlemas mallis oli ainult üks käändevorm, jätsin need statistiliste testide tegemiseks valimist välja (vt tabel 1 ja 2). Nii jäi pronoomeni *keegi* puhul alles 466 vormi (vt tabel 9) ja *miski* puhul 416 vormi (vt tabel 10). Tabelis 9 olen esitanud pronoomeni *keegi* ja tabelis 10 pronoomeni *miski* käändevormide esinemise sagedused ja osakaalud vormimallide järgi käänete alusel.

Tabel 9. Pronoomeni *keegi* käändevormide esinemise sagedused ja osakaalud vormimallide järgi käänete alusel

Käändevorm	Ablatiiv	Adessiiv	Allatiiv	Komitatiiv	Kokku
TLgi	21 (75%)	58 (36%)	125 (67,6%)	41 (44,6%)	245 (100%)
TgiL, TLgiL	7 (25%)	103 (64%)	60 (32,4%)	51 (55,4%)	221 (100%)
Kokku	28 (100%)	161 (100%)	185 (100%)	92 (100%)	466 (100%)

Pronoomeni *keegi* puhul paikneb *-gi/-ki* ablatiivis ja allatiivis pigem käändelõpu järel, vastavalt 75% ja 67,6% (vt tabel 9). Vastupidine on olukord adessiivi ja komitatiivi puhul, kus käändevormid on moodustatud pigem mallide TgiL ja TLgiL järgi, vastavalt 64% ja 55,4%. Ka hii-ruut-testi tulemusel sõltub *-gi/-ki* asukoht käändest. Testi tulemus on statistiliselt oluline: χ^2 (3, N = 466) = 42,379, p = 3,334e-09, Craméri V = 0,3.

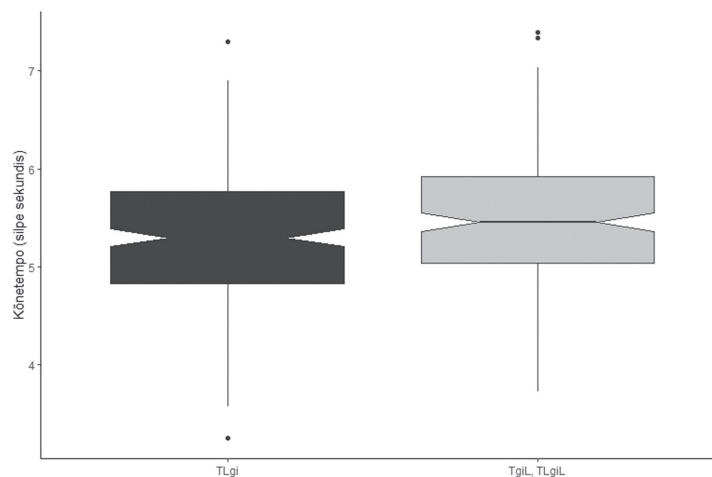
Tabel 10. Pronoomeni *miski* käändevormide esinemise sagedused ja osakaalud vormimallide järgi käänete alusel

Käändevorm	Allatiiv	Elatiiv	Komitatiiv	Translatiiv	Kokku
TLgi	37 (78,7%)	190 (97,4%)	81 (61,4%)	39 (92,9%)	347 (100%)
TgiL, TLgiL	10 (21,3%)	5 (2,6%)	51 (38,6%)	3 (7,1%)	69 (100%)
Kokku	47 (100%)	195 (100%)	132 (100%)	42 (100%)	416 (100%)

Tabelist 10 nähtub, et pronoomenil *miski* on kõigi nelja käände puhul ülekaalus vormid, kus *-gi/-ki* paikneb käändelõpu järel: allatiivis 78,7%, elatiivis 97,4%, komitatiivis 61,4% ja translatiivis 92,9%. Siiski esineb allatiivis (21,3%) ja komitatiivis (38,6%) märkimisväärselt ka selliseid vorme, kus *-gi/-ki* paikneb käändelõpu ees või kahe käändelõpu vahel. Sarnaselt pronoomeniga *keegi* sõltub ka pronoomeni *miski* puhul hii-ruut-testi tulemusel *-gi/-ki* asukoht käandest: $\chi^2(3, N = 416) = 77,555$, $p < 2,2e-16$, Craméri $V = 0,43$.

5.6. Kõnetempo

Mida kiirem on kõnetempo, seda vähem aega on kõnelejal öeldut planeerida ja kontrollida, mistõttu võib oletada, et mida kiiremini kõneleja räägib, seda suurema tõenäosusega paigutab ta *-gi/-ki* käändelõpu ette. Kõnetempot mõõtsin lokaalselt, mis tähendab, et kõneleja kõnetempo võib vestluse jooksul näiteks vastavalt teemale või rõhutamise vajadusele muutuda (vt Lippus jt 2024: 149). Kõnetempo arvutamiseks arvestasin sama kõneleja kõnet 20 sekundit enne ja 20 sekundit pärast käändevormi ütlemist, seda iga juhuvalimis oleva käändevormi puhul. Selle järgi arvutasin, kui mitu silpi sekundis ütleb kõneleja keskmiselt konkreetset vormi ümbritseva(te)s kõnevooru(de)s. Mõõtmise tulemusel selgus, et *-gi/-ki* asukoht sõltus kõnetempost ainult pronoomeni *keegi* puhul. Joonisel 1 olen esitanud pronoomeni *keegi* käändevormide kasutajate keskmise kõnetempo (silpe sekundis).



Joonis 1. Kõnelejate kõnetempo pronoomeni *keegi* puhul

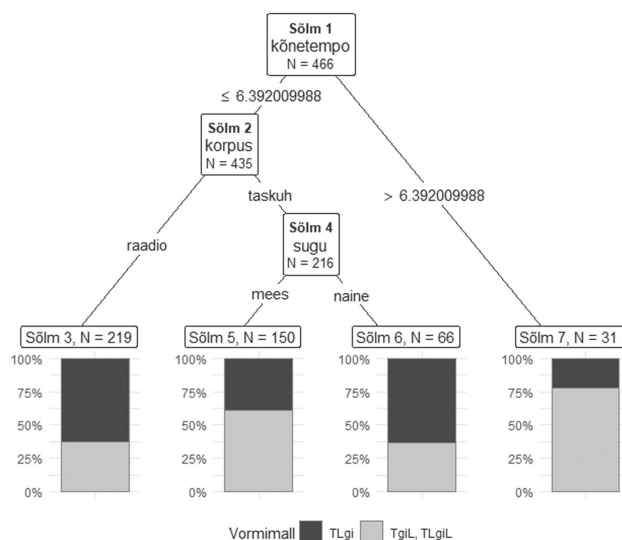
Pronoomeni *keegi* puhul oli kõnetempo keskmiselt 5,28 silpi sekundis neil kõnelejatel, kes paigutasid *-gi/-ki* käändelõpu järele, ja keskmiselt 5,47 silpi sekundis neil kõnelejatel, kes paigutasid *-gi/-ki* käändelõpu ette või kahe käändelõpu vahele. See tähendab, et kiirema kõnetempoga kõnelejad paigutavad *-gi/-ki* käändevormides pigem käändelõpu ette või kahe käändelõpu vahele, mis viitab ka nende vormide suuremale loomulikkusele. Ka t-testi tulemusel sõltub *-gi/-ki* paiknemine käändevormis kõnetempost. Testi tulemus on statistiliselt oluline: $p = 0,0018$, Coheni $d = 0,32$.

6. Pronoomenite *keegi* ja *miski* käändevormide varieerumist mõjutavate tegurite mitmetunnuseline analüüs

Et hinnata *-gi/-ki* asukoha varieerumist mõjutavaid tegureid visuaalselt ja näha nende koosmõjusid, tegin tingimusliku otsustuspuu ja binomiaalse logistilise regressiooni segamudeli.

6.1. Tingimuslik otsustuspuu pronoomeni *keegi* käändevormide varieerumise kohta³

Statistiliselt on *-gi/-ki* asukoha varieerumisega kõige tugevamalt seotud kõnetempo (joonis 2). Kõnetempo järgi hargneb puu kaheks: kõnelejad, kelle kõnetempo on kiirem kui 6,392 silpi sekundis, jagunevad ühte harusse, ja kõnelejad, kelle kõnetempo on võrdne või aeglasem kui 6,392 silpi sekundis, jagunevad teise harusse. Kõnelejad, kelle kõnetempo on kiirem kui 6,392 silpi sekundis, kasutavad suurema tõenäosusega neid käändevorme, kus *-gi/-ki* paikneb käändelõpu ees või kahe käändelõpu vahel. Selliseid kõnesegmente on aga andmestikus vähe (vaid 31). Kõnelejate puhul, kelle kõnetempo on võrdne või aeglasem kui 6,392 silpi sekundis, sõltub *-gi/-ki* paiknemine korpusest (sõlm 2). Raadiosaadete puhul paigutavad kõnelejad liite tõenäolisemalt käändelõpu järele kui kuhugi mujale. Taskuhäälingute puhul sõltub *-gi/-ki* paigutus omakorda kõneleja soost (sõlm 4): kui räägivad mehed, paigutavad nad *-gi/-ki* suurema tõenäosusega käändelõpu ette või kahe käändelõpu vahele, ning kui naised, paigutavad nad *-gi/-ki* suurema tõenäosusega käändelõpu järele kui kuhugi mujale. Tingimusliku otsustuspuu klassifitseerimistäpsus on *keegi* puhul 64% ja C-indeksi väärtus samuti 0,64.

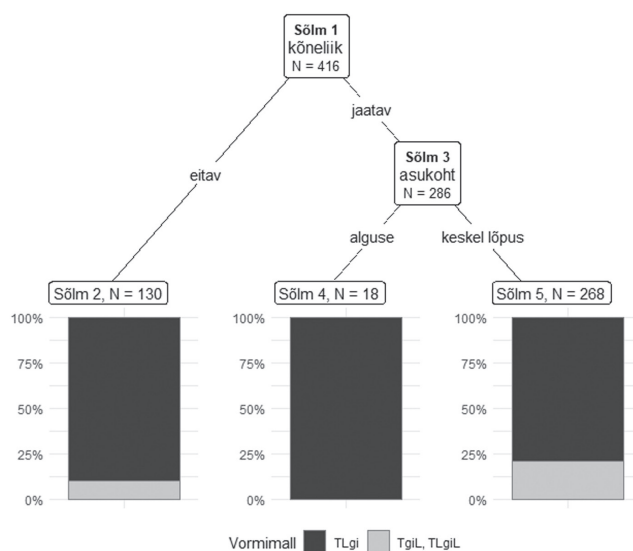


Joonis 2. *-gi/-ki* asukoha varieerumist seletavad olulised tegurid pronoomeni *keegi* puhul

³ puu <- ctree(vormimall ~ asukoht + laiendina_esinemine + kõneliik + mitte_ees + rõhumäärsõna_ol + korpus + rõhutatud + sugu + kõnetempo, data = andmed_k[andmed_k\$ääne %in% c("ablatiiv", "adessiiv", "allatiiv", "komitatiiv"),])

Kuigi ühetunnuselisest analüüsist tuli välja, et *-gi/-ki* asukoha varieerumisega on pronoomeni *keegi* puhul statistiliselt oluliselt seotud korpus, pronoomeni asukoht osalauses, kõneleja sugu, kääne ja kõnetempo, selgub tingimusliku otsustuspuu mudeli põhjal, et kõiki tegureid koos vaadates mõjutavad varieerumist enim kõne-situatsioon ja kõnelejate individuaalsed omadused.

6.2. Tingimuslik otsustuspuu pronoomeni *miski* käändevormide varieerumise kohta⁴



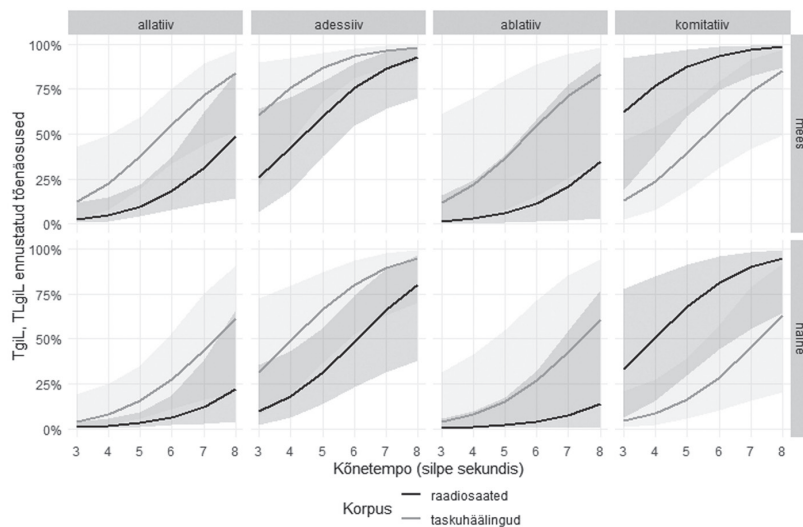
Joonis 3. *-gi/-ki* asukoha varieerumist seletavad olulised tegurid pronoomeni *miski* puhul

Et pronoomeni *miski* käändevormide varieerumise ulatus on tugevalt kallutatud vormide poole, kus *-gi/-ki* paikneb käändelõpu järel (vt ptk 4), tegin ka tingimusliku otsustuspuu mudeli vaid nende tegurite põhjal, mis osutasid ühetunnuselises analüüsis statistiliselt oluliseks (vt ptk 5). Joonise 3 järgi on *miski* puhul *-gi/-ki* asukoha varieerumisega statistiliselt kõige tugevamalt seotud kõneliik, mille järgi jaguneb puu kaheks: lausungid, mille kõneliik on eitav, jagunevad ühte harusse, ja lausungid, mille kõneliik on jaatav, teise harusse. Olenemata nimetatud jagunemisest on pronoomeni *miski* puhul igal juhul suurem tõenäosus kasutada vorme, kus *-gi/-ki* paikneb käändelõpu järel. Tõenäosus paigutada *-gi/-ki* käändelõpu ette või kahe käändelõpu vahele on eitava kõneliigiga lausungites 10% (sõlm 2) ja jaatava kõneliigiga lausungites 20,9% (juhul kui käändevorm paikneb osalause keskel või lõpus) (sõlm 5). Tingimusliku otsustuspuu klassifitseerimistäpsus on *miski* puhul 84% ja C-indeksi väärtus 0,61.

⁴ `puu_m1 <- ctree(vormimall ~ asukoht + kõneliik, data = andmed_m[andmed_m$kääne %in% c("allatiiv", "elatiiv", "komitatiiv", "translatiiv"),])`

6.3. Binomiaalne logistilise regressiooni segamudel⁵

Andmete analüüsiks kasutasin ka binomiaalset logistilise regressiooni segamudelit, sest see arvestab asjaoluga, et andmed ei ole sõltumatud, st selle uurimuse valimis on ühest saatest mitu episoodi ning ka eri vaatluste puhul korduvad samad kõnelejad, mis tähendab, et valimis on ka sõltuvaid vaatlusi. Seega võimaldab kõnelejaspetsiifilise juhusliku mõju kaasamine vaadata (niivõrd, kui kõnelejaid oli võimalik korpuse metaandmete ja salvestuste kuulamise põhjal kindlaks teha), kas analüüsitavad tegurid on statistiliselt olulised ka siis, kui võtta arvesse vaatluste klasterdumist kõnelejati.



Joonis 4. *-gi/-ki* paiknemise tõenäosus pronoomeni *keegi* puhul käändelõpu ees või kahe käändelõpu vahel, võttes arvesse kõnetempo, sugu ning käänat koosmõjus korpusega

Jooniselt 4 on näha, et kõigi käänete puhul, eriti aga adessiivis ja komitatiivis, on meestel suurem tõenäosus moodustada mallidele TgiL ja TLgiL vastavaid käändevorme kui naistel. See tõenäosus kasvab veelgi kõnetempo tõustes. Taskuhäälingutes kasutavad mõlemast soost kõnelejad suurema tõenäosusega liidet käändelõpu ees või kahe käändelõpu vahel. Erinev on kasutus komitatiivis, kus nii mehed kui naised kasutavad liidet käändelõpu ees või kahe käändelõpu vahel pigem raadiosaadete korpuses. Kõiki jooniseid kokku võttes saab öelda, et olenemata käändest, korpusest või kõneleja soost, paigutavad kõnelejad pronoomeni *keegi* puhul kiiremini rääkides *-gi/-ki* tõenäolisemalt käändelõpu ette või kahe käändelõpu vahele. Ka mudeli rühmasisese korrelatsioonikordaja (ICC) järgi on 65% kogu varieerumisest seletatav erinevustega individuaalsete kõnelejate vahel, st kõnelejate vormieelistustega. Kuna pronoomeni *miski* käändevormide varieerumise ulatus on tugevalt kallutatud vormide poole, kus *-gi/-ki* paikneb käändelõpu järel, ei saanud *miski* puhul binomiaalset logistilise regressiooni segamudelit teha.

⁵ `m7 <- glmr(vormimall ~ kõnetempo + kääne*korpus + sugu + (1|kid), data = subset.andmed_k, family = binomial)`

7. Kokkuvõte ja arutelu

gi-/ki-liitelistes indefiniitpronoomenites esineb varieeruvust, kus liide paikneb käändelõpu järel (nt *kellelegi*), ees (nt *kellegile*), kahe käändelõpu vahel (nt *kellelegile*) või käändelõpu ees ja järel (nt *kellegilegi*). Artikli eesmärk oli selgitada kahe suulise keele korpuse (raadiosaated ja taskuhäälingud) põhjal välja pronoomenite *keegi* ja *miski* käändevormide varieerumise ulatus ja uurida varieerumist mõjutavaid tegureid.

Uurimusest selgus, et nii *keegi* kui ka *miski* puhul oli korpuses enim vorme, kus *-gi/-ki* paiknes käändelõpu järel, vastavalt 54,2% ja 85,3%. Vormid, kus liide paiknes käändelõpu ees või kahe käändelõpu vahel, moodustasid 45,8% ja 14,7% kõigist vormidest. Kirjalikus keeles olid *-gi/-ki* paiknemise osakaalud mujal kui käändelõpu järel kaks korda väiksemad: 21,3% (*keegi*) ja 7,5% (*miski*) (Kängsepp 2024: 1020, 1021). Sellistest tulemustest on selgelt näha, et kui keelekasutus on vähem planeeritud ja kontrollitud, tuleb ka loomulik varieerumine rohkem esile. Pronoomeni *keegi* puhul on vormid, kus *-gi/-ki* paikneb käändelõpu ees, keelekasutajatele peaaegu sama loomulikud kui need, kus liide paikneb käändelõpu järel.

Kõiki tegureid koos vaadates selgus, et *keegi* puhul mõjutas varieerumist enim kõnetempo, millele järgnesid kõnetempost lähtuvalt korpus ja sugu, st kõnesituatsioon ja kõnelejate individuaalsed omadused. Pronoomeni *miski* puhul oli igal juhul suurem tõenäosus kasutada vorme, kus *-gi/-ki* paiknes käändelõpu järel. Siiski esildusid varieerumist mõjutavate teguritena enim morfosüntaktilised tegurid: kõneliik, millele järgnes kõneliigist lähtuvalt pronoomeni asukoht osalause. Kõnelejate individuaalsete omaduste olulisust pronoomeni *keegi* varieerumisel kinnitas ka binomiaalne logistilise regressiooni segamudel, mille järgi 65% varieerumisest on seletatav kõnelejate individuaalsete erinevustega, st kõnelejate vormieelistustega. Võrreldes siinset uurimust kirjaliku keele uurimusega (vt Kängsepp 2024), selgus, et kirjaliku keele puhul mõjutas *keegi* käändevormide varieerumist statistiliselt kõige tugevamalt teksti žanr, st perioodika- ja blogitekstides paiknes *-gi/-ki* tõenäolisemalt käändelõpu järel, foorumitekstides käändelõpu ees, kahe käändelõpu vahel või käändelõpu ees ja järel.

Nende kahe uurimuse tulemusi võrreldes on näha, et pronoomeni *keegi* puhul mõjutavad varieerumist kõige enam suhtlussituatsioon (ametlik või mitteametlik suhtlus; toimetatud või toimetamata tekst) ja kõnelejate individuaalsed eripärad. See tähendab, et ühelt poolt laseb vähem planeeritud ja kontrollitud keelekasutus avalduda loomulikult varieerumisel, mida näitab ka varieerumise murdeline taust (Saareste 1955: 16). Teiselt poolt on keelekasutajatel kinnistunud individuaalsed vormieelistused; need võivad omakorda sõltuda eri teguritest (nt vanus, päritolu, haridus), mida selles uurimuses täpsemalt ei vaadatud.

See uurimus näitab võimalusi suulise keele käändevormide varieerumise kirjeldamiseks statistiliste meetodite abil, täiendades ühtlasi varasemaid kirjaliku keele põhjal tehtud uurimusi (vt nt Kängsepp 2024). Uurimus annab terviklikuma ülevaate *gi-/ki*-liiteliste indefiniitpronoomenite vormide varieerumise ulatusest ja põhjustest ning selle tulemusi saab rakendada ka keelekorraldusotsuste tegemisel.

Viidatud kirjandus

- Agresti, Alan 2007. An Introduction to Categorical Data Analysis. 2nd ed. Wiley Series in Probability and Statistics. New Jersey: John Wiley & Sons. <https://doi.org/10.1002/0470114754>
- Alumäe, Tanel; Tilk, Ottokar; Asadullah 2018. Advanced rich transcription system for Estonian speech. – *Frontiers in Artificial Intelligence and Applications*, 307, 1–8. <https://doi.org/10.3233/978-1-61499-912-6-1>
- Auer, Peter 2009. On-line syntax: Thoughts on the temporality of spoken language. – *Language Sciences*, 31 (1), 1–13. <https://doi.org/10.1016/j.langsci.2007.10.004>
- Bates, Douglas; Maechler, Martin; Bolker, Ben; Walker, Steve 2015. Fitting linear mixed-effects models using lme4. – *Journal of Statistical Software*, 67 (1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Cramér, Harald 1946. Mathematical Methods of Statistics. Princeton Landmarks in Mathematics and Physics, 9. Princeton: Princeton University Press. <https://doi.org/10.1515/9781400883868>
- Cohen, Jacob 1988. Statistical Power Analysis for the Behavioral Sciences. 2nd ed. Hillsdale: Lawrence Erlbaum.
- Cowan, Nelson 2001. The magical number 4 in short-term memory: A reconsideration of mental storage capacity. – *Behavioral and Brain Sciences*, 24 (1), 87–114. <https://doi.org/10.1017/S0140525X01003922>
- EG 2023 = Metslang, Helle; Ereht, Mati; Habicht, Külli; Hennoste, Tiit; Kasik, Reet; Teras, Pire; Viht, Annika; Asu, Eva Liina; Lindström, Liina; Lippus, Pärtel; Pajusalu, Renate; Plado, Helen; Rääbis, Andriela; Veismann, Ann 2023. Eesti grammatika. Tartu: Tartu Ülikooli Kirjastus. <https://doi.org/10.12697/EG>
- Ereht, Mati 2017. Eitus. – Mati Ereht, Helle Metslang (Toim.), Eesti keele süntaks. Eesti keele varamu, III. Tartu: Tartu Ülikooli Kirjastus, 181–189.
- Hothorn, Torsten; Zeileis, Achim 2015. partykit: A modular toolkit for recursive partytioning in R. – *Journal of Machine Learning Research*, 16, 3905–3909. <https://jmlr.org/papers/v16/hothorn15a.html> (23.9.2024).
- Hothorn, Torsten; Hornik, Kurt; Achim Zeileis 2006. Unbiased recursive partitioning: A conditional inference framework. – *Journal of Computational and Graphical Statistics*, 15 (3), 651–674. <https://doi.org/10.1198/106186006X133933>
- Jacewicz, Ewa; Fox, Robert A.; O'Neill, Caitlin; Salmons, Joseph 2009. Articulation rate across dialect, age and gender. – *Language Variation and Change* 21, 233–256. <https://doi.org/10.1017/S0954394509990093>
- Kängsepp, Annika 2024. Käänevormide varieerumine ja seda mõjutavad tegurid kirjalikus keeles indefiniitpronoomeni *keegi* näitel. – *Keel ja Kirjandus*, 11, 1016–1037. <https://doi.org/10.54013/kk803a3>
- Labov, William 1990. The intersection of sex and social class in the course of linguistic change. – *Language Variation and Change*, 2, 205–254. <https://doi.org/10.1017/S0954394500000338>
- Lindström, Liina 2017. Lause infostruktuur ja sõnajärg. – Mati Ereht, Helle Metslang (Toim.), Eesti keele süntaks. Eesti keele varamu, III. Tartu: Tartu Ülikooli Kirjastus, 537–565.
- Lippus, Pärtel; Alumäe, Tanel; Orasmaa, Siim; Tsepelina, Katrin; Lindström, Liina 2023a. Eesti Rahvusringhäälingu raadiosaadete korpus. <https://doi.org/10.23673/re-441>
- Lippus, Pärtel; Alumäe, Tanel; Orasmaa, Siim; Pilvik, Maarja-Liisa; Lindström, Liina 2023b. Eesti taskuhäälingukorpus. <https://doi.org/10.23673/re-445>
- Lippus, Pärtel; Pilvik, Maarja-Liisa; Lõo, Kaidi; Lindström, Liina 2024. Kõnetempo ja -soravuse varieerumine eesti keeles. – *Eesti Rakenduslingvistika Ühingu aastaraamat*, 20, 149–163. <https://doi.org/10.5128/ERYa20.09>
- Lüdecke, Daniel; Patil, Indrajeet; Ben-Shachar, Mattan S.; Wiernik, Brenton M.; Waggoner, Philip; Makowski, Dominique 2021. see: An R package for visualizing statistical

- models. – Journal of Open Source Software, 6 (64), a3393. <https://doi.org/10.21105/joss.03393>
- Navarro, Danielle 2015. Learning statistics with R: A tutorial for psychology students and other beginners. Version 0.6. Sydney: University of New South Wales. <https://learningstatisticswithr.com/lsr-o.6.pdf> (23.9.2024).
- Nevis, Joel A. 1984. A non-endoclititic in Estonian. – *Lingua*, 64 (2–3), 209–224. [https://doi.org/10.1016/0024-3841\(84\)90017-2](https://doi.org/10.1016/0024-3841(84)90017-2)
- Pant, Annika 2018. Asesõnade *keegi*, *miski*, *kumbki* käändevormide varieerumine eesti kirjakeeles. BA. Tartu Ülikool. <http://hdl.handle.net/10062/60560>
- Pant, Annika 2020. Pronoomenite *keegi*, *miski*, *kumbki*, *ükski* käändevormide kasutus tänapäeva eesti keeles. MA. Tartu Ülikool. <http://hdl.handle.net/10062/68143>
- Robin, Xavier; Turck, Natacha; Hainard, Alexandre; Tiberti, Natalia; Lisacek, Frédérique; Sanchez, Jean-Charles; Müller, Markus 2011. pROC: An open-source package for R and S+ to analyze and compare ROC curves. – *BMC Bioinformatics*, 12, a77. <https://doi.org/10.1186/1471-2105-12-77>
- Rull, Aado 1917. Kaaslõpp *-gi* ja ta kidurad kaimud. – *Eesti Kirjandus*, 2, 82–88.
- Saareste, Andrus 1923. *Kellegile, mingit, kuskil*. – *Eesti Keel*, 4, 116–117.
- Saareste, Andrus 1936. Eesti õigekeelsuse päevaküsimustest. – *Eesti Kirjandus*, 12, 548–556.
- Saareste, Andrus 1955. *Petit atlas des parlers estoniens*. Väike eesti murdeatlas. Uppsala: Almqvist & Wiksells.
- Sang, Joel 1983. *Eitus eesti keeles*. Tallinn: Valgus.
- Trudgill, Peter 1972. Sex, covert prestige and linguistic change in the urban British English of Norwich. – *Language in Society*, 1 (2), 179–195. <https://doi.org/10.1017/S0047404500000488>
- Wickham, Hadley 2016. *ggplot2: Elegant Graphics for Data Analysis*. 2nd ed. Use R! Cham: Springer. <https://doi.org/10.1007/978-3-319-24277-4>

Võrguviited

- ggplot2. Create Elegant Data Visualisations Using the Grammar of Graphics. <https://CRAN.R-project.org/package=ggplot2> (23.9.2024).
- lsr. Companion to “Learning Statistics with R”. <https://CRAN.R-project.org/package=lsr> (23.9.2024).
- Lüdecke, Daniel 2023. sjPlot: Data visualization for statistics in social science. R package version 2.8.14. <https://cran.r-project.org/web/packages/sjPlot/index.html> (23.9.2024).
- partykit. A modular toolkit for recursive partytioning in R. <https://CRAN.R-project.org/package=partykit> (23.9.2024).
- pROC. Display and analyze ROC curves. <https://CRAN.R-project.org/package=pROC> (23.9.2024).
- R Core Team 2022. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. <https://www.R-project.org/>
- see. Model Visualisation Toolbox for ‘easystats’ and ‘ggplot2’. <https://CRAN.R-project.org/package=see> (23.9.2024).
- sjPlot. Data Visualization for Statistics in Social Science. <https://cran.r-project.org/web/packages/sjPlot/index.html> (23.9.2024).

VARIATION IN THE CASE FORMS OF THE INDEFINITE PRONOUNS *KEEGI* 'SOMEONE' AND *MISKI* 'SOMETHING' IN SPOKEN ESTONIAN

Annika Kängsepp

University of Tartu

In the case forms of the indefinite pronouns *keegi* 'someone', *miski* 'something', *kumbki* 'either', and *ükski* 'none', the *-gi/-ki* can be placed after the case ending (e.g., *kellelegi*), before the case ending (e.g., *kellegile*), between two case endings (e.g., *kellelegile*) or before and after the case ending (e.g., *kellegilegi*) (Rull 1917, Saareste 1923). This variation has a strong dialectal background: forms with *-gi/-ki* after the case ending have historically been common only in Southern and Northeast Estonia (Saareste 1955: 16).

In this article, I used data from Estonian Public Broadcasting's Radio Corpus (Lippus *et al.* 2023a) and Estonian Podcast Corpus (Lippus *et al.* 2023b) to provide an overview of the extent of variation and to describe the factors influencing this variation. Results indicate that for *keegi*, *-gi/-ki* appears after (54.2%) and before or between two case endings (45.8%) at nearly equal frequencies, while for *miski*, the proportions are 85.3% and 14.7%, respectively. The primary factors influencing this variation for *keegi* were speech tempo, and for *miski* polarity.

Keywords: language usage, variation, corpus linguistics, morphosyntax, morphemics, statistical methods, Estonian

Annika Kängsepp (Tartu Ülikool) uurib *gi/-ki*-liiteliste indefiniitpronoomenite käändevormide varieerumist ning tema peamine uurimisvaldkond on eesti keele morfoloogiline varieerumine. Jakobi 2, 50090 Tartu, Estonia
annika.kangsepp@ut.ee