

# KAS NOORED ON INGLISE KEELEGA OBSESSED? MILLEST RÄÄGIVAD KORPUSANDMED?

Virve-Anneli Vihman, Kristiina Praakli,  
Maarja-Liisa Pilvik, Mari-Liis Korkus  
Tartu Ülikool

**Annotatsioon.** Artiklis analüüsime inglise keele kasutust projekti „Teismeliste keel Eestis“ (TeKE)<sup>1</sup> raames 9–18-aastaste noorte suulisest ja täätisuhtlusest kogutud keeleandmetes. Andmeid koguti rahvateaduse meetodil. Nagu TeKE korpuse sagedusandmed ja neist kooruvad mustrid viitavad, tasub teismeliste keelekasutuses pöörata tähelepanu inglise keelele. Artiklis esitame esmase analüüsi inglise keele kasutusest ja osakaalust projekti kaasatud noorte suulise ja täätisuhtluse andmetes. Tulemused osutavad, et inglise keele osakaal varieerub kõnelejati, ei ole seostatav soo ega vanusega, vaid sõltub pigem muudest vestlusega seotud asjaoludest, sh vestlusteemast ja registrist. Küll aga kooruvad soolised ja vanuselised erinevused välja noorte ingliskeelse sõnavara analüüsist.

**Võtmesõnad:** noortekeel, suuline keel, täätikeel, korpusanalüüs, eesti keel, inglise keel

## I. Sissejuhatus

Dominantse maailmakeelena on inglise keel võtnud viimastel aastakümnetel ka Eestis varasemast teistsuguse rolli. Suhtluskeskkondi, kus inglise keel on nähtav ja kuuldav, on oluliselt juurde tulnud veebis, meedias ja avalikus ruumis. Inglise keel on enim õpitud võõrkeel üldhariduskoolides; kõrgharidustasemel peetakse

---

<sup>1</sup> Projekti „Teismeliste keel Eestis“ (EKKD3) rahastab riiklik programm „Eesti keel ja kultuur digiajastul“ (EKKD).

inglise keelt pigem baasoskuseks kui võõrkeeleks (vt nt Tartu ülikooli keelepõhimõtted 2009<sup>2</sup>). Nii veebis kui väljaspool on inglise keel atraktiivne ning produktiivne suhtlust vahendav keel (vt lähemalt Zabrodskaia, Kask 2017; Praakli, Koreinik 2020). Lisaks inglise keele ainetundidele üldhariduskoolis saavad noored mitmekeelset keelelist sisendit neile tähenduslikest suhtlusvõrgustikest, sh sotsiaalmeediast. Nende lähedane suhe inglise keelega annab neile juurdepääsu üleilmale.

Eestlaste ja eriti noorte väga head inglise keele oskust tõstetakse esile eri raportites (nt EF English Proficiency Index<sup>3</sup>). Teisalt kostub Eesti avalikkuses arvamusi inglise keele negatiivsest mõjust emakeele oskusele ning keelelisele lohakusele. Nii väitis kultuuriajakirja Akadeemia peatoimetaja 2019. aasta arvamusefestivalil, et inglise keele õpetamine koolides tuleks üleüldse lõpetada, sest „see on kui umbrohi, see tuleb inimestele ise külge“ (Voltri 2019). Muu hulgas on meedias hämmeldust väljendatud noorte inglise keelest mõjutatud keelekasutuse üle (nimetatud ka *Estonenglish*'iks), nt: „*Estonenglish* jäi igal juhul kõrva kriipima ja ilusast emakeele kasutamisest on selline kõnepruuk kaugel“ (Kull 2019). Eesti keele seisundi uuring (Lukk *et al.* 2017) jällegi vahendas emakeeleõpetajate muret noorte eestikeelse sõnavara muutumise ja kahanemise ning täiskasvanute keelepruugist eemaldumise üle, teisalt tõdedes, et põhjalikud uurinud selle kohta siiski puuduvad.

Noorte spontaanset suhtlust käsitlevaid uuringuid Eesti keeleteaduses napib. Noorte keelepruugile pole teadlastel vanuselistel ja eetilistel asjaoludel ka sugugi lihtne ligi pääseda. Sestap on oluline ja teretulnud noorte kui enda keeles vaieldamatult asjatundjate aktiivne kaasalöömine noortekeele uurimisel, sh nii andmekorjes kui analüüsimisel. Eesti keeleteaduses ongi suurem osa noortekeele uuringuist kirjutatud eeskätt (üli)õpilaste teadustööde vormis

---

<sup>2</sup> Inglise keele oskus on eristatud teiste võõrkeelte oskusest: „[...] akadeemilises maailmas on inglise keel kujunenud *lingua franca*'ks.“ Vt <https://www.ajakiri.ut.ee/artikkel/2923> (04.05.2022).

<sup>3</sup> EF Education First. <https://www.ef.com/wwen/epi/> (04.05.2022).

(nt Ruven 1999; Rannamäe 2006; Happonen 2010; Lehtpuu 2016). Ka õpilaste teadustööde riiklikule konkursile<sup>4</sup> (2002–2022) esitatud keelealastest uurimistöödest nähtub, et üks noori kõnetavaid teemasid on eakaaslaste sõnavara, mida on vaadeldud nii suulise suhtluse, veebikeskkondade kui kiirsuhtlusäppide keeleandmete põhjal.

„Teismeliste keel Eestis“ (edaspidi TeKE; 2019–2022)<sup>5</sup> projekt on eesti noortekeele uurimisloos esimene, mis analüüsib laiapõhjaliselt korpusandmete toel inglise keele kasutust noortevahelises suhtluses. Projekti raames kogusime keeleandmeid 9–18-aastaste noorte omavahelisest suulisest ja tähtsuhtlusest<sup>6</sup>. Siinses artiklis otsime vastuseid järgmistele küsimustele:

1. Kui palju kasutavad eesti emakeelega teismelised ingliskeelseid sõnu omavahelises eestikeelses vestluses?
2. Milliseid sõnu kasutatakse enim?
3. Kas ja kuidas erineb inglise keele kasutus suulisest ja tähtsuhtluses?
4. Kas ja kuidas erineb teismeliste inglise keele kasutus vanuse ja soo lõikes?

Järgmises peatükis võtame vaatluse alla noorte keele uurimise ning koodivahetuse kui keskseima mitmekeelse suhtluse nähtuse noortekeeles, eriti praeguses meediastunud maailmas, kus just tehnilistel platvormidel on noorte suhtluses üha suurem roll. Kolmandas peatükis kirjeldame ainekku ja meetodit ning loodavat korpus. Neljas peatükk tutvustab uuringu tulemusi: anname ülevaate

<sup>4</sup> Eesti Teadusagentuur. <https://etag.ee/tegevused/konkursid/opilaste-teadustoode-konkurss/varasemad-konkursid/> (08.08.2022).

<sup>5</sup> Vt <https://sisu.ut.ee/teke/>.

<sup>6</sup> Siin ja edaspidi räägime kitsamalt *tähtkeelest* ja *-suhtlusest* – see on üks veebikeele variante, mille aines pärineb kiirsuhtlusäppide vahendusel peetud vestlustest (vrd netikeele mõistega, mis pakuti välja ajal, kui tehniliselt vahendatud suhtlus toimus jututubades, e-kirjades ja/või kommentaarides; vt Hennoste 2000). Veebikeel on heterogeenne keelevariant, mis osaliselt sõltub rakenduste tehnoloogilistest funktsioonidest ning teksti- ja vooruvahetuse kiirusest (vt McCulloch 2019; ka Androutsopoulos 2015).

inglise keele osakaalust suulistes ja tähtvestlustes, koodivahetuse avaldumisest ning sagedasematest ingliskeelsetest sõnadest. Järgnevad arutelu ja kokkuvõte.

## 2. Millest kõneleb noorte keelekasutus?

Teismelistelt kogutud keeleandmed kui väljavõtted noorema vanuserühma keelelisest käitumisest peegeldavad keele varieerumist, varasemate uurimisandmetega kõrvutavalt ka võimalikke keele muutusi, nt sõnavaras; teisalt ka ühiskonnas toimuvaid protsesse. Noorte keelekasutust kirjeldatakse märksõnadega nagu innovaatiivne, loov, trenditeadlik (vt Eckert 1997, 2000; Jonsson *et al.* 2019). Noorte keeleline minapilt saab olulisi täiendusi ja mõjutusi murdekeele, mil muutub olulisemaks just eakaaslastelt saadav sisend. Nende keelekasutuse põhjal saab aimu neid kõnetavate prestiižsete keelte, suhtluses aktsepteeritud mitmekeelsete suhtlusnähtuste ja noori kõnetavate vestlusteemade kohta.

Ühiskondlikus kontekstis peegeldavad noorte keelelised repertuaarid nii kohalikke kui üleilmselt haaratud muutusi ja protsesse. Praegusaja noorte puhul hõlmab meedia ja meediastumine (Hepp 2018) nende kõiki igapäevapraktikaid: kooliskäimist, vaba aja veetmist ning omavahelist läbikäimist. Kooliealistel on COVID-19 pandeemia ja sellega kaasnenud e-õppe rohkus seda tendentsi veelgi süvendanud. Eeskätt meediastumisest ning aktiivsest sotsiaalmeedias toimetamisest tulenevalt saavad noored keelelist sisendit ka üle ajaaliste ja ruumiliste vahemaade neile tähenduslikest võrgustikest ning vahendavad seda enda võrgustikesse. Noorte hea inglise keele oskus toetab võrgustikele juurdepääsu, ühtlasi saades neilt toetust ja sisendit. Meediastumine ja platvormide kättesaadavus toob noorte keelepruugi nähtavamale ja kuuldamamale kui eales varem. Siiski pole noortekeele puhul põhjust kõnelda diskreetselt keelekujust, kuna see on varieeruv ja kiiresti muutuv ning sisaldab erinevaid kontaktmõjusid.

Teismeliste keelepruugis avalduv suhtlusvahendite mitmekesisus nähtub noortekeele uurimustes (vt nt Rampton 1995; Kern,

Selting 2011; Nortier, Svendsen 2015; Tagliamonte 2016). Suhtluse sotsiolingvistika raamistikus on eeskätt Euroopa suurlinnade näitel uuritud noorte mitmekeelseid keelekujusid ning suhtlusnähtusi (nt *code-switching*, *crossing*), sageli ühendatult teise keele õppimise ja omandamise küsimustega (vt nt Jørgensen 2002, 2008; Källström, Lindberg 2011; Lehtonen 2015; Jonsson *et al.* 2019). Omaette valdkonna moodustavad nn slängisõnavara uuringud, sh ka tabu- ja vandesõnade uuringud (vt Stenström 2020). Digiajastu on suulise keele kõrval toonud fookusesse suhtluspraktikad kiirsuhtlus- ja mängurakendustes (vt Durkin *et al.* 2010).

Eesti noortekeele kogumine ja teaduslik uurimine sai alguse eelmise sajandi esikümnenditel, kuid järjepidev uurimistöö selles vallas on siiski lünklik, põhiosa moodustavad gümnasistide ja üliõpilaste uurimused (vt ülevaated Tender 1994; Hennoste 2000; Praakli 2022). Uurijate rõhuasetus on põhiliselt lasunud noorte sõnavaral (lähemalt Praakli, ilmumas), kus keelekontaktidest johtuvaid keelemuutusi on näha laensõnades ja mitmekeelsetes suhtluspraktikates (nt Loog 1988, 1990). Alates nullindatest on üha rohkem tähelepanu pälvinud inglise keele kasutus, seda nii suuliste (intervjuud, argivestlused, vlogid; nt Lehtpuu 2016; Ratt 2017; Korkus 2021) kui digisuhtluse andmete (tšätid, blogid; nt Kask 2021) valguses.

## 2.1. KOODIVAHETUSE UURIMISEST

Mitmekeelsusuuringute fookus lasub kahe või enama keele(variandi) kooskasutamise ja suhtluskontaktidest johtuvate muutuste kirjeldamisel. Üks mitmetest kontaktsituatsioonidest ilmnevatest ja noortekeelele iseloomulikest nähtustest, mis ka siinse artikli keskmes, on koodivahetus. Põhitähenduses tähistab koodivahetus mitmekeelset suhtlusviisi, kus lausungis või vestluses kasutatakse vähemalt kahte keelt või keelevarianti (Verschik 2004; Myers-Scotton 2017).

Koodivahetusuuringud on üldiselt keskendunud kas selle sotsio-pragmatilistele funktsioonidele (nt Auer 1998) või lausungisiselt eri keele elementide morfosüntaktilise integreerimise küsimustele

(nt Myers-Scotton 1993; Deuchar 2020). Artiklis käsitleme koodivahetusilminguid eeskätt leksikaalsest vaatest, keskendudes sisestatud ingliskeelsetele elementidele eestikeelses diskursuses.

Koodivahetuse kirjeldamisel on oluline eristada, milline on vestluskontekstis põhi-, milline sisestuskeel (vt nt Myers-Scottoni 1993 maatrikskeele mudel). Siinses artiklis loeme keeleanndmetest lähtuvalt vestluste põhikeeleks vaikumisi eesti keele. Poplack (1980) eristab koodivahetusilmingute puhul kolme põhitüüpi. Esimeseks on lausesisene koodivahetus (ingl *intrasentential CS*), kus keeltevaheline üleminek toimub ühe lausungi või lause sees (näide 1<sup>7</sup>). Teise tüübina eristab Poplack lauseülest koodivahetust (ingl *intersentential CS*, ka *code alternation*), kus keeltevaheline üleminek toimub lausungite või lausete vahel (näide 2<sup>8</sup>). Viimasena eristab Poplack süntaksivälise diskursuspragmaatiliste üksuste (nt partiklite, hüüdsõnade) ülekannet (ingl *tag-switching*). Need paiknevad tavaliselt lausungi või lause alguses (näide 3) või lõpus (näide 2, *right* 'õigus, eks') mistahes positsioonis.

1. **CHA10M:** *right clicki* mu nime peale
2. **CHA14M\_1:** su emakeel on eesti keel *right?*  
**CHA14M\_2:** *nah* hispaania keel  
**CHA14M\_1:** *sirius*  
**CHA14M\_2:** *ofc its* eesti keel  
**CHA14M\_1:** *what languses u speak and how bad or good*  
**CHA14M\_2:** inglise keel suht hea  
**CHA14M\_2:** vene keel suht sitt
3. **SUU14N\_1:** ja siis kõik arvasid et me k- ee nagu meeldime üksteisele ja käime  
**SUU14N\_1:** aga me (.) me ei käinud

<sup>7</sup> Igas näites sisaldab lausungile eelnev kood korpus (SUU/CHA), kõneleja vanust, sugu ja (vajadusel) järjenumbrit. Kaldkiri tähistab sisestuskeele elemente.

<sup>8</sup> Siinne näide ilmestab muuhulgas vajadust noorte keelehoiakute väljendamise lähemaks analüüsiks.

SUU14N\_2: *vau[\_wow\_]* tõesti vä[või]

SUU14N\_1: jaa (...) jah

SUU14N\_2: *oo\_mai\_kaad[\_oh\_my\_god\_]* ma arvasin et sa nagu täiega käisid temaga

Kvantitatiivne lähenemine võimaldab eristada suure hulga inglise keele näidete põhjal ka kasutatavate keelte määra. Kuna lühikeste lausungite puhul on väiksem võimalus kasutada kahte keelt koos, siis eristame andmeanalüüsis kuni kahesõnalisi (lühikesi) lausungeid pikematest.

Morfofonoloogilise tasandi kirjeldamisel on mitmekeelsusuringuis rakendatud ka Lars Johanson'i koodikopeerimise mudelit (Johanson 2002; Verschik 2008; Kask 2021). Mudel eristab kolme kopeerimistüüpi: täielikud (ingl *global copy*; näide 4, milles ingliskeelsed sõnad on üle kantud terviklikul kujul, häälduse ja tähendusega), valikulised (ingl *selected copy*; näide 5a–b, milles kõnelejad kasutavad ingliskeelsete sõnade omadusi valikuliselt ning ortograafia viitab eestipärasele hääldusele) ja segakooptiad (ingl *mixed copy*; näide 6, milles liitsõna sisaldab mõlema keele elemente):

4. SUU16N: see oli päris *lol[\_LOL\_]* ikka  
SUU16N: aga no ma olin et *jolo[\_YOLO\_]*
5. a. CHA17M: valgus *is veri good*  
b. CHA17N: *Yesyes iz uki dunt võrri* [it's ok don't worry]
6. CHA16N: nagu ma realselt tegelen õppimisega ja siis unustan täiesti ära, et *livetunnid* on kaa

Koodivahetuses esinevate keelte eristamine toetub traditsioonilisele vaatele, mille kohaselt keeled on eraldiseisvad süsteemid. Uuemaid kontseptsioone (nt *metrolingualism*, *translanguaging*, *superdiversity*, *polylingualism*; vt tähenduslikest erinevustest Pennycook 2016) ilmestab arusaam, et mitmekeelse kõneleja keeled moodustavad ühe süsteemi, samuti ei sõltu koodivahetus kõneleja ühe või teise keele oskusest, vaid tema suhtluseesmärkidest (Androutsopoulos 2015). Seda väidet toetab nähtus, kus noored kasutavad eri keelte sõnu või

väljendeid ilma, et nad keelt valdaksid, nt inglise keelt kõnelevad ameerika noored hispaaniakeelset sõna *mamacita* või eesti noortekees vene keelest mugandatud *taipohh*.

Artiklis loome ülevaatliku kvantitatiivse aluse, millest hiljem liikuda lähemale kvalitatiivsele analüüsile. Kõik ülal nimetatud koodivahetuse avaldumisvormid on kaasatud ka siinse artikli kvantitatiivsesse analüüsi, koodivahetuse struktuuralse integreerimise ja eri tüüpide analüüs järgneb edasistes uurimustes.

## 2.2. (SOTSIAAL)MEEDIASTUNUD KEELE REGISTRID<sup>9</sup>

Kaasaja noortekultuuris mängib olulist rolli sotsiaalmeedia.<sup>10</sup> Noored on selle regulaarsed tarbijad eelkõige nutiseadmete kaudu (Lauricella *et al.* 2014). Selle valguses nimetab Siibak tänapäeva noori mitte digi- vaid sotsiaalmeediapõlvkonnaks (2020: 30), kuna nooremates vanuserühmades on aktiivne sotsiaalmeedia kasutus kujunenud sotsiaalseks normiks ning põlvkondliku identiteedi vahendiks.

Just üleilmastumisel (Bucholtz, Skapoulli 2009; Otsuji, Pennycook 2010) ja meediastumisel (Leppänen 2007) on noorte keelekasutuses keskne roll. Nende mõju nähtub mitmekihiliselt:

1. Mitmeregistrilisuus: kasutatakse akronüüme jm veebipõhiseid keelendeid (nt *OMG* ehk 'oh my god', *XD* 'iksdee' ehk visuaalselt tekkinud naerunägu, *DM* 'dii emm' ehk 'direct message'; vt ka Crystal 2008);
2. Mitmekeelsus: kasutatakse noortekees üleilmselt tuntud sõnu või väljendeid (nt *hot girl summer* 'kuuma tüdruku suvi', *on fleek* 'tasemel, ideaalne', *vibe* 'hõng, meeleolu') ning jälgendatakse nende häädust (vrd nt sõna *meme* erinevaid

---

<sup>9</sup> Mõistame *registri* all kasutuskeskseid varieeruvaid keelekujusid, keelevariante; siinse artikli raamides suulist argikõnet ning tšäitikeelt.

<sup>10</sup> Võib ka väita, et sotsiaalmeedia mõju on laiem ning läbinähtav eri ühiskonna kihtides, seetõttu räägitakse ka tänapäeval platvormiühiskonnast (ingl *Platform Society*; vt lähemalt van Dijck *et al.* 2018).

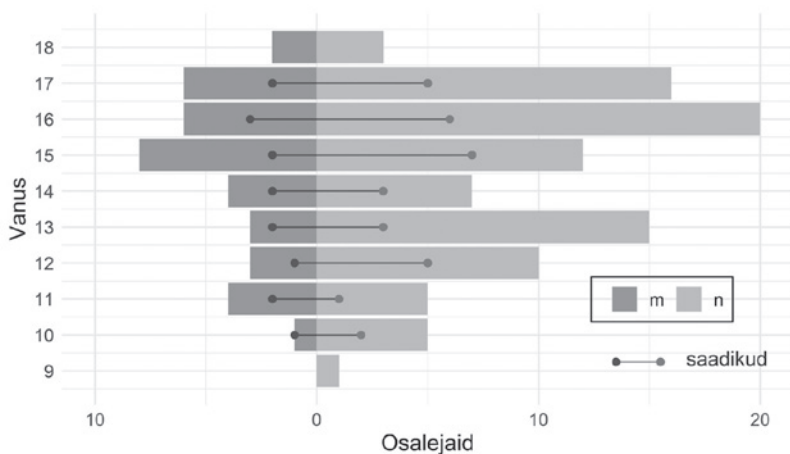


hääldusvariante: eestipärane ‘meem’, ortograafiaühine ‘meme’, ingliskeelne ‘miim’);

3. Mitmeallikalisus: viidatakse erinevatele allikatele ja tsiteeritakse filmides/raamatutes/sarjades/veebikeskkondades kuulnud-nähtut (vt Sierra 2021).

Veebivõrgu laiemasse kasutusse jõudmine ning infoajastu tehnilised võimalused on laiendanud suhtluskeskkondi, neis rakendatavaid vahendeid (nt e-kirjad, foorumipostitused, kommentaarid, *online*-mängude sõnumivahetused jms) ning mitmekesistanud suhtlus- ja tekstiloome viise. Ühtlasi on tehniliselt vahendatud suhtlus loonud keelekontaktidele uue pinnase ning muutnud seeläbi kontaktis olemise kättesaadavust, viise ja võimalusi. Veebikontaktid haaravad laiemat kõnelejaskonda, kontaktid on intensiivsemad, suhtlusesse kaasatakse erinevaid keeli ning muid elemente (emotikonid, videolingid jm). Teaduslik huvi veebisuhtluse ning selle keelevariantide vastu sai alguse 1990ndatel (nt Ferrara *et al.* 1991; Donath *et al.* 1999; Crystal 2001). Ka eesti veebisuhtluse kohta kirjutati esimesed uuringud samal ajal (nt Salla 1999), hilisemalt on – peamiselt üliõpilastööde vormis – vaadeldud nii eri keskkondade suhtlusvahendeid kui ka mitmekeelset suhtlust nt blogides-vlogides, tsättides (nt Igav 2013; Kilp 2017; Ratt 2017; Kask 2021).

Nii nagu suulises keeles, esineb ka tsätikeeles üsna palju varieeruvust (vt lähemalt McCulloch 2019). Tagliamonte ja Denis (2008) toovad välja, et kiirsuhtlusvestlustes kombineerivad noored formaalseid (nt kirjakeelseid vorme *shall* ja *must*) ja mitteformaalseid (nt emotikonid, lühendid) keelelemente. Platvormide mitmekülgne funktsionaalsus võimaldab kasutajatel korraga tegutseda mitmes rakenduses ning sellest tulenevalt on nende eneseesitlus ja tähendusloome platvormiti erinev (Deakin, Wakefield 2014; Boczkowski *et al.* 2018; Tagg, Lyons 2021). Uurimishuvi veebikeele (ja kitsamalt ka tsätikeele) vastu on jõudsalt kasvamas, eriti võib seda märgata (üli)õpilasuuringustes (ehk noored uurivad noori; vt ka 2).



**Joonis 1.** Kõigi projektis osalenud noorte vanus ja sugu (keelesaadikud ja kaasatud sõbrad)

### 3. Ainestik ja meetod

Analüüsivad keeleandmed kogusime projekti „Teismeliste keel Eestis“ raames. Selle põhieesmärk on luua noortekeele korpus, mis võimaldab uurida noorte spontaanset keelekasutust suulistes ja tähtivestlustes. Selleks kutsusime eri piirkondade üldhariduskoolide 4., 6., 8. ja 10. klassides õppivaid õpilasi projekti nn keelesaadiku-teks.<sup>11</sup> Vastava koolituse läbinuna kaasasid keelesaadikud projekti enda sõpru, salvestasid nendega suulisi vestlusi (keskmiselt kaks tundi saadiku kohta), lisaks kogusid ja vahendasid uurimisrühmale tähtides peetud jutujamisi. Kõik artiklis kasutatavad keeleandmed ja -näited pärinevad TeKE korpusest.

Kokku osales andmekorjes 131 noort vanuses 9–18 eluaastat. Joonis 1 annab ülevaate korpuse vanuselisest ja soolisest koosseisust. Kuigi sihtisime vanuselist, soolist ja piirkondlikku tasakaalu, on valimis ülekaalukalt naissoost osalejaid (94 tüdrukut, 37 poissi).

<sup>11</sup> Projekti rahvateaduse meetodil põhinevast andmekorjest lähemalt vt Mandel *et al.* (2022).

Keelesaadikud pärinesid Tallinna, Tartu, Saaremaa ja Võrumaa koolidest, kuid nende kaasatud sõprade kaudu on projektis esindatud 11 maakonda (lisaks üks välisriigis elav õpilane).

### 3.1. KORPUSTE KIRJELDUS

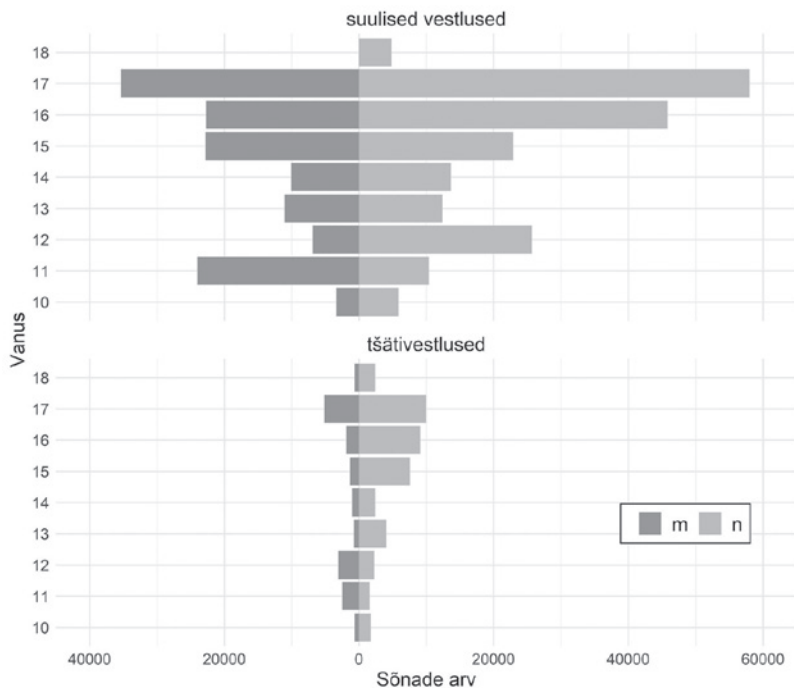
Andmekorjega kogusime 97 tundi helisalvestusi ning 111 521 sõna mahus tšätte (valdavalt Messengerist, mõnevõrra Discordist). Lisaks küsisime noortelt veebiküsitluse kaudu nende (võõr)keeleoskuse ja -kasutuse ning internetitarbimise kohta. Kogutud materjalist on artikli kirjutamise hetkel (mai 2022) suulistest andmetest transkribeeritud kujul analüüsivalmis 64 kõneleja suulised vestlused (kokku 39 tundi, 335 822 sõna<sup>12</sup>) ning tšätivestlused 97 osalejalt (mahus 59 688 sõna<sup>13</sup>). Keskmise sõnade arv kõneleja kohta suulistes vestlustes on 5247 ( $sd = 3058$ ), tšätivestlustes 585 sõna ( $sd = 531$ ), ent kõnelejatevaheline varieeruvus on väga suur. Valimi sõnade arvu vastavalt kõneleja soole ja vanusele kuvab joonis 2.

Suuliste vestluste helifailide transkribeerimiseks kasutasime annotatsioonitarkvara ELAN<sup>14</sup>. Iga lausungi puhul märgistasime selle piirid, transkribeerisime iga kõneleja kõnet eraldi transkriptsiooniridadel ning märgendasime kõnekeelsusi ja kasutatud keele. Kõik isikuandmed ja muu isiku tuvastamist võimaldav materjal on muudetud. Keelejuhtide kõne transkribeerimisel lähtusime kohendatud kirjakeele ortograafia põhimõtetest: kirjutasime üles kõik keeleüksused, sh üneemid, häälightsused, poolikud sõnad, oletatavad sõnad (st sõnad, mille täpne kuju pole eristatav, kuid on kontekstist võimalik oletada). Kirjakeele ortograafiat ei kasutatud juhtudel, kui kõneleja keelepruuk sellest süsteemselt erines: näiteks üldises kõnekeeles laialt levinud (nt *keka*, *õps*, *põhimõtst*, *kakskend*) või kirja-keelest lahknevad (nt *videosid*, *praegult*) vormid. Sellistel puhkudel

<sup>12</sup> Sõnadeks on arvatud ka (nt eneseparanduste tõttu) pooleli jäänud sõnad, üneemid ja partiklid.

<sup>13</sup> Sõnadeks on arvatud ka lingid, pildid, emotikonid ja (sõnumi)reaktsioonid.

<sup>14</sup> ELAN (versioon 6.2). <https://archive.mpi.nl/tla/elan> (03.05.2022).



**Joonis 2.** Valimi sõnade arv vastavalt kõneleja soole ja vanusele

lisasime sõna järele nurksulgudesse sõna kirjakeelse vormi (nt *kakskend*[*kakskümmend*], *praegult*[*praegu*]), seda eeskätt järgmises etapis rakendatava automaatse morfoloogilise analüüsi tarvis. Nurksulud ja alakriipsud eristavad võõrkeelseid sõnu ja pikemaid võõrkeelseid fraase (nt *dõup*[*\_dope\_*], *õu\_mai\_gaad*[*\_oh\_my\_god\_*]). Eraldi keelereal on märgendatud teise keele kasutust (nt inglise, vene, hispaania; ka eesti murded). Kõik transkribeeritud failid vaatas üle, parandas ja ühtlustas veel vähemalt üks transkribeerija.

Tšätiaimestiku edastasid keelesaadikud Wordi-failidena. Nendest eemaldasime sõnumisaatja nime sisaldavad read (nt *Maarja sent*), samuti muutsime ära tekstides esinevad isikunimed ning isiku tuvastamist võimaldava sisu. Kõnelejate voorud eristasime failis eri taustavärvidega. Wordi-failid konverteerisime xml-failideks, mis

võimaldas html-värvikoodide abil eristada üksteisest eri kõnele-  
jate voorud ning sisuteksti (sh pildid, lingid, emotikonid, (sõnumi)  
reaktsioonid) igasugusest metainfost (kuupäevad, profiilipildid, tea-  
vitustekstid; nt *You received a call from*). Eraldatud info koondasime  
tabelkujul tsv-failidesse, mille ridadel oli iga tšäti tekstirida ning  
tulpades seda iseloomustav metainfo (sisutekst/metatekst, kuupäev  
ja kellaeg, osaleja kood). Sisutekstile lisasime märgendustööriista  
INCEpTION<sup>15</sup> (Klie *et al.* 2018) abil info võõrkeelte, lühendite  
(nt *mdea* ‘ma ei tea’, *idk* ‘I don’t know’) kasutuse ning konfident-  
siaalse teabe kohta.

### 3.2. ÜLEVAADE OSALEJATEST

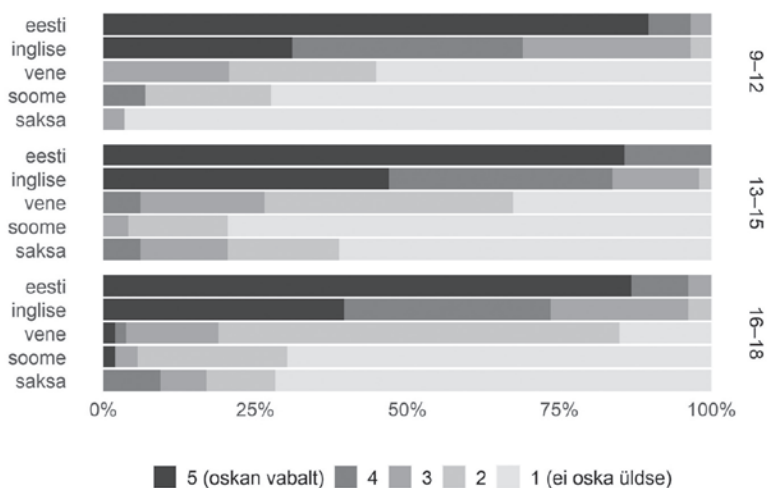
Projektis osalenud noorte taustainfo kogusime veebiküsimustiku  
kaudu. Siinses artiklis on relevantset noorte antud hinnangud oma  
keeleoskuse kohta (vt joonis 3) ning kiirsuhtlusäppide kasutushar-  
jumused (vt joonis 4).

Keeleoskuse andmed põhinevad vastajate hinnangutel. Selleks  
paluti märkida viiepalliskaalal, kuidas nad iga keele puhul oma kee-  
leoskust hindaksid. Ootuspäraselt hindavad kõik vanuserühmad  
kõige kõrgemaks eesti keele oskust. Ka inglise keele oskusele antud  
hinnang on väga kõrge (viiepalliskaalal hinded 4 või 5), st noored  
on arvamusel, et nad oskavad inglise keelt „vabalt“ või „väga hästi“. Ühtlasi näeme, et inglise keele oskuse hinnangud sarnanevad pigem  
eesti keele kui teiste küsitud võõrkeelte (vene, soome, saksa) oskuste  
hinnangule. Inglise keele väga hea oskus võib olla seotud mitme  
teguriga, ent põhilisena nähtub tugev seos osalusmeediaga (vt Siibak  
2020), kus märkimisväärne osa argisuhtlusest ning rahvusvahelisest  
noortekultuurist on ingliskeelne.

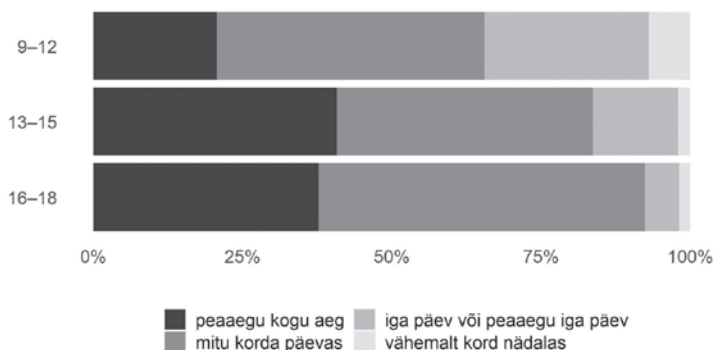
TeKE projektis osalenud teismeliste vastustest nähtus, et ena-  
mik suhtleb kiirsuhtlusäppides kas peaaegu kogu aeg või mitu korda  
päevas, seejuures on kasutus tihedam vanemate osalejate (13–15- ja

---

<sup>15</sup> INCEpTION (versioon 22.4). <https://inception-project.github.io> (03.05.2022).



**Joonis 3.** Osalejate hinnangud enda keeleoskusele



**Joonis 4.** Osalejate vastused küsimusele „Kui sageli suhtled kiirsuhtlusäppides?“

16–18-aastaste) seas. Vaid marginaalne osa (alla 5%) vastanuist kasutab suhtlusäppe harvem kui kord päevas. Populaarseimad kiirsuhtlusäpid on Messenger (54%), Snapchat (27%) ja Discord (10%), viimane eelkõige poiste hulgas. TeKE projektile edastatud sõnumi-vestlused peeti valdavalt Messengeris, väiksem osa Discordis.

## 4. Tulemused

### 4.1. INGLISE KEELE OSAKAAL SUULISTES JA TŠÄTIVESTLUSTES

Kvantifitseerisime koodivahetuse intensiivsust noortekeeles võrdlemisi lihtsal viisil ning hindasime seda kui ingliskeelsete sõnade osakaalu<sup>16</sup> kõigist kõneleja sõnadest. Olgu öeldud, et see, mis on korpuses märgitud inglise keeleks, on kohati vaieldav, kuna hulk ingliskeelseks loetud sõnu on eesti keeles võrdlemisi hästi kohanevad (nt *sorri*, *tšätt*, *vaib*). Korpuses on märgendatud võõrkeeli pigem inklusiivselt, et koodivahetuse uurijal oleks võimalik luua andmete alamhulki enda vajadustest lähtuvalt.

Ingliskeelsete sõnade protsenti võib siinjuures pidada võrdlemisi konservatiivseks hinnanguks koodivahetuse intensiivsusele, kuna suulistes vestlustes on sõnadeks loetud kõik semantilised ja grammatilised sõnad, lisaks üneemid, poolikud sõnad jm suulise kõne komponendid, millest osadel keelemärgendit ei ole. Tšättides omakorda lähevad sõnadena arvesse ka muud semantilist sisu kandvad elemendid, nagu pildid, emotikonid ja lingid, millel samuti keelemärgendid puuduvad.

Leidsime, et suulistes vestlustes on ingliskeelsete sõnade protsent keskmiselt 3,3, tšätivestlustes aga märgatavalt kõrgem – 7,3. Eesti- ja ingliskeelsete sõnade osakaalu soo ja vanuserühmade lõikes illustreerib tabel 1.

Ingliskeelsete sõnade esinemise kohta selgeid mustreid ei soo ega vanuserühmade lõikes välja ei joonistu. Võib eeldada, et noored omandavad inglise keelt juba varakult (nagu viitab joonis 3) ning näeme, et ingliskeelsete sõnade osakaal vanusega ei kasva. Vanusest olulisemat rolli näivad mängivat kõneleja individuaalsed valikud

---

<sup>16</sup> Edasised analüüsid põhinevad suuliste vestluste transkriptsioonide lausungi- ja sõnakihil, kus võõrkeelsed sõnad on eristatud eestikeelsetest, ent milles konkreetse keele kohta infot pole. Transkriptsioonide keelekihi põhjal oleme aga teinud kindlaks, et võõrkeele kasutusjuhtudest üle 99% moodustab just inglise keel, seega räägime siin ja edaspidi vaid inglise keelest. Sama kehtib tšätivestluste kohta.

**Tabel 1.** Eesti- ja ingliskeelsete sõnade osakaal soo ja vanuserühmade võrdluses

Vanus aastates ja sugu		Suulised vestlused		Tšätivestlused	
		ee % (N)	en % (N)	ee % (N)	en % (N)
9–12	m	97% (33 317)	3% (941)	90% (5586)	10% (587)
	n	95% (39 699)	5% (2198)	91% (5103)	9% (480)
	<b>Kokku</b>	96% (73 016)	4% (3139)	91% (10 689)	9% (1067)
13–15	m	95% (41 653)	5% (2307)	82% (2877)	18% (649)
	n	98% (48 037)	2% (991)	95% (13 359)	5% (656)
	<b>Kokku</b>	96% (89 690)	4% (3298)	93% (16 236)	7% (1305)
16–18	m	97% (56 081)	3% (1934)	93% (7144)	7% (518)
	n	97% (105 847)	3% (2817)	93% (21 232)	7% (1497)
	<b>Kokku</b>	97% (161 928)	3% (4751)	93% (28 376)	7% (2015)
<b>Kokku</b>		96,7% (324 634)	3,3% (11 188)	92,7% (55 301)	7,3% (4387)

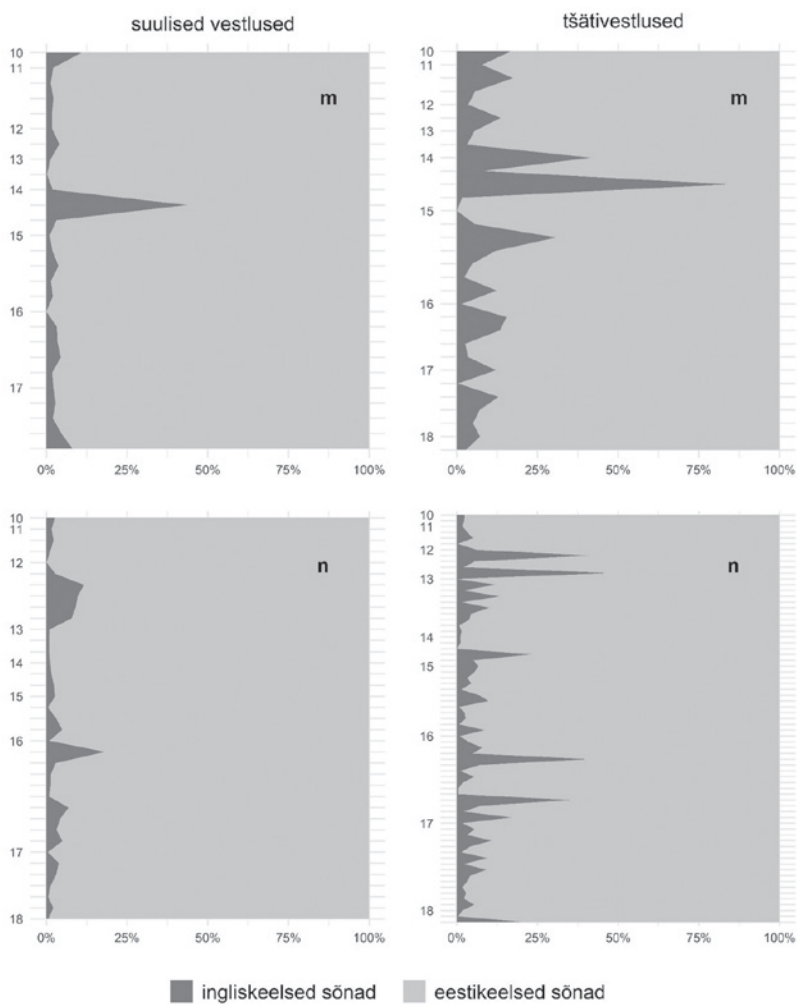
(nt veebikeskkondades veedetud aeg, suhtlusvõrgustik jne), vestlusteema ja -olukord. Joonis 5, kus on esitatud eesti- ja ingliskeelsete sõnade suhtelised sagedused, illustreerib võõrkeelse keeleainese osakaalu ulatuslikku varieerumist üksikkõnelejate lõikes, aga ka registriti.

#### 4.2. LAUSESISENE JA LAUSEÜLENE KOODIVAHETUS

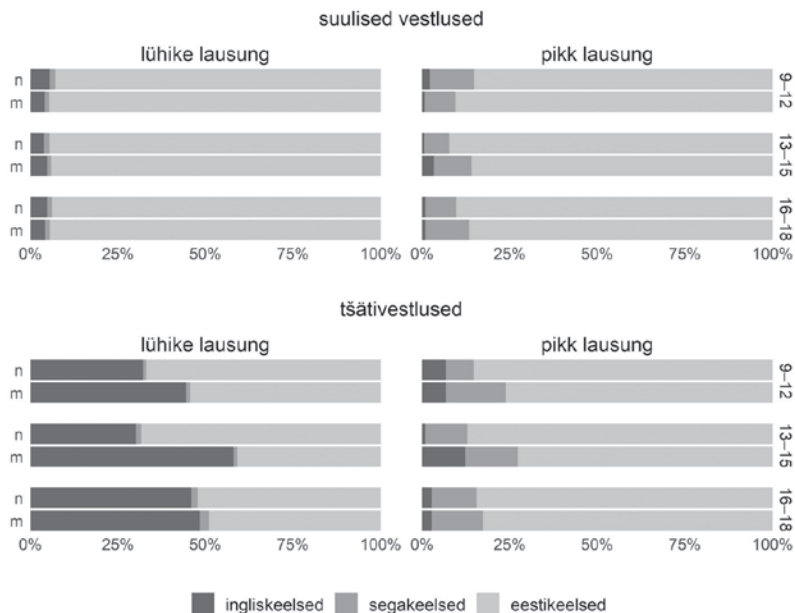
Lisaks ingliskeelsete sõnade osakaalule huvitusime ka sellest, kui võrd ulatuslikult esineb teismeliste suhtluses lausesisest<sup>17</sup> ja lauseülest koodivahetust. Selleks eristasime täielikult eestikeelsed lausungid, täielikult ingliskeelsed lausungid ning lausungid, milles esineb mõlema keele elemente. Suulistes vestlustes on eestikeelseid lausungeid 90,3%, ingliskeelseid 2,3% ja segakeelseid 7,4%;

<sup>17</sup> Siia alla liigitub ka diskursuspragmatiliste üksuste ülekanne, nn sildivahetus (ingl *tag-switching*), mida meie valitud meetod ei võimaldanud eristada.





**Joonis 5.** Inglis- ja eestikeelsete sõnade suhtelised sagedused individuaalsete kõnelejate keelekasutuses



**Joonis 6.** *Inglisekeelsete, eestikeelsete ja segakeelsete lausungite suhteline sagedus suulistes ja tšätivestlustes. Kuni kahesõnalised (lühikesed) lausungid on eristatud pikematest*

tšätivestlustes on vastavad protsendid 85,1%, 7% ja 7,9%<sup>18</sup>. Seega on täiesti võõrkeelsete lausungite osakaal tšätikorpuses märgatavalt suurem ja koodivahetuse intensiivsus üldisemalt kõrgem, olgugi et suuliste ja tšätivestluste lausungid ei ole omavahel üks ühele võrreldavad. Tšätivestlustes tulevad selgemini esile ka sugudevahelised erinevused: nooremates vanuserühmades on poisid märksa aktiivsemad koodivahetajad, mis peegeldab paljuski nende vestlusteemasid (ainestiku näitel on selleks sport ja arvutimängud).

<sup>18</sup> Keskmine lausungi pikkus on suulistes vestlustes 6,26 sõna ( $sd = 5,3$ ) ja tšätivestlustes 4,13 sõna ( $sd = 8,26$ ). Viimaste puhul on kahe korpus võrreldavuse huvides lausung võrdsustatud ühe kõneleja ühe saadetud sõnumiga.

### 4.3. SAGEDAMAD INGLISKEELSE SÕNAD

Lõpetuseks uurisime, milliseid sõnu TeKE korpuse keelejuhid kõige enam kasutavad ning kas noorte ingliskeelse sõnavara kasutus võimaldab eristada vanuse ja soo põhjal ka eri kõnelejagruppe. Joonis 7 kuvab soo ja vanuserühma põhjal suulistes vestlustes kõige sagedamad ingliskeelsed sõnad. Teksti suurus väljendab sõna suhtelist sagedust (st mida suurem sõna, seda sagedam), tumedus aga selle normaliseeritud levikut iga grupi kõnelejate hulgas (st mida tumedam sõna, seda enam eri kasutajaid).

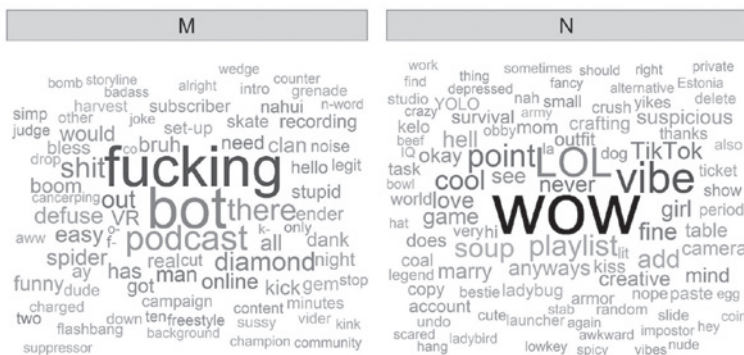
Ei ole üllatav, et korpuse kõige sagedamad ning levinumad sõnad on enamasti grammatilised sõnad, mis esinevad pigem pikemates fraasides (nt *I, the, oh, my, god*). Küllalt üldised ja levinud on vestluses suhtluslikke funktsioone täitvad (erinevad diskursusmarkerid, pöördumised, nt *wow, yeah, bye, yo, LOL, well*) ning hinnanguid



**Joonis 7.** Suuliste vestluste sõnad, mis esinevad vähemalt viis korda. Sõna suurus väljendab selle suhtelist sagedust, tumedus levikut iga grupi eri kõnelejate hulgas

(sageli omadussõnad, nt *cool, nice, creepy, basic*) või emotsioone väljendavad sõnad (mh intensiivistajad ja tugevamaks peetud ropp-sõnad, nt  *fucking, shit, bitch, damn*), viimased eriti poiste kõnes. Suur osa nooremate kõnelejate ingliskeelsest sõnavarast seondub mängukeskkondadega (nt *block, skin, armor, bot, defuse*), kusjuures kõige nooremas vanuserühmas on seda tüüpi sõnavara levinud nii poiste kui tüdrukute keeles. Vanemates vanuserühmades annab eriti tüdrukute sõnavaras tooni erinevate sotsiaalmeedia platvormidega seostatav sõnavara (nt *selfie, playlist, chat, story*). Sotsiaalmeedia ja mängudega seotud ingliskeelne sõnavara näibki moodustavat suure osa noorte nn erialaterminoloogiast, milles peegelduvad nende igapäevapraktikad ja eelistatud ajaveetmisviisid (vt ka Praakli *et al.* samas kogumikus). Huvitava tähelepanekuna nähtub asjaolu, et kui korpuse nooremates vanuserühmades vestlevad enamasti poisid poistega ja tüdrukud tüdrukutega, siis vanemas vanuserühmas vestlevad poisid rohkem ka tüdrukutega, seetõttu on ka vanema vanuserühma ingliskeelne sõnavara sugude lõikes sarnasem.

Selleks, et võrrelda ingliskeelse sõnavara kasutust soopõhiselt veidi lähemalt, eristasime omakorda sõnad, mille suhteline sagedus ühes rühmas oli vähemalt kolm korda suurem kui teises rühmas. Joonis 8 kuvab vasakul ingliskeelsed sõnad, mis on vanuserühmade



**Joonis 8.** Ingliskeelsete sõnade kasutus suulises korpuses, eri soost kõnelejate hulgas

üleselt omasemad poistele, ning paremal sõnad, mida kasutavad pigem tüdrukud. Sarnase suhtelise sagedusega sõnad (nt *oh my god*) joonisel ei kajastu.

Nn mängusõnavara ning negatiivseid emotsioone väljendavad sõnad (nt *shit, fucking, dank*) on omased pigem poistele, positiivsed emotsioonid ja hinnangud (nt *cool, fine, love*) ning sotsiaalmeedia kasutamiseiga seonduv sõnavara (nt *playlist, survival, crafting*) pigem tüdrukutele. Poistele ainuomane on ka erinevate n-ö maskuliinsete pöördumismarkerite (nt *dude, man, bruh*) kasutamine.

Sagedased ingliskeelsed sõnad tšättides (joonis 9) jaotuvad oluliselt ühtlasemalt. Lisaks funktsioonisõnadele (nt *to, the, a, than*) annab tooni tšätile kui registrile iseloomulik lühendite kasutus (nt *lol* 'laughing out loud', *plz* 'please', *gn* 'good night', *idk* 'I don't know', *wym* 'what [do] you mean'). Tuleb ka arvestada, et sõnade



**Joonis 9.** Tšätivestluste sõnad, mis esinevad vähemalt kolm korda. Sõnaloendites on ühtlustatud suur- ja väiketähe kasutust, et hajuvas andmestikus sagedusi koondada

kirjapilt on tähtselt oluliselt varieeruvam kui transkriptsioonides, kus ei tähistatud kogu häälduslikku varieerumist, mistõttu sagedaste tähtsõnade loendisse jõuab võrdlemisi juhuslik hulk sama keelelist funktsiooni täitvatest sõnadest (nt *yay*, *yayyy*, *jei*, *jeei*, *yei*).

## 5. Arutelu ja kokkuvõte

Siinne uuring annab esmase ülevaate eesti teismeliste keelekorpusandmetest. Artiklis tutvustame kvantitatiivseid andmeid teismeliste kakskeelsetest suhtluspraktikatest nendevaheliste suuliste ja tähtsõnaste näitel. Inglise keel on küll laialt levinud (inglisekeelseid sõnu leidub pea kõigilt kõnelejatelt), kuid selle osakaal ei ole ülemäära suur. Suulises suhtluses moodustab inglise keel kõikidest sõnadest *ca* 3%, sh ei sõltu inglise keele osakaal vanusest ega soost. Erinevused ilmnevad aga registritevahelises võrdluses: tähtsõnades moodustab ingliskeelsete sõnade maht ainekust 7,3%; ühes vanuserühmas (poisid vanuses 13–15 a) koguni ligi viiendiku (18%) kõigist sõnadest.

Sõnade sagedusloend näitab inglise keele esinemise temaatilisi ja funktsionaalseid erinevusi eri vanuses ja soost kasutajatel. Inglise keele atraktiivsus seisneb prestiižis ja valdkondade üleses ning eri suhtluseesmärkidega kasutusulatuses. Ühtlasi on inglise keel kaasaja noortekultuuri üks tunnuslikke jooni; sestap võib ka öelda, et selle justkui märkamatu kasutus eestikeelses suhtluses on osa grupikuuluvust tähistavast koodist. Teisest küljest viitavad andmed sellele, et inglise keele esinemine pole sõnamahult sedavõrd dominantne, nagu võib jääda mulje meedias avaldatud arvamustest.

Noorte sõnavaras ilmnevad selged seosed tänapäeva internetikultuuriga, mille kaudu nende keelerepertuaari jõuab sageli mitmesuguseid ingliskeelseid sõnu ja väljendeid. Võib eeldada, et sotsiaalmeedia on vahendaja, mille kaudu omandatakse varakult inglise keelt (vrd võrkeelsete telekanalite roll möödunud aastakümnetel). Sõnavara näitab, kas ja kui palju tarbitav sisu eri vanuserühmade lõikes sarnaneb või erineb. Kuigi sagedusloenditest nähtub, et

mitmesuguseid meemiväljendid ja -sõnad on üsna levinud (nt *dank*, *vibe*, *lol*, *basic*), ei tähenda see, et noorte inglise keele kasutust iseloomustavad üksnes trendikad keelendid. Kõige sagedasemad üksused (nt *oh my god*, *wow*, *fuckin*) on kasutusel ka väljaspool sotsiaalmeediat. Noorte keelekasutuses esineb rohket individuaalset varieerumist, mistõttu tulevikus on oluline lisada sõnapilvedele detailsemat analüüsi, et kirjeldada ingliskeelse sõnavara kasutust ning keelelist ja sotsiaalset konteksti.

Nagu taustaandmetest nähtub, peavad noored enda inglise keele oskust väga heaks. See on ühiskondlikus mõttes väärt teadmine, sõltumata sellest, kas nende tegelik keeleoskus vastab keeletestide standarditele või mitte. Noorte kõrge enesehinnang võõrkeelte oskusele, keeleteadlikkus ning keelelised kogemused on oluline ressurss, millega võiks haridussüsteemis eri pädevuste (nt kultuuri- ja väärtus-, enesemääratlus-, suhtluspädevus) kujundamisel ka laiemalt arvestada. Ehk tuleks laiemat tähelepanu pöörata noorte mitmekeelsusele, nende keelelisse repertuaari kuuluvate suhtusvahendite ja viiside paljumeelsele rakendamisele laiemalt. Inglise keeles ei peaks nägema ohtu eesti keelele, vaid vahendit ja võimalust noorte mitmekeelse teadlikkuse kujundamiseks.

Ühtlasi vihjavad sõnavara erinevused eri registrites ning poiste ja tüdrukute vahel sellele, et noorte sõnavara ja suhtluspraktikad tuleks vaatluse alla võtta laiemalt. TeKE korpus võimaldab lisaks mitmekeelsetele suhtluspraktikatele lähemalt uurida nii sõnade koosinemise sagedusi, morfosüntaktilist integreerimist kui ka muid nähtusi. Plaanis on uurida ka kõneleja tasemel registri mõju keelekasutusele ja idiolektile. Tulevikus lubab korpus uurida keele varieerumise ja muutumise protsesse laiemalt.

## KIRJANDUS

- Androutsopoulos, Jannis 2015. Networked multilingualism: Some language practices on Facebook and their implications. – *International Journal of Bilingualism* 19 (2), 185–205.
- Auer, Peter 1998. *Code-Switching in Conversation. Language, Interaction and Identity*. London: Routledge.
- Boczkowski, Pablo. J.; Matassi, Mora; Mitchelstein, Eugenia 2018. How young users deal with multiple platforms: The role of meaning-making in social media repertoires. – *Journal of Computer-Mediated Communication* 23 (5), 245–259. <https://doi.org/10.1093/jcmc/zmy012>.
- Bucholtz, Mary; Skapoulli, Elena 2009. Youth language at the intersection: From migration to globalization. – *Pragmatics* 19 (1), 1–16. <https://doi.org/10.1075/prag.19.1.01buc>.
- Crystal, David 2001. *Language and the Internet*. Cambridge: Cambridge University Press.
- Crystal, David 2008. *Txtng: the Gr8 Db8*. Oxford: Oxford University Press.
- Deakin, Hannah; Wakefield, Kelly 2014. Skype interviewing: reflections of two PhD researchers. – *Qualitative Research* 14 (5), 603–616. <https://doi.org/10.1177/1468794113488126>.
- Deuchar, Margaret 2020. Code-Switching in Linguistics: A Position Paper. – *Languages* 5 (2), 22. <https://doi.org/10.3390/languages5020022>.
- Dijk, José van; Poell, Thomas; Waal, Martijn de 2018. *The Platform Society: Public Values in a Connective World*. Oxford: Oxford University Press.
- Donath, Judith; Karahalios, Karrie; Viégas, Fernanda 1999. Visualizing conversation. – *Journal of Computer-Mediated Communication* 4 (4), JCMC442. <https://doi.org/10.1111/j.1083-6101.1999.tb00107.x>.
- Durkin, Kevin; Conti-Ramsden, Gina; Walker, Allan J. 2010. Computer-mediated communication in adolescents with and without a history of specific language impairment (SLI). – *Computers in Human Behavior* 26, 176–185. <https://doi.org/10.1016/j.chb.2009.10.007>.
- Eckert, Penelope 1997. Why ethnography? – *Ungdomsspråk i Norden*. Eds. Ulla-Britt Kotsinas, Anna-Brita Stenström, Anna-Malin Karlsson. Stockholm: Stockholm University, 52–62. <https://web.stanford.edu/~eckert/PDF/whyethnography.pdf> (06.10.2022).
- Eckert, Penelope 2000. *Language Variation as Social Practice: The Linguistic Construction of Identity in Belten High*. Hoboken: Wiley-Blackwell.



- Ferrara, Kathleen; Brunner, Hans; Whittemore, Greg 1991. Interactive written discourse as an emergent register. – *Written Communication* 8 (1), 8–34. <https://doi.org/10.1177%2F0741088391008001002>.
- Happonen, Natali 2010. Tähendusülekanded kui sõnavara täiendamise viisi eesti, soome ja vene koolinoorte kõnekeeles ja slängis – inimkesksed metafoorid. Magistritöö. Tartu Ülikool, eesti ja üldkeeleteaduse instituut. Tartu.
- Hennoste, Tiit (toim.) 2000. Eesti keele allkeeled. Tartu Ülikooli eesti keele õppetooli toimetised 16. Tartu: Tartu Ülikool.
- Hepp, Andreas 2018. Meediakultuur. Meediastunud maailmade kultuur. Tlk Katrin Kaugver. (= Bibliotheca mediorum et communicationis.) Tallinn: Tallinna Ülikooli Kirjastus.
- Igav, Reet 2013. Inglise-eesti koodikopeerimine Facebooki vestlustes. Magistritöö. Tallinna Ülikool, humanitaarteaduste instituut. Tallinn.
- Johanson, Lars 2002. Do languages die of ‘structuritis’? On the role of code-copying in language endangerment. – *Italian Journal of Linguistics* 14, 249–270.
- Jonsson, Rickard; Årman, Henning; Milani, Tommaso M. 2019. Youth language. – *The Routledge Handbook of Linguistic Ethnography*. Ed. Karin Tusting. London: Routledge, 259–272.
- Jørgensen, Jens Normann (toim.) 2002. De unges sprog: Artikler om sproglig adfærd, sproglige holdninger og flersprogethed hos unge i Danmark. København: Akademisk.
- Jørgensen, Jens Normann 2008. Polylingual languaging around and among children and adolescents. – *International Journal of Multilingualism* 5 (3), 161–176. <https://doi.org/10.1080/14790710802387562>.
- Kask, Helin 2021. English-Estonian code-copying in Estonian blogs and vlogs. Väitekirj. Tallinna Ülikool, humanitaarteaduste instituut. Tallinn.
- Kern, Friederike; Selting, Margret (toim.) 2011. *Ethnic Styles of Speaking in European Metropolitan Areas*. Amsterdam: John Benjamins Publishing. <https://doi.org/10.1075/silv.8>.
- Kilp, Geidi 2017. Eesti-inglise-jaapani mitmekeelne suhtlus Facebooki keskkonnas. Magistritöö. Tallinna Ülikool, humanitaarteaduste instituut. Tallinn.

- Klie, Jan-Christoph; Bugert, Michael; Boullosa, Beto; Eckart de Castilho, Richard; Gurevych, Iryna 2018. The INCEPTION Platform: Machine-Assisted and Knowledge-Oriented Interactive Annotation. – Proceedings of System Demonstrations of the 27th International Conference on Computational Linguistics (COLING 2018). Santa Fe, New Mexico, USA, 5–9. [https://inception-project.github.io/publications/INCEPTION-COLING2018-Demo.pdf](https://inception-project.github.io/publications/INCEP<small>T</small>ION-COLING2018-Demo.pdf) (06.10.2022).
- Korkus, Mari-Liis 2021. Sotsiopragmaatiline sissevaade koodivahetusele: näiteid rootsieesti teismeliste keelest. – Eesti Rakenduslingvistika Ühingu aastaraamat 17, 141–157. <http://dx.doi.org/10.5128/ERYa17.08>.
- Kull, Kristiina 2019. Repliik. Räägid estonenglish'it? – Sakala, 9. august. <https://sakala.postimees.ee/6748860/repliik-raagid-estonenglish-it> (06.10.2022).
- Källström, Roger; Lindberg, Inger (toim.) 2011. Young Urban Swedish: Variation and Change in Multilingual Settings. Gothenburg: University of Gothenburg, Department of Swedish Language.
- Lauricella, Alexis R.; Cingel, Drew P.; Blackwell, Courtney; Wartella, Ellen; Conway, Annie 2014. The mobile generation: Youth and adolescent ownership and use of new media. – Communication Research Reports 31 (4), 357–364. <https://doi.org/10.1080/08824096.2014.963221>.
- Lehtonen, Heini 2015. Tyylytellen: Nuorten kielelliset resurssit ja kielen sosiaalinen indeksisyys monietnisessä Helsingissä. Väitöskirja. Helsinki: Helsingin yliopisto.
- Lehtpuu, Kairit 2016. Teismeliste suulise keelekasutuse leksikaalsed jooned. Magistritöö. Tartu Ülikool, eesti ja üldkeeleteaduse instituut. Tartu.
- Leppänen, Sirpa 2007. Youth language in media contexts: Insights into the functions of English in Finland. – World Englishes 26 (2), 149–169. <https://doi.org/10.1111/j.1467-971X.2007.00499.x>.
- Loog, Mai 1988. Soome mõjusid Tallinna noorte keelekasutuses. – Keel ja Kirjandus 2, 83–89; 3, 144–148.
- Loog, Mai 1990. Laensõnad Tallinna koolinoorte slängis ehk ida pole ilma-kaar. – Keel ja Kirjandus 3, 167–169.
- Lukk, Maarika; Koreinik, Kadri; Kaldur, Kristjan; Vihman, Virve-Anneli; Villenthal, Anneli; Kivistik, Kats; Jaigma, Martin; Pertšjonok, Anastasia 2017. Eesti keeleseisund. Tartu: Tartu Ülikool ja Balti Uuringute Instituut (IBS). [https://www.hm.ee/sites/default/files/aruanne\\_2017.pdf](https://www.hm.ee/sites/default/files/aruanne_2017.pdf) (06.10.2022).

- Mandel, Aive; Praakli, Kristiina; Vihman, Virve-Anneli; Koreinik, Kadri 2022. Rahvateadusega noortekeelt kogumas. – *Oma Keel* 44 (1), 16–21. <https://www.emakeeleselts.ee/wp-content/uploads/2022/06/Mandel-jt.pdf> (06.10.2022).
- McCulloch, Gretchen 2019. *Because Internet: Understanding the New Rules of Language*. New York: Riverhead Books.
- Myers-Scotton, Carol 1993. *Duelling Languages: Grammatical Structure in Codeswitching*. Oxford: Clarendon Press.
- Myers-Scotton, Carol 2017. Code-Switching. – *The Handbook of Sociolinguistics*. Ed. Florian Coulmas. Hoboken: Wiley-Blackwell, 217–237.
- Nortier, Jacomine; Svendsen, Bente Ailin (toim.) 2015. *Language, Youth and Identity in the 21st Century: Linguistic Practices across Urban Spaces*. Cambridge: Cambridge University Press.
- Otsuji, Emi; Pennycook, Alastair 2010. Metrolingualism: Fixity, fluidity and language in flux. – *International Journal of Multilingualism* 7 (3), 240–254. <https://doi.org/10.1080/14790710903414331>.
- Pennycook, Alastair 2016. Mobile times, mobile terms: The trans-superpoly-metro movement. – *Sociolinguistics: Theoretical Debates*. Ed. Nikolas Coupland. Cambridge: Cambridge University Press, 201–216. <https://doi.org/10.1017/CBO9781107449787.010>.
- Poplack, Shana 1980. Sometimes I'll start a sentence in Spanish Y TERMINO EN ESPAÑOL: Toward a typology of code-switching. – *Linguistics* 18, 581–618. <https://doi.org/10.1515/ling.1980.18.7-8.581>.
- Praakli, Kristiina; Koreinik, Kadri 2020. Keelemuutus vs. sotsiolingvistiline muutus: eesti keele sotsiooperioodid *re-revisited*. – *Keel ja Kirjandus* 11, 915–934.
- Praakli, Kristiina 2022. Sissevaade noortekeele uurimisse Eestis: üldist ja üksikasjalikku. – *Keel ja Kirjandus* 11, 1014–1030. <https://doi.org/10.54013/kk779a3>
- Praakli, Kristiina; Korkus, Mari-Liis; Mandel, Aive; Kaukonen, Elisabeth; Kängsepp, Annika; Aasa, Triin; Algvare, Kristel; Eriksoo, Helen; Mägi, Marion; Tomson, Getri; Lindström, Liina 2022. „Mis keeles ma räägin, I don't know“. Eesti sisuloojate inglise keele kasutusest YouTube'is. – *Philologia Estonica Tallinnensis* 7, 263–291. DOI: <https://doi.org/10.22601/PET.2022.07.10>.

- Rampton, Ben 1995. *Crossing: Language and Ethnicity among Adolescents*. London, New York: Longman.
- Rannamäe, Kerttu 2006. Koolinoorte släng. Bakalaureusetöö. Tallinna Ülikool, humanitaarteaduste instituut. Tallinn.
- Ratt, Silja 2017. Inglise-eesti koodikopeerimine Maria Rannavälja vlogides. Bakalaureusetöö. Tallinna Ülikool, humanitaarteaduste instituut. Tallinn.
- Ruven, Annika 1999. Antsla, Põlva, Värskä õpilaste slängist. Bakalaureusetöö. Tartu Ülikool, eesti ja üldkeeleteaduse instituut. Tartu.
- Salla, Sigrid 1999. Jututubade arvutisläng. – Artikleid ja arhivaale II. Keeleuenduse kirjastik B3. Toim Helgi Vihma. Tallinn: Johannes Aaviku Selts, 60–70.
- Sierra, Sylvia 2021. *Millennials Talking Media: Creating Intertextual Identities in Everyday Conversation*. Oxford, New York: Oxford University Press.
- Siibak, Andra 2020. Digipõlvkonnast sotsiaalmeedia põlvkonnaks. Põlvkondlikku enesemääratlust kujundavad trendid Eesti noorte internetikasutuses. – Methis. *Studia Humaniora Estonica* 21 (26), 17–34. <https://doi.org/10.7592/methis.v21i26.16908>.
- Stenström, Anna-Brita 2020. English-and Spanish-speaking teenagers' use of rude vocatives. – *Swearing and Cursing: Contexts and Practices in a Critical Linguistic Perspective*. Eds. Nico Nassenstein, Anne Storch. Berlin, Boston: De Gruyter Mouton, 281–302.
- Tagg, Caroline; Lyons, Agnieszka 2021. Polymedia repertoires of networked individuals. A day-in-the-life approach. – *Pragmatics and Society* 12 (5), 725–755. <https://doi.org/10.1075/ps.20051.tag>.
- Tagliamonte, Sali A.; Denis, Derek 2008. Linguistic ruin? LOL! Instant messaging and teen language. – *American Speech* 83 (1), 3–34. <https://doi.org/10.1215/00031283-2008-001>.
- Tagliamonte, Sali A. 2016. *Teen Talk: The Language of Adolescents*. Cambridge: Cambridge University Press.
- Tender, Tõnu 1994. Eesti släng: olemus ja uurimislugu. – *Keel ja Kirjandus* 5–6, 293–297, 346–355.
- Verschik, Anna 2004. Koodivahetus meil ja mujal. – *Keel ja Kirjandus* 1, 25–45.

- Verschik, Anna 2008. *Emerging Bilingual Speech: From Monolingualism to Code-Copying*. London: Bloomsbury Publishing.
- Voltri, Johannes 2019. Üleskutse Arvamusfestivalilt: inglise keel on kui umbrohi, mida koolis ei peaks õpetama. – Postimees, 10. august. <https://www.postimees.ee/6750278/uleskutse-arvamusfestivalilt-inglise-keel-on-kui-umbrohi-mida-koolis-ei-peak-opetama> (06.10.2022).
- Zabrodskaja, Anastassia; Kask, Helin 2017. Eesti keele kontaktid. – Eesti rändeajastul. Eesti inimarengu aruanne 2016/2017. Toim Tiit Tammaru. Eesti Koostöö Kogu, 177–185. <https://www.2017.inimareng.ee/ranne-identiteedi-ja-keelemuutused/eesti-keele-kontaktid/> (06.10.2022).

## SUMMARY

# ARE ESTONIAN TEENAGERS OBSESSED WITH ENGLISH? WHAT DO CORPUS DATA SAY?

Virve-Anneli Vihman, Kristiina Praakli,  
Maarja-Liisa Pilvik, Mari-Liis Korkus  
University of Tartu

**Abstract.** This article presents findings from the project “Teen speak in Estonia” (TeKE; 2019–2022), in which a dual corpus was compiled of spoken and (online) chat messaging language used by Estonian speakers aged 9–18. The data was collected through a citizen science approach. We analysed frequency data from the corpus to investigate patterns in the use of English words in teenagers’ language choices. The results show that the amount of English used is not predicted by the speakers’ age or gender; instead, language choices show a great deal of individual variation and reflect features of speakers’ interactional context and register (spoken or chat conversations). A snapshot of their most frequently used vocabulary, however, shows differences by age and gender.

**Keywords:** youth language, spoken language, online language, corpus analysis, Estonian, English

Autorid kuuluvad Tartu ülikooli “Teismeliste keel Eestis” töörühma (EKKD3), mida rahastab Haridus- ja Teadusministeeriumi teadus- ja arendustegevuse programm „Eesti keel ja kultuur digiajastul 2019–2027“.

**Virve-Anneli Vihman** on Tartu ülikooli psühholingvistika kaasprofessor. Ta peamised uurimissuunad on seotud keele omandamise, kakskeelsuse ja keele varieerumisega. [virve.vihman@ut.ee](mailto:virve.vihman@ut.ee)

**Kristiina Praakli** on Tartu ülikooli rakenduslingvistika kaasprofessor. Ta uurib teismeliste keelt ja mitmekeelset suhtlust. [kristiina.praakli@ut.ee](mailto:kristiina.praakli@ut.ee)

**Maarja-Liisa Pilvik** on Tartu ülikooli eesti keele teadur. Tema uurimisteemad seonduvad keele varieerumisega eesti keele murretes ja situatiivsetes allkeeltes.  
maarja-liisa.pilvik@ut.ee

**Mari-Liis Korkus** on Tartu ülikooli doktorant, kelle uurimisteemadeks on teismeliste mitmekeelne suhtlus ja eesti-rootsi koodivahetus.  
mari-liis.korkus@ut.ee