

TARTU ÜLIKOOL
HUMANITAARTEADUSTE JA KUNSTIDE VALDKOND
EESTI JA ÜLDKEELETEADUSE INSTITUUT

Sofia Kriuchkova

ERINEVUSED EESTI KEELT KÕNELEVATE MEESTE JA NAISTE
SUULISES KEELEKASUTUSES

Magistritöö

Juhendajad professor Liina Lindström ja teadur Kristiina Vaik

TARTU 2025

Autorsuse kinnitus

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

Sofia Kriuchkova

Lühikokkuvõte

Vaatamata sellele, et eesti keeleteaduses on olemas mõned uuringud, mis käsitlevad sugu keelelise varieerumise mõjutegurina, on sooga seotud keeleerinevuste uurimine jäänud killustatuks ega ole kujunenud iseseisvaks teadusharuks. Siinne magistritöö täidab seda lünka, uurides, millised keelelised erinevused eksisteerivad eesti keelt kõnelevate meeste ja naiste keelekasutuses ja kuivõrd need vastavad teistes keeltes tehtud sarnaste uurimuste tulemustele.

Analüüs viidi läbi Eesti taskuhäälingukorpuse materjalil, millest moodustati kaks temaatiliselt tasakaalustatud alamkorpust - ainult meeste (umbes 43 tundi) ja ainult naiste (umbes 33,5 tundi) kõnega. Statistilised testid (Mann-Whitney U-test ja χ^2 -test koos Holm-Bonferroni korrigeerimisega) näitasid olulisi erinevusi leksikaalsete ja grammatiliste tunnuste kasutamises. Naised kasutasid sagedamini omadussõnu, adverbe, partikleid, intensiivistajaid, viisakuskonstruktsioone ja deminutiive. Mehed näitasid veidi suuremat sõnavara mitmekesisust (TTR), kuid nende kõnes ei täheldatud käskiva kõneviisi, eituse ja umbisikulise tegumoe oodatud ülekaalu.

Tulemused on üldiselt vastavuses inglise, vene ja teistes keeltes tehtud uurimuste andmetega, kuigi mõned kõrvalekalded üldistest suundumustest võivad olla tingitud kasutatud materjali žanrilistest omadustest. Uurimus näitab, et sugu võib mõjutada keeleliste tunnuste valikut eesti keeles, mis rõhutab edasiste sammude vajalikkust soolingvistika valdkonnas.

Võtmesõnad: keelekasutus, soolingvistika, sugu, suuline keel, varieerumine

Sisukord

Sissejuhatus	6
1. Sotsiaalne sugu	9
2. Soouurimused ja keelekasutus	11
2.1. Soolingvistika areng	11
2.2. Meeste ja naiste keelekasutust eristavad jooned varasemate tööde põhjal	12
2.2.1. Emotsionaalsus	13
2.2.2. Viisakus	15
2.2.3. Keelenormid ja vandesõnad	16
2.3. Meeste ja naiste keelekasutuse erinevused eesti keeles	18
3. Taskuhäälingud kui suulise keele erivorm	21
3.1. Suulise keele erijooned	21
3.2. Taskuhäälingud	22
3.3. Automaatse kõnetuvastuse kasutamine eesti keele analüüsimiseks	
24	
4. Materjal	26
5. Meetod	28
5.1. Keelelised tunnused	28
5.2. Korpuse töötlus	31
5.3. Statistilised meetodid	33
6. Tulemused ja analüüs	36
6.1. Sõnaliigid	36
6.2. Keelelised tunnused	42

6.3. Partiklid.....	51
6.4. Intensiivistajad.....	54
6.5. KP-markerid	58
6.6. TTR.....	60
6.7. Tulemused hüpoteeside valguses.....	60
Kokkuvõte	62
Kirjandus	64
Differences in the Spoken Language Use of Estonian-Speaking Men and Women.....	73

Sissejuhatus

Uurimused soo ja keelekasutuse seostest on tänapäeval üha suurema tähtsusega, kuna keel mitte üksnes ei peegelda olemasolevaid sotsiaalseid reaalsusi, vaid kujundab ka keelekasutajate arusaamu, käitumist ja mõtlemist (Coates 2004). Lääne kultuurides, kus on traditsiooniliselt olnud patriarhaalsed ühiskonnad, on naistel ja meestel olnud erinevad sotsiaalsed rollid (Lakoff 1973). Sellest tulenevalt eeldatakse, et meeste ja naiste käitumine ja ka suhtlemine on samuti erinev.

Juba üle viiekümne aasta on uuritud, kuidas sugu ja soolised normid mõjutavat kõnet ja selle tajumist ning mille poolest naiste ja meeste keelekasutus erineb (Posio jt 2024). Eestis hakati sellele teaduslikku tähelepanu pöörama aga suhteliselt hilja, mis on suuresti seotud nõukogude mineviku mõjuga — see kujundas sugupoolte norme ja stereotüüpe viisil, kus ametlikult kuulutatud võrdsus kaasnes tegelikkuses patriarhaalsete hoiakute püsimise ja traditsiooniliste soorollide kinnistumisega igapäevaste praktikate ja keele kaudu (Annuk & Seigel 2020). Selle tulemusena jäid soouurimused üldiselt, aga ka keeleteaduses, Eestis pikka aega teadusvälja äärealale ning on alles hiljuti hakanud akadeemilises kogukonnas süstemaatilist tähelepanu pälvima (Marling & Annuk 2023). Välisest soolisest võrdsusest hoolimata mõjutavad ühiskonda endiselt patriarhaalsed hoiakud (Ross 2018).

Kaasaegsed uurimused näitavad, et meeste ja naiste keelekasutus erineb mitmete keeleliste tunnuste poolest. Nende hulka kuuluvad näiteks kõnetempo, mis on statistiliselt eri keeltes meestel kõrgem (Pépiot 2014, Fuscone jt 2021), erinevate sõnaliikide kasutamise sagedus (Newman jt 2008, Litvinova jt 2017, Cremades 2024), kirjakeele normi järgimine (Dibrova 2013, Posio jt 2024), intensiivistajate ja deminutiivide sagedase kasutamisega väljendatud emotsionaalsus või roppuste sagedase kasutamisega väljendatud emotsionaalsus (Argamon jt 2003, Konstantinovskaia 2020), erinevate pragmaatiliste konstruktsioonide kasutamine viisakuse markeritena (Coates 2004, Posio jt 2024) jne. Kuigi eesti keeles puudub grammatilise soo kategooria ning seda peetakse võrreldes paljude teiste Euroopa keeltega vähem androtsentriliseks (Kaukonen jt 2025), näitavad mitmed uurimused siiski, et sugu mõjutab keeleliste vahendite valikut. Näiteks Lippus jt (2024) on näidanud, et sugu võib mõjutada kõnetempot ja soravust, kuigi see mõju on nõrk ja ebajärjekindel. Aasa (2024) magistritöö põhjal esineb noorte seas soolisi erinevusi intensiivistajate kasutamises, eelkõige leksikaalsete valikute tasandil. Teras (2019) leidis, et naised järgivad sagedamini hääldusnorme, näiteks foneemi /h/ säilitamist formaalses kõnes. Lisaks osutavad Kaukonen

jt (2025), et eesti keeles leidub endiselt sooliselt markeeritud sõnu – eriti ametinimetustes – mille kasutus aitab keeles stereotüüpe kinnistada. Siiski ei ole eesti keeleteaduses seni käsitletud soolist keelekasutust omaette süstemaatilise uurimisvaldkonnana, mistõttu on meie arusaam sellest, kuidas sotsiaalne sugu igapäevases suulises suhtluses tegelikult avaldub, endiselt killustatud ja puudulik.

Nende tunnuste uurimine on oluline nii ühiskondlikust kui ka tehnoloogilisest vaatenurgast. Ühest küljest soodustab meeste ja naiste keelekasutuse erinevuse ja sellega seotud sooliste stereotüüpide mõistmine soolist võrdõiguslikkust ning aitab võidelda nende stereotüüpide ja diskrimineerimise vastu (nt Baker 2014, Posio jt 2024). Teisest küljest võimaldab see ka näiteks parandada automaatse teksti autorite tuvastamise tehnoloogiaid ja vähendada sooga seotud eelarvamusi suurtes keelemudelites (nt Devinnay jt 2022, Soundararajan & Delany 2024).

Tähtis on märkida, et siinses töös käsitletakse peamiselt andmetel põhinevaid tendentse, mitte rangeid reegleid. Loomulikult ei ole kõik naised stereotüüpselt „naiselikud“ ja kõik mehed ei ole stereotüüpselt „mehelikud“. Iga inimene on erinev ja kõnekäitumist võivad mõjutada paljud tegurid. (Cheng jt 2011: 79) Sellegipoolest kasutan analüüsi ühtsuse huvides siin ja edaspidi termineid naine/mees tähenduses „naiselik“ ja „mehelik“, pidades silmas eelkõige tüüpiliselt naiste või meeste grupiga seotud keelelisi valikuid.

Magistritöö põhieesmärgiks on automaatse tekstitöötlemise ja statistilise analüüsi kasutamise abil tuvastada eesti keelt kõnelevate meeste ja naiste suulise keelekasutuse peamised tunnused. Materjalina kasutan olemasolevat Eesti taskuhäälingukorpust (Lippus jt 2023). Magistritöös olen püstitanud järgmised uurimisküsimused:

- 1) Kas ja kuidas erineb naiste ja meeste keelekasutus?
- 2) Kui erinevused on olemas, kas need erinevused sarnanevad varasemates, teiste keelte kohta tehtud uurimustes ilmnenutega?

Nende uurimiseks olen püstitanud järgmised hüpoteesid, mis põhinevad teiste keelt vastavate uuringute tulemustel:

- 1) Eesti meeste ja naiste suulises keelekasutuses esinevad leksikaalsed ja morfosüntaktilised erinevused.

- 2) Meeste suulises keelekasutuses on sõnavara mitmekesisem (Argamon jt 2003, Litvinova jt 2017).
- 3) Naiste keelekasutus on isikulisem: naised kasutavad sagedamini esimese ja teise isiku pronoomeneid. Meeste keelekasutus on distantseeritum: mehed kasutavad sagedamini passiivi või umbisikuline tegumoe vorme. (Argamon jt 2003)
- 4) Naiste kõne on emotsionaalsem, mis väljendub sagedamas omadussõnade, partiklite ja intensiivistajate kasutamises, ja on suunatud kirjeldamisele, mis väljendub ka adverbide sagedasemas kasutamises. Meeste kõne on rohkem ülesandele ja eesmärgile orienteeritud, mis väljendub nimisõnade ja verbi sagedasemas kasutamises. (Konstantinovskaia 2020, Newman jt 2008, Stratton & Sundquist 2022, Stückler 2022)
- 5) Naised kasutavad rohkem viisakusmarkereid, sealhulgas tingivat kõneviisi ja komplementausega predikaatidel põhinevad markereid (siin ja edasi KP-markerid) nagu *ma arvan, et...*, *ma kardan, et...*, *tundub, et...* (Brown & Levinson 1978, Czerwionka jt 2023, Kirilina 2005)
- 6) Meeste kõne on otsekohesem, mistõttu nad kasutavad sagedamini eitust ja käskivat kõneviisi (Czerwionka jt 2023, Dibrova 2013).
- 7) Naised kasutavad sagedamini deminutiive (Konstantinovskaia 2020, Senkina jt 2018)

Magistritöö koosneb kuuest peatükist. Esimeses peatükis annan bioloogilise ja sotsiaalse soo määratlused. Käsitlen, kuidas sotsiaalne sugu on seotud sooliste normidega ning miks traditsioonilised patriarhaalsed hoiakud on Eesti ühiskonnas endiselt levinud. Teises peatükis annan ülevaate soolingvistika arengust ja uurimustest, kus soo mõju keelekasutusele on analüüsitud kvantitatiivsete meetodite abil. Kolmandas peatükis kirjeldan suulise keele eripärasid, selgitan, miks peetakse taskuhäälinguid sobivaks materjaliks suulise keele uurimisel, ning annan ülevaate automaatse kõnetuvastussüsteemi toimimisest, mille abil korpust transkribeeriti. Neljandas peatükis kirjeldan materjali kogumise ja ettevalmistamise etappe ning arutlen kitsakohtade üle. Viiendas peatükis kirjeldan kogutud korpuse töötlemise etappe ja analüüsis kasutatud statistilisi meetodeid. Kuuendas peatükis esitan ja arutlen tulemuste üle.

Eesti keel ei ole minu emakeel, seega kasutasin veebipõhist masintõlke rakendust DeepL (2025), mis võimaldab tõlgitud teksti kohe kasutajaliideses kontrollida ja toimetada. Lisaks kasutasin eesti keele sõnastuse parandamiseks või sobivate terminite leidmiseks ChatGPT (OpenAI, 2025).

1. Sotsiaalne sugu

Sooga seotud küsimused on pikka aega olnud akadeemiliste arutelude teemaks, eriti sotsiaal-kultuuriliste erinevuste kontekstis. Selles peatükis arutlen, kuidas soonormid ja stereotüübid kujunevad ja püsivad ning millist mõju avaldavad need Eesti ühiskonnas.

Eristatakse bioloogilist sugu ja sotsiaalset sugu (ingl. *sex* ja *gender*). Bioloogiline sugu on bioloogiliselt põhinevad erinevused meeste ja naiste vahel (anatomilised, geneetilised ja füsioloogilised) (Mills 2018). Sotsiaalne sugu ehk *gender* on performatiivne konstrukt, mis kujuneb teatud sotsiaalsete praktikate kordamise kaudu. Need praktikad on tingitud sotsiaalsetest normidest, rollidest ja ootustest, mille alusel ühiskond määratleb, mis on "naiselik" või "mehelik". (Butler 2002)

Soolised normid määravad meeste ja naiste käitumise ühiskonnas ning võivad kultuuriti erineda, sest need on kujunenud konkreetsete ajalooliste, religioosete ja sotsiaalmajanduslike tingimuste kontekstis (Eagly jt 2012). Euroopa ühiskondades on sajandeid püsinud binaarne soorollide mudel, mille kohaselt on mehed seotud ettevõtlikkuse, juhtimise ja saavutustega ning naised hooldamise, meelelahutuse ja koduga (Eagly jt 2012, ka Butler 2002, Kasik 2006). Need hoiakud on pikka aega säilinud religioosete, haridus- ja õiguslike institutsioonide kaudu ning kajastuvad ka igapäevases suhtluses (Butler 2002, Eagly jt 2012).

Vaatamata sellele, et arusaamad sellest, mida mõeldakse soo all, on juba ammu muutunud (Connell 2020) ning viimasel kümnendil on toimunud muutused suurema soolise võrdsuse suunas, on stereotüübid meeste ja naiste traditsiooniliste rollide kohta ühiskonnas ikka veel tugevalt juurdunud ja seetõttu mängivad igapäevases suhtluses suurt rolli. (European Commission 2025)

Sooliste normide ja stereotüüpide kujunemisele Eesti ühiskonnas avaldas märkimisväärset mõju nõukogude periood. Kuigi ametlik nõukogude ideoloogia kuulutas soolist võrdsust, erines tegelikkus sageli deklaratiivsetest loosungitest (Annuk & Seigel 2020). Sooline võrdsus Nõukogude Liidus tähendas naiste ulatuslikku kaasamist tootmisse ja ühiskondlikku tegevusse, kuid erasektoris ja igapäevaelus säilisid traditsioonilised patriarhaalsed hoiakud ning selgelt määratletud soorollid (Annuk 2015: 71–72). See tõi kaasa naiste topeltkoormuse. Nad pidid täitma professionaalseid kohustusi samaväärselt

meestega, samas kandes endiselt peamist vastutust majapidamise ja laste kasvatamise eest. Selline vastuolu kajastus nõukogude propagandakirjanduses, mis pidas seda tugevalt avalikkuse teadvuses. (Annuk 2015)

Iseseisvuse taastamine ja nõukogudejärgne periood tõid kaasa olulisi muutusi, mis siiski ei kõrvaldanud soolisi stereotüüpe, vaid pigem kujundasid neid ümber. Feministlikke ideid ja soolise võrdõiguslikkuse küsimusi võeti Eesti ühiskonnas vastu ettevaatlikkusega, kuna neid seostati nõukogude ideoloogiaga, millest taheti distantseeruda (Marling & Annuk 2023: 138–140). See on toonud kaasa olukorra, kus konservatiivsed arusaamad soorollidest tugevnesid (Marling & Annuk 2023: 141-143). Naist hakati üha enam määratlema eelkõige tema reproduktiivse ja perekondliku funktsiooni kaudu (Ross 2018: 123–125). Ühiskondlikus teadvuses kinnistus ideaalne naisekuvand kui hoolitsev, emotsionaalne ja perekonnale orienteeritud. Meest aga hakati määratlema kui majanduslikku ülalpidajat ja kaitsjat. Mehelikkus hakkas tähendama iseseisvust, jõudu, otsustavust ja emotsionaalset vaoshoitust. (Ross 2018)

Neid soostereotüüpe toetatakse tänapäevalgi meedias ja igapäevases suhtluses (Kasik 2006: 10, Ross 2018: 124–126). Mehi peetakse sagedamini kõrge staatusega ametite esindajateks, mis on seotud juhtimise ja füüsilise jõuga, samas kui naisi seostatakse valdkondadega, mis eeldavad hoolitsust ja suhtlemisostust (Ross 2018). Sellest tulenevalt oodatakse Eesti ühiskonnas endiselt, et mehed ja naised järgiksid traditsioonilisi soolisi norme (Ross 2018), mis seletab, miks vaatamata viimaste aastakümnete olulistele sotsiaalsetele ja majanduslikele muutustele sotsiaalne sugu jätkuvalt mõjutab eesti keelt kõnelevate inimeste keelelist ja sotsiaalset käitumist.

2. Soouurimused ja keelekasutus

2.1. Soolingvistika areng

Kuivõrd sugu mõjutab meeste ja naiste käitumist, põhjustab see ka erinevusi meeste ja naiste keelekasutuses. Kusjuures soolised erinevused avalduvad nii kirjalikus kui ka suulises keeles (Mulac 1998) ja nii argi-, kui ka avalikus-ametlikus vestluses (Argmon jt 2003).

Üheks tuntuimaks kirjutiseks soolingvistikas oli Robin Lakoffi (1973) artikkel „Language and Woman's Place“, kus uuriti naiste ja meeste suulise keelekasutuse erinevusi. Lakoffi sõnul kasutavad naised ühiskonnas piiratud õiguste tõttu vähem agressiivset keelt, niinimetatud „naiste keelt“, mida iseloomustavad viisakusväljendid (siin nt *would you please...* ('kas sa/te palun...'), *I really appreciate it if...* ('Oleksin väga tänulik, kui'), intensiivistajad¹ (eriti nt *so* 'nii' ja *very* 'väga') ja ebamäärasust väljendavad sõnad ja väljendid (nt *kinda* 'nagu') ja *I guess* 'ma arvan'), otsene tsiteerimine ning „tühised“ omadussõnad² (nt *sweet* 'kena', *divine* 'jumalik', *charming* 'võluv', *adorable* 'armas') (Lakoff 1973: 53-57). Samuti peeti „naiste keelt“ alaväärseks ja seda seostati peamiselt laste keelekasutusega kuna arvati, et naiste kõnes nagu laste kõnes on palju deminutiive³ ja pehmeid⁴ (Lakoff 1973 ja Liivak 2023:6).

Meeste ja naiste erinevat keelekasutust on lingvistikas tõlgendatud mitme erineva teoreetilise mudeli abil. Näiteks Lakoffi vaadet sooga seotud erinevustele keeles nimetatakse defitsiidi mudeliks (ingl. *deficit approach*), sest naiste viisakas ja tähelepanelik keelekasutus tugevdab ühiskondlikke uskumusi ja ootusi, et naised on nõrgemad ja ebakindlamad, mistõttu naiste keelekasutus on „defitsiitne“ võrreldes meeste keelekasutusega (Coates 2004: 5-6; Kirilina 2005: 17). Samaaegselt defitsiidi mudeliga

¹ Intensiivistajad on sõnad, mis tugevdavad või nõrgendavad teise sõna omaduse astet (Aasa 2024: 20).

² Tühised omadussõnad on omadussõnad, mis kasutatakse tähistamiseks kõneleja heakskiitu või imetlust millegi suhtes (Lakoff 1973: 53).

³ Deminutiivid on vähendusliitega sõnad, mis tähistavad väiksust, väljendades samal ajal kiindumust, emotsionaalset värvingut, ironiat, palvete pehendamist või halvustamist (Liivak 2023: 6).

⁴ Pehmeid on keelelised vahendid, mis võimaldavad kõnelejal lisada ütlusele oma hinnangu, vähendada näohtu, väljendada otseselt kindlust või ebakindlust, anda märku teiste seisukohtade võimalikkusest ja rõhutada oma hinnangute subjektiivsust (Teiva 2024: 13)

töötati välja domineerimise mudel (ingl. *dominance approach*) (Fishman 1980), mille kohaselt meeste ja naiste keelekasutuste erinevused ei ole seotud naiste nõrkuse ja ebakindlusega, vaid meeste liigse enesekindlusega, ning erinevuste mudel (ingl. *difference approach*) (Malz & Borker 1982), mille kohaselt meeste ja naiste sotsiaalsed normid erinevad sarnaselt sellele, kuidas erinevad sotsiaalsed normid erinevates kultuurides, mis kajastub ka keelekasutuses (Coates 2004: 5-6; Kirilina 2005: 98, 101, 165, 202). Kuigi põhjused, miks mehed ja naised räägivad erinevalt, on erinevad, on nende kolme mudeli abil kirjeldatud erinevused samad, näiteks mainitakse naiste kõne suuremat viisakust, ebamäärasust ja kohanemisvõimet ning meeste otsekoheust ja enesekindlust. Nimetatud tähelepanekud olid vastavuses teiste, peamiselt inglise keeles läbi viidud empiiriliste uurimuste tulemustega (nt Berryman ja Wilcox 1980, Edelsky 1976, Haas 1979, Kramer 1977). Seega on soolingvistikas kujunenud üldised arusaamad sellest, mis on meeste ja naiste keelekasutusele iseloomulik. Arvatakse, et naiste keelekasutus on tõenäolisemalt seotud viisakuse, õrnuse ja solidaarsuse väljendamisega ning on emotsionaalsemalt värvitud. Seevastu mehelikku keelekasutust kirjeldatakse traditsiooniliselt kui otsekoheemat, kategoorilisemat ning autoriteedi kinnitamisele ja kõneleja staatuse demonstreerimisele orienteeritud keelekasutust. (Posio jt 2024).

2.2. Meeste ja naiste keelekasutust eristavad jooned varasemate tööde põhjal

Soolingvistika on pikka aega olnud kirjeldav ja paljud valdkonna tööd põhinesid peamiselt autorite isiklikel kogemustel ning sõltusid seega ühiskonnas eksisteerivatest soostereotüüpidest. Selle tõttu on seda keeleteaduse suunda sageli kritiseeritud (Posio 2024). Kuid alates 20. sajandi lõpust on tänu tehnoloogia arengule ja kättesaadavate keelekorpusete tekkimisele hakatud soolingvistika valdkonnas aktiivselt kasutama kvantitatiivset lähenemisviisi. Teadlastel tekkis võimalus töötada suurte andmekogumitega (vt Argamon et al. 2003; Newman et al. 2008) ja kasutada statistilisi analüüsimeetodeid (vt Gauthier 2017; Posio 2024). Üleminek kvantitatiivsetele uurimustele võimaldas täpsemalt kindlaks teha, millised keelelised tunnused on omased meeste ja naiste keelekasutusele ning kuidas neid mõjutavad sotsiaalsed rollid võrreldes individuaalsete keelekasutuse

eripäraga. Järgnevalt annan ülevaate neist meest ja naiste keelekasutust eristavatest joontest, mis on kvantitatiivsetest uurimustest välja tulnud.

2.2.1. Emotsionaalsus

Jakobsoni (2012: 5-11) järgi on üks keele peamistest funktsioonidest emotiivne (ekspressiivne) funktsioon. Selle eesmärk on väljendada kõneleja suhtumist avalduse sisusse ja vestluspartnerisse endasse. Emotsionaalsust väljendatakse keeles erinevate keeleliste vahenditega, näiteks emotsionaalse sõnavara, partiklite, intensiivistajate, hüüatuste või intonatsiooni kaudu. Arvatakse, et naiste keelekasutuses on rohkem emotsionaalsuse markereid kui meeste kõnes, sest naised räägivad rohkem tunnetest ja mehed faktidest (Newman jt 2008).

Newman jt (2018) analüüsisid üle 14 000 teksti eri žanritest, kasutades LIWC (ingl. *Linguistic Inquiry and Word Count*) keeleanalüüsi tööriista, mis automaatselt arvutab, mil määral inimesed kasutavad eri kategooriatesse kuuluvaid sõnu laias valikus tekstides, näites meilides, kõnedes, luuletustes või transkribeeritud igapäevakõnes. Tulemused näitasid, et naised kasutasid sagedamini isikulisi pronomeneid, funktsioonisõnu (artiklid, partiklid) ja emotsionaalse sisuga omadussõnu (sõltuvalt LIWC kontekstist - sageli on see *happy* 'õnnelik', *sad* 'kurb', *angry* 'vihane' jms.), mehed aga nimisõnu ja leksikaalsi verbe (ehk sisulised, tegevust või seisundit väljendavad verbid, nt *run* 'jooksma', *speak* 'rääkima', *believe* 'uskuma' jms.). Argamon jt (2003) leidsid, et naiste tekstides on personaalpronoomenite osakaal suurem, mis muudab nende keelekasutuse isiklikumaks.

Naiste ja meeste keelekasutuse erinevusi uurides käsitletakse sageli partikleid⁵ eraldi. Esiteks tuleb partiklid grupeerida vastavalt nende funktsioonile kõnes ja seejärel vaadata üksikute rühmade kasutusmustreid. Näiteks Stückler (2022) uuris saksa keele modaalspartikleid ning käsitles eelkõige seda, kuidas vanus ja sugu mõjutavad modaalspartiklite kasutamist. Oma analüüsi tulemusena jõudis ta järeldusele, et seos soo

⁵ Partiklid on (mitmesõnalised) üksused, mille peamised funktsioonid on tagasiside väljendamine, hoiakute-hinnangute väljendamine ja dialoogi kõnevoorude suhetele osutamine. Partiklid täidavad olulist rolli suulisel suhtluses, aidates säilitada selle loomulikkust ja voolavust. Seetõttu esinevad nad kõnes sageli. (EKG II 1993; Hennoste jt 2022: 3)

ja modaalpartiklite kasutamise vahel on olemas. Ta leidis, et naised kasutavad sotsiaalse distantssi kontekstis veidi rohkem modaalpartikleid (Stückler 2022: 135), samas kui mehed kasutasid neid rohkem sotsiaalse läheduse kontekstis. Seda seost tuleks siiski veel uurida, sest meeste ja naiste rühmad ei olnud vanuselt võrdsed. Naised olid keskmiselt nooremad ja võisid tajuda konteksti sotsiaalset lähedust meestest erinevalt. (Stückler 2022) Teiseks on mõistlik uurida iga partikli kasutamist eraldi. Näiteks venekeelsete partiklite *tipa* ja *vrode*, millel on sama tähendus 'nagu', võrdluses leiti, et meeste ja naiste seas ei ole partikli *vrode* kasutamises olulisi erinevusi (Kolyaseva 2022: 188). Samal ajal oli vene keele grammatika normide poolt stigmatiseeritud partikkel *tipa* naiste kõnes oluliselt alaesindatud (Kolyaseva 2022: 184). Samuti on pragmaatiliste markerite jaotuse uurimine vene igapäeva kõnes näidanud, et mehed ja naised eelistavad erinevaid partikleid. *Tak* 'no nii' ja *korotše* 'no nii' esines sagedamini meeste kõnes, samas kui *tam* 'siis' ja *kak bõ* 'nagu' on pigem naiste keelekasutusele omased (Bogdanova-Beglarian jt 2019: 441).

On keeli, milles partiklid viitavad algselt soo tunnustele. Näiteks jaapani keele lõpupartikleid *zo* ja *ze* 'sa tead' kasutavad ainult mehed. Vastavalt sellele seostatakse ühiskonnas keelekasutust, milles need partiklid esinevad, meeste sooliste normidega (Konstantinovskaia 2020: 175). Tänapäeval kasutavad jaapanlannad oma kõnes partikleid *zo* ja *ze* sihipäraselt, et nende kõne kõlaks enesekindlamalt, selgelt väljendatud otsustega viia läbi erinevaid tegevusi (Konstantinovskaia 2020: 183).

Intensiivistajad nagu partiklid on seotud suurema emotsionaalsusega ja seega ka naiste keelekasutusega. Seda kinnitavad ka statistilised uurimused. Ito ja Tagliamonte (2003: 264-276) järelavad, et sugu mõjutab ingliskeelsete intensiivistajate kasutamist. Keskmise vanuserühma (umbes 35-66-aastased) naised kasutasid intensiivistajat *really* 'tõesti' sagedamini kui sama vanuserühma mehed. Teistes vanuserühmades (alla 35-aastased ja üle 66-aastased) ei leitud aga seost intensiivistajate kasutamise ja kõneleja soo vahel. Stratton ja Sundquist (2022: 395-403) leidsid norra keele intensiivistajate uurimisel, et naised kasutavad kõikides vanusekategoriates sagedamini intensiivistajaid (nt *veldig* 'väga', *litt* 'natuke', *helt* 'täiesti') kui mehed. Analüüsi tulemuste kohaselt on aga sugu kõigi arvesse võetud tegurite (omadussõna semantiline kategooria ja süntaktiline positsioon) seas intensiivistaja kasutust kõige vähem mõjutanud (Stratton & Sundquist : 403-408).

Soolingvistika uurimisel on varasemates töödes sageli järelatud, et teatud emotsionaalsuse markerid, nagu emotsionaalse sisuga omadussõnad, partiklid või intensiivistajad, esinevad naiste kõnes sagedamini kui meeste omas. Kuid suuremahuliste keelekorpusete põhjal tehtud analüüsid (nt Baker 2014) näitavad, et need erinevused on sageli väga väikesed ning nende tõlgendamine sõltub suuresti uuritavast suhtlussituatsioonist ja andmestiku laadist. Seetõttu on oluline vältida liigseid üldistusi ning analüüsida keelelisi tunnuseid täpsemalt, arvestades muu hulgas ka suhtlusolukorra tüüpi (nt kas tegemist on formaalse või argise suhtlusega, kahe era- või avaliku isiku vahel jms) ning seda, kas uuritakse suulist või kirjalikku keelt.

2.2.2. Viisakus

Brown ja Levinson (1978) määratlevad viisakust mõiste „nägu“ (*face*) kaudu. See termin on võetud igapäevasest keelekasutusest sellistes fraasides nagu *to lose face*, st muutuda teiste poolt vähem austatud inimeseks. Nad toovad välja kaks põhilist soovi, mida me püüame suhtlemisel rahuldada. Esimene on soov mitte olla koormatud (negatiivne nägu). Inimestele ei meeldi, kui neile peale surutakse, seega pehmendatakse palveid sageli vabanduste (nt *Vabandust, kas see koht on vaba?*) ja viisakust väljendava sõnaga *palun*. Teine on soov olla tunnustatud ja heaks kiidetud (positiivne nägu). Et säilitada vestluspartneri positiivne nägu, ütleme komplimente, kiidame, pehmendame kriitikat (nt *üldiselt on hea, kuid on paar kommentaari*). Seega on viisakus lugupidamine teise inimese soovide suhtes, mida väljendatakse keeles erinevalt. Brown ja Levinson toovad näite, et fraasid nagu *I think* ‘ma arvan’, *I suppose* ‘ma oletan’, *I believe* ‘ma usun’ toimivad positiivse ja negatiivse viisakuse strateegiatena sõltuvalt kontekstist. Positiivse viisakuse korral väljendavad kaasatust ja soovi arvestada kuulajaga, negatiivse viisakuse puhul aga aitavad need pehmendada väidet ja vältida otsest pealesurumust, võimaldades distantsi ja adressaadi autonoomia austamist. Namaziandost ja Shafiee (2018) võrdlesid 40 iraani üliõpilast (20 meest ja 20 naist) akadeemilises suulises kõnes ja loendasid fraaside *I think*, *I mean*, *as I probably know* jt sagedust ning jõudsid järeldusele, et esiteks kasutavad naised neid sagedamini ja teiseks suurema mitmekesisusega. Mehed kasutavad neid fraase pauside sisustamiseks kõnes ja naised oma arvamuse viisakamaks väljendamiseks. (Namaziandost & Shafiee : 69-78)

Ka viisakas palve võib olla väljendatud tingiva kõneviisi abil (Ackermann 2023). Palved, mis on väljendatud käskiva kõneviisi abil, tajutakse karmimana (Czerwionka jt 2023). Madridi üliõpilaste seas läbi viidud uurimust selgus, et mehed kasutavad oma palvetes märkimisväärselt imperatiivseid (käskiva kõneviisi) vorme, st otseseid ja kõige vähem viisakaid, samas kui naised eelistavad pehmemaid vorme (tingiv kõneviis). Lisaks kasutasid mehed madala võimsuse ja madala sotsiaalse läheduse olukorras imperatiivi 66,7% ja naised ainult 47,3% juhtudest. See näitab, et mehed kalduvad valima otsesemaid kõneforme (käskiv kõneviis), samas kui naised kasutavad pigem distantseeritumaid või viisakamaid vorme (tingiv kõneviis). (Czerwionka jt 2023: 14). Samuti on leitud, et naiste keelekasutus on kaudsem ja viisakam, kuna naised eelistavad vältida otseseid keeldumisi ja vastasseisu. Siregar jt (2023: 4655) viisid läbi uurimuse Indoneesia turumüüjate seas, mille eesmärk oli analüüsida meeste ja naiste keeldumisstrateegiaid igapäevases suhtluses. Selgus, et mehed olid altimad keelduma otseselt, kasutades selgelt sõnastatud eituspartiklit ilma pehmemusteta, samas kui naised eelistasid pehmemaid keeldumisvorme, näiteks vabandusi või selgitusi.

Mõnes keeles, näiteks vene ja saksa keeles, võib viisakust väljendada deminutiivsete järelliidete abil (Dressler & Barbaresi 1994: 193, 222, 232, 242, 267). Näiteks kui venekeelne palve sisaldab deminutiivseid pöördumisvorme (*sõnok* 'poeg, pojake', *Sašen'ka, družok* 'sõber, sõbrake'), tajutakse neid viisakatena, isegi kui need ei sisalda sõnu *palun* või *vabandust* (Dementieva 2021: 43-45). Senkina jt (2018: 7, 13) viisid läbi uurimuse, milles analüüsisid, kuidas vanemad ja lapsed üksteisele pöörduvad. Tulemuste järgi kasutasid tüdrukud sagedamini deminutiive (nt *mamočka, mamulja* 'ema'), samas kui poisid kasutasid lühendeid (nt *ma* 'ema'), mitteametlikku *batja* 'isa' ja formaalset *otec* 'isa'.

2.2.3. Keelenormid ja vandesõnad

Viisakuse mõistega on tihedalt seotud keelenormi mõiste. Labovi (1973) järgi on keelenormid ühiskonna sotsiaalsete hoiakute kogum, mis reguleerib, milliseid keelelisi valikuid peetakse antud sotsiaalses olukorras vastuvõetavaks ja korrektseks.

Arvatakse, et mehed järgivad üldiselt vähem keelenorme kui naised. Siin keskendun vandesõnade kui keelenormide rikkumise kasutamisele, et näidata, kuidas need erinevused sooperspektiivis avalduvad.

Kõik vandesõnad on ametlikes-avalikes suhtlusolukordades seotud keelelise tabuga ja mõningaid peetakse ebaviisakaks ka privaatses suhtluses, kuid sellest hoolimata on vandesõnad meeste keelekasutuses kõigis registrites⁶ sagedasemad kui naiste keelekasutuses (Lakoff 1973; Zemskaja 1993).

Idee eristada meeste ja naiste vandesõnu⁷ on olemas juba ammu (Tucker 1961: 86). Lakoff (1973) kirjeldas “naiste keelt” järgmiste näidete abil:

- **Oh dear!** *You've put the peanut butter in the refrigerator again.* ‘Oh issand! Sa panid jälle maapähklivõi külmkappi.’
- **Shit!!** *You've put the peanut butter in the refrigerator again.* ‘Kurat! Sa panid jälle maapähklivõi külmkappi.’

Ta väidab, et esimene lause esineb tõenäolisemalt naise keelekasutuses ja teine meeste keelekasutuses. Mehed kasutavad sagedamini vandesõnu, samas kui naised kasutavad rohkem sotsiaalselt aktsepteeritud väljendeid või asendavad tabu seotud vandesõnad eufemismidega (Lakoff 1973: 55). Zemskaja jt (1993: 122, 131) jõuab samadele järeldustele, uurides venekeelsete meeste ja naiste keelekasutuste erinevust. Ka venekeelsetele meestele on iseloomulik sagedamini kasutada vandesõnu, samas kui naised kasutavad selle asemel sagedamini eufemisme. Näiteks kui sõna *hren'* ‘pask’ võiks esineda pigem meeste keelekasutuses, siis naiste keelekasutuses võib see esineda kujul *hrenovinka* (sama tähendusega, aga deminutiivse järelliitega).

Kvantitatiivsed uurimused on näidanud, et tänapäeva inglise keeles on selline suundumus endiselt olemas. Gauthier (2017: 135-138) järgi avaldavad mehed Twitteris veidi rohkem vandesõnu sisaldavaid postitusi kui naised. Vanusega see erinevus suurenes. Noorte (12-18-aastaste) seas on erinevus minimaalne: 9,4% meeste

⁶ Register on keeleliste vahendite ja normide kogum, mille kõnelejad valivad sõltuvalt konkreetsest situatsiooni- ja suhtluskeskkonnast (Pajusalu 2000: 16-17).

⁷ Vandesõna on lekseem või sõnaühend, mis viisakas või neutraalses vestluses on võimeline tekitama adressaadis solvangut või negatiivset reaktsiooni. Sõna vandumine ei sõltu ainult selle semantikast (mis on tavaliselt seotud tabuteemadega, nagu seks, usund, rahvuse ja sooga seotud stereotüübid jne), vaid eelkõige lausungi pragmaatilisest funktsioonist ja sotsiokultuurilisest kontekstist (suhtlemise aeg ja koht, kõneleja ja kuulaja staatus, lausungi eesmärk jne). (Gauthier 2017: 14-16) Minu töös on vandesõna defineeritud kui lekseemi, mis teatud kontekstides väljendab tugevaid emotsioone (nt viha, ärritus, alandamine või solidaarsus).

säutsudest ja 8,4% naiste postitustest sisaldab vandesõnu,äiskasvanutel (31-45-aastased) oli see 6,2% meestel ja 2,8% naistel. Samuti on täheldatud, et sõnu *fuck*, *shit*, *cock*, *pussy* jms on meeste postitustes sagedamini, aga naiste postitustes bitch ja lühendid *fuck* (nt *fk*, *fck* jms) (Gauthier 2017: 120-121).

Vene keeles kuulub vandesõnade jagunemine meeste ja naiste vandesõnadeks pigem keeleliste stereotüüpide kategooriasse. Konstantinovskaja (2020: 189) uurimus näitas, et tänapäeva vene keeles kasutavad naised samu vandesõnu (nt *tšert* 'kurat', *blja* 'lits, libu' jne) kui mehed, ja sama sagedusega. Ainult sõna *blin* 'kurat' mis erinevalt teistest on kõige vähem tabuline, esineb statistiliselt sagedamini naiste keelekasutuses. Tänu vandesõnade kasutamisele tunnete ja emotsioonide väljendamise vahendina koos teiste keeleliste vahenditega (deminutiivide, pehmendajatega jne) tundub naiste kõne rikkalikum ja hõlbustab paremini suhtlemist, eriti isikutevahelist suhtlus.

Teiste keelte kontekstis võib olla olukord vandesõnade kasutamisega teistsugune. See sõltub vandesõnade funktsioonist ja ühiskonna sotsiaalsetest normidest, kus keelt kasutatakse.

2.3. Meeste ja naiste keelekasutuse erinevused eesti keeles

Praeguseks ei ole eraldi uurimust, mis oleks pühendatud konkreetsetelt pragmaatilistele või morfosüntaktilistele erinevustele eesti keelt kõnelevate meeste ja naiste kõnes, kuid mitmed eesti keele varieerumist käsitlevad tööd käsitlevad sugu kui olulist keele ja kõne varieerumist mõjutavat tegurit. Need tööd on eelkõige keskendunud häälalusnähtustele: näiteks Aare ja Lippus (2017) tuvastasid meeste ja naiste vahel erinevusi nii põhitooni, kui ka häälelaadi ja kõnetempo kasutamises. Näidati, et mehed kasutavad sagedamini madalamat põhitooni ja kärisevat häält (ingl *creaky voice*), samuti võtavad nad vestluses rohkem kõneaega enda alla. See võib olla seotud soorollide mõjuga. Eriti võivad madal hääli ja kärin olla maskuliinsuse markerid, ning see, et mehed võtavad rohkem kõneaega, võib viidata domineerivamale rollile vestluses.

Väikest soo mõju eesti keele kõne tempole märgiti artiklis „Kõnetempo ja -soravuse varieerumine eesti keeles“ (Lippus jt 2024). Autorid jõudsid järeldusele, et sugu mõjutab kõnetempot eesti keeles minimaalselt ja on ebajärjekindel. Tulemuste järgi räägivad tüdrukud veidi kiiremini, kuid see erinevus ühtlustub, kui võtta arvesse muid tegureid.

Täiskasvanud osalejatel ei mõjutanud sugu kõnetempot üldse, kõnetempo langes vanusega võrdselt nii meestel kui naistel. Seega ei toeta eesti andmed teiste keelte (eelkõige inglise keele) põhjal tehtud järeldusi, et mehed räägivad kiiremini. Autorid omistavad selle erinevuse kultuurilistele eripäradele.

Eestikeelsete uurimuste tulemused, milles sugu on üks mõjutavatest teguritest, kattuvad teiste keelte puhul leitud üldiste muustritega. Ka eesti keeles järgivad naised pigem keelenorme. Näiteks foneemi /h/ korrektne hääldamine tähendab selle selget artikulatsiooni sõna alguse positsioonides ja vokaalide vahel (EKK 2007). Teras (2019) järgi järgivad naised neid suuniseid rohkem. Ametlikus kontekstis säilitavad nad foneemi /h/ 86% juhtudest, samas kui mehed säilitavad selle 80% juhtudest Mitteametlikus kontekstis langeb foneemi /h/ osakaal naistel 52%-ni, meestel 66%-ni. (Teras 2019: 217) Naiste suurem erinevus näitab, et naised tunnetavad registri muutust paremini.

Ka kirjutamises järgivad naised rohkem kirjakeele norme. Näiteks jõuab Pedaja-Ansen (2015) artiklis, mis käsitleb võrdlusvormide kasutamist tänapäeva eesti keeles, järelduseni, et naised järgivad sagedamini kirjakeele norme, mis on kooskõlas inglise keelel põhinevate uurimuste tulemustega.

Oma magistritöös uuris Aasa (2024) eesti noorte keeles esinevate intensiivistajate kasutamise eripärasid ning arutles võimalike sooliste erinevuste üle. Vanema korpuse (SEKK, 1996-2004) materjaliga tulemused ei näidanud sooga seotud erinevusi intensiivistajate kasutamise sageduses. Siiski erinesid meeste ja naiste leksikaalsed eelistused. Naised kasutasid peamiselt intensiivistajat *nii*, mehed *väga*. Statistiline analüüs, mis viidi läbi uuema korpuse põhjal (TeKE, 2020-2021), näitas, et tüdrukud kasutavad intensiivistajaid sagedamini, kuigi mõju suurus oli nõrk (Cramer's V = 0,059). (Aasa 2024: 54-65)

Eesti keeles on üks viisakuse väljendusviis sinatada või teietada (Pajusalu 2010). Pajusalu jt (2010) korraldasid küsitluse vene- ja eestikeelsete üliõpilaste seas, et uurida, millised tegurid mõjutavad pronoomeni *sina* või *teie* valimist teise isiku poole pöördumisel. Väike sooga seotud erinevus ilmnis olukordades, kus klassis olev õpilane peab pöörduma tundmatu klassikaaslase poole. Tulemused näitavad, et mehed olid rohkem valmis minema otse sinatama kui naised (meeste keskmine hinne (2,4-2,6) oli veidi kõrgem kui naistel (2,1-2,3) (Pajusalu jt 2010: 211). Üldiselt näitas analüüs siiski, et sugu on vaid täiendav tegur

teiste kontekstiliste tegurite (vanus, tutvusaste jne) hulgas ning ei määra iseenesest valikut, vaid võib ainult rõhutada teatud nüansse (Pajusalu jt 2010: 219).

Kuigi eesti keele kohta puuduvad suuremahulised uurimused, mis oleksid suunatud konkreetselt meeste ja naiste kõne erinevuste uurimisele, võimaldavad eri keeleteaduse valdkondade kogutud andmed rääkida teatud sooga seotud suundumuste olemasolust. Eesti keelt käsitlevate uurimuste autorite tähelepanekud on suures osas kooskõlas teistes keeltes läbi viidud sooliste keeleuurimuste tulemustega. See näitab, et on olemas universaalsed sotsiolingvistilised reeglipärasused, mis ilmnevad ka eesti keeles. Käesolevas kontekstis tundub oluline jätkata eesti keelt kõnelevate meeste ja naiste kõne erinevuste sihipärast uurimist, kasutades korpus- ja kvantitatiivseid meetodeid, mis võimaldaks selgitada nende erinevuste olemust ja nende seost sotsiaalsete rollidega. Oma töös jätkan meeste ja naiste kõne erinevuste uurimist, keskendudes grammatilistele ja pragmaatilistele tunnustele.

3. Taskuhäälingud kui suulise keele erivorm

Siinses töös kasutan materjalina taskuhäälingute korpust, mida kirjeldan üksikasjalikult 4. peatükis. Taskuhäälingud on audioformaadid, enamasti monoloogide või vestluste kujul, ning nende sisu analüüsimiseks tuleb see teisendada kirjalikku vormi. Seetõttu on oluline selgitada, kuidas suuline keel erineb kirjalikust keelest, miks sobivad taskuhäälingud selle uurimiseks ja kuidas automaatne kõnetuvastus selles protsessis aitab. Esmalt kirjeldan alapeatükis 4.1. suulise kõne põhijooni, nagu protsessuaalsus, lineaarsus ja situatiivsus, samuti suulise ja kirjaliku kõne erinevusi ning argivestluse ja institutsionaalsete vestluste vahelist erinevust. Alapeatükis 4.2. käsitlen taskuhäälinguid kui digitaalse sekundaarse suulisuse näidet, st spontaanse, kuid vahendatud kommunikatsiooni uut vormi, mis ühendab endas elava kõne ja meedia tootmise elemente. Lõpuks kirjeldan alapeatükis 4.3. automaatset kõnetuvastussüsteemi, mis võimaldab suure hulga eestikeelse kõnematerjali tõhusat transkribeerimist ja uurimist.

3.1. Suulise keele erijooned

Suuline keel erineb oluliselt kirjalikust keelest. Seda iseloomustavad mitmed erijooned, mis tulenevad elava suhtluse tingimustest ja protsessist. Nende hulka kuuluvad protsessuaalsus, lühiajalisus, lineaarsus, väljaõeldu kustutamatus ning seotus kõnelemise tempoga ja suhtlejate mälupeirangutega. (EG 2023: 983, 987; Hennoste 2000: 49)

Protsessuaalsus tuleneb sellest, et suuline keel kujuneb vahetult suhtlusprotsessis. Kõnelejal ei ole alati aega fraase ette välja mõelda ega neid toimetada. Lühiajalisus tähendab, et lausung ei ole kuskil täielikult olemas. Sõnad kaovad kohe pärast lausumist ja jäävad ainult kõneleja ja kuulaja mällu. Lineaarsus tähendab, et teave edastatakse järjestikku ja reaalselt. Kuulaja tajub sõnumit selle esitamisel, kuid tal puudub võimalus minna tagasi algusesse. See seab piirangud sõnastamise kiirusele ja lausungite pikkusele. Korruga saab mällu talletada vaid piiratud hulga teavet, nii et suulised lausungid on tavaliselt lühemad ja lihtsama süntaksiga. Õeldut ei saa kustutada, seetõttu muutuvad vead ja keelevääratused lausungi osaks ja neid saab parandada ainult enesekorrigeerimise või täiendavate selgitustega. (EG 2023: 987-988; Hennoste 2000: 54) Suulise keele süntaksi tunnuseks on ka lausepiiride hägusus, sissejuhatavate konstruktsioonide paljusus ja vähem mitmekesine sõnavara ehk korduste kasutamine sünonüümide asemel (EG 2023: 1016-1018). Lausungi lõppu tähistatakse tavaliselt intonatsiooni muutusega (Hennoste 2000: 53).

Suuline keel on valdavalt dialoog, mis põhineb vahelduvatel kõnevoorudel, kus kõnelejad aktiivselt omavahel suhtlevad. See ilmneb küsimuste ja reaktsioonide (*jah?*, *mhmh*), emotsionaalsete vastureaktsioonide ja pragmaatiliste partiklite (*noh*, *ei=no*, *ododot*) kasutamises, mis toetavad vestlust. Partiklitel nagu *noh*, *vot*, *eks*, *vä*, *mhmh* on oluline kommunikatiivne roll. Need aitavad kõnelejal struktureerida kõnet, täita pause ja selgitada öeldut. Suulises keeles on levinud ka lühendid ja ebastandardised sõnavormid, nagu *sis*, *vä*, *nüid*, *sin*, *nimodi*, *vel*, *kule*, *ota*, *vata*, mis lihtsustavad hääldamist ja kiirendavad suhtlemist. (EG 2023: 989-990; Hennoste 2000: 50-54)

Veel üks oluline erinevus suulise ja kirjaliku keele vahel on multimodaalsus. Mõtete väljendamiseks kasutatakse mitte ainult sõnu, vaid ka intonatsiooni, žeste, kehahoiakuid, silmaliigutusi ja näoilmeid, mille valik sõltub vestluse teemast ja väliskontekstist. (EG 2023: 988)

Suuline vestlus jaguneb argivestluseks ja institutsionaalseteks vestlusteks (EG 2023: 1177). Argivestlus kasutatakse mitteametlikes olukordades, näiteks sõprade, perekonna ja tuttavatega suhtlemisel. Seda iseloomustab suurem spontaansus ja emotsionaalsus. Selles kasutatakse sageli lühendatud vorme, suhtluspartikleid ja reaktsioone ning vähem kirjakeele normide järgimist, mis muudab selle paindlikuks ja situatsiooniga kohandatavaks. (EG 2023: 1177-1178, 1180) Institutsionaalne vestlus esineb seevastu ametlikes ja formaalsetes olukordades - koosolekutel, kohtumenetlustes, intervjuudes, loengutes või telesaadetes. See on lähemal kirjaliku keelele ja allub kirjakeele normidele, väldib žargooni ja seda iseloomustavad rohkem kontrollitud süntaktilised struktuurid, täielikud laused ja murde- või kõnekeele vormide piiratud kasutamine. Ka ametlikus vestluses on rohkem planeerimist, isegi kui see on vormilt spontaanne. (EG 2023: 1179-1180)

3.2. Taskuhäälingud

Kiiresti areneva digitaalse meediakultuuri kontekstis on taskuhäälingutest (ingl. *podcast*) saanud üks populaarsemaid massikommunikatsiooni formaate. Kaasaegsete määratluste kohaselt on taskuhäälingud episoodiline, enamasti kõnepõhine audiomaterjal, mida saab alla laadida ja voogedastada interneti kaudu. Erinevalt traditsioonilisest raadioringhäälingust võimaldavad taskuhäälingud igatüüpi saada meediatootjaks ja hakata levitama teavet ja ideid ülemaailmsele kuulajaskonnale. (Rime jt 2022: 1270)

Taskuhäälingud hõlmavad mitmeid žanre, näiteks intervjuud, uudised, loengud, isiklikud päevikud ja eriilmelised lood. Žanrite mitmekesisus võimaldab igal kuulajal leida sisu, mis vastab tema huvidele, harjumustele ja emotsionaalsetele vajadustele. (Rime jt 2022: 1274)

Taskuhäälingu edu sõltub aga mitte žanrist, vaid autori ja publiku vahelise suhtluse iseloomust. Sama taskuhäälingu regulaarne kuulamine soodustab parasotsiaalsete suhete tekkimist, st kuulaja hakkab saatejuhti tajuma kui inimest, kellega tal on emotsionaalne side (Mayer jt 2024: 3-4). Elava vestluse formaadi säilitamine taskuhäälingu loomisel aitab seda sidet tugevdada. Robert MacDougall (2009: 2-3) väidab, et taskuhäälingud on sekundaarse suulisuse (ingl. *secondary orality*)⁸ otsene ilming, sest nad edastavad digitaalselt elavat keelt, luues näost-näkkude vestluse, intiimsuse ja isikliku kaasatuse. Taskuhäälingut ei tajuta mitte infovooluna, vaid vestlusena, millesse kuulaja on omamoodi kaasatud. Selline keeleline lähedus tekitab kuulajates püsiva soovi taskuhäälingu juurde tagasi pöörduda, seega peaksid taskuhäälingud kõlama loomulikult, mitte nagu ettevalmistatud kõne. (Mayer jt 2024)

Taskuhäälingu loomise protsess on mitmetasandiline ja hõlmab mitmeid etappe: planeerimine, salvestamine, järeltootmine ja avaldamine. Planeerimise käigus kujundavad autorid üldise kontseptsiooni, mõtlevad struktuuri üle, määravad kindlaks temaatilise raamistiku, mõnikord kirjutavad stsenaariume või lühikavasid, valivad külalisi ja asukohti. See etapp on oluline, et saada üldine arusaam sellest, kuidas valmis meediatoode eetris kõlab, kuid see hõlmab pigem stsenaariumi vaid üldjoontes loomist kui tekstide jääka fikseerimist. Taskuhäälingute salvestamine võib toimuda kõikjal alates professionaalsest stuudiost kuni riidekapini. Dialoog, kõrvalepõiked, naer, vead, pausid, keelevääratused, sekkumised jäetakse tavaliselt lõplikku versiooni, et luua ehtsuse efekt. Järeltöötus hõlmab montaaži, heliefektide lisamist ja heli kvaliteedi parandamist. Avaldamisetapis majutatakse taskuhääling ja tehakse see kuulajatele kättesaadavaks RSS⁹-voo ja platvormide (Spotify,

⁸ Walter Ongi definitsiooni järgi on sekundaarne suulisuus digitaalselt vahendatud tagasipöördumine suulise kultuuri omaduste juurde (MacDougall 2009:3).

⁹ RSS-voog on struktureeritud XML-fail, mis levitab automaatselt digitaalset sisu (näiteks taskuhäälingu) uuendusi. Taskuhäälingute puhul sisaldab RSS-voog metainfot taskuhäälingu kohta (pealkiri, kirjeldus, kaanekujundus, episoodid, lingid helifailidele ja avaldamiskuupäevad) ning võimaldab platvormidel (näiteks Spotify, Apple Podcasts, Google Podcasts) automaatselt uusi episoodide leida ja kuulajatele kuvada. (Rime jt 2022: 1272)

Apple Podcasts jne) kaudu. Oluline on, et taskuhääling oleks hõlpsasti kättesaadav ja et selle elavat dialoogi simuleeriv mõju ei kaoks üheski tootmisetapis. (Rime jt 2022)

Seega võib öelda, et taskuhääling on elava suhtluse vorm, mis on kohandatud digitaalsesse keskkonda. Saatejuhi suhtlus külalistega säilitab suulise keele peamised erijooned (protsessuaalsus, linearsus, väljaõeldu kustutamatus) ja elava suhtluse omadused (spontaansus, dialoogilisus, intonatsiooniline väljendusrikkus), mis võimaldab käsitleda taskuhäälinguid mitte ainult meediaformaadina, vaid ka allikana kaasaegse suulise keele uurimiseks. Taskuhäälingute žanride mitmekesisus omakorda võimaldab uurida nii argivestlustele kui ka avalik-ametlikele vestlustele lähedast keelekasutust.

3.3. Automaatse kõnetuvastuse kasutamine eesti keele analüüsimiseks

Suulist keelt on võimalik süstemaatiliselt analüüsida alles pärast selle transkribeerimist (EG 2023: 984-985). Suurte korpuste puhul on käsitsi transkribeerimine väga aeganõudev ja energiamahukas, sellisel juhul lihtsustab ülesannet automaatne kõnetuvastus (Gorisch & Schmidt 2024).

Automaatne kõnetuvastus on protsess, mille käigus kõnesignaali teisendatakse kõnetöötlusalgoritmide abil tekstimärkide jadaks (Tampel & Karpov 2016: 5). Automaatne kõnetuvastus koosneb mitmest etapist: esmalt salvestatakse ja töödeldakse helisignaali, seejärel analüüsitakse selle sageduslikku koostist, et eraldada iseloomulikud tunnused. Need tunnused võrreldakse akustiliste mudelite abil võimalike kõnehelidega, ning keelemudelid aitavad leida kõige tõenäolisema ja sobivama sõnade järjestuse. Süsteemi lõplik ülesanne on määrata kõige tõenäolisem tekst, mis vastab kõneldud fraasile. (Tampel & Karpov 2016: 36-38, 112-113)

Eesti keele automaatne kõnetuvastussüsteem, mida on kasutatud minu töös kasutatava korpuse (vt peatükk 4) tuvastamiseks, on välja töötatud Aivo Olev ja Tanel Alumäe (2022) poolt. See on mõeldud poolsontaanse kõne töötlemiseks žanrites nagu loengud, intervjuud ja raadiosaated jne. Lisaks kõne teisendamisele tekstiks teostab süsteem ka kõneleja tuvastamist, kirjavahemärkide taastamist ja keele tuvastamist, et vältida moonutusi võõrsõnade töötlemisel. Süsteem põhineb närvivõrkudel, mis on treenitud suure hulga eestikeelse kõne põhjal. Keelemudel aitab vähendada vigade arvu. Erinevate tuvastamise

meetodite kombineerimisega on saavutatud paremad tulemused, kusjuures keskmine viga kasutajate andmetes oli 11,9%.

Kuna automaatne kõne tuvastamine põhineb tõenäosuslikel algoritmidel, sõltub selle täpsus salvestuse kvaliteedist, kõne selgusest ja sisendandmete vastavusest treeningmaterjalile. Eriti suur on vigade oht spontaanse, mitme kõnelejaga või segakeelse kõne analüüsimisel. Automaatse kõnetuvastuse praktilise hinnangu eesti taskuhäälingute korpusel (Lippus jt 2023) annab Len Toots (2023). Ta analüüsis käsitsi kahe taskuhäälingute automaatset kõnetuvastust. Tulemused näitasid kõrget veamäära – 23,7%. Kõik tuvastatud vead olid klassifitseeritud 15 kategooriasse, millest kõige sagedasemad olid järgmised:

- sõna puudumine (22% kõigist vigadest ehk 5,2% kahes taskuhäälingutes),
- valesti tuvastatud kõneleja voores (20% kõigist vigadest ehk 4,7% kahes taskuhäälingutes),
- valesti tuvastatud lause või lauseosa (15% kõigist vigadest ehk 3,6% kahes taskuhäälingutes),
- valesti tuvastatud eestikeelne sõna (13% kõigist vigadest ehk 3% kahes taskuhäälingutes).

Ülejäänud veatüübid (õigekiri, akronüümid, võõrsõnad jne) olid alla 7% kõigist vigadest ehk 1,7% kahes taskuhäälingutes. (Toots 2023: 23-28)

Need vead seavad piiranguid materjali ja meetodi valikul ning tulemuste tõlgendamisel. Seega automaatsete transkriptsioonide kasutamisel analüüsides tuleb neid arvesse võtta ja saadud andmete üle kriitiliselt järele mõelda.

4. Materjal

Uurimismaterjali allikana kasutasin eesti taskuhäälingukorpust (Lippus jt 2023). Korpus koosneb eestikeelsetest taskuhäälingusaadetest ja nende transkriptsioonidest, mis on salvestatud JSON-failidesse. Salvestused on transkribeeritud Tallinna Tehnikaülikooli automaatse kõnetuvastusega ning automaatselt morfanalüüsitud EstNLTK-ga (Orasmaa jt 2016). Kokku sisaldab korpus 10 633 episoodi 184 erinevast taskuhäälingutest kogukestusega 10 918 tundi. Korpuse põhjal koostasın edasiseks analüüsiks oma valimi.

Kuna minu magistritöö eesmärk on uurida, kas ja kuidas erineb eesti keelt kõnelevate meeste ja naiste keelekasutus, valisin analüüsiks taskuhäälingud, kus osalesid ainult mehed või ainult naised. Selle lahendusega vältisin ühtlasi kõneleja soo valesti tuvastamist, kuna üks tüüpiline viga korpuses oli valesti tuvastatud kõneleja voorus. TTaskuhäälingute eraldamiseks kahte (vaid meeste ja vaid naiste kõnet sisaldavasse) alamkorpusesse pidin kindlaks määrama kõnelejate soo. Algselt plaanisin seda teha kõnelejate nimede põhjal, mis on loetletud kas metaandmetes või kõnelejate sõnavõttudes. See lähenemine osutus ebaõnnestunuks, kuna paljudes taskuhäälingutes ei olnud kõnelejaid mainitud ei metaandmetes ega ka tekstis. Nendel juhtudel, kui mul õnnestus nimi leida kas metaandmetest või tekstist, oli suur tõenäosus, et nimi tuvastati valesti. Toots (2023: 25) leidis, et nimetuvastus on üks tüüpilisi vigu eesti keele automaattuvastuses. Lõpuks kontrollisin Pythoni skriptiga, kas kõnelejate nimed on märgitud episoodi transkriptsioonifaili nimes. Sageli olid seal ära toodud külaliste nimed. Seejärel otsisin käsitsi episoodi *podcast.ee* või *podcastid.ee* keskkondadest, kus vaatasin, kas kõikide kõnelejate (nii saatejuhi kui ka külaliste) nimed on toodud välja episoodi kirjelduses. Vajadusel kuulasin sissejuhatavaid kõnekatkeid, et veenduda, kas määrasin sood õigesti. Iga episoodi alguses nimetati saatejuhi kui ka külaliste nimed.

Samuti aitasid episoodide kuulamine ja kirjelduste lugemine mul kontrollida, et vastavas taskuhäälingu episoodis ei oleks kuulajate telefonikõnesid, kelle sugu ma ei saa kindlaks teha.

Keelekasutus varieerub sõltuvalt vestluse teemast. Suhtluse teemad mõjutavad nii kõne struktuurilisi tunnuseid (nt süntaktiliste konstruktsioonide valik) kui ka pragmaatilisi aspekte, näiteks küsimuste sõnastamise viisid või jutustuse juhtimine. (Labov 1972) Selleks et välistada võimalus, et meeste ja naiste keelekasutuse erinevused tulenevad mitte

soost, vaid taskuhäälingute temaatikast, tuli tagada, et mõlemas alamkorpuses oleks esindatud võrdselt erinevaid teemasid käsitlevaid salvestisi.

Tabel 1. Taskuhäälingute teemade jaotus (tundides).

	Mehed	Naised	Kokku	%
Poliitika	6,6	5,2	11,8	15,4%
Kinnisvara	8,1	6,3	14,4	18,8%
Meditiin	4,1	3,2	7,3	9,6%
Muusika ja Kino	9,6	7,5	17,1	22,3%
Igapäevaelu	11,1	8,8	19,9	25,9%
Esoteerika	1,7	1,3	3	4%
Sport	1,7	1,3	3	4%

Liigitasin taskuhäälingud käsitsi kaheksasse kategooriasse episoodi pealkirjade ja kirjelduste põhjal podcast.ee või podcastid.ee. Tuli arvesse võtta iga episoodi kestust. Episoodid, kus rääkisid ainult mehed, olid sageli keskmiselt pool tundi pikemad kui need, kus rääkisid ainult naised. Seetõttu oli võrdluseks oluline, et kui näiteks 10% meeste kõnest oli pühendatud poliitikale, siis sama osakaal – 10% – oleks poliitikaemadel ka naiste alamkorpuses. Tabelis 1 on esitatud teemade jaotus taskuhäälingute kaupa, väljendatuna kõne tundides. Esimesed kolm rühma esindavad eeldatavalt ametlikumat keelekasutust, ülejäänud informaalsemat keelekasutust. Kõik väärtused on ümardatud ühe kümnendkohani.

Ma eeldasin, et ühe kõnelejaga episoodides on kõne tõenäolisemalt ette valmistatud ja seal puuduvad dialoogijooned, nagu nõustumine, mittenõustumine, täpsustamine või katkestamine. Seetõttu tagasin, et episoodides oleks rohkem kui üks kõneleja.

Lõpuks sain kokku 49 taskuhäälinguepisoodi, kus kõnelejateks olid ainult mehed – see teeb umbes 43 tundi suulist kõnet – ning 71 episoodi, kus kõnelejateks olid ainult naised, mis vastab ligikaudu 33,5 tunnile suulisele kõnele.

5. Meetod

Selles peatükis kirjeldan analüüsi metodoloogiat. Esmalt kirjeldan, milliseid keelelisi tunnuseid analüüsimiseks võtan ning miks just neid võtan. Alapeatükis 5.2. kirjeldan, kuidas korpust automaatselt töötlesin ning kus ja miks oli vaja täiendavat käsitsi kontrollimist. Lõpuks alapeatükis 5.3. kirjeldan andmete analüüsimiseks kasutatavaid statistilisi meetodeid.

5.1. Keelelised tunnused

Suulise kõne eripärad (vt alapeatükk 3.1.) ning selle automaatsel kõne-tekstiks transkribeerimisel esinevad vead (vt peatükk 4) piiravad korpusest automaatselt eraldatavate tunnuste valikut. Näiteks lause struktuuri uurimine sellise korpuse põhjal on keeruline, sest ei ole võimalik määratleda selgeid lause- ja osalause piire. Ka grammatiliselt korrektsete kirjakeele lausetega esineb automaatsete analüsaatorite jaoks sageli väljakutseid, spontaanselt suulisest kõnest rääkimata probleemid tekivad nii pealerääkimise kui üksteisele vahele segamisega, mille tõttu võib süsteem nende teksti segi ajada, ja kui kõneleja kasutab neologisme, anglitsime, russisme ja muid uusi, süsteemile tundmatuid sõnu, siis on suurem tõenäosus, et need sõnad tuvastatakse valesti. Seega olen valinud analüüsiks sellised tunnused, mille esinemise sagedust meeste ja naiste kõnes saab minu uurimuse raames võrrelda, võttes arvesse korpuse eripära.

Tekstidest eraldasin järgmised keelelised tunnused:

- 1) sõnaliigid (nimisõnad, omadussõnad, leksikaalsed verbid, abiverbid¹⁰ (*olema, saama, võima, tohtima, pidama, nägema, paistma, tunduma*), adverbid, pronoomenid, kaassõnad);
- 2) esimese ja teise isiku pronoomenid;
- 3) partiklid;
- 4) omadussõnade, adverbide, verbide intensiivistajad;
- 5) mõningad verbi grammatilised kategooriad (umbisikuline tegumood, tingiv kõneviis, käskiv kõneviis);

¹⁰ Siin ja edasi mõeldakse kõiki verbe, mis universaalsete sõltuvuste (Universal Dependencies) kohaselt on tähistatud aux.

- 6) eitus (eituspartiklite *ei, ära, ärge, ärgem* kasutamine);
- 7) komplementausega predikaatidel põhinevad markerid (siin ja edasi KP-markerid) nagu *ma arvan, et..., ma kardan, et..., tundub, et...* jne;
- 8) deminutiiv *-kene*;
- 9) sõnavara mitmekesisus (ingl *type-token ratio*) (siin ja edasi TTR)¹¹

Keelelisi tunnuseid valisin tuginedes inglise keele jaoks tehtud kirjeldustele meeste ja naiste keelekasutuse kohta ning teiste keelte kohta tehtud uurimustele. Erinevate sõnaliikide (tunnused 1, 3, 4) sagedasem kasutamine on seotud suurema emotsionaalsusega (adjektiivide, partitsiipide ja intensiivsõnade puhul) või vastupidi suurema konkreetsuse ja objektikesksusega (leksikaalsete verbide ja substantiivide puhul) (nt Newman jt 2008, Cremades 2024, Posio jt 2024). Esimese ja teise isiku pronoomenite (2) sagedast kasutamist naiste kõnes seostatakse tavaliselt naiste sooviga luua isiklikumat, usalduslikumat suhtlust. Soovist hoida vestluspartnerist distantsi kasutavad mehed kõnes sagedamini umbisikulist tegumoodi (5). (Posio jt 2024) Mehed väidavad ja vaidlustavad sagedamini vestluspartneri arvamust ning kasutavad seetõttu sagedamini eitust (6) (nt Pilkington 1992). KP-markerid (7) peetakse naiste kõnes üldiselt viisakuse markeriks (nt Brown & Levinson 1978, Coates 2004, Namaziandost & Shafiee 2018) ja see hõlmab ka tingiva kõneviisi (5) kaudu esitatud palveid. Meeste kõne peaks olema otsekohesem, seega peaks olema rohkem palveid käskiva kõneviisiga (5) (nt Czerwionka jt 2023, Kirilina 2005). On ka arvamus, et deminutiivid (8) esinevad naiste kõnes sagedamini ka emotsionaalsuse ja viisakuse markerina (nt Argamon jt 2003, Konstantinovskaia 2020, Senkina jt 2023). TTR (9) on teksti sõnavara mitmekesisuse lihtne kvantitatiivne näitaja (Brezina 2018: 58). Meeste kirjalike venekeelsete tekstide puhul leiti kõrge TTR (nt Litvinova jt 2017), see võib olla seotud ka sellega, et meeste kõne sisaldab rohkem harvaesinevat ja erialast sõnavara, samas kui naiste kõne on üldiselt lihtsam ja kuulajale arusaadavam (nt Kirilina 2005).

Soolingvistilistes uurimustes (nt Argamon jt 2003; Czerwionka jt 2023; Litvinova jt 2017; Newman jt 2008; Posio jt 2024) on juba üksikasjalikult kirjeldatud sõnaliike (nt nimisõnad, omadussõnad, verbid, mäarsõnad, asesõnad ja kaassõnad) ning verbi grammatilisi kategooriaid, nagu umbiskuline tegumood, tingiv ja käskiv kõneviis. Need kategooriad on

¹¹ TTR on koefitsient, mis mõõdab teksti sõnavara mitmekesisust kui unikaalsete sõnade arvu ja sõnade koguarvu suhet (Brezina 2018: 57-59).

osutunud sobivaks ka eesti keele analüüsil, mistõttu ei ole neid käesolevas töös eraldi ümber defineeritud ega kohandatud. Ainsaks erandiks oli see, et jagasin tegusõnad abiverbideks ja leksikaalseteks verbideks, kuna mind huvitas just leksikaalsete verbide kasutus, sest nende kaudu väljendatakse otseselt tegevust või seisundit, mida varasemad uuringud on seostanud suurema sagedusega meeste keelekasutuses (Newman jt 2008, Posio jt 2024). Partiklite, intensiivistajate ja muude tunnuste loendid vajasisid aga kohandamist eesti keele grammatikaga.

Eituse all käsitlen ainult üldeitust (lauseeitust), mis moodustatakse eituse partikliga *ei* või eitava abiverbi *ära*, *ärge* ja eitab kogu lauset korraga (EKK 2007: 466), nt *Ma ei tea seda. Ära kõnni murul!*. Erieitust, mis eitab lause konkreetset elementi (EKS 2017: 190), analüüsi ei kaasatud. Kuna on arvatud, et eituse sagedane kasutamine meeste kõnes võib tuleneda sellest, et mehed kipuvad otsesemalt vastu vaidlema ja vaidlustama vestluspartneri väiteid, mis väljendub kogu väite eituses (nt Edelsky 1981, Coates 2004), siis käsitlen oma töö raames ainult üldeitust.

Siinses magistritöös lugesin partikliteks kõik morfanalüsaatori poolt tuvastatud partiklid (nt *no*, *vot*, *jah*, *kurat*), välja arvatud *tere* ja *aitäh*, mille kasutamine taskuhäälingutes võib olla seotud formaadiga. Lisaks lisasin loetellu partiklid, mida analüsaator ei tuvasta, kuid mis võivad funktsioneerida partiklitena, nt *jumal*, *nagu*, *eks ole*, *vaata*, *oota* ja *kuule*. (EKG II 1993: 107, 172; Jürine jt 2016; Hennoste jt 2022)

Morfanalüsaator analüüsib intensiivistajaid enamasti adverbideks. Kuna kõik adverbid ei ole intensiivistajad, koostasin juba olemasolevate uurimuste põhjal (Petron 2023; Aasa 2024) loendi konkreetsetest lemmadest, mille kasutamine koos omadussõnade, adverbide või verbidega muudaks need intensiivistajateks (nt *absoluutselt*, *väga*, *eriti* jms). Täielik intensiivistajate loend on leitav GitHubist¹². Intensiivistajate kasutamine erineb funktsionaalselt sõltuvalt sellest, millise sõnaliigi kõrval nad esinevad (Aasa 2024), seega eraldan siinses töös omadussõnade, verbide ja adverbide intensiivistajad.

Komplementlausega predikaadil põhinevad markerid (KP-markerid) on konstruktsioonid, mille keskmes on verb (nt *ma arvan*, *et*; *ma ütlen*, *et* jms). Sellised markerid pehmendavad väiteid, markeerivad eelkõnelejast erinevat seisukohta ja kaasavad teda enam suhtlusesse,

¹² <https://github.com/SofiaKriuchkova/ma-diploma-project-kriuchkova/blob/main/intensiivistajad.txt>

avardades nii dialoogilist ruumi (Hennoste jt 2022: 130). KP-markerite piiramiseks koostasin bigrammide ja trigrammide loendid ja valisin sealt käsitsi välja kõige sagedasemad markerid (üle 50 esinemist korpuses). Bigrammide ja trigrammide loendid on leitavad GitHubist¹³.

Eesti keeles on mitmeid deminutiiviliite, nende hulgas on *-ke* (nt *vennike*, *armsake*), *-kene* (nt *natukene*, *emakene*) ja *-u* (nt *poisu*) (Liivak 2023: 7). Morfanalüsaator ei tuvasta neid automaatselt, nii et ainus võimalus on leida neid *-ke-*, *kene-* ja *-u*-lõpuliste nimisõnade ja omadussõnade hulgast. Probleem on selles, et eesti keeles on sõnu, mis lõpevad lõpuga *-ke* (nt *päike*, *rõske*) või *-u* (nt *kuu*, *tal*, *mõru*) ka ilma järelliideta. Vähem on sõnu, mis lõpevad lõpuga *-kene* (võib-olla ainult *skene*). Kui ma oleksin otsinud kõik *-ke* ja *-u* lõpuga lemmasid, oleks lõplik käsitsi kontrollimine olnud väga aeganõudev, seega kaasasin analüüsi ainult suffiksi *-kene* kui deminutiivi liite.

5.2. Korpuse töötlus

Andmete töötlemiseks kasutasin programmeerimiskeelt Python (versioon 3.12.2) andmete visualiseerimise keskkonnas Jupyter Notebook (versioon 7.2.2).

Kõigepealt koostasin tekstidest kõikide lemmade loendid meeste ja naiste alamkorpuste jaoks. Lemmade loendite põhjal koostasin uuritud KP-markerite piiritlemiseks bigrammide ja trigrammide loendid. Pärast arvutasin TTR-i ka lemmade loendi põhjal. Kuna meeste korpuses oli vähem episoodi, kuid ajalisel rohkem spontaanset kõnet, võis see mõjutada TTR-i tulemusi. Näiteks suurema episoodide arvu tõttu võis naiste korpuses korduda tervitus- ja hüvastijätufraasid, mis oleks alandanud TTR-i. Selle probleemi lahendamiseks arvutasin TTR-i ka sarnase pikkuse ja temaga episoodide põhjal, valides ühe kinnisvarateemalise ja ühe armastusteemalise episoodi nii meeste kui ka naiste alamkorpusest.

Edasi koondasin iga uuritud tunnuse kohta episoodi indeksi, kõneakti¹⁴ numbri, kõne kestuse minutites, lause (et hiljem oleks võimalik kontrollida konteksti) ja sõna lemma.

¹³ https://github.com/SofiaKriuchkova/ma-diploma-project-kriuchkova/tree/main/n_grammid

¹⁴ siin ja edasi kõneakt viitab JSON-failis olevatele section-üksustele, mis tähistavad järjestatud kõnefragmente

Samuti määrasin soo tähise (M/F) vastavalt alamkorpusele. Tulemused salvestasin tabelite kujul CSV-failidesse.

Sõnaliikide eraldamiseks kasutasin morfanalüsaatori automaatset märgendust. Ülejäänud tunnuste eraldamiseks vaatasin järgmiseid tingimusi:

- Eituse puhul valisin ainult eitusepartikleid enne verbi (lemmad *ei* ja *ära*);
- 1. ja 2. isiku pronoomenite puhul valisin ainult lemmad *mina* ja *sina*;
- Tingiva kõneviisi puhul salvestasin kõik verbid, mille vormidena olid märgitud verbi lõpud *-ks*, *-ksin*, *-ksid*, *-ksime*, *-ksite*, *-taks*, *-nuksin*, *-nuksid*, *-nuks*, *-nuksime*, *-nuksite*, *-tuks*;
- Käskiva kõneviisi puhul salvestasin kõik verbid, mille vormidena olid märgitud verbi lõpud *null*, *-ge*, *-gem*;
- Umbisikulise tegumoe puhul salvestasin kõik verbid, mille vormidena olid märgitud verbi lõpud *-takse*, *-ti*, *-ta*, *-tud*, *-taks*, *-tuks*;
- Intensiivistajate loendamiseks kasutasin intensiivistajate loendit, koondatud varasematest eesti keele uurimustest (Petron 2023; Aasa 2024);
- Partiklite puhul kogusin kõik morfanalüsaatori poolt leitud partiklid, v.a. lemma *tere* ja *aitäh* ja lekseemid *nagu*, *jumal* ja *eks* (viimase puhul ainult siis, kui sellele järgnes verbi *olema* vorm, nt *eks ole*);
- KP-markerite loendamiseks kasutasin *n*-grammide abil leitud KP-markereid ((*ma*) *arvan*, (*ma*) *usun*, (*ma*) *kardan*, (*mulle*) *tundub*, *paistab*, *näib*, (*ma*) *loodan*, (*ma*) *ütle(n/ks)*);
- Eraldi vaatlesin verbe *ootama*, *vaatama*, *kuulama*, *kuulma*, sest morfoloogiline analüsaator märgendab need ühtemoodi nii käskiva kõneviisi verbi (nt *et kui taha üle saada tema neid asju, siis vaata, kellega sa suudled, piltlikult öeldes...*) kui ka kõnekeelse partiklina (nt *...seda noh, vaata, ongi need küsimused, et mis see siis nagu see point on, et igal inimesel tõesti on...*).
- Tegin eraldi tabelid partiklite, intensiivistajate ja KP-markerite jaoks, mis esinevad korpuses rohkem kui 50 korda. See võimaldas edaspidi kindlaks teha, milliseid partikleid, intensiivistajaid ja KP-markereid naised või mehed sagedamini kasutavad.

Kõik tabelid kontrollisin käsitsi üle.

Seejärel arvutasin iga tunnuse jaoks selle absoluutse ja normaliseeritud sageduse (sagedus = (absoluutne sagedus / kogu sõnade arv) × 1000) kohta ühes kõneaktis ning lõin uued tabelid.

Kõik kokkuvõtlikud tabelid on ühtses formaadis:

episode_id	speech_act_index	gender	minutes	absolut_freq	freq_per_1000_tokens

Tabelid salvestasin failidesse, mis nimetasin uuritava tunnuse või lemma järgi. Tabelid on kättesaadavad Githubis¹⁵.

Ka andmete kättesaamiseks, töötlemiseks ja analüüsimiseks kasutatud Pythoni kood on leitav Githubi lehelt <https://github.com/SofiaKriuchkova/ma-diploma-project-kriuchkova>

5.3. Statistilised meetodid

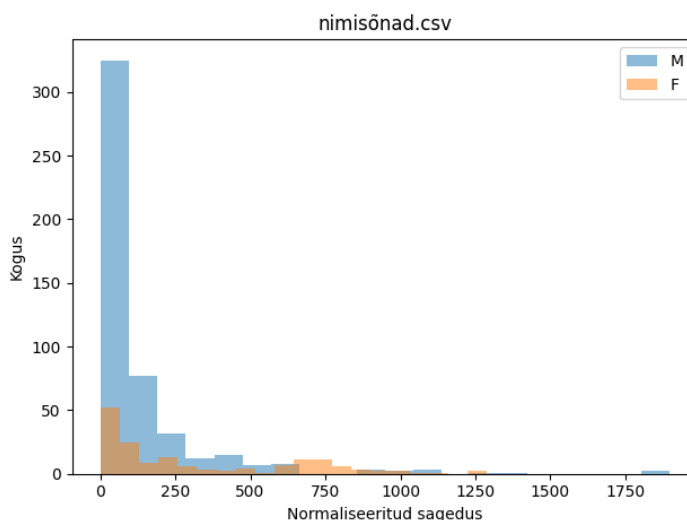
Sooga seotud erinevuste testimiseks võrdlesin iga vaadeldud kategooria normaliseeritud sagedusi meeste ja naiste seksioonide vahel, normaliseerimise baasiks oli 1000 sõned. Viisin läbi testid mitme tunnuste jaoks korraga, mis olid grupeeritud järgmiselt:

- 1) sõnaliigid (nimisõnad, omadussõnad, leksikaalsed verbid, abiverbid, adverbid, pronoomenid, kaassõnad);
- 2) keelelised tunnused (intensiivistajad verbidega, intensiivistajad omadussõnadega, intensiivistajad adverbidega, eitus, deminutiiv, esimese ja teise isiku pronoomenid, partiklis, umbisikuline tegumood, tingiv kõneviis, käskiv kõneviis, partiklid *vaata/oota/kuule*, KP-markerid);
- 3) partiklite lemmad;
- 4) intensiivistajate lemmad;
- 5) KP-markerite lemmad.

Käesolevas analüüsis võrreldakse meeste ja naiste kõnes kasutatavate keeleliste tunnuste esinemissagedust. Kahe sõltumatu rühma võrdlemiseks rakendati dispersioonianalüüsi.

¹⁵ <https://github.com/SofiaKriuchkova/ma-diploma-project-kriuchkova/tree/main/sõnaliigid>
<https://github.com/SofiaKriuchkova/ma-diploma-project-kriuchkova/tree/main/tunnused>

Esmalt keelt iga tunnuse kohta igas rühmas xandmete jaotust histogrammide abil. Kõigil juhtudel ilmnis selge kõrvalekalle normaaljaotusest: andmed olid tugeva vasakpoolse kaldega ja väljendunud pikliku sabaga (vt joonis 1). Seetõttu rakendasin Mann-Whitney U-testi, mis sobib kahe rühma võrdlemiseks ka siis, kui andmed ei ole normaaljaotusega (Brezina 2018: 195).



Joonis 1. Tunnuste jaotuse näide. Nimisõnade jaotus korpusel.

Mitme paralleelse testimise korral suureneb tõenäosus saada vähemalt üks valepositiivne tulemus. Selle riski kontrollimiseks korrigeeriti esialgseid p-väärtusi Holm–Bonferroni meetodi abil. Meetod tagab, et vähemalt ühe eksliku järelduse kogutõenäosus tunnuste kogumi kohta ei ületa 5% (Cabin & Mitchell 2000).

Kõik p-väärtused järjestati kasvavas järjekorras. Iga p-väärtust võrreldi vastava künnisega, mis arvutati valemiga $\alpha / (m - k + 1)$, kus α on olulisuse tase (0,05), m on tehtud võrdluste koguarv ja k on järjestuses p-väärtuse järjekorranumber. Kui p-väärtus jäi alla arvutatud künnise, loeti tunnus oluliseks. Esimese künnise ületamise korral analüüs peatati.

Selleks, et näha tunnuse ja soo vahelist tugevust ja suunda, arvutasin astakute biserialsete korrelatsioonide (ingl. *rank biserial correlation*) koefitsiendi, mis arvutatakse valemi abil:

$$r_{rb} = \frac{\bar{R}_{\text{rühm 1}} - \bar{R}_{\text{rühm 2}}}{N},$$

Väärtused jäävad vahemikku -1 ja 1. Mida lähemal 1 või -1, seda suurem on erinevus. Märk näitab suunda: miinus tähendab, et teise rühma keskmine koht on kõrgem. (Brezina 2018: 196-197)

Andmete suure hulga tõttu olid Mann-Whitney U-testiga saadud p-väärtused isegi pärast Holm-Bonferroni korrektsiooni väga madalad ja viitasid peaaegu kõigil juhtudel statistilist olulistele erinevustele. Tulemuste kehtivuse täiendavaks kontrolliks kasutasin Pearsoni χ^2 -testi, mis hindab, kas kahe näitajate – sugu ja keelelise tunnuse esinemine – vahel esineb seos. Iga tunnuse kohta tehakse eraldi test ning iga test põhineb tunnuste koosinemiste sagedusi sisaldava 2x2 risttabeli vaadeldud ja oodatud sageduste võrdlemisel (Pilvik, ilmumas: 141). Selle tulemusena väljastatakse hiiruit-statistik ja selle põhjal arvutatud p-väärtus. Ka χ^2 -testi tulemused korrigeeriti Holm-Bonferroni meetodi abil. Siinses uurimuses määrasin olulisuse tasemeks 0,05. Kui p-väärtus oli pärast korrigeerimist väiksem (nt $p = 0,031$), lugesin erinevuse statistiliselt oluliseks. Kui aga p-väärtus oli suurem kui 0,05 (nt $p = 0,061$), lugesin, et erinevus ei ole statistiliselt oluline.

Kuna χ^2 -test ei kajasta seose tugevust, arvasin lisaks Crameri V seose koefitsiendi iga tunnuse jaoks, et hinnata mõju suurust. See põhineb χ^2 -statistikal ja vaatluste koguarvul. V väärtused jäävad vahemikku 0 ja 1, kus $\approx 0,10$ näitab nõrka seost, $\approx 0,30$ näitab keskmist seost ja $\approx 0,50$ ja rohkem näitab tugevat seost (Brezina 2018: 115).

Seega annab Mann-Whitney U, χ^2 -testi, Holm-Bonferroni korrektsiooni ja kahe sõltumatu mõju suuruse (biseriaalne korrelatsioon ja Crameri V) kombinatsioon usaldusväärse ja tõlgendatava ülevaate soo seotud erinevuste kohta kõigis uuritud keelelises tunnustes.

6. Tulemused ja analüüs

Selles peatükis kirjeldan ja tõlgendan erinevate testide tulemusi. Alapeatükkides 6.1. - 6.7. esitan statistiliste testide tulemused sõnaliikide, keeleliste tunnuste, partiklite, intensiivistajate ja KP-markerite kohta tõlgendan neid ning võrdlen varasemate uurimuste tulemustega. Partiklite, intensiivistajate ja KP-markerite puhul kirjeldan lisaks iga keeleüksuse funktsiooni vestluses ning tõlgendan tulemusi selle põhjal. Alapeatükis 6.6. analüüsin ja tõlgendan TTR-näitajat. Alapeatükis 6.7. võtan analüüsi põhijäreldused kokku.

6.1. Sõnaliigid

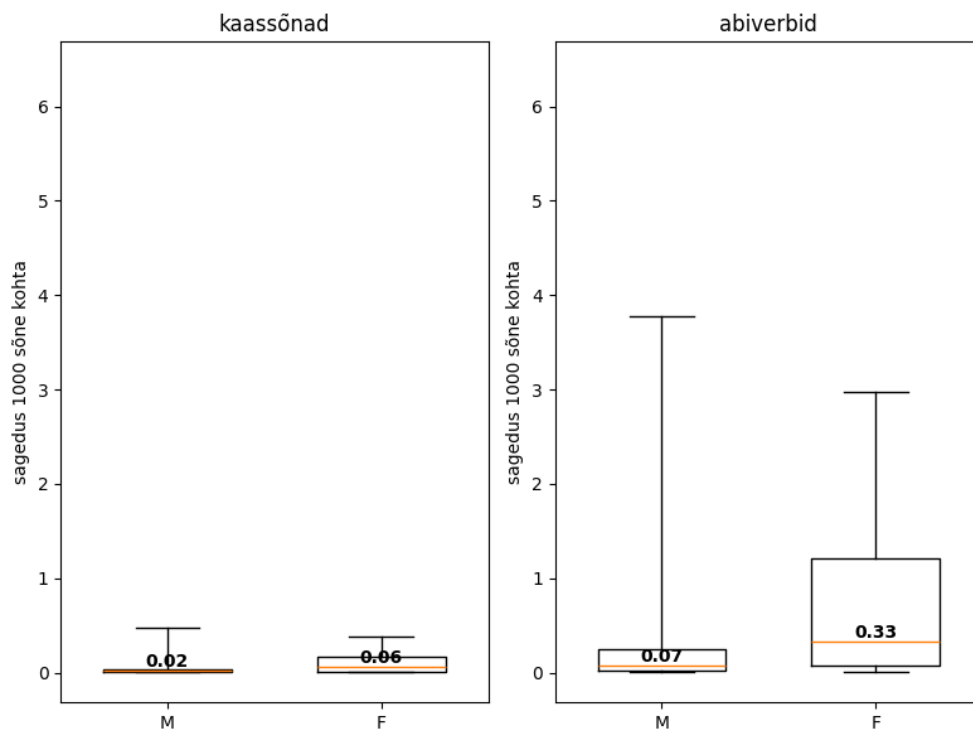
Esimeses grupis vaatlesin seitsme sõnaliigi kasutamise sagedust naiste ja meeste alamkorpuses: nimisõnad, omadussõnad, verbid, adverbid, pronoomenid ja kaassõnad. Verbid jagasin abiverbideks (*olema, saama, võima, tohtima, pidama, nägema, paistma, tunduma*) ja leksikaalseks verbiks (kõik ülejäänud verbid) (vt alapeatükk 5.1.). Joonised 2-4 illustreerivad kasutussagedusi. Karpdiagrammidel on kujutatud mediaan (keskmine esinemissagedus, joon karbi sees), interkvartiilivahemik (karp, mille sisse jääb täpselt pool ehk 50% kõigist andmetest: alumine serv tähendab, et veerand väärtustest on sellest madalamad, ja ülemine serv, et kolmveerand väärtustest on madalamad) ja vurrud (jooned, mis näitavad kõige suuremaid ja kõige väiksemaid väärtusi või üksikuid tavapärasest vahemikust välja jäävaid väärtusi). Iga sõnaliigi puhul kajastab vasakpoolne diagramm meeste andmeid (M), parempoolne diagramm naiste andmeid (F). Ka ülejäänud selles peatükis olevad diagrammid on loodud samal viisil.

Kõik diagrammid on kättesaadavad ka GitHubis¹⁶.

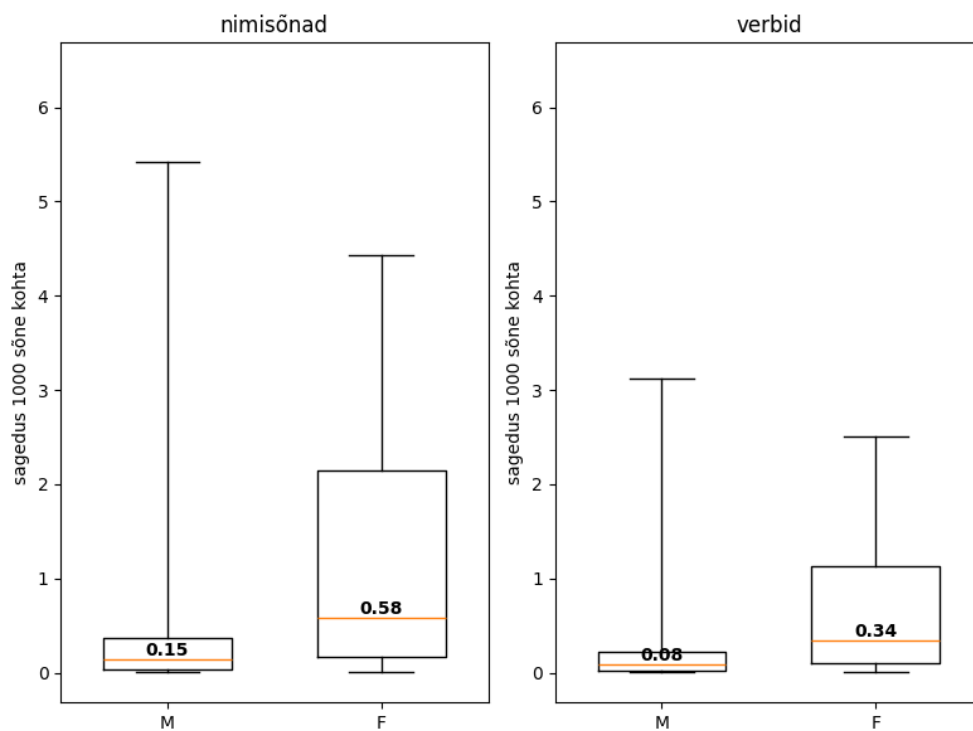
Diagrammid (joonised 2-5) näitavad, et kõigil sõnaliikidel on keskmine kasutussagedus (mediaanid) naistel kõrgem kui meestel. Näiteks pronoomenite, leksikaalsete verbide, omadussõnade ja nimisõnade mediaan on naiste alamkorpuses keskmiselt 3-4 korda kõrgem kui meeste alamkorpuses. Abiverbide ja adverbide puhul on erinevus veelgi suurem: naiste mediaan on peaaegu 5 korda suurem (0,33 vs. 0,07 abiverbide puhul; 0,61

¹⁶ <https://github.com/SofiaKriuchkova/ma-diploma-project-kriuchkova/tree/main/boxplots>

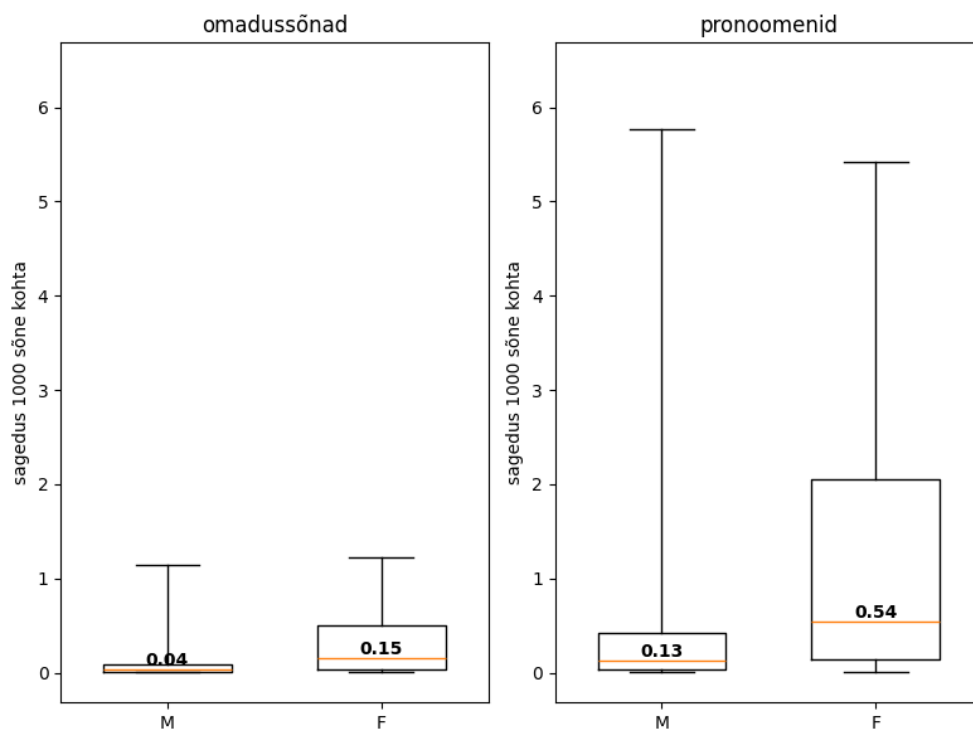
vs. 0,13 adverbide puhul). Isegi kaassõnade puhul, kus üldine kasutus on madal, on naiste mediaan kolm korda kõrgem (0,06 vs. 0,02). Interkvartiilvahemike (diagrammide karpide suurus) võrdlus näitab, et naistel on iga sõnaliigi tüüpilise kasutuse vahemik laiem: see tähendab, et varieeruvus on naiste alamkorpuse piires suurem. See on eriti märgatav pronoomenite, leksikaalsete verbide, nimisõnade, abiverbide ja adverbide puhul. Diagrammide vurrud, mis kajastavad äärmuslikke väärtusi (väljaspool 50% keskväärtuse andmeid), kipuvad olema meestel pikemad. See näitab, et kuigi suurem osa meeste alamkorpusest on koondunud mediaanile lähemale (madal ja stabiilne kasutus), on üksikuid juhtumeid, kus konkreetse sõnaliigi kasutamine erineb oluliselt üldisest suundumusest. Näiteks kasutavad mõned mehed pronoomenit, nimisõnu, adverbe ja abiverbe märgatavalt rohkem kui teised.



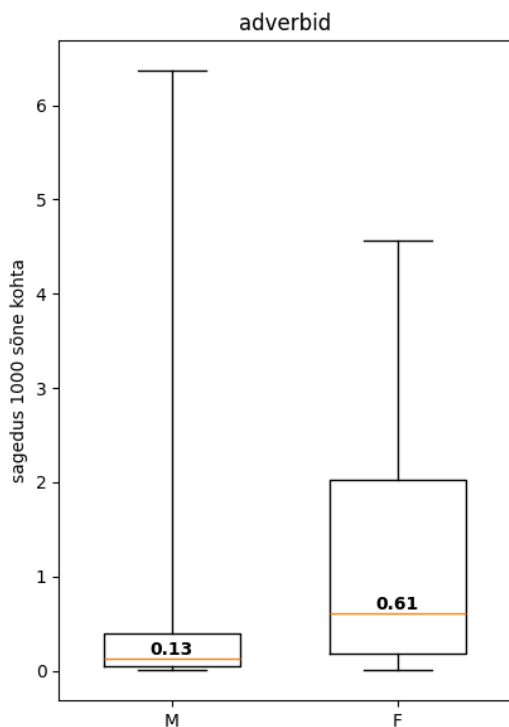
Joonis 2. Sõnaliigid. Kaassõnade ja abiverbide kasutussagedus meeste ja naiste kõnes (1000 sõne kohta).



Joonis 3. Sõnaliigid. Nimisõnade ja verbide kasutussagedus meeste ja naiste kõnes (1000 sõne kohta).



Joonis 4. Sõnaliigid. Omadussõnade ja pronoomenite kasutussagedus meeste ja naiste kõnes (1000 sõne kohta).



Joonis 5. Sõnaliigid. Adverbide kasutussagedus meeste ja naiste kõnes (1000 sõne kohta).

Üldiselt näitavad tulemused, et naised kasutavad vaadeldavaid sõnaliike sagedamini ja suurema mitmekesisusega. Sellele viitavad laiemad kastid diagrammidel. Meestel seevastu on jaotus kontsentreeritum - karbid on kitsamad, kuid vurrud on pikemad. See näitab, et korpuses esinevat meeste keelekasutust iseloomustab suurem ühtlus, kuid see sisaldab ka mõningaid silmatorkavaid kõrvalekaldeid, samas kui naiste keelekasutust iseloomustab laiem ja ühtlasem varieeruvus.

Tabel 2 kajastab sõnaliikide kohta tehtud statistiliste testide tulemused, mis annavad kinnituse, millised erinevused on olulised ja kui suur on mõju.

Veerg `p_mann_whitney` viitab Mann-Whitney U-testi teststatistikule. Veerg `p_whitney_holm` näitab tulemust pärast Holm-Bonferroni meetodi abil tehtud korrektsiooni, veerg `signif_holm` tabeli väärtuste lihtsaks tõlgendamiseks näitab kohe, kas tulemus pärast korrektsiooni on statistiliselt oluline (True) või mitte (False). Veerg `p_chi2` näitab Pearsoni χ^2 -testi tulemust ja veerg `p_chi2_holm` näitab selle tulemust pärast

korrektsiooni. Veerg `cramer_v` näitab mõju suurust¹⁷. Veerg `r_rb` viitab korrelatsiooni tulemustele. Kui korrelatsioon on miinusmärgiga, siis see, et näitab, et sõnaliigi esinemissagedus on iseloomulik naiste keelekasutusele, ning mida lähemal on väärtus 1 või -1, seda tugevam on sõnaliigi ja soo vaheline seos. Tabeli väärtuste lihtsaks tõlgendamiseks on veerus `signif_chi2` kohe märgitud, kas tulemus on pärast korrektiooni statistiliselt oluline (*True*) või mitte (*False*). Lõpptulemusena võtsin `signif_chi2` veerus esitatud tulemuse. Näiteks kaassõnade puhul on näha, et kuigi Mann–Whitney testi järgi on erinevus meeste ja naiste alamkorpuste vahel pärast korrektiooni statistiliselt oluline ($p < 0,001$), siis χ^2 -testi järgi ei ilmne pärast korrektiooni olulist erinevust ($p > 0,05$). Seetõttu tõlgendan selliseid tulemusi kui statistiliselt ebaolulisi ja ei pea neid aluseks järelduste tegemiseks süstemaatiliste erinevuste kohta selle sõnaliigi kasutamises meeste ja naiste vahel. Näiteks omadussõnade puhul on aga tulemused üheselt statistiliselt olulised nii Mann–Whitney testi ($p = 9,56 \times 10^{-16}$) kui ka χ^2 -testi ($p = 1,02 \times 10^{-14}$) järgi pärast korrektiooni. See ei viita juhuslikule erinevusele, vaid süstemaatilisele mustrile. Crameri V väärtus (0,00997) näitab küllaltki nõrka, kuid siiski olemasolevat mõju. Negatiivne biseriaalne korrelatsioon ($r = -0,4354$) kinnitab veelgi, et omadussõnade suurem kasutamine on iseloomulik naiste alamkorpusele.

Muud käesoleva peatüki tabelid on koostatud samal viisil. Need tabelid kajastavad ainult analüüsi jaoks olulist statistikat, täielikumad tabelid, mis sisaldavad ka nt absoluutseid sagedusi, on leitavad GitHubist¹⁸.

¹⁷ V väärtused jäävad vahemikku 0 ja 1: umbes 0,10 näitab nõrka seost, umbes 0,30 näitab keskmist seost ja 0,50 ja rohkem näitab tugevat seost nt soo ja keelelise tunnuse vahel. (Brezina 2018: 115).

¹⁸ <https://github.com/SofiaKriuchkova/ma-diploma-project-kriuchkova/tree/main/tulemused>

Tabel 2. Sõnaliigid. Statistiliste testide tulemused.

POS	p_mann_whitney	p_whitney_holm	signif_holm	chi2_stat	p_chi2	p_chi2_holm	cramer_v	r_rb	signif_chi2
kaassõnad	1.71×10^{-13}	1.71×10^{-13}	True	1.70×10^{-1}	6.78×10^{-1}	9.03×10^{-1}	5.15×10^{-4}	-0.4111	False
nimisõnad	5.63×10^{-17}	3.38×10^{-16}	True	2.18×10^0	1.39×10^{-1}	5.58×10^{-1}	1.85×10^{-3}	-0.4397	False
verbid	2.22×10^{-16}	8.88×10^{-17}	True	5.67×10^{-1}	4.51×10^{-1}	9.03×10^{-1}	9.41×10^{-4}	-0.4359	False
abiverbid	1.77×10^{-16}	8.86×10^{-16}	True	7.11×10^0	7.68×10^{-3}	3.84×10^{-2}	3.33×10^{-3}	-0.4364	True
omadussõnad	4.34×10^{-16}	9.56×10^{-16}	True	6.37×10^1	1.46×10^{-15}	1.02×10^{-14}	9.97×10^{-3}	-0.4354	True
pronoomenid	3.19×10^{-16}	9.56×10^{-16}	True	1.01×10^3	1.37×10^{-1}	6.93×10^{-1}	9.18×10^{-2}	-0.4313	False
adverbid	6.21×10^{-18}	4.35×10^{-17}	True	1.44×10^0	2.31×10^{-14}	2.71×10^{-13}	1.5×10^{-3}	-0.4584	True

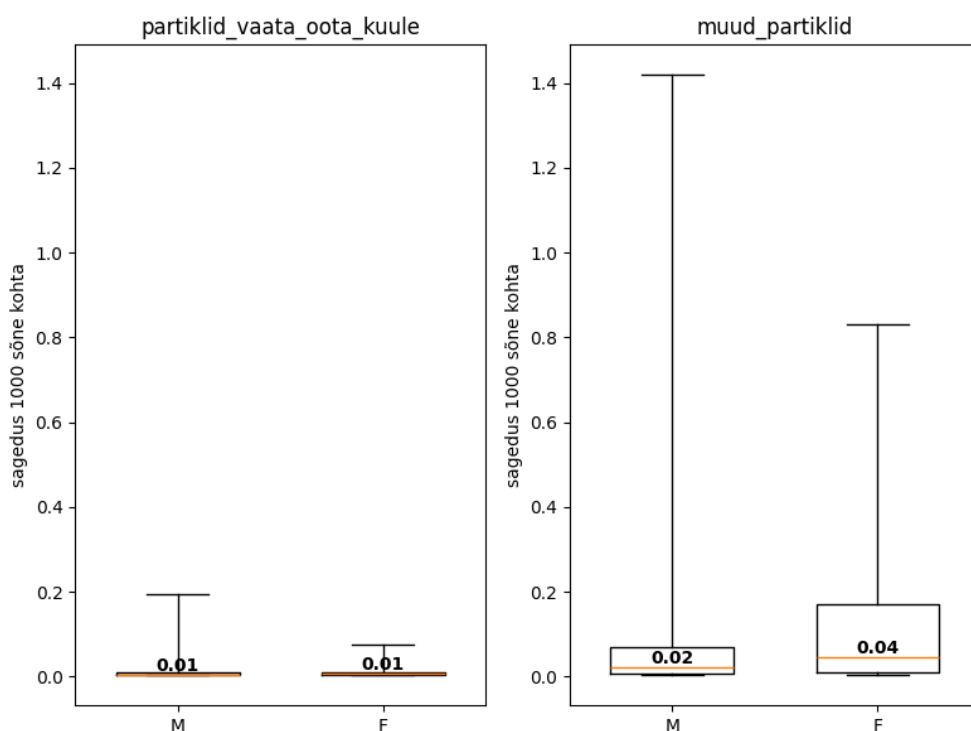
Statistilise analüüsi tulemused viitavad, et meeste ja naiste keelekasutuses esineb statistiliselt oluline erinevus omadussõnade, adverbide ja abiverbide kasutuses. Vastavalt biseriaalsele korrelatsioonile (veerg r_rb) kasutavad naised neid sõnaliike sagedamini ja seejuures oli sõnaliigi ja naiste keelekasutuse vaheline seos keskmise tugevusega ($r \approx -0,44$ omadussõnade ja abiverbide puhul ja $r \approx -0,46$ adverbide puhul). Samas Crameri V järgi on mõju suurus kõigil kolmel juhul väga väike, mis tähendab, et kuigi meeste ja naiste erinevus on statistiliselt kinnitatud, on selle praktiline tähendus siiski väike.

Seega on tulemused osaliselt kooskõlas varasemate uurimistulemustega (nt Argamon jt 2003, Litvinova jt 2017, Newman jt 2008). Naised kasutavad sagedamini omadussõnu, adverbe ja abiverbe, mis on kooskõlas arusaamaga, et naiste keelekasutuses on rohkem kirjeldavaid sõnu. Kuid vastupidist tendentsi, et meeste keelekasutuses esinevad sagedamini nimisõnad ja verbid, ei leitud. Samuti on oluline meeles pidada, et sõnaliigid on siinse töö raames kõige vähem kontrollitud keeleline tunnus. Kuigi erinevused võivad olla statistiliselt olulised suure andmehulga korral, on soo ja sõnaliigi sageduse vahelise seose tegelik tugevus väike, seetõttu tuleks tulevikus arvestada ka teiste teguritega, nagu vanus ja haridustase, kõneleja register jms.

6.2. Keelelised tunnused

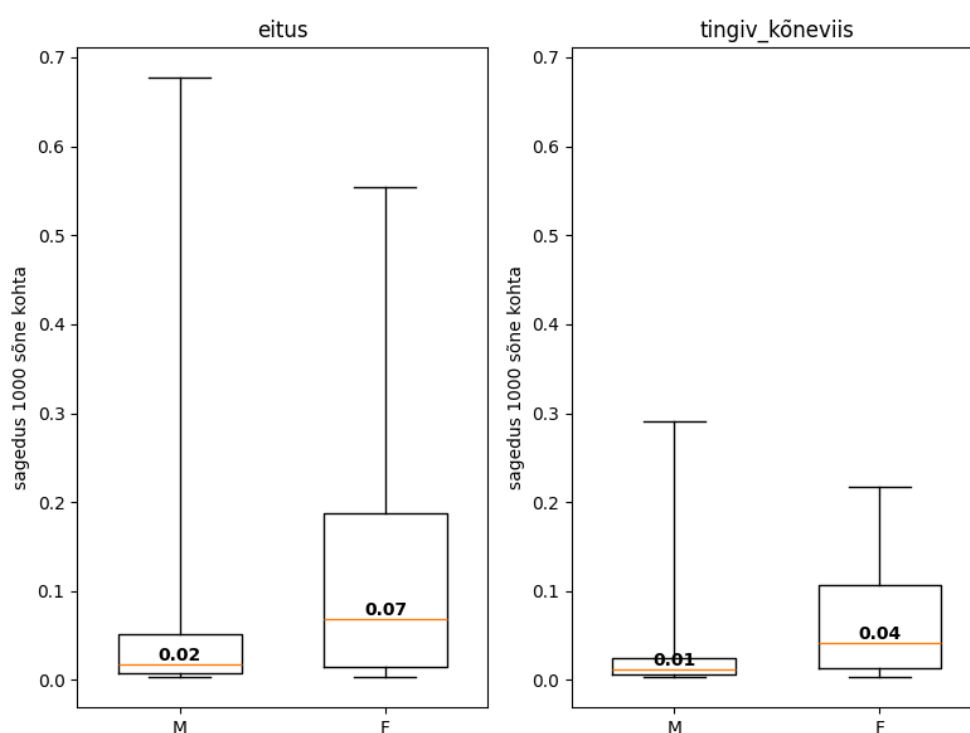
Teises grupis vaatlesin üheksa keeleliste tunnuste kasutamise sagedust naiste ja meeste alamkorpuses: eitus, deminutiiv *-kene*, intensiivistajad, esimese ja teise isiku pronoomenid, partiklid, käskiv kõneviis, tingiv kõneviis, umbisikuline tegumood ja KP-markerid). Intensiivistajaid käsitlen eraldi sõltuvalt sellest, millise sõnaliigiga nad koos esinesid (vt alapeatükk 5.1.). Samuti käsitlen partikleid *vaata / oota / kuule* eraldi muudest partiklitest, kuna kogusin neid eraldi (vt alapeatükk 5.2.).

Joonisel 6 on näidatud erinevate partiklite esinemissagedused meeste ja naiste alamkorpustes. Diagrammid näitavad, et naised kasutavad partikleid keskmiselt sagedamini kui mehed. Muud partiklid kategoorias on naiste alamkorpuses mediaanväärtus 0,04, meeste alamkorpuses aga 0,02. Kategoorias partiklid *vaata/oota/kuule* on mediaanid samad (0,01), kuid meestel on väärtuste hajuvus suurem, mida näitavad diagrammi pikemad vurrud. Muude partiklite puhul on interkvartiilvahemikud naistel laiemad, mis viitab mitmekesisemale kasutamisele.

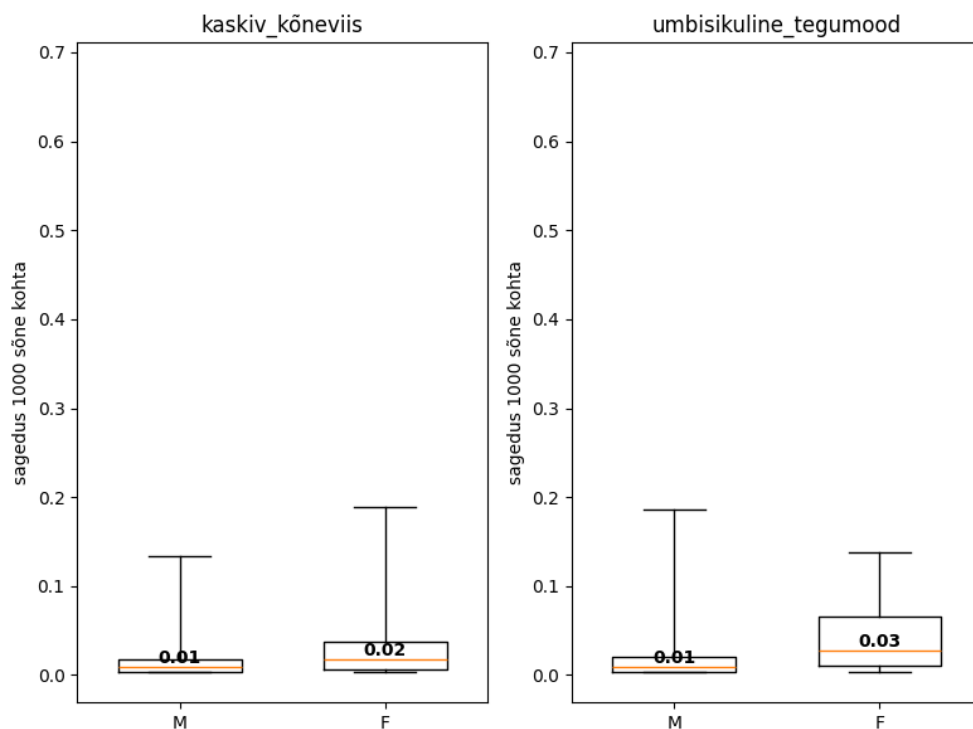


Joonis 6. Partiklite kasutussagedus meeste ja naiste kõnes (1000 sõne kohta).

Joonised 7 ja 8 illustreerivad erinevate verbi grammatiliste kategooriate esinemissagedusi naiste ja meeste alamkorpustes. Vastavalt joonistele on kõigi verbi verbi grammatiliste kategooriate puhul keskmine sagedus (mediaanid) naistel kõrgem kui meestel. Näiteks eituse puhul on mediaan naiste alamkorpuses 0,07, meestel aga 0,02, ning tingiva kõneviisi puhul vastavalt 0,04 ja 0,01. Käskiva kõneviisi ja umbisikulise tegumoe puhul on erinevus vähem märgatav, kuid siiski olemas: vastavalt 0,02 vs 0,01 ja 0,03 vs 0,01. Interkvartiilivahemike võrdlus näitab, et kasutuse vahemik on naiste alamkorpuses laiem, eriti eituse ja tingiva kõneviisi kategooriates. See viitab suuremale varieeruvusele naiste tekstides. Samal ajal eituse ja tingiva kõneviisi kategooriates on vurrud pikemad meestel. See viitab üksikjuhtumitele, kus neid vorme kasutatakse meeste tekstides oluliselt sagedamini kui keskmiselt.

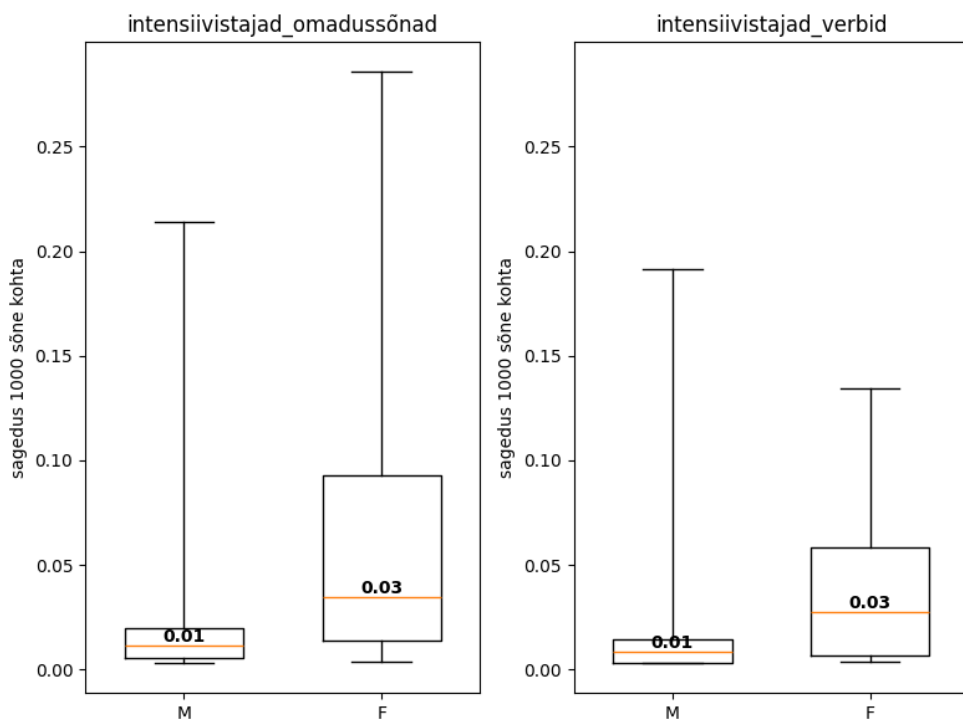


Joonis 7. Verbi grammatilised kategooriad. Eituse ja tingiva kõneviisi kasutussagedus meeste ja naiste kõnes (1000 sõne kohta).

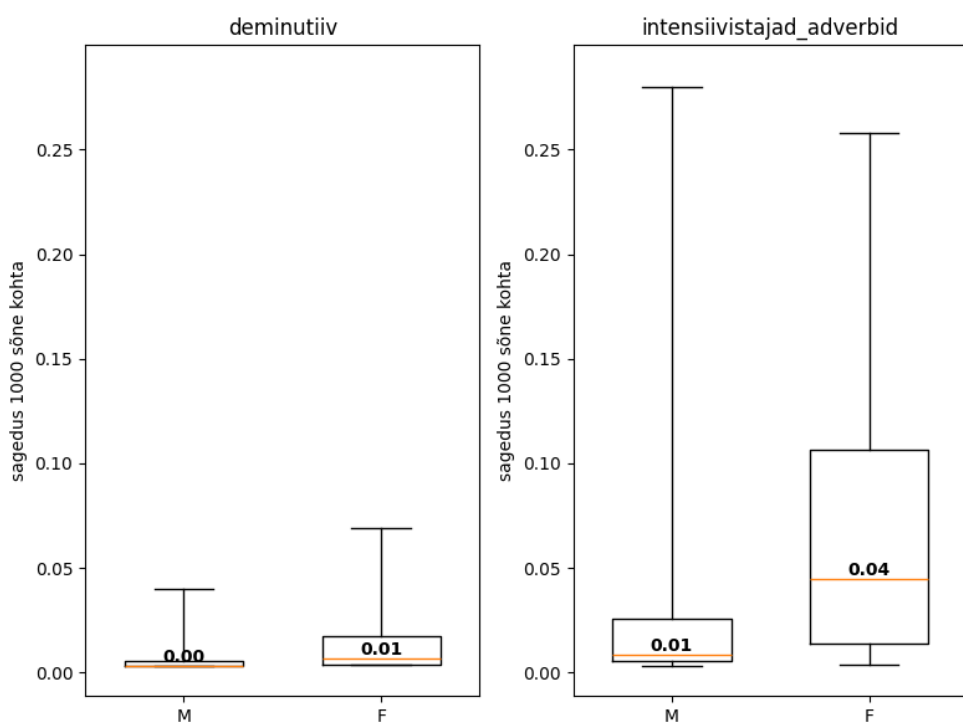


Joonis 8. Verbi grammatilised kategooriad. Käskiva kõneviisi ja umbisikulise tegumoe kasutussagedus meeste ja naiste kõnes (1000 sõne kohta).

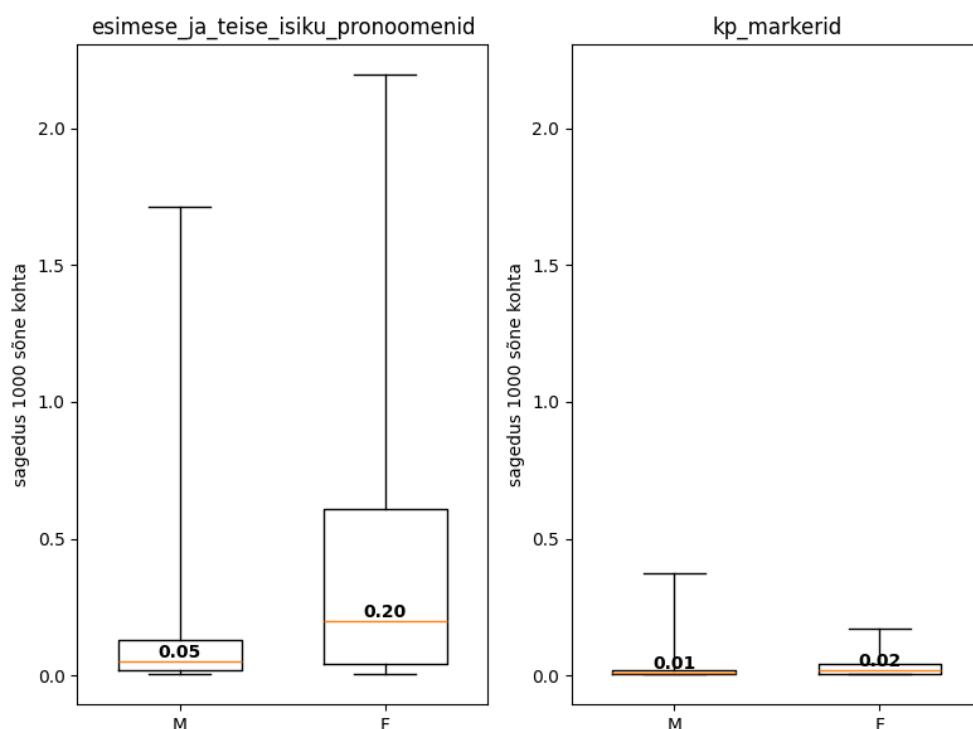
Joonised 9 ja 10 illustreerivad intensiivistajate ja deminutiivi esinemissagedusi naiste ja meeste alamkorpustes. Diagrammid näitavad, et intensiivistajate ja deminutiivi kasutamine on naistel järjepidevalt veidi suurem kui meestel. Näiteks omadussõnade ja verbide intensiivistajate kategooriates on naiste mediaanväärtused veidi kõrgemad kui meestel (0,03 vs. 0,01). Adverbide puhul püsib see erinevus samuti (0,04 naistel ja 0,01 meestel), kuigi mõlemal juhul jäävad absoluutväärtused madalaks. Deminutiivi *-kene* kasutatakse mõlemas alamkorpuses harva, kuid naissoost tekstides on mediaan siiski veidi kõrgem (0,01 vs. 0,00). Kõigi esitatud kategooriate puhul on interkvartililivahemik naistel laiem, mis viitab nende keeleliste tunnuste mitmekesisemale kasutamisele. Diagrammide vurrud näitavad, et äärmuslike väärtuste erinevused meeste ja naiste vahel on enamikes kategooriates väikesed. Erandiks on omadussõnade intensiivistajad, kus väärtuste hajuvus on märgatavalt suurem naiste puhul, ning verbide intensiivistajad, kus äärmuslikud väärtused on kõrgemad meeste seas. See viitab sellele, et just neis kahes kategoorias esineb üksikjuhtumeid sagedasemast kasutusest vastavalt naiste ja meeste tekstides.



Joonis 9. Intensiivistajad. Omadussõnade ja verbide intensiivistajate kasutussagedus meeste ja naiste kõnes (1000 sõne kohta).



Joonis 10. Intensiivistajad. Adverbide intensiivistajate ja deminutiivi *-kene* kasutussagedus meeste ja naiste kõnes (1000 sõne kohta).



Joonis 11. Muud tunnused. Esimese ja teise isiku pronoomenite ja KP-markerite kasutussagedus meeste ja naiste kõnes (1000 sõne kohta).

Joonisel 11 on näidatud esimese ja teise isiku pronoomenite ja KP-markerite esinemissagedused meeste ja naiste alamkorpustes. Esimese ja teise isiku pronoomenite puhul on naiste mediaan 0,20, võrreldes 0,05-ga meeste puhul, mis näitab, et naised kasutavad neid vorme oluliselt sagedamini võrreldes meestega. Lisaks on naiste puhul täheldatav laiem interkvartiilvahemik ja pikemad vurrud, mis viitab suuremale varieeruvusele. Meeste puhul jääb nende asesõnade kasutus suhteliselt madalaks ja vähem mitmekesiseks. KP-markerite puhul on erinevused vähem väljendunud: naiste mediaan on 0,02 ja meestel 0,01. Kuigi absoluutväärtused on madalad, on naiste kasutusvahemik veidi laiem, mis viitab pisut sagedasemale ja mitmekesisemale kasutusele.

Üldiselt näitavad diagrammid, et naiste keelekasutuses esineb enamik uuritud keelelisi tunnuseid sagedamini ja mitmekesisemalt. Nende mediaanväärtused on kõrgemad ja vahemikud laiemad, mis viitab suuremale varieeruvusele. Meeste puhul esinevad seevastu

mõnes kategoorias, nagu eitus või verbide intensiivistajad, üksikjuhud väga sagedasest kasutusest, kuid üldiselt on nende keelekasutus nende tunnuste osas vähem küllastunud.

Tabelis 3 on kujutatud nende keelelise tunnuse statistiliste testide tulemused.

Statistilise analüüsi tulemused näitavad, et meeste ja naiste keelekasutuse vahel on enamiku uuritud tunnuste puhul statistiliselt olulised erinevused. Ainult umbisikulise tegumoe puhul olid χ^2 -testi tulemus statistiliselt ebaoluline ($p = 0,45$). Vastavalt biseriaalse korrelatsiooni väärtustele (veerg r_{rb}) kasutavad naised neid keelelisi elemente sagedamini, millele viitavad koefitsiendi negatiivsed väärtused. Kõigil olulistel juhtudel on täheldatud keskmise tugevusega seos - kõige enam väljenduvad erinevused verbide intensiivistajate ($r \approx -0,51$), adverbide intensiivistajate ($r \approx -0,50$), samuti deminutiivi *-kene* ($r \approx -0,55$) kasutamisel. Crameri V-kriteeriumi järgi jääb mõju suurus siiski väikeseks, hoolimata statistilisest olulisusest. Kõrgeimad Crameri V väärtused on omadussõnade intensiivistajate ($\approx 0,01$), *vaata/oota/kuule* partiklite ($\approx 0,02$) ja muude partiklite ($\approx 0,03$) kategooriates, mis siiski viitab nõrgale mõjule. See tähendab, et kuigi meeste ja naiste vahelised erinevused on statistiliselt usaldusväärsed, on nende praktiline tähtsus piiratud.

Tabel 3. Muud tunnused. Statistiliste testide tulemused.

Tunnus	p_mann _whitne y	p_whitne y_holm	signif _holm	chi2_sta t	p_chi2	p_chi2_h olm	cramer_ v	r_rb	signif_ chi2
intensiivistaja d verbidega	1.62×10^{-16}	1.78×10^{-15}	True	1.37×10^1	2.09×10^{-4}	6.28×10^{-4}	4.36×10^{-3}	-0.5057	True
partiklid (vaata, oota, kuule)	4.30×10^{-3}	4.30×10^{-3}	True	2.25×10^2	7.15×10^{-51}	7.86×10^{-50}	1.87×10^{-2}	-0.2526	True
intensiivistaja d adverbidega	1.16×10^{-16}	1.40×10^{-15}	True	6.29×10^1	2.17×10^{-15}	1.95×10^{-14}	9.91×10^{-3}	-0.4975	True
eitus	8.37×10^{-1}	5.02×10^{-12}	True	2.48×10^1	6.39×10^{-7}	3.19×10^{-6}	6.22×10^{-3}	-0.4035	True
tingiv kõneviis	1.58×10^{-15}	1.42×10^{-14}	True	3.46×10^1	4.15×10^{-9}	2.90×10^{-8}	7.34×10^{-3}	-0.4724	True
intensiivistaja d omadussõnade ga	2.28×10^{-16}	2.28×10^{-15}	True	6.79×10^1	1.74×10^{-16}	1.74×10^{-15}	1.03×10^{-2}	-0.4901	True
esimese ja teise isiku pronoomenid	8.89×10^{-14}	6.22×10^{-13}	True	5.63×10^0	1.77×10^{-2}	3.54×10^{-2}	2.96×10^{-3}	-0.4011	True
käskiv kõneviis	3.44×10^{-9}	1.03×10^{-8}	True	2.87×10^1	8.58×10^{-8}	5.15×10^{-7}	6.69×10^{-3}	-0.3707	True
umbisikuline tegumood	6.39×10^{-14}	5.11×10^{-13}	True	5.58×10^{-1}	4.55×10^{-1}	4.55×10^{-1}	9.33×10^{-4}	-0.4791	False
teised partiklid	3.31×10^{-6}	6.63×10^{-6}	True	6.01×10^3	1.20×10^{-132}	1.44×10^{-131}	3.06×10^{-2}	-0.2660	True
kp-markerid	5.15×10^{-10}	2.06×10^{-9}	True	4.43×10^1	2.78×10^{-11}	2.22×10^{-10}	8.32×10^{-3}	-0.3955	True
deminutiiv	6.72×10^{-10}	3.36×10^{-11}	True	2.56×10^1	4.1×10^{-7}	2.46×10^{-6}	6.33×10^{-3}	-0.5525	True

Analüüsis käsitlesin erinevaid keelilisi tunnuseid, millest igaüks täidab kõnekeeles teatud funktsioone. Inensiivistajad näiteks rõhutavad lausungi emotsionaalset värvingut, väljendavad kõneleja subjektiivset hinnangut ja kaasatust. Eesti keeles võib neid jagada

rühmadesse sõltuvalt sellest, millise sõnaliigiga nad kaasnevad, näiteks verbidega (nt *väga armastama*), omadussõnadega (nt *tüiesti vale*) ja adverbidega (nt *üsna kiiresti*). (Aasa 2024) Tulemuste kohaselt kasutavad naised sagedamini kõiki intensiivistajate tüüpe, mis võib viidata suuremale kalduvusele subjektiivsele hindamisele ja emotsionaalsele kaasatusele, nagu on märkinud Czerwionka jt (2023) ja Newman jt (2008). Samas näitavad uurimuse andmed, et vurrud on diagrammides pikemad maskuliinsetes alamtähendustes just verbide intensiivistajate puhul (joonis 9), mis võib viidata nende haruldasemale, kuid intensiivsemale kasutamisele meeste poolt.

Deminutiiv *-kene* täidab pehmendavat, hoolivust või sõbralikkust väljendavat funktsiooni ja loob emotsionaalse läheduse õhkkonna. Suulises kõnekeeles kasutatakse neid vorme soojema, mitteametlikuma suhtluse loomiseks ja sageli väljendavad need subjektiivselt positiivset suhtumist objekti või isiku suhtes. (Liivak 2023) See on kooskõlas varasemate tähelepanekutega (Posio jt 2024, Senkina jt 2018), kus rõhutatakse, et naiste kõne võib olla rohkem orienteeritud emotsionaalsele kontaktile ja viisakusele.

Eitus on kasutusel eriarvamuse, selgituse või paranduse väljendamiseks (EKG II 1993: 155). Varasemas soolingvistika kirjanduses on märgitud, et eituse sagedane kasutamine on iseloomulik meeste keelekasutusele (nt Eagly jt 2012, Pilkington 1992, Siregar jt 2023). Uuritavas korpuses on aga täheldatud, et naised kasutavad eitamist sagedamini. See võib olla seotud sellega, et nagu näitasid tulemused sõnaliikide lõikes, kasutavad naised sagedamini abiverbe, millega eitus korpuses tavaliselt realiseerub. Näiteks korpuses on:

- *ma ei olnud kuni kuuenda klassini kindel, mida ma tahan...*
- *ta ei saa rääkida, sest ta kardab...*

Tingiv kõneviis on mõeldud avalduse pehmendamiseks, hüpoteetilise ja viisakuse väljendamiseks. See vorm võimaldab kõnelejal vältida otsekohesust ja väljendada delikaatsust. (EG 2023: 555-558, 1169-1171). Tulemuste kohaselt kasutavad naised seda veidi sagedamini, mis on kooskõlas järeldustega soo seotud erinevuste kohta keele viisakusstrateegiates (nt Czerwionka jt 2023; Posio jt 2024).

Käskiv kõneviisil on dialoogi juhtimise funktsioon: seda kasutatakse üleskutseteks, juhisteks ja tegutsemisele õhutamiseks (EG 2023: 1165-1176). Vaatamata ootustele, et selliseid vorme kasutavad sagedamini mehed (nt Czerwionka jt 2023; Posio jt 2024), kasutasid uuritavas korpuses just naised rohkem käskivat kõneviisi. Üks võimalik seletus

on see, et naiste taskuhäälingute saatejuhid lähenesid oma külalistele sageli kutsega istuda maha või alustada vestlust mingil konkreetsel teemal. Näiteks korpuses on:

- ...*aga tegelikult lihtsalt istu korraks siia diivani peale...*
- **Rüügi** natukene nendest lähemalt, kust sa tuled, kes sa oled...

Umbisikuline tegumood võimaldab keskenduda tegevusele, täpsustamata selle tegijat, nihutades seega tähelepanu eemale kõnelejalt ja andes lausungile neutraalsuse. Seda kasutatakse ka üldiste järelduste väljendamiseks või otsese vastutuse väljendamise vältimiseks. (EG 2023: 442) Kuigi varasemate ootuste põhjal võis eeldada, et mehed kasutavad seda vormi sagedamini (Posio jt 2024), ei leitud uurimiskorpuses statistiliselt olulist erinevust.

Partiklid *vaata, oota* ja *kuule* on mõeldud tähelepanu äratamiseks ja hoidmiseks, dialoogi tempo reguleerimiseks ja suhtlemisvalmiduse väljendamiseks (Keevallik 2003, Keevallik 2008). Nende sagedasem kasutamine naiste poolt võib peegeldada naiste kõne kaasavat ja dialoogilist orientatsiooni (Newman jt 2008). KP-markerid struktureerivad dialoogi, aitavad korraldada üleminekuid teemade vahel ja loogiliselt siduda väiteid. Need aitavad kaasa vestluse selgusele ja täpsusele. (EG 2023: 1109-1110) Naised kasutasid neid uurimuses veidi sagedamini, mis võib näidata suuremat suhtumist korrastatud vestlusesse (Newman jt 2008, Posio jt 2024).

Esimese ja teise isiku pronoomenid aitavad luua isiklikku kontakti, kaasata vestluspartnerit ja rõhutavad kõneleja subjektiivset positsiooni. Nende sagedasem kasutamine naiste poolt võib olla seotud naiste keelekasutuse interpersonaalse rõhuasetusega (Newman jt 2008, Posio jt 2024).

Seega täidavad kirjeldatud keelelised tunnused mitmesuguseid kommunikatiivseid funktsioone ning paljudel juhtudel on erinevused nende kasutamises meeste ja naiste poolt kooskõlas varasemate empiiriliste tähelepanekute ja teoreetiliste järeldustega, mis on esitatud näiteks Czerwionka jt (2023), Newman jt (2008), Posio jt (2024) ning Senkina jt (2018) artiklites. Eriti naiste keelekasutuses on järjepidevalt veidi rohkem keelelisi tunnuseid, mida kasutatakse kaasamise, emotsionaalsuse, viisakuse ja kontakti otsimise eesmärgil. Seevastu mõned tulemused, näiteks eituse ja imperatiivse meeoleu sagedasem kasutamine naiste poolt, on vastuolus varasemate uurimuste tulemustega ning vajavad täiendavat uurimist.

6.3. Partiklid

Partiklid on muutumatud sõnad, mis ei kuulu lause grammatilisse struktuuri. Need juhivad vestluse kulgu, väljendavad kõneleja hoiakut ning aitavad luua sidusust vestluse käigus. (EG 2023: 1066-1067) Kõige sagedasemate, st korpuses üle 50 korra esinevate partiklite hulgas oli 10 partiklit: *jah, jumal, okei, nagu, vot, noh, eks ole, oh, no* ja *kurat*. Kuna partiklid on üldiselt seotud emotsionaalsuse või ebakindlusega, on neid sageli seostatud naiste keelekasutusega (nt Bogdanova-Beglarian jt 2019, Kolyaseva 2022, Konstantinovskaia 2020, Newman jt 2008), ent kuna partiklite kasutamine on üldisemalt suulisele suhtlusele iseloomulik nähtus, kasutavad neid kõneldes kõik. Erinevused võivad olla partiklite kasutussageduses, teatud partiklite eelistamises või mitte-eelistamises ning üldises leksikaalses koostises. Partiklite üldist kasutussagedust käsitlesin eelmises alapeatükis. Siin analüüsin kõige sagedasemate partiklite kasutamist eraldi.

Kuna partiklite puhul osutusid mediaanide väärtused enamikul juhtudel peaaegu võrdseks nulliga ja erinevused ei ületanud 0,05, näib illustratiivne analüüs olevat üleliigne (GitHub¹⁹). Seetõttu lähen kohe statistiliste testide tulemuste kirjelduse juurde.

Tabelis 4 on kujutatud partiklite statistiliste testide tulemused. Need näitavad, et partiklite *no, noh, nagu, okei, jumal* ja *eks ole* kasutamisel on statistiliselt olulised erinevused. Vaatamata Mann-Whitney testide madalatele p-väärtustele, χ^2 -testide järgi ei ole statistilist erinevust partiklite *oh, vot* ja *jah* kasutamisel, mistõttu on need edasistest analüüsides välja jäetud. Vastavalt biseriaalse korrelatsiooni väärtusele (veerg *r_{rb}*) kasutavad naised partikleid *no, noh, nagu, okei, jumal* ja *eks ole* sagedamini. Seos soo ja nende partiklite kasutamise vahel ulatub nõrgast kuni mõõdukalt tugevani: kõige tugevam seos on täheldatud partikli *jumal* puhul ($r \approx -0,60$). Crameri V väärtused on kõigi partiklite puhul väga madalad ($V < 0,01$), mis tähendab, et kuigi meeste ja naiste erinevused on statistiliselt märkimisväärsed, on nende praktiline tähtsus marginaalne.

¹⁹ https://github.com/SofiaKriuchkova/ma-diploma-project-kriuchkova/tree/main/boxplots/boxplots_partiklid

Tabel 4. Partiklid. Statistiliste testide tulemused.

partikkel	p_mann_whitney	p_whitney_holm	signif_norm_holm	chi2_stat	p_chi2	p_chi2_holm	cramer_v	r_rb	signif_chi2
no	2.15×10^{-3}	2.15×10^{-3}	True	7.08×10^1	3.87×10^{-17}	3.10×10^{-16}	1.05×10^{-2}	-0.2603	True
oh	3.71×10^{-5}	2.23×10^{-4}	True	3.70×10^0	5.43×10^{-2}	1.63×10^{-1}	2.41×10^{-3}	-0.7270	False
vot	2.50×10^{-4}	7.80×10^{-4}	True	5.13×10^{-2}	8.21×10^{-1}	8.21×10^{-1}	2.83×10^{-4}	-0.4959	False
nagu	3.56×10^{-4}	7.80×10^{-4}	True	5.72×10^2	1.87×10^{-126}	1.67×10^{-125}	2.99×10^{-2}	-0.2074	True
okei	6.80×10^{-5}	3.40×10^{-4}	True	1.36×10^1	2.31×10^{-4}	1.39×10^{-3}	4.60×10^{-3}	-0.4046	True
jah	1.25×10^{-5}	8.8×10^{-5}	True	1.46×10^0	2.28×10^{-1}	4.55×10^{-1}	1.51×10^{-3}	-0.5358	False
jumal	1.23×10^{-6}	1.0×10^{-5}	True	1.35×10^1	2.44×10^{-4}	1.39×10^{-3}	4.58×10^{-3}	-0.6017	True
eks ole	1.20×10^{-4}	4.83×10^{-4}	True	3.44×10^1	4.47×10^{-9}	3.13×10^{-8}	7.33×10^{-3}	-0.4052	True
noh	2.51×10^{-7}	2.00×10^{-6}	True	6.62×10^0	1.01×10^{-2}	4.03×10^{-2}	3.22×10^{-3}	-0.4454	True
kurat	NA	NA	NA	NA	NA	NA	NA	NA	NA

Partikkel *kurat* ei kuulu testitulemuste hulka, kuna see ei esinenud naiste alamkorpuses üldse. Meeste kõnes ilmus partikkel eri vanuses kõnelejate kõnes, kes arutasid erinevaid teemasid (nt meditsiin²⁰, muusika²¹, armastus²²), neutraalses ja mitteametlikus registris.

Seega kasutavad naised keskmiselt veidi sagedamini partikleid *no*, *noh*, *nagu*, *okei*, *jumal* ja *eks ole*, kuid selle mõju suurus jääb väikeseks. Mehed kasutavad arvatavasti sagedamini partiklit *kurat*, kuna see esines ainult meeste korpuses. Keeles täidavad need partiklid erinevaid funktsioone.

Tavapärastelt eristab grammatika funktsionaalselt partikleid *no* ja *noh*. Venitatud *noh* annab aega fraasi üle järele mõelda, lühike *no* töötab käivitajana. (EG 2023: 1099-1101) Kuid arvestades, et korpus on automaatselt tuvastatud ja nende partiklite tuvastamisel võis

²⁰ Episood

20220106_239_marius_vahter_ja_sven_kabanen_mikrobioomi_olulisusest_detoksist_ja_probiotikumidest

²¹ Episood 20210413_episood_4_kas_vanasti_oli_muusika_parem

²² Episood 20190908_bipolaarne_armastus_2

esineda vigu, käsitlen siinses töös *no* ja *noh* sünonüümidenä. Partikli *no(h)* peamine funktsioon on vestluskaaslast suunata, näiteks kiirendada järgmisele tegevusele üleminekut, täita pausi, väljendada valmisolekut jätkata. Vastavas positsioonis võivad need leevendada nõustumatusi. (EG 2023: 1099-1101; Keevallik 2016: 217-218) Näited korpusest:

- *no* siis hakkame kohe rääkima...
- aga *noh* aga keegi...
- - aga kuule!
-*noh*...
- ei, *no* ma arvan, et see on nii väike

Partikkel *nagu*, mis on kujunenud võrdlevast sidesõnast, on muutunud hüpoteese ja enesekorrigeerimisi väljendavaks partikliks. See võib esineda suvalises kohas lauses ja võib näidata kõneleja kindlusastet ning pehmemdada väite kategoorilisust (EG 2023: 1108) Näide korpusest:

- *Praegu tundub ikka väga... nagu... uskumatu on summa.*

Partikkel *okei* on laenatud inglise keelest (*okay*) ja täidab sarnaseid funktsioone nagu inglise keeles. Kolm peamist on positiivne nõustumine, neutraalne üleminekumärgistus ja pehmemdamine enne eriarvamuse väljendamist. (EG 2023: 1100; Kolsar 2017: 20-33) Näited korpusest:

- ...*et okei*, nii ma teen
- *okei*, räägime nüüd meestest...
- *okei*, aga ma ei oska, ma ei tee, kuidas mis asi on näha

Eks ole toimib kinnitusküsimusena, kui kõneleja esitab hüpoteesi ja ootab kinnitust, vähendab väite kategoorilisust ning näitab kõneleja ebakindlust ehk demonstreerib, et vestluskaaslased omavad mingis valdkonnas peaaegu võrdset teadmist. Üldjoontes muudab küsimus väite ühise kinnituse aktiks, jättes adressaadile formaalse õiguse viimasel sõnal olla. (EG 2023: 1115-1119) Näited korpusest:

- *Vaata kus siis NATO on, NATO on teises kohas, eks ole...*

- *Ja kõik kõik need asjad, eks ole, ja tuleb tõusta*
- *me hindame patsienti tervikuna, eks ole....*

Partikleid *no(h)*, *nagu*, *okei* ja *eks ole* ühendab nende roll lauses. Nende kaudu reguleerib kõneleja diskursuse voolu, pehmenab kategoorilisust või demonstreerib ebakindlust. Statistiliste testide tulemuste kohaselt kasutavad naised neid partikleid sagedamini, mis näitab sarnasust teiste keelte naiste keelekasutuse omaduste kirjeldustega, kus naistele on iseloomulik sagedasem pehmenamist ja ebakindlust väljendavate markerite kasutamine (nt Kirilina 2005, Konstantinovskaia 2020, Newman jt 2008).

Sõnad *jumal* ja *kurat* moodustavad eesti keeles sümboolse seose sakraalse ja demoonilise vahel, kuid nende tänapäevased funktsioonid on täiesti ilmalikud. Neid võib kasutada vandesõnadena sünonüümsete tähendustega astme- ja emotsioonitugevdajatega. (Jürine jt 2016: 11-13, 20-21; Kehayov 2009: 150-152; Pajusalu 2006: 25-30) Suulises keeles ei sisalda sõnad enam oma otsest tähendust, vaid aitavad kõnelejal väljendada oma suhtumist öeldusse. Kuid *jumal* tähistab pigem üllatust, ärevust või kergendust suhteliselt neutraalses või positiivses emotsionaalses kontekstis (Jürine jt 2016: 11–21), samas kui *kurat* väljendab ärritust, viha või pahameelt ja säilitab jõhkra värvingu (Pajusalu 2006: 20-27).

Testide tulemused näitasid, et mehed ja naised kasutavad sõnu *jumal* ja *kurat* erinevalt. Mehed kipuvad sagedamini kasutama sõna *kurat* (nt *Tahtis midagi öelda, aga kurat, ma ei tea...*), naised aga sõna *jumal* (nt *Sa ei jää sinna sisse, issand jumal, kui õudne, issand, kui mul on paha.*). See on kooskõlas varasema teadmisega, et meeste ja naiste vandesõnad on erinevad (Gauthier 2017). Mehed kasutavad sagedamini ühiskonna seisukohast jõhkramad väljendeid, mistõttu tundub nende kõne vastavalt karmim, samas kui naiste kõne on vastupidiselt pehmem ja viisakam (Zemaskaja 1993).

6.4. Intensiivistajad

Intensiivistajad on sõnad, mis tugevdavad või nõrgendavad teise sõna omaduse astet (nt *väga*, *veidi*, *päris*) (Aasa 2023: 20). Kõige sagedasemad intensiivistajad, st intensiivistajad, mis esinesid korpuses rohkem kui 50 korda, on *päris*, *üsna*, *absoluutselt*, *eriti*, *palju*, *natukene*, *suhteliselt*, *täitsa*, *suht*, *liiga*, *üldse*, *täiesti*, *natuke*, *hästi*, *tõesti*, *kui*, *nii* ja *väga*. Nagu partiklid, on ka intensiivistajad seotud emotsionaalsusega, nii et nende sagedast kasutamist on seotud naiste keelekasutuse eripäraga (nt Kirilina 2005, Newman jt 2008).

Kuna intensiivistajate puhul osutusid mediaanide väärtused enamikul juhtudel peaaegu võrdseks nulliga ja erinevused ei ületanud 0,05, näib illustratiivne analüüs olevat üleliigne (GitHubis²³). Seetõttu lähen kohe statistiliste testide tulemuste kirjelduse juurde.

Tabelis 5 on kujutatud intensiivistajate statistiliste testide tulemused. Need näitavad, et intensiivistajate *nii*, *üldse*, *tõesti*, *kui*, *palju*, *täesti*, *hästi*, *natukene* ja *väga* kasutamise erinevused meeste ja naiste vahel on statistiliselt olulised. Vastavalt biseriaalse korrelatsiooni väärtustele (veerg *r_{rb}*) kasutavad naised neid sagedamini, mis kajastub koefitsientide negatiivsetes väärtustes. Seose tugevus ulatub keskmisest ($r \approx -0,52$ *nii* puhul) kuni üsna tugevani ($r \approx -0,63$ *tõesti* puhul). Cramer V kriteeriumi kohaselt jääb mõju suurus kõigil olulistel juhtudel väga nõrgaks ($V < 0,1$). Suurimad väärtused on registreeritud *hästi* ($\approx 0,01$) ja *väga* ($\approx 0,01$) puhul, mis näitab leitud erinevuste minimaalset praktilist olulisust.

Seega kasutavad naised keskmiselt veidi sagedamini intensiivistajaid *nii*, *üldse*, *tõesti*, *kui*, *palju*, *täesti*, *hästi*, *natukene* ja *väga*, kuid selle mõju suurus jääb väikeseks. Need intensiivistajad saab paigutada skaalale, mis näitab, kui palju nad sõna tähendust tugevdavad või nõrgendavad. Eesti keele grammatika kirjeldab seda skaalat mitme osaliselt kattuva klassi kaudu (EG 2023: 727–728).

Skaala ülemises, maksimeerivas otsas on sõnad *väga* ja *täiesti* ehk tüüpilised augmentatiivsed intensiivsusadverbiaalid, mis määravad omaduse täieliku astme (EG 2023: 727). Näiteks korpusest:

- *Maailm on **väga** kirju...*
- *..et meil on vaja **täiesti** uut lähenemist selles uues maailmas...*

Skaala vastasküljel asuvad detraktiivid (nt *natukene*), mis teenivad astme minimeerimist (EG 2023: 728). Näiteks korpusest:

- *et sealt võib **natukene** julgemalt neid suuremaid kalu ka süüa...*

²³ https://github.com/SofiaKriuchkova/ma-diploma-project-kriuchkova/tree/main/boxplots/boxplots_intensiivistajad

Nende vahel paiknevad hulgalised tugevust näitavad sõnad nagu *palju*, mida grammatika eristab alamklassi kvantumimäärus, kuna need vastavad küsimusele *kui palju?* (EG 2023: 668; 761). Näiteks korpusest:

- *palju* rohkem plaane on palju selles mõttes...

Lõpuks nihutavad pragmaatilised partiklid nagu *üldse* ja *nii* fookuse kvantitatiivselt skaalalt kõneleja suhtumisele öeldusse (EG 2023: 698, 1068). Näiteks korpusest:

- *kuidas sul on võimalik sellist asja üldse teha?*
- *vaata Eestis on ju meediaturg on ka nii väike*

Nende sõnade roll ei ole ainult semantilise astme väljendamine, vaid nad ühendavad ka laused, annavad märku sõnastuse otsimisest, pehmendavad erimeelsusi ja tagavad sotsiaalse läheduse dialoogis osalejate vahel (EG 2023:766-767, 1068). Intensiivistajate valdav kasutamine naiste kõnes on kooskõlas varasemate sotsiolinguistiliste uurimustega, et naised kasutavad intensiivistajaid sagedamini (Newman 2008, Stratton & Sundquist 2022) ning on samuti rohkem kalduvad keelekasutuse strateegiatele, mis loovad viisakuse ja sotsiaalse läheduse õhkkonna (Posio jt 2024).

Tabel 5. Intensiivistajad. Statistiliste testide tulemused.

Intensii vistaja	p_mann _whitne y	p_whitne y_holm	signif_n orm_hol m	chi2_stat	p_chi2	p_chi2_hol m	cramer_v	r_rb	signif_chi 2
nii	2.00×10^{-15}	3.19×10^{-14}	True	9.40×10^0	2.17×10^{-3}	2.38×10^{-2}	3.83×10^{-3}	-0.5232	True
üldse	1.25×10^{-11}	1.74×10^{-10}	True	1.58×10^1	6.87×10^{-5}	8.93×10^{-4}	4.97×10^{-3}	-0.5413	True
tõesti	8.28×10^{-11}	1.08×10^{-9}	True	3.50×10^1	3.38×10^{-9}	4.73×10^{-8}	7.39×10^{-3}	-0.6261	True
kui	3.67×10^{-17}	6.61×10^{-16}	True	8.02×10^0	4.62×10^{-3}	4.16×10^{-2}	3.54×10^{-3}	-0.5357	True
palju	3.04×10^{-10}	3.65×10^{-9}	True	4.18×10^1	1.03×10^{-10}	1.54×10^{-9}	8.06×10^{-3}	-0.4791	True
eriti	1.52×10^{-8}	1.68×10^{-7}	True	5.69×10^0	1.71×10^{-2}	1.37×10^{-1}	2.98×10^{-3}	-0.6741	False
absoluutselt	1.32×10^{-3}	3.95×10^{-3}	True	3.79×10^0	5.14×10^{-2}	3.07×10^{-1}	2.43×10^{-3}	-0.5029	False
natuke	1.20×10^{-6}	8.14×10^{-6}	True	8.14×10^{-1}	3.67×10^{-1}	1	1.13×10^{-3}	-0.4616	False
päris	2.19×10^{-8}	2.19×10^{-7}	True	2.59×10^{-1}	6.16×10^{-1}	1	6.26×10^{-4}	-0.4477	False
täitsa	1.34×10^{-4}	6.72×10^{-4}	True	4.66×10^0	3.08×10^{-2}	2.16×10^{-1}	2.70×10^{-3}	-0.5340	False
täiesti	7.00×10^{-8}	5.60×10^{-7}	True	8.47×10^0	3.60×10^{-3}	3.60×10^{-2}	3.64×10^{-3}	-0.5065	True
üsna	4.31×10^{-3}	8.62×10^{-3}	True	2.28×10^0	1.31×10^{-1}	5.24×10^{-1}	1.89×10^{-3}	-0.3871	False
hästi	5.57×10^{-13}	8.35×10^{-12}	True	7.19×10^1	2.25×10^{-17}	3.60×10^{-16}	1.06×10^{-2}	-0.5575	True
suht	5.66×10^{-2}	5.66×10^{-2}	False						False
natukene	3.22×10^{-8}	2.90×10^{-7}	True	1.42×10^1	1.63×10^{-4}	1.95×10^{-3}	4.71×10^{-3}	-0.5730	True
liiga	3.12×10^{-6}	1.86×10^{-5}	True	3.16×10^0	7.52×10^{-2}	3.76×10^{-1}	2.22×10^{-3}	-0.5211	False
suhteliselt	4.53×10^{-4}	1.81×10^{-3}	True	9.11×10^{-2}	7.62×10^{-1}	1	3.77×10^{-4}	-0.5333	False
väga	7.00×10^{-16}	1.19×10^{-14}	True	1.14×10^2	1.55×10^{-26}	1.64×10^{-25}	1.33×10^{-2}	-0.4981	True

6.5. KP-markerid

KP-markerid on kirjeldatud kui liikuvate partiklite eriline alamklasskonstruktsiooniga (*ma/mulle*) + verb + (*et*), mida kõneleja kasutab, et näidata, millisest vaatenurgast kuulata järgnevat sisu (EG 2023: 1109–1110). Korpuses esines üle 50 korra markereid nagu (*mulle*) *tundub (et)*, (*ma*) *loodan (et)*, (*ma*) *ütle(n/ks) (et)*, (*ma*) *usun (et)* ja (*ma*) *arvan (et)*.

Kuna KP-markerite puhul osutusid mediaanide väärtused enamikul juhtudel peaaegu võrdseks nulliga ja erinevused ei ületanud 0,05, näib illustratiivne analüüs olevat üleliigne (GitHubis²⁴). Seetõttu lähen kohe statistiliste testide tulemuste kirjelduse juurde.

Tabelis 6 on kujutatud KP-markerite statistiliste testide tulemused. Need näitavad, et (*ma*) *ütle(n/ks) (et)* ja (*ma*) *arvan (et)* markerite kasutamise erinevust naiste ja meeste alamkorpustes on statistiliselt oluline. Vastavalt biseriaalse korrelatsiooni väärtustele (veerg *r_rb*) kasutavad naised neid väljendeid sagedamini, mida tõendavad negatiivsed koefitsiendid. Samas on korrelatsiooni tugevus nõrk ($\approx 0,3$). Cramer V kriteeriumi kohaselt jääb mõju isegi statistiliselt olulistel juhtudel väga nõrgaks ($V < 0,1$). Näiteks (*ma*) *ütle(n/ks) (et)* puhul on see väärtus $\approx 0,007$ ja (*ma*) *arvan (et)* puhul $\approx 0,005$, mis näitab, et nende erinevuste praktiline tähtsus on minimaalne.

Tabel 6. KP-markerid. Statistiliste testide tulemused.

Marker	p_mann_whitney	p_whitney_holm	signif_norm_holm	chi2_stat	p_chi2	p_chi2_holm	cramer_v	r_rb	signif_chi2
(Ma) loodan (et)	1.26×10^{-5}	2.53×10^{-5}	True	3.68×10^{-1}	5.43×10^{-1}	8.28×10^{-1}	7.59×10^{-4}	-0.5828	False
(Mulle) tundub (et)	7.60×10^{-7}	3.04×10^{-6}	True	6.67×10^{-1}	4.14×10^{-1}	8.28×10^{-1}	1.02×10^{-3}	-0.4627	False
(Ma) arvan (et)	3.01×10^{-4}	3.01×10^{-4}	True	3.50×10^1	3.35×10^{-9}	1.68×10^{-8}	7.39×10^{-3}	-0.2716	True
Ütle me	2.66×10^{-6}	7.99×10^{-6}	True	1.67×10^1	4.82×10^{-5}	1.93×10^{-4}	5.08×10^{-3}	-0.3631	True
(Ma) usun (et)	2.57×10^{-9}	1.29×10^{-8}	True	1.25×10^0	2.64×10^{-1}	7.93×10^{-1}	1.39×10^{-3}	-0.6943	False

²⁴ https://github.com/SofiaKriuchkova/ma-diploma-project-kriuchkova/tree/main/boxplots/boxplots_kp

KP-markerid võimaldavad väljendada kõneleja kindlusastet öeldud väite suhtes (EG 2023: 1109–1110). Vähim kindlus on väljendatud märgendiga *mulle tundub (et)*: see distantseerib autorit ja pehmenab kategoriseerimist, sest *tundub* lubab alati eksimise võimalust. Näide korpusest:

- *Ja nüüd arenguga **tundub**, et kõik on palju keerulisem...*

Veidi rohkem kindlust näitab *ma arvan (et)* marker, see näitab arvamuse individuaalsust, aga avab võimaluse vastuväidete esitamiseks. Näide korpusest:

- ***Ma arvan**, et see on nagu üks nendest võtmetes, aga mitte ainult kindel...*

Kui *arvan* asemel esineb *ma usun (et)*, suureneb kindlusaste. See ei ole enam oletus, vaid veendumus, järeldus, mille eest autor on valmis vastama. Näide korpusest:

- *Et **ma usun**, et see kõige parem viis teha...*

Pööre emotsiooni suunas toimub markeriga *ma loodan (et)*: sama tähendust väljendub verb *loodan*, nii et propositioon saab soovitud, kuid veel mitte saavutatud seisundi värvingu. Näide korpusest:

- ***Ma loodan**, et see ei ole selline.*

Lõpuks viib marker *ma ütlen (et)* väljenduse kõne toimimise selgitamise tasandile: kõneleja annab ette teada, et ütleb nüüd midagi olulist, haarates sellega initsiatiivi ja tugevdades järgneva sõnumi kaalu. Näide korpusest:

- *Ja aga **ütlen** siis aitäh juba äkki või?*

Üldjoontes pehmenavad need markerid väidet. Naised kasutavad KP-markereid (*ma*) *ütlen(n/ks) (et)* ja (*ma*) *arvan (et)* keskmiselt veidi sagedamini, kuid selle mõju suurus jääb väikeseks. See on siiski kooskõlas teiste keelte puhul tehtud uurimustega, et naised kalduvad rohkem KP-markereid oma väidete pehmendamiseks (nt Kirilina 2005, Namaziandost & Shafiee 2018).

6.6. TTR

TTR (ingl. *type-token ratio*) on teksti sõnavara mitmekesisuse lihtne kvantitatiivne näitaja, mis on kujunenud üheks sagedasti kasutatavaks näitajaks sõnavara varieeruvuse hindamisel (Brezina 2018: 57-58). Mitmes keeles, sealhulgas vene ja inglise keeles, on täheldatud, et meeste tekstide TTR on tavaliselt kõrgem kui naistel. Seda nähtust on seletatud meeste sagedasema neologismide ja erialase leksika kasutamisega. (Litvinova jt 2017)

Siinses töös võrdlesin ka alamkorpuste sõnavara mitmekesisust. Meeste alamkorpuse TTR on 0,06, samas kui naiste alamkorpuse TTR on 0,05. Meeste kõne üldmaht on suurem (43 tundi vs. 33,5 tundi). Pikemate tekstide TTR peaks olema madalam korduvate funktsioonisõnade arvu suurenemise tõttu (Brezina 2018: 58), kuid siin täheldasin vastupidist. Täiendavad kontrollid kahe episoodi puhul meeste korpusest ja kahe episoodi puhul naiste korpusest, mis olid teemade ja kestuse poolest identsed, näitasid samuti meeste suuremat TTR-i. Kinnisvara episoodides on TTR meeste puhul 0,26 ja naiste puhul 0,19. Armastuse episoodides on TTR meeste puhul 0,17 ja naiste puhul 0,16.

Tulemused näitasid seega mitmekesisemat sõnavara meeste korpuses, kuid konkreetsemaid järeldusi on veel liiga vara teha. Tuleb uurida, mis konkreetset mõjutab tulemust, kuidas korpused leksikaalselt erinevad, ning teha täiendavaid teste teksti suuruse suhtes vähem tundlike mõõdikutega. Täiendavaid teste ei tehtud siinses töös uurimuse mahu tõttu, kuid tulevased testid selgitavad täheldatud erinevuse olemust.

6.7. Tulemused hüpoteeside valguses

Kokkuvõtteks kinnitavad statistilise analüüsi tulemused enamikku sissejuhatuses esitatud hüpoteese. Statistiliselt olulised erinevused esinesid enamiku uuritud keeleliste tunnuste puhul (hüpotees 1). Naised kasutasid märgatavalt sagedamini omadussõnu, adverbe, partikleid, intensiivistajaid ja deminutiive ning viisakusega seotud konstruktsioone (sh tingiv kõneviis ja KP-markerid). See vastab hüpoteesidele 4, 5 ja 7. Samuti leidis kinnitust, et naised kasutasid sagedamini 1. ja 2. isiku pronoomenit, mis võib viidata isiklikumale ja kaasavamale keelekasutusele (hüpotees 3). Siiski said hüpoteesid 3 ja 4 ainult osaliselt kinnitust. Ootus, et mehed kasutavad sagedamini umbisikulist tegumoodi ning opereerivad rohkem verbide ja nimisõnadega, ei leidnud veenvat kinnitust - statistilisi erinevusi ei leitud.

Hüpotees suuremast sõnavara mitmekesisusest meeste keelekasutuses (hüpotees 2) on saanud esialgset kinnitust. TTR oli meestel kõrgem nii üldiselt kui ka teemalt ja kestuselt võrreldavate episoodide puhul. Arvestades selle näitaja tundlikkust teksti pikkuse suhtes, on siiski vaja täiendavaid teste. Meeste kõne suurema otsekoheuse hüpotees (hüpotees 6) ei leidnud kinnitust. Mehed kasutavad mõnikord rohkem eitust, kuid keskmiselt esineb eitust oluliselt sagedamini naiste korpuses. Sarnane on pilt ka käskiva kõneviisiga: vastupidiselt ootustele osutus käskiv kõneviis iseloomulikumaks naiste keelekasutusele.

Üldiselt on tulemused suures osas võrreldavad varasemate soolingvistika tulemustega. Naiste keelekasutust iseloomustab rikkalikum ja varieeruvam stiililiste ja pragmaatiliste vahendite kasutamine, samas kui meeste keelekasutus on ühtlasem ja sisaldab üksikuid tippusid mõnede konstruktsioonide kasutamisel. Eriti märgatav on eraldatus ekspressiivsete ja modaalsete elementide, nagu intensiivistajate, deminutiivi *-kene*, partiklite ja KP-markerite kasutamisel. (Brown & Levinson 1978; Kirilina 2005; Konstantinovskaia 2020; Newman jt 2008; Senkina jt 2018; Stratton & Sundquist 2022) Tulemuste kohaselt kasutavad sõna *jumal* peamiselt naised ja *kurat* eranditult mehed. See on kooskõlas mõttega, et mehed kasutavad pigem karmimaid ja järsemaid vandesõnu, samas kui naised kasutavad neutraalsemaid variante (Jürine jt 2016; Pajusalu 2006; Gauthier 2017).

Oluline on märkida, et vaatamata leitud erinevuste statistilisele olulisusele jääb mõju suurus (Crameri V) enamikul juhtudel väga väikeseks. See piirab nende praktilist asjakohasust ja osutab vajadusele edasiste analüüside järele, milles võetakse arvesse selliseid tegureid nagu vanus, register ja suhtluskontekstid (Argamon jt 2003; Litvinova jt 2017).

Kokkuvõte

Siinses magistritöös oli eesmärgiks välja selgitada, kas ja millised on erinevused meeste ja naiste keelekasutuses. Lisaks püüdis uurimus välja selgitada, kas leitud erinevused on kooskõlas teiste keelte kohta tehtud varasemate uurimuste tulemustega.

Nendele küsimustele vastamiseks koostasid Eesti Taskuhäälingukorpuse põhjal (Lippus jt 2023) kaks temaatiliselt tasakaalustatud alamkorpust, mis sisaldasid ainult neid episoodide, milles osalesid ainult mehed (umbes 43 tundi kõnet) või ainult naised (umbes 33,5 tundi). See võimaldas minimeerida vead soo määramises ja tagada tingimuste võrreldavus. Tagasin temaatilise tasakaalu meeste ja naiste alamkorpuse vahel, et kõrvaldada žanri ja temaatiliste tegurite mõju.

Esimesele uurimisküsimusele, kas meeste ja naiste kõnekäitumises on erinevusi ja millised need on, sain jaatava vastuse. Analüüsid (Mann-Whitney U-test ja χ^2 -test koos Holm-Bonferroni korrigeerimisega) näitasid, et meeste ja naiste keelekasutuse vahel on tõepoolest statistiliselt olulised erinevused mitmete leksikaalsete ja grammatiliste tunnuste osas. Naised kasutasid sagedamini omadussõnu, adverbe, partikleid, intensiivistajaid, viisakuskonstruktsioone (sh tingivat kõneviisi ja KP-markereid) ning deminutiivi. Kuigi mõju suurus jäi kõigil eespool nimetatud juhtudel väikeseks, on tulemused siiski kooskõlas hüpoteesidega, et naiste kõne on väljendusrikkam, emotsionaalsem ja isikupärasem. Meestel oli sõnavara mitmekesisus (TTR) veidi suurem, kuid umbisikulise tegumoe, käskiva kõneviisi ja eituse oodatud ülekaal nende keelekasutuses ei leidnud kinnitust. Veelgi enam, need vormid olid naiste keelekasutuses veidi sagedasemad, mis võib olla seotud taskuhäälingutele iseloomulike kommunikatiivsete rollidega (nt saatejuhi roll).

Teisele uurimisküsimusele, kas tuvastatud tunnused on kooskõlas teiste keelte kohta tehtud varasemate uurimuste tulemustega, võib samuti vastata jaatavalt. Tulemused kinnitavad suures osas inglise, vene ja teiste keelte kohta tehtud uurimuste järeldusi (Argamon jt 2003, Czerwionka jt 2023, Gauthier 2017, Newman jt 2008, Konstantinovskaia 2020, Stratton & Sundquist 2022, Stückler 2022). Mõned tunnused (nt käskiva kõneviisi ja eituse kasutamine) ei lange siiski kokku muude keelte tulemustega, mis võib viidata žanriliste tegurite mõjule.

Neid tulemusi tuleb siiski tõlgendada ettevaatusega, sest uurimisel on mitmeid piiranguid. Esiteks on valim žanripetsiifiline: taskuhäälinguid ei saa täielikult laiendada igapäevastele

dialogidele või avalike-ametlikele vestlusele. Teiseks, automaatse transkriptsiooni puhul peame arvestama tuvastusvigu, eriti haruldaste lekseemide ja neologismide puhul, mis võivad mõjutada sageduste arvutamist.

Edasiste uurimuste väljavaadete hulka kuuluvad korpuse laiendamine igapäevaste dialogide, raadiosaadete ja muude suuliste žanrite lisamine väljaspool meediakeskkonda, mitmemõõtmeliste statistiliste mudelite kasutamine, mis võimaldavad samaaegselt kontrollida teemat, žanri, vanust ja sotsiaalset staatust, ning kultuuridevaheliste võrdluste tegemine teiste soome-ugri ja Euroopa keelte materjalidega.

Kirjandus

Aare, K. & Lippus, P. 2017. Some gender patterns in Estonian dyadic conversations. In *Nordic prosody. Proceedings of the XIIth conference*, Trondheim (pp. 29-38).

Aasa, T., 2024. *Omadussõnade intensiivistamine eesti noortekeeles kahe korpuse võrdluses*. Tartu Ülikool, eesti ja üldkeeleteaduse instituut. Magistritöö.

Ackermann, T. 2023. Mitigating strategies and politeness in German requests. *Journal of Politeness Research*, 19(2), 355-389. <https://doi.org/10.1515/pr-2021-0034>

Annuk, E. 2015. Soodiskursustest Eestis nõukogude perioodil. *Ariadne Lõng*, 15(1-2), 70-89.

Annuk, E., & Seigel, M. 2020. Pregnancy, Motherhood And/as/or Dissent: The Soviet Micro-rhetorics of Gender, *Rhetoric Review*, 39:4, 502-521, <https://doi.org/10.1080/07350198.2020.1805575>

Argamon jt = Argamon, S., Koppel, M., Fine, J., & Shimon, A. R. 2003. Gender, genre, and writing style in formal written texts. *Text & talk*, 23(3), 321-346. <https://doi.org/10.1515/text.2003.014>

Baker, P. 2014. *Using corpora to analyze gender*. Bloomsbury Publishing. 20-45

Berryman, C. L. & Wilcox, J. R. 1980. Attitudes toward male and female speech: Experiments on the effects of sex-typical language. *Western Journal of Speech Communication* 44(1). 50–59.

Bogdanova-Beglarian jt = Bogdanova-Beglarian, N., Sherstinova, T., Blinova, O., & Martynenko, G. 2019. Pragmatic markers distribution in Russian everyday speech: Frequency lists and other statistics for discourse modeling. In *International Conference on Speech and Computer* (pp. 433-443). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-99579-3_5

Brezina, V. 2018. *Statistics in corpus linguistics: A practical guide*. Cambridge University Press.

Brown, P. & Levinson, S. 1978. Universal in language usage: politeness phenomena, pp. 56—289 in *Goody, E. (ed.) Questions and Politeness*. Cambridge University Press.

Butler, J. 2002. *Gender trouble*. Routledge. <https://doi.org/10.4324/9780203824979>

Cabin, R. J. & Mitchell, R. J. 2000. To Bonferroni or not to Bonferroni: when and how are the questions. *Bulletin of the Ecological Society of America*, 81(3), 246–8.

Cheng jt = Cheng, N., Chandramouli, R., & Subbalakshmi, K. P. 2011. Author gender identification from text. *Digital investigation*, 8(1), 78-88. <https://doi.org/10.1016/j.diin.2011.04.002>

Coates, J. 2004. *Women, men and language*. New York: Pearson/Longman. <https://doi.org/10.4324/9781315645612>

Cremades, E. 2024. "Authorship Analysis, Social Networks and Catalan," *Tertium Linguistic Journal*, 9(1).

Czerwionka jt = Czerwionka, L., Staszkiwicz, B., & Shamloo, F. 2023. Contextual variables as predictors of verb form: An analysis of gender and stance in Peninsular Spanish requests. *Languages*, 8(3), 202. <https://doi.org/10.3390/languages8030202>

DeepL. 2025. DeepL Translator. Kasutatud 26. mai 2025, aadressil <https://www.deepl.com>

Dementieva, T. 2021. Lingvokulturoložičeskaja spetsiifika funktsionirovanija lingvističeskikh sredstv vyraženiija kategorii vežlivosti v russkom i kitajskom jazykah [Linguocultural specificity of functioning of linguistic means of expressing the category of politeness in Russian and Chinese languages]

Devinney jt = Devinney, H., Björklund, J., & Björklund, H. 2022. Theories of “gender” in nlp bias research. In *Proceedings of the 2022 ACM conference on fairness, accountability, and transparency* (pp. 2083-2102). <https://doi.org/10.1145/3531146.3534627>

Dibrova, E. V. 2013. Identifikacija gendernoj prinadležnosti avtora po lingvističeskim osobennostjam produciruemogo im teksta. [Identifying an author's gender from the linguistic features of the text he or she produces.] *Meždunarodnyi naučno-issledovatel'skij žurnal*, (11-2 (18)), 102-103.

Dressler, W. & Barbaresi L. M. 1994. Morphopragmatics. Diminutives and Intensifiers in Italian, German and Other Languages. – Berlin: Mouton de Gruyter.
<https://doi.org/10.1515/9783110877052>

Edelsky, C. 1976. Subjective reactions to sex-linked language. *The Journal of Social Psychology* 99(1). 97–104. <https://doi.org/10.1080/00224545.1976.9924752>

Eagly jt = Eagly, A. H., Woo, W., & Diekman, A. B. 2012. Social role theory of sex differences and similarities: A current appraisal. In *The developmental social psychology of gender* (pp. 123-174). Psychology Press.

EG = Metslang, Helle, Mari Erelt, Külli Habicht, Tiit Hennoste, Reet Kasik, Pire Teras, Annika Viht, Eva Liina Asu, Liina Lindström, Pärtel Lippus, Renate Pajusalu, Helen Plado, Andriela Rääbis, Ann Veismann 2023. Eesti grammatika. Tartu Ülikooli Kirjastus.

EKG II 1993 = Erelt, Mati, Reet Kasik, Helle Metslang, Henno Rajandi, Kristiina Ross, Henn Saari, Kaja Tael, Silvi Vare 1993. Eesti keele grammatika II. Süntaks. Lisa: kiri. Tallinn: Eesti Teaduste Akadeemia Keele ja Kirjanduse Instituut.

EKK = Erelt, Mati, Tiiu Erelt, Kristiina Ross 2007. Eesti keele käsiraamat. Kolmas, täiendatud trükk. Tallinn: Eesti Keele Sihtasutus.

European Commission 202. Directorate-General for Justice and Consumers, 2025 report on gender equality in the EU, Publications Office of the European Union.
<https://data.europa.eu/doi/10.2838/5262357> (vaadatud 01.05.2025)

Fishman, P. M. 1980. Conversational insecurity. In *Language* (pp. 127-132). Pergamon.

Fuscone jt = Fuscone, S., Favre, B., & Prévot, L. 2021. Reproducibility in speech rate convergence experiments. In *Language Resources and Evaluation*, 55 (3), 817–832. <https://doi.org/10.1007/s10579-021-09528-6>

Gauthier, M. 2017. *Age, gender, fuck, and Twitter: A sociolinguistic analysis of swear words in a corpus of British tweets*. PhD diss., Université Lumière Lyon 2.

Gorisch, J., & Schmidt, T. 2024. Evaluating workflows for creating orthographic transcripts for oral corpora by transcribing from scratch or correcting ASR-output. In *Proceedings of the*

2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024) (pp. 6564-6574).

Haas, A. 1979. Male and female spoken language differences: Stereotypes and evidence. *Psychological Bulletin* 86(3). 616–626. <https://doi.org/10.1037/0033-2909.86.3.616>

Hennoste, T. 2000. Sissejuhatus suulisse eesti keelde II. Suulise kõne erisõnavara I. *Akadeemia*, 6, 1343-1374.

Hennoste, T., Klumpp, G., & Metslang, H. 2022. Diskursusemarkerid ja pragmaatika. *Keel ja Kirjandus*, 2022(1).

Hennoste jt = Hennoste, T., Prillop, K., Habicht, K., Metslang, H., Laanesoo, K., Pärismaa, L., ... & Simmul, C. E. 2022. Komplementausega predikaatidel põhinevate diskursusemarkerite kasutus eri registrites. *Keel ja Kirjandus*, 2022(1).

Ito, R. & Tagliamonte S. 2003. Well Weird, Right Dodgy, Very Strange, Really Cool: Layering and Recycling in English Intensifiers. In *Language in Society* 32, 2, 257–279. <https://doi.org/10.1017/S0047404503322055>

Jakobson, R. 2012. Lingvistika ja poeetika. Tõlk. Neeme Lopp ja Arne Merilai.)–*Akadeemia*, 10, 1731-1773.

Jürine jt = Jürine, A., Pajusalu K., Pajusalu R., Tragel I., & Veismann A. 2016. Estonian Jumal „God“. – Folklore: *Electronic Journal of Folklore* 64. Toim Mare Kõiva, Andres Kuperjanov, Hille Pajupuu, Renate Pajusalu, Mari Uusküla. Tartu, 9–32.

Kasik, Reet 2006. Feministlik lingvistika. *Oma Keel* 13, 5–12

Kaukonen jt = Kaukonen E, Oskolskaia P, Lindström L & Marling R 2025. Gender, language and labour: gender perception of Estonian and Russian occupational titles. *Front. Commun.* 9:1454022. doi: 10.3389/fcomm.2024.1454022

Keevallik, L. 2003. *From interaction to grammar: Estonian finite verb forms in conversation.* PhD diss., Uppsala University, Uppsala.

Keevallik, L. 2008. Conjunction and sequenced actions: The Estonian complementizer and evidential particle et. In *Crosslinguistic studies of clause combining: The multifunctionality of conjunctions* (pp. 125-152). John Benjamins Publishing Company.

Keevallik, L. 2016. Estonian no (o)(h) in turns and sequences: Families of function. *NU/NÄ: A family of discourse markers across the languages of Europe and beyond*, 58, 213.

Kehayov, P. 2009. Taboo intensifiers as polarity items: evidence from Estonian. *Language Typology and Universals*, 62(1-2), 140-164. <https://doi.org/10.1524/stuf.2009.0009>

Kirilina, A.V. 2005. *Gender i Jazõk*. [Gender and Language] Litres.

Kluyver = Kluyver, T., Ragan-Kelley, B., Fernando P., Granger, B., Bussonnier, M., Frederic, J., ... & Willing, C. 2016. Notebooks – a publishing format for reproducible computational workflows. In *F. Loizides & B. Schmidt (Eds.), Positioning and Power in Academic Publishing: Players, Agents and Agendas* (pp. 87–90). <https://doi.org/10.3233/978-1-61499-649-1-87>

Kolsar, K. 2017. *Partikkel okei suulises argisuhtluses*. Tartu Ülikool, eesti ja üldkeeleteaduse instituut. Bakalaureusetöö.

Kolyaseva, A. 2022. The divergent paths of pragmaticalization: The case of the Russian particles *tipa* and *vrode*. *Journal of pragmatics*, 201, 181-196. <https://doi.org/10.1515/ling-2021-0004>

Konstantinovskaia, N., Konstantinovskaia, N. 2020. Russian and Japanese Women's Real Language Practices. *The Language of Feminine Beauty in Russian and Japanese Societies*, 163-191.

Kramer, C. 1977. Perceptions of female and male speech. *Language and Speech* 20(2). 151–161. <https://doi.org/10.1177/002383097702000207>

Labov, W. 1973. *Sociolinguistic patterns* (No. 4). University of Pennsylvania press.

Lakoff, Robin. 1973. *Language and woman's place*. *Language in Society* 2(1). 45–79.

Liivak, M. 2023. *ke(ne)-liitelised deminutiivid eesti suulises argisuhtluses*. Tartu Ülikool, eesti ja üldkeeleteaduse instituut. Magistritöö.

Lippus jt = Lippus, P., Alumäe T., Orasmaa S., Pivik M.-L., & Lindström L. 2023. Eesti taskuhäälingukorpus. <https://doi.org/10.23673/re-445>

Lippus jt = Lippus, P., Pilvik, M. L., Lõo, K., & Lindström, L. 2024. Kõnetempo ja-soravuse varieerumine eesti keeles. *Eesti Rakenduslingvistika Ühingu aastaraamat*, 20, 149-163.

Litvinova jt = Litvinova, T., Seredin, P., Litvinova, O., & Zagorovskaya, O. 2017. Differences in type-token ratio and part-of-speech frequencies in male and female Russian written texts. In *Proceedings of the Workshop on Stylistic Variation* (pp. 69-73). <https://doi.org/10.18653/v1/W17-4909>

Maltz, D. N., & Borker, R. A. 2018. A cultural approach to male-female miscommunication. In *The matrix of language* (pp. 81-98). Routledge.

Marling, R., & Annuk, E. 2023. Gender Perspectives in the Study of Estonian Society and Culture: Current State and Challenges. *Folklore: Electronic Journal of Folklore*, 90, 137-154.

Mayer jt = Mayer, F., Auschel, E., Baumbach, M., & Schindler, K. 2024. Corporate Communication Through Podcasts: The Impact of Voice-Fit and Attention During Listening on Parasocial Interaction with the Host and Persuasive Effects. *Journal of Radio & Audio Media*, 1-28. <https://doi.org/10.1080/19376529.2024.2396840>

MacDougall, R. 2009. Our new aural ecologies: Podcasting, publicity, and secondary orality. Paper presented at *Media in Transition 6: Stone and Papyrus, Storage and Transmission*, Massachusetts Institute of Technology, Cambridge, MA.

McKinney, W., ja teised. 2010. Data structures for statistical computing in python. In *Proceedings of the 9th Python in Science Conference* (Vol. 445, pp. 51–56).

Mills, A. 2018. *Biology of sex*. University of Toronto Press.

Mulac A. 1998. The gender-linked language effect: do language differences really make a difference?;

Namaziandost, E., & Shafiee, S. 2018. Gender differences in the use of lexical hedges in academic spoken language among Iranian EFL learners: a comparative study. *International Journal of Research in English Education*, 3(4), 63-80. <http://dx.doi.org/10.35313/jbit.v6i2.2303>

Newman jt = Newman, M. L., Groom, C. J., Handelman, L. D., & Pennebaker, J. W. 2008. Gender Differences in Language Use: An Analysis of 14,000 Text Samples, *Discourse Processes*, 45(3), pp. 211–236. <https://doi.org/10.1080/01638530802073712>

Olev, A. & Alumäe, T. 2022. Estonian Speech Recognition and Transcription Editing Service. In *Baltic Journal of Modern Computing*, kd 10, nr 3, lk 409–421.

OpenAI. 2025. ChatGPT (GPT-4o) [LLM]. <https://chat.openai.com>

Orasmaa jt = Orasmaa, S., Petmanson, T., Tkachenko, A., Laur, S., & Kaalep, H.-J. 2016. EstNLTK – NLP toolkit for Estonian. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation*

Pajusalu, K. 2000. Eesti keele allkeeled (Vol. 16). T. Hennoste (Ed.). Tartu Ülikool.

Pajusalu, R. 2006. jumal ja kurat argikeeles. – Kuradi tark jumala loll. Toim Enn Kasak, Anne Kull. Tallinn: *Eesti Päevalehe Kirjastus*, 17–32.

Pajusalu, R. 2010. Teie või sina?. *Oma keel*, 5-10.

Pajusalu jt = Pajusalu, R., Vihman, V., Klaas, B., & Pajusalu, K. 2010. Eestlaste ja venelaste suhtluskäitumine: sina, teie ja keegi veel. *Eesti Rakenduslingvistika Ühingu aastaraamat*, 6, 207-224.

Pedaja-Ansen, A., 2015. Võrdevormid tänapäeva eesti keeles. *Oma Keel*, 2, 9–17. Copy.

Pépiot, E. 2014. Male and female speech: A study of mean f₀, f₀ range, phonation type and speech rate in Parisian French and American English speakers. – Proceedings: Speech Prosody 7, May 2014, Dublin, Ireland, 305–309. <https://doi.org/10.21437/SpeechProsody.2014-49>

Petron, E. 2023. *Omadussõnad ja intensiivistajad eesti eakate keeles*. Tartu Ülikool, eesti ja üldkeeleteaduse instituut. Magistritöö.

Pickering, M. J., & Garrod, S., 2004. The interactive-alignment model: Developments and refinements. *Behavioral and Brain Sciences*, 27(2), 212-225.

Pilkington, J. 1992. ‘‘Don’t try and make out that I’m nice!’ The Different Strategies Women and Men Use when Gossiping’ in *Wellington Working Papers in Linguistics*. 5, pp.37-60.

Pilvik, M.-L. (ilmumas). Korpusandmete statistiline analüüs. Teoses L. Lindström, A. Veismann, M.-L. Pilvik, K. Muischnek & J. Padrik (Toim.), *Korpused lingvistilises uurimistöös. Praktiline käsiraamat*. Tartu Ülikooli Kirjastus.

Posio jt = Posio, P., Kachel, S., & Uclés-Ramada, G. 2024. Morphosyntactic stereotypes of speakers with different genders and sexual orientations: an experimental investigation. *Linguistics*, 62(6), 1581-1625. <https://doi.org/10.1515/ling-2022-0143>

Rangel, F., Rosso, P. 2013. Use of language and author profiling: identification of gender and age. In *Proceeding of the 10th workshop on natural language processing and cognitive science (NLPCS 2013)*.

Rime jt = Rime, J., Pike, C., & Collins, T. 2022. What is a podcast? Considering innovations in podcasting through the six-tensions framework. *Convergence*, 28(5), 1260-1282. <https://doi.org/10.1177/135485652211044>

Ross, J. 2018. Aira Kaalust Mari Saadini. Nõukogude Eesti naisarenguromaan ja selle lugemisviisid. [From Aira Kaal to Mari Saat. Soviet Estonian Women’s Bildungsroman and Its Reading Modes.] Tartu: Tartu Ülikooli Kirjastus.

Seabold, S., & Perktold, J. 2010. statsmodels: Econometric and statistical modeling with python. In *9th Python in Science Conference*.

Senkina, D. 2018. How Mothers and Fathers Address their Sons and Daughters in Russian and Vice Versa: A Quantitative Study. Higher School of Economics Research Paper No. WP BRP, 69.

Siregar jt = Siregar, L. S., Manurung, L. W., Silitonga, H. H., & Naibaho, T. 2023. Refusal Speech Act Used By Male And Female Sellers At Pasar Sambu Medan. *Innovative: Journal Of Social Science Research*, 3(2), 4648-4657.

Soundararajan, S., & Delany, S. J. 2024. Investigating Gender Bias in Large Language Models Through Text Generation. In *Proceedings of the 7th International Conference on Natural Language and Speech Processing (ICNLSP 2024)* (pp. 410-424).

Stratton, J. M. & Sundquist J.D. 2022. A Variationist Sociolinguistic Analysis of Intensifiers in Oslo Norwegian. – *Journal of Germanic Linguistics* 34, 4, 385–419. <https://doi.org/10.1017/S1470542722000022>.

Stückler, L. 2023. „Das is ja wohl eh klar.“: Eine Fallstudie zu Abtönungspartikeln und ihrer Verwendung in gesprochener Sprache. *Linguistik online*, 123(5).

Zemskaja jt = Zemskaja, E. A., Kitajgorodskaja, M. V., & Rjazanova, N. N. 1993. Osobennosti mužskoj i ženskoj reči [Feauters of feminine and masculine speech']. In E. A. Zemskaja & D. N. Shmelev (Eds.), *Russkij jazyk v ego funktsionirovanii* (pp. 90-136). Moscow: Nauka.

Tampel, I. B., & Karpov, A. A. 2016. Avtomatičeskoe raspoznavanie reči. [Automatic Speech Recognition] *Učebnoe posobie. Saint Petersburg: Universitet ITMO*.

Teras, P. 2019. Sõnaalgulise/h/hääldus samadel kõnelejalatel avalikus ja argisuhtluses. Eesti ja soome-ugri keeleteaduse ajakiri. *Journal of Estonian and Finno-Ugric Linguistics*, 10(1), 211-231.

Toots, L. 2023. *LGBT+ kogukonnaga seotud sõnavara Eesti LGBT+ aktivistide taskuhäälingutes* Tartu Ülikool, eesti ja üldkeeleteaduse instituut. Bakalaureusetöö.

Tucker, S. 1961. *English Examined*. Cambridge University Press.

Van Rossum, G., & Drake, F. L. 2009. *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace.

Differences in the Spoken Language Use of Estonian-Speaking Men and Women

Although some research in Estonian linguistics addresses gender as a factor in linguistic variation, investigations into gender-related language differences remain fragmented and have yet to develop into a distinct field of study. This master's thesis seeks to address this gap by examining the linguistic differences in the language use of Estonian-speaking men and women, and by assessing the extent to which these differences align with findings from similar studies in other languages.

In this master's thesis, I have formulated the following research questions:

- 1) Do men and women differ in their language use, and if so, how?
- 2) If such differences exist, do they resemble those found in previous studies conducted on other languages?

To address these questions, I propose the following hypotheses, based on the findings of comparable studies in other languages:

- 1) Lexical and morphosyntactic differences can be observed in the spoken language use of Estonian men and women.
- 2) Male speakers tend to use a more varied vocabulary (Argamon et al., 2003; Litvinova et al., 2017).
- 3) Female speech is more personal, characterized by more frequent use of first- and second-person pronouns. Male speech, on the other hand, is more impersonal or distanced, with more frequent use of the passive voice or impersonal verb forms (Argamon et al., 2003).
- 4) Female speech tends to be more emotional, as reflected in the more frequent use of adjectives, particles, and intensifiers, and more descriptive, as indicated by the greater use of adverbs. Male speech is more task- and goal-oriented, reflected in the higher frequency of nouns and verbs (Konstantinovskaia, 2020; Newman et al., 2008; Stratton & Sundquist, 2022; Stückler, 2022).
- 5) Women use more politeness markers, including conditional mood and predicate-based complement clause constructions (hereafter referred to as PM-markers), such as "I think that...", "I'm afraid that...", "It seems that..." (Brown & Levinson, 1978; Czerwionka et al., 2023; Kirilina, 2005).

- 6) Men's speech is more direct, with more frequent use of negation and the imperative mood (Czerwionka et al., 2023; Dibrova, 2013).
- 7) Women tend to use diminutives more frequently (Konstantinovskaia, 2020; Senkina et al., 2018).

The study was conducted using the Estonian podcast corpus, from which two thematically balanced sub-corpora were created: one consisting exclusively of male speech (approximately 43 hours) and the other of female speech (approximately 33.5 hours). The results of statistical tests (Mann-Whitney U-test and χ^2 -test with Holm-Bonferroni correction) showed significant differences in the use of lexical and grammatical features. Women were more likely to use adjectives, adverbs, particles, intensifiers, polite constructions and diminutives. While men demonstrated a slightly greater lexical variety, as indicated by a higher type-token ratio (TTR), they did not exhibit a notable overuse of imperatives, negation, or impersonal constructions.

The results are broadly in line with studies in English, Russian and other languages, although some deviations from general trends may be caused by the genre of the data used. The study demonstrates that gender can influence the use of linguistic devices in Estonian and highlights the need for further research in this area.

Lihtlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks

Mina, Sofia Kriuchkova,

1) annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) enda loodud teose „Erinevused eesti keelt kõnelevate meeste ja naiste suulises keelekasutuses“, mille juhendajad on Liina Lindström ja Kristiina Vaik, 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 Commonsi litsentsiga CC BY NC ND 3.0, 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.

Sofia Kriuchkova

26.05.2025