

NOORTEKEELE ANDMEKORJE JA KOGUD: EMAKEELE SELTSI JA EESTI KEELE INSTITUUDI KEELEVARA

KRISTIINA PRAAKLI

Annotatsioon. Noorte keeleandmed on oluline allikas keele varieerumise uurimisel. Teadusprojektide ja üliõpilastööde raames on noortelt kogutud suulist kõnet, tähtvestlusi ja sõnaloendeid. Ülevaates võtan vaatluse alla Emakeele Seltsi eestvedamisel noortekeele kohta kogutud andmeid, lähtudes dokumendianalüüsi põhimõtetest. Enamik andmetest on saadud kogumisvõistluste teel, sõnasedelite ja -loendite või uurimistööde vormis. Andmed dokumenteerivad eeskätt noortekeele sõnavara kujunemist ajas ja ruumis, kajastavad ühiskondlikke muutusi ning annavad väärtuslikku teavet andmekorje metodoloogiasse kohta.¹

Võtmesõnad: andmed, andmekorje, noortekeel, Emakeele Selts, Eesti Keele Instituut

Noorte² keelekasutust on kirjeldatud kui dünaamilist ja uuenduslikku (Vihman jt 2022: 295), viidates eelkõige sõnavara mitmekesisusele, kiirele muutumisele ning uute väljendusvahendite loomisele ja omaksvõtule. Just see teebki noortelt kogutu keele varieerumise uurimisel väärtuslikuks. Noored on keeleuuendajatena ja muutuste algatajatena uurijatele olulised infoallikad (Eckert 1997: 52). Seetõttu on igasugune teave noorte keelekasutuse kohta vajalik keele varieerumise ning noortekeele kultuuriliste ja sotsiaalsete aspektide uurimiseks.

Viimastel aastatel on noortekeele andmekorje ja uurimine tänu riiklikele programmidele saanud sisse uue hoo. Näiteks on projekti „Teismeliste

¹ Artikkel valmis riikliku programmi „Eesti keel ja kultuur digiajastul“ projekti „Digiajastu keel: suulised ja tähtvestlused“ (EKKD-TA16) toel.

² Pean „noorte“ ja „noortekeele“ all silmas üldhariduskooli II ja III kooliastme õpilasi ja gümnasiste ja nende keelekasutuse viise, vanuseliselt 12–19aastasi, kellelt ongi enim keeleandmeid kogutud (vt Praakli 2022: 1015–1024).

keel Eestis“ (TÜ, 2019–2023) raames kogutud suulist kõnet (u 97 tundi) ja tähtvestlusi (u 60 000 sõna), loodud noortekeele korpus (Vihman jt 2022; Vihman jt 2024), avaldatud mitmeid artikleid ning kaitstud lõputöid. Andmekorje ja analüüs jätkub projektis „Digiajastu keel: suulised ja tähtvestlused“ (TÜ, 2024–2027). Mõlemas projektis koguvad andmeid noored ise. Kogutud andmed annavad rohkesti uusi teadmisi praegusaja noorte keeleliste ressursside, keelevalikute ja suhtlusviiside kohta, eriti digisuhtluses, avardades seeläbi arusaamu noorte keelekasutuse mitmekesisusest ja dünaamikast.

Heites pilgu minevikku, näeme, et mitmesuguste algatuste raames (nt sõnakogumisvõistlused, olümpiaadid, õpilastööde konkursid) on noorte keeleandmeid aktiivselt talletatud varemgi ning just noori endid andmekorjesse kaasatud. Kogutud andmeid pole küll alati jõutud tulemusteks vormistada, kuid talletatuna on need näiteks sõnavara kirjeldamiseks (vorm, tähendus, kasutus) vägagi olulised. Eriti rikkalik on Emakeele Seltsi eestvedamisel aastakümnete jooksul kogutud materjal. Ülevaates tutvustangi lähemalt seda, kuidas ja milliseid andmeid on seltsile edastatud, mida kogud sisaldavad ning milliste võimalike piirangutega tuleb kogutu puhul arvestada. Meetodina rakendan dokumendianalüüsi põhimõtteid. Seltsi kogud (sedel- ja vihikkogu) on deponeeritud Eesti Keele Instituudi eesti murrete ja soome-ugri keelte arhiivi EMSUKA.

Eesti noortekeele lünklikku, ent ligi saja aasta pikkust uurimislugu, andmekorje- ja analüüsimeetodeid ning tulemusi olen käsitlenud mõne aasta taguses artiklis (Praakli 2022), sestap jäävad need küsimused siinsest ülevaatest kõrvale.

1. Andmekorje algus

Noorte keele dokumenteerimine algab murdekogumise tuules 1920ndatel, kui luuakse Emakeele Selts (23.03.1920). Kogumistööd asub juhtima Andrus Saareste (1922), kümnendi lõpus valmivad tema juhendamisel esimesed üliõpilasuurimused (vt Tender 1994; Viikberg 1997; Praakli 2022). Sarnaselt praegusajaga pälvib noortekeel sadakond aastat tagasigi rohkesti meedia tähelepanu, ka Emakeele Selts on pidevalt „pildil“. Eriti suurt tähelepanu äratav Clarissa Alliku uurimus (1928) Tallinna tüdrukute sõnavarast. Allik peab seltsis ettekande, millele järgneb mitu sisukat meediakajastust (nt Postimees 1928). Eri sotsiaalsete rühmade (õpilased

ja üliõpilased, kaitseväelased) keelepruugi käsitlemise kõrval läbib tolla-seid kirjutisi teinegi teema: noorte tegevusvõimekus ning asjatundlikkus andmete kogumisel. Olgu siinkohal näiteks Üliõpilaslehes ilmunud üleskutse (Vaigla 1928: 97): „Meie erikeel, me oma keel, me kodune keel nõnda öelda – peaks olema huvitav eriti meile enestele – ka kogumis- ja uurimisesemena. [...] Seda erikeelset materjali – väikeses kui rohkem ulatuses – võtab vastu tänuga Akadeemiline Emakeele Selts Tartus, Lutsu 10.“

Noorte keeleandmed ongi seltsi kogudesse jõudnud kas üliõpilaste kogutuna (enne teist maailmasõda Saarest, hiljem Paul Ariste eestvedamisel), murdekogumisvõistlustele esitatud kogude ja uurimustena (vt ülevaateid Kodumurdes ja Viikberg 2008, 2021). Jättes kõrvale üliõpilaste esitatu, võib kõigi teiste algatuste andmekorjes näha rahvateaduse põhimõtete järgimist (lähemalt Mandel jt 2022), kus uued andmed ja teadmised luuakse uurijate ja vabatahtlikest andmekogujate, mittelingvistide koostöös. Eesti teadusmaailmast leiab tavakodanike kaasamisest näiteid rohkesti: kas või Jakob Hurda (1888) omaaegne üleskutse „Paar palwid ...“ või loodusteadlaste ilma-, linnu- ja taimevaatlused. Mitteprofessionaale, kuid omas vallas asjatundjaid kutsutakse andmekorjesse eri põhjustel. Noorte puhul on põhjus ilmne: uurijail endil on üsnagi keeruline noorte võrgustikele ning keeleandmetele juurde pääseda. Sestap usaldatakse juhendatud andmekorje noortele, nagu on teinud ka Emakeele Selts.

2. Kogumisvõistlused kui tee noortekeele juurde

Noortekeele andmekogude kujunemisel mängivad suurimat rolli seltsi korraldatud murdekogumisvõistlused, mille sõja-aastail katkenud traditsioon taastati 1957. aastal (Viikberg 2008: 97–98; eriti 2021: 362–368). Algselt keskendutakse vanapärase murdeainese kogumisele, kuid alates 1960ndatest hakatakse võistluste üleskutsetes selgelt rõhutama ka õpilaste sõnavara andmete kogumise tähtsust, sest neid andmeid on väga vähe (vt nt Ühistöö 1964: 4). Seltsi nimel esitatud üleskutsed ilmuvad üleriigilistes ning maakonnalehtedes, jõuavad ka väliseesti lehtedesse. Kevaditi avaldatud üleskutsed jätavad kogumisest huvitatuile pool aastat aega andmete esitamiseks. Kuna tegu on võistlusega, mille paremaid töid premeeritakse rahaliste autasude või kirjandusega, selgitatakse üleskutsetes sõnade kirjaneku nõudeid ning hindamiskriteeriume. Esimese üleskutse peale (1964) laekubki kaks võistlustööd.

Võistlusele esitatav sõnavara tuleb dokumenteerida sõnasedelite vormis. Ajalooliselt on sõnasedelid olnud peamised andmekandjad ning sõnalise info talletamise vahendid. Täpsemate juhiste ja nõuetekohase paberi (formaadis 8 x 10) saamiseks palutakse võistlusest osavõtjail pöörduda seltsi (eeskätt Helju Kaalu kui võistlejate usaldusisiku) poole (nt Harju Elu 1968: 3). Sõnasedelite vormistamise kohta on selts koostanud mitmeid juhendmaterjale (nt Viires 1985: 3–4, noortekeele puhul eriti 9–11), milles palutakse tähelepanu pöörata järgmisele: igale sedelile kantud sõnale tuleb juurde lisada kirjakeelne vaste, tähendus, vorminäited (kolm põhivormi), kontekst ja kasutuskoht ning andmed koguja kohta. Tegelikult on sedelite vormistamise detailsusaste kogujati muidugi erinev.

Noorte (ja teiste sotsiaalsete rühmade) sõnavara kogumine toimub aastakümneid üsna ühte moodi. Suurim muudatus toimub 1989. aastal, kui selts korraldab üheskoos Keele ja Kirjanduse Instituudiga üleriigilise slängi kogumise võistluse (lähemalt Viikberg 2021: 366–367). 1989. aasta märtsis ilmub meedias (vt nt Sirp ja Vasar 1989: 7) vastav üleskutse, milles andmekorjet selgitatakse sooviga talletada Eesti linnaslängi ning eri ametialade ja suhtlusvõrgustike keelekasutust. Võistlusele laekub 13 tööd, mida on oluliselt rohkem kui varem võistlusaastati laekunud töö või paar (üksikutel kordadel neli-viis). 1989. aasta võistlus jääbki viimaseks suuremaks aktsiooniks, hilisemad (2002, 2008) varasematega võrreldavat huvi enam ei ärata ning sõnavara avalike üleskutsete vormis kogumine lõpeb.

3. Sedelkogu: sõnasedelitel dokumenteeritud andmed

Minul kasutada olnud EMSUKA kogude ja muude arhiivimaterjalide põhjal on ajavahemikul 1964–2002 seltsile saadetud 127 sõnasedelitena vormistatud kogu (kokku 42 443 sedelit), mis pole murdekeel selle traditsioonilises tähenduses. Kogude pealkirjade järgi sisaldavad materjalid õpilaste, gümnasistide ja üliõpilaste (nimetatud *erikeeleks*, *argooks*, *slängiks*) ning teiste suhtlus- ja ametirühmade (nt kaitseväelaste, vangide, raudteelaste jt) sõnavara, ka *slängi/argood* laiemalt. Kogusid, mille pealkirjas sisaldub otsene viide õpilaste keelele, on 84 (kokku 34 747 sedelit).

Esimene võistlustöö-sedelkogu („Tallinn. Õpilassõnavara“, 474 sedelit) laekub seltsile 1964. aastal, saatjaks Annemari Karm. Kogu pälvib auhinna ja rohkesti kiidusõnu. Õpilaskeele kogusid saadetakse seltsile pea igal võistlusaastal, tavaliselt töö või paar aastas (vt ülevaateid

Kodumurdest). Kõige viljakamad on 1980ndad, mil võistlused kõnetavad noori rohkem kui varem: näiteks 1983., 1985. ja 1987. aastal laekub igal võistlusaastal viis, 1989. aasta slängivõistlusele kogunisti 13 tööd. Huvi on aktiivsem ka 1990ndate alguses, misjärel hakkab osalejate arv tasapisi vähenema.

Sedelite hulk saadetud kogudes on vägagi erinev: mõnes on üksikud sõnanäited, teises mõnikümmend või mõnisada, suurimates tuhandeid. Puhtalt sedelite (mitte sõnade) arvust lähtuvalt on enim andmeid esitanud Aita Maasik (seitse kogu Tartust, 1966–1988; 4589 sedelit) ning Tõnu Tender (viis kogu Tallinnast ja Võhmast, 1983–1987; 3247 sedelit). Maasiku ja Tenderi kõrval on teisedki kogujad võistlustest korduvalt osa võtnud. Enamik saadetud andmetest on kogutud Tallinnast (27 kogu), Tartust (17), Viljandist (9) ja Pärnust (3), kaks kogu Antslast, Valgast, Elvast, Haapsalust, Põltsamaalt ja Sindist, mujalt üks (nt Loksa, Järva-Jaani, Kuusalu, Vändra, Hargla jt). Torkab silma, et ühtki kogu pole Hiiu- ja Saaremaalt ega Ida-Virumaalt.

Suurem osa kogudest on pealkirjastatud mingi linna või asula õpilas-sõnavara, õpilasargoo või õpilaskeelena. Slängi mõiste jõuab kogude pealkirjadesse 1980ndate alguses, kümnendi teises pooles on see pealkirjades juba tavaline, kuid kasutusele jäävad ka mõisted *argoo* ja *õpilassõna*. Seevastu 1989. aasta võistlusele saadetud pealkirjades domineerib mõiste *släng*, küllap slängivõistluse pealkirjast tulenevalt.

Vaadakem põgusalt sedagi, mida seltsile saadetud sõnasedelid sisaldavad. Õpilaskeele nime all on esitatud mitmekesist sõnavara: igapäevast kõnekeelt, spetsiifilisemat oma sõpruskonna, klassi või kooli sõnavara, murdekeelendeid ning tsiteeringuid popkultuurist. Näeme ka, et noored loovad uut sõnavara neid kõnetavatel teemadel (nt kool, omavahelised suhted, meelelahutus jt) ning selleks kasutatavate vahendite võttestik on rikkalik. Nii mõnegi keelendi puhul võib näha nende esmakasutusele tulekut ning kasutuse kujunemist ajas (nt *näts*, *džiinsid*). Väärtuslikku teavet pakuvad ka kogude kaaskirjad: seegi on relevantne aines, kuna aitab paremini mõista kogujate arusaamu sellest, mida on koguma mindud või mida kogumistööst arvatakse.

4. Vihikkogu: käsikirjalised sõnaloendid ja uurimused

EMSUKA vihikkogu näol on tegu seltsile saadetud või seltsi halduses olevate mitmesuguste käsikirjadega. Neist eri ajastutest pärit vihikkogudest, mis pole traditsioonilises tähenduses murdekogud, leiab kümme õpilaste ja gümnaasistide sõnavara dokumenteerivat käsikirja: Künnapuu ja Tõldsepa kogutu 1920ndatel, Allan Roosilehe ja Aimar Lauliku kogutu ja uurimistöö 1970ndatel, lisaks veel Tartu Riikliku Ülikooli (andmete esitaja Paul Ariste) ning Tallinna Pedagoogilise Instituudi üliõpilaste kogutu 1970.–1980. aastatel. Võtan lähemalt vaatluse all kaks kogu.

Ariste esitatud materjalid sisaldavad aastail 1977–1985 TRÜ eesti filoloogia eriala esmakursuslastelt arvestustöö „Ühiskondlikud murded“ tarbeks kogutud keeleandmeid. Sisult koondab see kümnete üliõpilaste kogutud õpilaskeele sõnavara kas oma kodukohast või endisest koolist. Materjalid on sisu ja vormi poolest ebaühtlased, nii nagu üliõpilased neid kogunud ja Aristele esitanud on – mõni on esitanud üksikuid sõnu, teine lähenenud süsteemselt ning loetlenud lehekülgede kaupa sõnu ja selgitusi. Lõpuni pole siiski selge, kui suur on Ariste (ja teiste käsikirjade) sõnaline maht. Praegune maht on teada käsikirjade lehekülgede arvuna.

Allan Roosilehe ja Aimar Lauliku (1979) käsikirjas esitatud andmed (kokku 2548 sõna) on kogutud küsimustiku vormis u 500 õpilaselt kümnest Tallinna koolist. Töö vääribki tähelepanu nii andmekorje metodoloogiliste küsimuste tõstatamise kui ka kogutud sõnavara poolest. Nagu autorid täheldavad (Roosileht, Laulik 1979: 5–6), kipuvad noored andmetega manipuleerima, nt sõnu välja mõtlema ning esitama harvaesinevaid keelen-deid. Samuti tõdevad nad, et mittekorduva sõnavarani oleks andmekorjes jõutud ka oluliselt väiksema vaevaga. Töös dokumenteeritud laensõnad näitavad, et kogumise ajahetkel pole soome keel Tallinna noorte seas veel aktuaalne ning töös esitatud soome laenude loetelu (nt *pikku sisko*, *rakkaus*, *pullo*, *pomo*, lk 70) erineb kvalitatiivselt sellest, mida on hiljem dokumenteerinud Tõnu Tender (1984) ja Mai Loog (1988).

Mainitagu ka, et nimetatud kogud on loodud nii individuaalse kui ka kollektiivse andmekorje tulemusena. Näiteks TRÜ õppejõud Paul Ariste kogus keeleandmeid kümnetelt üliõpilastelt nende kursusetööde raames, Allan Roosileht ja Aimar Laulik viisid kahekesi läbi andmekorje kümnes Tallinna koolis, Mai Loog emakeeleõpetajate ja õpilaste kaasabil vähemasti kuuest koolist, samas kui Tõnu Tender kogus materjali üksinda.

5. Tagasiside kogutule

Sõnavõistluste kokkuvõtted ilmuvad seltsi kogumikus Kodumurre (vt eriti numbreid 7 ja 21), kus antakse töödele tagasisidet ning nõu tulevastele kogujatele. Näiteks Annemari Karmi andmestikku (Viires 1965: 106) ise-loomustatakse sõnadega „uudne“, „terviklik“ ja „asjakohane materjal“, mis annab „hea ettekujutuse kaasaegsest õpilasargoost“. Kogutu hulgast leiab õppeainete nimetusi (nt *ven'na* 'vene keel', *bota* 'botaanika'), *kas-* tuletisi (*pingikas* 'pinginaaber', *ajakas* 'ajalugu'), metafoore (*vokk* 'jalgratas'), lühendamisi (*klassijuss* 'klassijuhataja'), paroole (*juhtmed käivad talla alt läbi* 'halva reageerimisega').

Kõige põhjalikuma tagasisivaate õpilaskeeles nime all seltsile saadetule leiab Kodumurdest nr 21 (Mirme jt 1989: 110–112). Autorid vaatavad siin kogutule mõnevõrra kriitiliselt tagasi, teevad metodoloogilisi ettepanekuid ning tutvustavad sõnavara kogumise näidisankeeti. Olgu ka öeldud, et ülevaateartikli kolmanda autori, Tõnu Tenderi kogutu on üks väheseid, mille analüüsi tulemused on artiklitenäe ilmunud.

Kodumurdes avaldatud kriitilisem tagasiside mainib mõne kogu puhul süsteemsuse puudumist ning puudulikult täidetud sõnaseideleid. Noortest kogujad ei pruugi tingimata eristada keelevariante ning võivad noortekeelena esitada ka andmeid, mis seda olla ei pruugi. Seda ei saa siiski neile ette heita, sest mittelingvistidest kogujad lähenevadki andmetele enda teadmiste ja isikliku kasutuskogemuse põhjal ning kui noore suhtlusvõrgustikus mingit keelendit aktiivselt kasutatakse, pole tal põhjust seda mitte noorte omaks pidada, kuigi keelend võib hoopiski pärineda kohapealsest murdepruugist.

6. Lõpetuseks

Mis tahes andmekogud on ajastutruud dokumendid, mis peegeldavad nii keelekasutust kui ka ühiskondlikke olusid. Noorte keelepruugist fikseeritud sõnad kajastavad nende jaoks olulisi teemasid, toovad esile tuumikvaldkonnad ja põlvkonnaülese sõnavara (nt koolisõnavara), annavad väärt teavet sõnavara struktuuri, vormi ja tähenduse muutumise kohta. Eri piirkondades kogutud sõnavara toob välja geograafilised erisused ning paikkondlike keelte funktsioonid noorte keelekasutuses. Eri aegadel kogutud andmetes kajastuvad ühiskondlikud muutused (nt inglise ja

soome keele jõudmine noorte sõnavarasse) ning kultuurilised, sotsiaalsed ja tehnoloogilised nähtused, mis on noortele vähegi kõneainet pakkunud.

Mõistagi tuleb kogude, eriti sedelitena vormistatud andmete puhul arvestada mitmete piirangutega. Esmajoones tuleb uurijail mees pidada, et kuigi kogumis- ja uurimisobjektina on silmas peetud suulist argikeelt, on seda kogutud kirjalikus vormis. Põhiprobleem lasubki selles, et kasutus-kontekstita dokumenteeritud sõnad ei suuda päriselt vahendada teavet nende tegeliku kasutuse kohta spontaanses suulises suhtluses. Suurimaks küsimuseks on, mil määral esindab kogutud sõnavara n-ö kuulnud sõnu, mil määral tegelikku (ja kelle?) kasutust. Need on aga mis tahes sedelkogude probleemid, millega tuleb andmeanalüüsis arvestada, nagu on oma väljakutsed mis tahes viisil kogutud andmete puhul. Sedelkogude digitaliseerimisega on Eesti Keele Instituudis algust tehtud: see on oluline samm ainestiku digitaalse säilitamise, andmete kergema kättesaadavuse, tõhusama andmetöötamise ning -analüüsi seisukohalt, mis aitab kaasaja keeletehnoloogiliste võimaluste toel edendada noorte keelekasutuse uurimist.

Ka seltsi tagasiside andmekorjele ja kogutule on hinnaline. Varasemate kogemustega arvestamine on praeguste välitööde korraldajatele igati oluline info, eriti selles osas, kuidas mittelingvistid nende võetud ülesannet ning noorte keelepruuki tõlgendavad. Kogujad vajavad täpset juhendamist, kuna pelk üleskutse ei pruugi olla piisav. Sajandi jooksul eesti noortelt kogutud andmed annavad üksikasjaliku ülevaate noorte sõnavara kujunemisest ajas ja ruumis. Lisaks keeleanalüüsidele pakuvad need väärtuslikku ainet kõigiks teisteks uuringuteks, kus uurimisobjektiks noored ning neilt kogutud andmed. Iga kogu loeb!

Kasutatud kirjandus

Allik, Clarissa 1928. Tallinna keskkooliõpilaste erikeelest. – Eesti Keel 1–2, 3–11.

Eckert, Penelope 1997. Why ethnography? – Ungdomsspråk i Norden: föredrag från ett forskarsymposium. (= Meddelanden från Institutionen för nordiska språk vid Stockholms universitet 43.) Toim. Ulla-Britt Kotsinas, Anna-Brita Stenström, Anna-Malin Karlsson. Stockholm: Stockholms universitet, 52–62.

EMSUKA MS = EMSUKA murdesõnavara koondkartoteek.

EMSUKA ES0028 = Allan Roosileht, Aimar Laulik 1979. Keskkooliõpilaste kõnepruuk.

EMSUKA ES0029 = Paul Ariste 1977. Argoosõnavara ja väljendeid Eesti NSV koolidest.

- Harju Elu 1968** = Uus sõnavarakogumise võistlus. – Harju Elu, 15. juuni.
- Hurt, Jakob 1888.** Paar palwid Eesti ärksamaile poegadele ja tütardele. – Eesti Postimees ehk Näddalaleht, 23. märts.
- Loog, Mai 1988.** Soome mõjusid Tallinna noorte keeleteaduses. – Keel ja Kirjandus 2, 83–89.
- Mandel, Aive, Kristiina Praakli, Virve-Anneli Vihman, Kadri Koreinik 2022.** Rahvateadusega noortekeelt kogumas. – Oma Keel 2, 16–21.
- Mirme, Mari, Lea Tammaru, Tõnu Tender 1989.** Põgus pilk slängikogudele KKI-s ja üldisele kogumispraktikale. – Kodumurre 21, 110–112.
- Postimees 1928.** „Roosinupukesed“ ja „karuohakad“. – Postimees, 6. märts.
- Praakli, Kristiina 2022.** Sissejuhatus noortekeele uurimisse. Üldist ja üksikasjalikku. – Keel ja Kirjandus 11, 1014–1030. <https://doi.org/10.54013/kk779a3>.
- Sirp ja Vasar 1989.** Kogume eesti slängi. – Sirp ja Vasar, 3. märts.
- Tender, Tõnu 1984.** Tänapäeva õpilasslängi. – Keel ja Kirjandus 3, 168–171.
- Tender, Tõnu 1994.** Eesti släng: olemus ja uurimislugu. – Keel ja Kirjandus 5, 291–297.
- Vaigla, A. 1928.** Mõningaid jooni üliõpilaskeskkonnast. – Üliõpilasleht 7, 94–97.
- Vihman, Virve-Anneli, Kristiina Praakli, Maarja-Liisa Pilvik, Mari-Liis Korkus 2022.** Kas noored on inglise keelega *obsessed*? Millest räägivad korpusandmed? – Philologia Estonica Tallinnensis 7, 292–322. <https://doi.org/10.22601/PET.2022.07.11>.
- Vihman, Virve-Anneli, Maarja-Liisa Pilvik, Aive Mandel, Annika Kängsepp, Mari Aigro, Kadri Koreinik, Kristiina Praakli, Liina Lindström 2024.** Estonian Teen Language Corpus. DOI: 10.23673/re-455.
- Viikberg, Jüri 1997.** Eesti slängiuurimise tänane seis. – Keel ja Kirjandus 10, 705–710.
- Viikberg, Jüri 2008.** Eesti keele kogud. – Eesti humanitaar- ja loodusteaduslikud kogud: seisund, kasutamine, andmebaasid. Toim. Erast Parmasto, Jüri Viikberg. Tartu: Tartu Ülikooli Kirjastus, 95–112.
- Viikberg, Jüri 2021.** Emakeele Selts ja murdekogumine. – Emakeele Selts 1920–2020. Koost. Jüri Valge. Tallinn: Eesti Teaduste Akadeemia Emakeele Selts, 349–369.
- Viires, Helmi 1965.** Märkmleid 1964. aasta võistlustöödest. – Kodumurre 7. Eesti NSV Teaduste Akadeemia Emakeele Selts. Tallinn, 102–109.
- Viires, Helmi 1985.** Juhend sõnasedelite täitmiseks. Tallinn: Eesti NSV Teaduste Akadeemia Emakeele Selts.
- Ühistöö 1964.** Rahvaliku sõnavara kogumise võistlus. – Ühistöö, 9. mai.

**The youth language collections and data collection
of the Estonian Mother Tongue Society and
the Institute of the Estonian Language**

KRISTIINA PRAAKLI

Data on youth language is an important resource for studying language variation and change. Through various research projects and student theses, samples of oral speech, chat conversations, and vocabulary have been collected from Estonian teenagers. This article focuses on the data compiled by the Estonian Mother Tongue Society and the Estonian Language Institute over the past century. The overview follows the principles of document analysis. Most of the data have been gathered through word competitions and are presented in the form of card files, word lists, copybooks, and research papers. These collected materials document the evolution of youth language, particularly its vocabulary, reflect societal changes, and offer valuable insights into data collection methodologies.

Keywords: data, data collection, youth language, Mother Tongue Society, the Institute of the Estonian Language

Kristiina Praakli
eesti ja üldkeeleteaduse instituut
Tartu Ülikool
Jakobi 2
51005 Tartu
kristiina.praakli@ut.ee