

TEISMEEA LOOVKIRJUTISTE SÕNAVARA JA SELLE HINDAMINE

Krista Kerge, Anne Uusen, Halliki Põlda

Ülevaade. Kõrvutatuna haritud täiskasvanud eestlastega, kes ei ole filoloogid, uuriti sõnavaramuutujaid eesti õppekeele kooli 5., 7., 9. ja 11. klassi õpilaste loovkirjutistes, sh 10% kakskeelsetelt. Kirjeldava statistika toel näidatakse kirjutiste sõnavara arengu suundumusi ea ja soo järgi, teisalt aga seda, kuidas õpetajad hindamisel sõnavara arvestavad. Tulemused panevad aluse kirjutiste ja nende hindamise edasise kvalitatiivse uurimise suundadele ja aitavad leida selleks sobivate parameetritega näidiskirjutisi. Ühtlasi vaetakse võimalusi parandada kirjutiste individuaalset hindamist üldhariduskoolis, eriti noorte elukäiku mõjutavatel eksamitel.*

Võtmesõnad: teismeiga, kognitiivne areng, sõnavara varieeruvus ja ulatus, leksikaalne tihedus, enamuskeel õppekeelena

1. Uurimisaine, eesmärgid ja materjal

Keeleareng ei lõpe lapse kooli minnes, vaid jätkub kooli ainetundides ning oluliselt ka täiseas (Nippold 2004: 1–3). Koolis pannakse alus kirjaoskusele sõna mitmekülgses tähenduses (vt Kern 2000: 16–17, Puksand, Kerge 2012), kuigi emakeeleõpetuse koht selles protsessis vajab uurimist ja läbirääkimist (Krogh 2012). Kirjutamine peegeldab õpilase üldist keelelist ja kognitiivset arengut (Brügelmann 1990: 21–25), mille seoseid on käsitlenud Edgar Krull (2000: 106–127) ja mille 29-aastase pikiuurimuse tulemusi Ingrid Schoon jt (2010). Kognitiivse arengu märgina nähakse varajase sõnavara suurenemist (Berk 2003: 357–361). Muude tegurite kõrval võib ka sõnavara varasest arengust sõltuda inimese hilisem sotsiaalne toimetulek (Schoon jt, samas). Ainult väga varane areng on sõltuvuses hoidekeelest (vt Goodman jt 2008).

Lapse sõnavara on suhteliselt ahas (vt Leiwo 1993: 98, Karlep 1998: 276) ja kasvab koos kogemusega. Kool toetab seda arengut kõigi ainete ja kogu tegevusega (vt artikleid Eesti keel 2011), kuid võrdtähtis on igasugune muu tekstikogemus (Corson 1995). Keeleareng on lahutamatu koos haridusega kasvavast keeletunnetusest (ingl *higher language cognition*, vt Hulstijn 2011).

* Uurimust on rahastanud Eesti Teadusfondi grant 8605 "Kirjutamise loomulikkuse mudel ja hindamine".

Oluliselt areneb sõnavara teismeeas; arenev kirjalik süntaks toob omakorda kaasa uusi funktsioonisõnu (Nippold 2004: 3–4). Klassist klassi peaks muutuma õpilasteksti sisu- ja funktsioonisõnade tasakaal, mis erineb kõnes ja kirjas, monoloogis ja dialoogis. Isegi omavaheliste teismemeilide vaba keel on suulisest täiskasvanudialoogist veidi täpsem ja ilmekam, ehkki sidendi- ja deiktikuvaesem. (Vt Kerge, Pajupuu 2010) Kõik see avaldub oletamisi ka õpilaste kirjutistes.

Eesti õpilaste emakeele kirjutusoskust on algkooli ulatuses doktoritööna uurinud Anne Uusen (2006a, vt ka 2006b). Gümnaasiumilõpetajate kirjandeid käsitleb Kersti Lepajõe väitekiri (2012a, vt ka 2012b). Praegu täieneva emakeeleõppija korpuse märgendusvõimalusi vaeb Kadri Sõrmus (2008). Täiskasvanute emakeelseid (L1) toimetamata kirjutisi ja sõnavara on uuritud ülikooli esmakursuslaste keeleoskuse osana (Ehala jt 2010), kirjaliike ka rööpselt teismelistega (vt Kerge 2010). Emakeele sõnavara arengut on peamiselt uuritud eelkoolieas, mil laps jõuab paari-kolme tuhande sõna valdamiseni (nt Urm 2012).

Enam tähelepanu on saanud eesti keel kui teine keel (L2). Näiteks on riikliku kõrgtasemeeksami kirjutised pakkunud võrdlust suulise keeleoskuse näitajatele (Pajupuu jt 2010), k.a sõnavara (Pajupuu jt 2009). Viimati on sõnavara uuritud eesti keele kui L2 eri taseme eksamite kirjutusosas (Alp jt 2013).

Allpool kajastub loomuliku kirjaliku keelekasutuse uurimuse üks tahk: eesti õppekeelega kooli loovkirjutiste individuaalsed ja ealised erinevused 5., 7., 9. ja 11. klassis. Artikli teema on teismeliste aktiivne sõnavara, mis lubab inimesel end reservatsioonideta väljendada (Corson 1995: 44–45), nagu seda siinkohal võimaldavad ülesanne ja teema (vt osa 2).

Uurimus on katseline ja alustav. Eesmärk on kirjeldada kirjaliku sõnavara arengut täiskasvanulikkuse suunas, andes lähte- ja võrdlusandmeid kognitiivse ja keelelise arengu ning teise keele omandamise uurimustele. Ühtlasi aitab tutvustatav statistiline ülevaade leida kirjutiste kvalitatiivse analüüsi suunad ja nendega sobivad kirjutised.

Tulemused toetavad ka emakeeleõpetuse praktikat. Oluline on teada, kui võrd kirjutise hindaja jälgib sõnavaramuutujaid, mida uurijal on võimalik mõõta, kuid õpetajal mitte, ehkki seda ametlikes juhendites nõutakse. Näiteks eeldab eesti keele riigieksami loovkirjutise maksimaalhinne muu kõrval, et sõnastus on täpne ja isikupärane ning sõnavara rikkalik. Hea hinde puhul on sõnavara mitmekülgne ja rahuldava puhul piisav; puuduliku ja nõrga hinde korral on kas tekst üldsõnaline või sõnavara piiratud. (Vt Riigieksam 2013: 6) Samas saab Hille Pajupuu jt (2009: 188) väitel juhendi järgi hinnata seda, kui ladus on jutt ja kui sidus tekst, kuid väga keeruline on hinnata sõnavara rikkust ja ulatust, mille objektiivne mõõtmine nõuab peale täpsete arvutuste ka sõnade keskmise kasutussageduse tundmist.

Sõnavara kui intuitiivselt hinnatava õpisihi mõtestamiseks on valitud loovkirjutistes uuritavad sõnavaramuutujad, mis objektiivselt mõõdetuna näitavad sõnavara mitmekülgust ja ulatust (vt osa 3). Ainus hüpotees on, et laps areneb kogu aeg (vt Krull 2000) – küsimus on, kuidas.

2. Uuritud kirjutiste laad ja maht

Kirjutised on loodud 2011/12. kuni 2013/14. õa eakohaselt ühetaolise ülesande alusel ning kindla kirjutusaja jooksul üheksas Tallinna, Harjumaa, Hiiumaa ja Lääne-Virumaa üldhariduskoolis, mille hulgas ei ole nn eliitkoole. Eetilistel põhjustel vabatahtlikke osalejakoole ei avaldata – leitud tendentse ei soovita kindlate koolide, õpilaste ega õpetajatega siduda.

Kirjutisi on omal harjunud viisil (holistiliselt ehk tervikuna ja püstitatud ülesandest lähtudes) hinnanud eesti keele ja kirjanduse õpetajad, kes kirjutuskatse oma õpilastega läbi viisid. Nad ei olnud teadlikud, mida kirjutistes uurima hakatakse ega saanud hindamiseks erijuhiseid.¹ Erandina hindas ühe klassi töid n-õ võõras emakeeleõpetaja, sest katse viis läbi matemaatik. Kaasatud õpilased pandi kirjutama sotsiaalse suunitlusega arvamust väärtuste kohta, manades neile ülesandes ette kujuteldava olukorra ja rolli ning piiritledes žanri ja kirjutusaja. Ülesande sõnastus on kohandatud eri vanuses laste jaoks.

5. ja 7. klassi ülesanne on sõnastatud järgmiselt:

Kujuta ette järgmist OLUKORDA.

Teie kool tahab korraldada konkursi kõige õnnelikuma ja elus hästi hakkama saava inimese leidmiseks. Kooli ajaleht kogub ja avaldab nüüd võimalikult palju arvamuskirjutisi selle kohta, milline on õnnelik ja elus hästi hakkama saav inimene.

KIRJUTA koolilehele ka oma arvamuse: milline inimene on sinu arvates õnnelik ja saab elus hästi hakkama?

Sinu arvamislugu peaks olema u. 150 sõna pikk ja mahub parasjagu alla kasti – ruumi peaks ülegi jääma. Kirjutamiseks on aega üks koolitund.

9. ja 11. klassi ülesanne erineb seepoolest, et ülesandega on seotud loomuliku kirjutusolukorra tavaline kaastegevus, s.o lugemine (üldsõnaline tekst väärtuskasvatuse tähtsusest Eestis eri aegadel; 421 sõnet).² Ülesanne on sõnastatud järgmiselt:

Käisid hiljuti

ÕPILASKONVERENTSIL, MILLE TEEMA OLI “VÄÄRTUSKASVATUS”.
Lugesid ka ühe ettekande põhjal kirjutatud teksti. Sinu ülesanne on kirjutada üritusel kuuldust ja loetust artikkel koolilehe jaoks. Lisaks üldisele ülevaatele peatu lähemalt kahel kolmest konverentsil käsitletud teemast.

1. Mis on hariduse peamine eesmärk – kas teadmised või tervikliku isiksuse kasvatamine?
2. Millised iseloomuomadused tagavad õnneliku elu ja ühiskonnas hakkamasaamise?
3. Kas kõik algab kodust või saab kool samuti lapse iseloomuomadusi kujundada?

Teksti nõutav pikkus on umbes 250 sõna. Aega kirjutamiseks on 60 minutit.

¹ Kirjutistes ei uurita ainult sõnavara, muud tulemused on avaldamata.

² Lugemispõhist ülesannet ei saanud kasutada kahes nooremas klassis, kus ollakse harjunud ümberjutustusi kirjutama (vt osa 4.3).

Ülesanded esitati 12 p kirjas. Mõlemal juhul järgnes joonitud ala, kuhu optimaalse pikkusega tekst mahutada.

Kokku uuriti 159 õpilaskirjutist arvestusliku kogumahuga u 30 000 sõnet. Kirjutise keskmine pikkus oli 5.-7. klassis 142 sõnet (eeldatud 150) ja 9.-11. klassis 230 (eeldatud 250). Juhisest sõltumata kirjutasiid õpilased väga erineva pikkusega tekste. (Vt tabel 1) Edgar Krulli järgi erinebki laste õpitempo tavaliselt kuni kuus, harvem isegi 10 korda (Krull 2000), aga individuaalsete mõjurite üle saab ühetaolise katse korral vaid spekuloida.

Valimist jäeti välja ja tabelis 1 ei kajastu paar tööd, millel puudusid teksti tunnused (eriti lühike ja selgelt pooleli). Valimisse kaasati nende 16 õpilase tööd, kes deklareerisid end kakskeelsena (11 eesti ja vene, 2 eesti ja inglise) või kelle kakskeelsus tuleneb hariduskeelest (3 vene emakeelega last). Selle rühma osatähtsus (u 10%) vastab ligikaudu eesti koolide üldisele olukorrale.³

Sugude osatähtsus valimis lubab võrdlusi – tüdrukuid on u 53%. Nn kakskeelsete kirjutisi vähese arvu tõttu soo järgi ei vaadelda. Kooli asukoha järgi näitajaid ei võrrelda, kuid lapsi valiti igas vanuses nii Tallinnast kui väljastpoolt.

Valim näitab seega vaid tendentse, mille olulisust statistiliselt kontrolliti.

Tabel 1. Uuritud kirjutiste arv ja mahuandmed

| Klass | Arv | Sh kakskeelsed | Maht sõnedes | | | | |
|-----------------------|-----|----------------|--------------|---------------|---------|-------|-------|
| | | | Keskmine | Standardhälve | Mediaan | Lühim | Pikim |
| 5. kl poisid | 20 | 1 | 136 | 36,78 | 140 | 79 | 236 |
| 5. kl tüdrukud | 23 | 3 | 157 | 32,69 | 151 | 102 | 269 |
| 5. klass | 43 | 4 | 147 | 34,73 | 146 | 91 | 253 |
| 7. kl poisid | 14 | 0 | 125 | 17,18 | 123 | 106 | 157 |
| 7. kl tüdrukud | 17 | 1 | 147 | 28,10 | 147 | 100 | 191 |
| 7. klass | 31 | 1 | 136 | 22,64 | 135 | 103 | 174 |
| 9. kl poisid | 25 | 4 | 179 | 80,42 | 154 | 81 | 345 |
| 9. kl tüdrukud | 24 | 1 | 241 | 77,62 | 244 | 99 | 448 |
| 9. klass | 49 | 5 | 210 | 79,02 | 199 | 90 | 397 |
| 11. kl poisid | 12 | 3 | 234 | 25,06 | 229 | 200 | 291 |
| 11. kl tüdrukud | 19 | 3 | 289 | 59,02 | 282 | 203 | 430 |
| 11. kl sugu märkimata | 5 | 0 | 205 | 36,72 | 219 | 150 | 243 |
| 11. klass | 36 | 6 | 247 | 47,87 | 251 | 177 | 337 |
| Kokku | 159 | 16 | | | | | |

Võrdluseks on kasutatud peamiselt eesti mittefiloloogidest haritlaste emakeelseid toimetamata kirjutisi, mis on loodud kõrgtaseme ametlikku eksamiolukorda ja kirjutusülesandeid jäljendades (vt Pajupuu jt 2010). Täiskasvanute korpuse kirjutisi on vaadeldud kui õpilase arengu loomulikku sihti ja hea keelekasutuse etaloni. Teine võrdluse suund seob emakeele ja teise keele (vt Alp jt 2013).

³ Oletus toetub veebis leitud artiklitele (kõikumine aastati ja paikkonniti 7% ja 13% vahel). HTM-i statistikas (<http://www.hm.ee/index.php?048055;30.9.2013>) neid andmeid ei leidunud.

3. Uuritavad muutujad ja metoodika

Sõnavaramuutujate nimetamisel on pilt kirju. Samastatakse ja eristatakse sõnavara rikkust (ingl *vocabulary richness*), leksikaalset varieeruvust (*lexical variation*) ja mitmekesisust (*lexical diversity*), mõõtes neid mitmesugustel meetoditel, mis kõik suhestavad eri sõnade arvu (N) sõnede ehk sõnaeksemplaride koguarvuga (n) tekstis või korpuses (vt Laufer, Nation 1995: 310, Vermeer 2000, Malvern jt 2004: 3 jm). Terminite ebahütlust analüüsib põhjalikult Guoxing Yu, kellele toetudes võiks sõnavara rikkuseks nimetada mitme uuritud muutuja kooskõla (Yu 2009: 238–241).

Uurisime kolme tüüpi muutujaid: 1) sõnavara üldine varieeruvus (kui palju eri sõnu kirjutaja kasutab); 2) sõnavara koostis ja ulatus (sagedaste ja harvade sõnade osatähtsus ja viimaste varieeruvus); 3) sisusõnade osatähtsus ehk leksikaalne tihe-
dus. Allpool esitatakse iga tunnuse mõõdikud (vrd L2-ga Alp jt 2013).

Kõigepealt SÕNAVARA ÜLDISEST VARIEERUVUSEST. Sõnade ja sõnede lihtne protsendi-
suhe (ingl *type-token ratio*) andis piloteerimisel kõige nõrgemaid korrelatsioone ega ole kõnekas (vt ka Vermeer 2000). Väiksemahuliste tekstide sõnavara varieeruvuse (rikkuse) mõõtmisel on Anne Vermeer (samam) pidanud sobivaks Uberi indeksi U , mis on osutunud heaks eristajaks nii eesti toimetamata täiskasvanutekstide (Pajupuu jt 2009) kui ka õpiteksti uurimisel (Puksand, Kerge 2012). Valem põhineb sõnede arvul n ja sõnade arvul N :

$$U = (\log n)^2 / (\log n - \log N)$$

Selle meetodi kõrvale valisime väga levinud Guiraud' indeksi G (Guiraud 1960, viidatud Mellor'i 2011 järgi), mis iseloomustab teksti sõnavara üldist varieeruvust sõnade arvu suhte järgi sõnearvu ruutjuuresse:

$$G = N / \sqrt{n}$$

Ehkki David Malvern jt (2004) samastavad sõnavara mitmekesisuse tema ulatusega (“the range of vocabulary and avoidance of repetition”), võib varieeruvus suur olla ka siis, kui kordamata kasutada sõnu paari tuhande kõige sagedama seast. Just nüüdistekstides harva esineva sõnavara valdamine on aga noortel kasin (vt Ehala jt 2010). Järelikult on oluline mõõta ka nn harva sõnavara (ingl *rare, advanced* või *sophisticated vocabulary*) rikkust (van Hout, Vermeer 2007), mida Soomes ja Eestis on nimetatud sõnavara ulatuseks (Tuokko 2007, Pajupuu jt 2009). Haritud sõnakasutus aitab süvenenumalt käsitleda teemat ning toetab kirjutise isikupära. Nimetame seda keerukaks sõnavaraks.

Näitamaks sõnavara ulatust, leiti iga osaleja LEKSIKAALNE PROFIIL (LFP; vt Laufer, Nation 1995, Meara 2005), kus kirjutise sõnad on sagedussõnastiku (Kaalep, Muischnek 2002) alusel jagatud 1000 sõna kaupa sagedusjärgkudesse: 1. järgus on 1000 sagedasimat sõna, edasi sagedusvahemikud 1001–2000, 2001–3000 9001–10 000 ning 11. järguna sagedussõnastikust väljapoole jäävad sõnad.

Üldistamiseks jagati iga kirjutise sõnastik tinglikult kaheks:

- *baassõnavara* ehk 4000 sagedasima sõna hulka kuuluvad sõnad (järgud 1–4), mis jagati jälgimise huvides omakorda kaheks: 1. järk (ülisagedad) ning koos 2.–4. järk;⁴
- *keerukas sõnavara* ehk kõik harvemad sõnad (järgud 5–11).⁵

KEERUKA SÕNAVARA OSATÄHTSUS (LS, ingl *lexical sophistication*; Laufer, Nation 1995: 309) näib eristavat olulisi arenguid, nt eesti keele kui L2 kirjutistes tasemeid A2 ja C1, olemata oluline lähemate tasemete võrdluses (vt Alp jt 2013: 281–282). Saamaks teada, mis toimub eesti koolis, mõõtsime seda näitajat teksti sõnede kohta:

$$LS = \text{keerukate sõnede arv} / n \times 100$$

KEERUKA SÕNAVARA VARIEERUVUST mõõtsime Advanced Guiraud' indeksiga AG (Daller jt 2003), mis suhestab sõnearvu n ruutjuurega n -ö harituma leksika:

$$AG = \text{keerukate sõnede arv} / \sqrt{n}$$

Laialt kasutatud LEKSIKAALSE TIHEDUSE indeks LD (ingl *lexical density*; Ure 1971) põhineb sisusõnade määral võrreldes funktsioonisõnadega, iseloomustades ühtaegu sõnavalikut ja süntaksit. Muutuja näitab teksti infotihedust ja võib olla teksti mõttelise keerukuse näitaja. (Vt arutlusi ja viiteid Kerge 2010: 45–48.) Lõtv tekst jääb paratamatult üldsõnaliseks. Ühtlasi räägib LD funktsioonisõnade määra muutumisest klassiti. Kuna arvutianalüüs (nt ka sagedussõnastik) ei erista funktsioonisõnu sisusõnade liigi piires (abimäärsõnu ja -verbe muudest), siis on leksikaalne tihedus arvatud järgmiselt:

$$LD = \text{nimisõnade} + \text{omadussõnade} + \text{verbide} + \text{adverbide sagedus}$$

Muutujat on mõõdetud ka õpitekstis kõrvu muud liiki tekstidega (Puksand, Kerge 2012: 189–191).

Hinnetega ja omavahel korreleeriti individuaalsete kirjutiste mediaannäitajad.

4. Tulemused ja diskussioon

Allpool peatutakse üldnäitajatel, hindamisse puutuval, sõnavara üksikmuutujatel ja lõpuks tervikpildil, diskuteerides pidevalt tulemuste üle. Arvutused, mida pole kommenteeritud, põhinevad tervikvalimil, sest õpetaja ei hinda töid kirjutaja emakeele järgi.

4.1. Üldised suundumused

TEKSTI PIKKUS. Nagu nähtus ülevaatest (tabel 1), on laste kirjutised erineva pikkusega, kuigi rühmiti oli seatud ühetaoline kirjutusaeg. Keskmiselt lühemad on igas vanuses poiste kirjutised – teksti pikkuse ja soo korrelatsioon on nõrk, kuid oluline ($r = 0,29$; $p < 0,001$). Teksti pikkus mõjutab oluliselt nii sõnavara üldist varieeruvust kui ka (küll nõrgalt) keeruka sõnavara näitajaid.

⁴ Ehkki Kaalepi-Muischneki (2002) sagedussõnastiku sissejuhatus järgi katavad 3000 sagedasimat sõna 80% kirjakeele korpuse sõnedest ning nende määr katab ka haritud täiskasvanute L1 esseedes u 66% tekstist (Pajupuu jt 2009: 192), oleme arvestanud ka võimalikke muid kontekste kui korpuses esindatud. Selline järgujaotus lubab õpilaste LFP-d võrrelda samas grandis uuritud L2 kirjutiste omaga (vt Alp jt 2013), kuid raskendab võrdlust haritlaste L1-ga (vt 4.4).

⁵ Google'iga saab otsida viidatud sagedussõnastiku (eesti kirjakeele korpuse ressursside 1 mln sõna baasil loodud sagedussõnastik). Siin harvaks loetud sõnu näeb seal pealkirja "10 000 sagedasimat lemmat" all, kust lingitud tabeli veerus "Kokku" on nende esinemiskordi 20 või vähem.

Uuringudisaini usaldusvääruse huvides kontrollisime ka seda, kui tugev on teksti pikkuse korrelatsioon eaga ja see osutus keskmiseks ($r = 0,5$). Tugeva korrelatsiooni korral ($r > 0,7$) moodustaksid nad koos tulemusi kallutava faktori. Kuna teksti pikkus ei mõjuta valimis hinnet, ei korreleeru eaga 5.-7. kl rühmas ja seos on nõrk ka 9.-11. kl osas ($r = 0,31$; $p = 0,004$), on tema mõju võrdlustele üldiselt väike.

Kirjutamise kulgu ja seeläbi teksti pikkust peaks mõjutama emakeel. Ann Chenoweth ja John Hayes (2001) on leidnud, et L2 puhul ilmnevad erinevused L1-st kirjutamise sujuvuses (sõnade arv minutis) ja mahus (sõnade hulk tekstiosa kirjutamisel ajaühikus). Siinses katses puudus teksti pikkusel deklareeritud emakeeltega vähimigi korrelatsioon.

KIRJUTAJA IGA. Ülevaate kirjutiste otsestest sõnavaranäitajatest annab klassiti tabel 2. (Leksikaalset tihedust vaatame allpool.)

Tabel 2. Olulisemate sõnavaramuutujate erinevused klassiti

| Klass | Möödik | | |
|--------|------------------------|---------------------------------|---|
| | Sõnavara varieeruvus G | Keeruka sõnavara varieeruvus AG | Keeruka sõnavara osatähtsus sõnedest LS |
| 5. kl | 5,73 | 0,40 | 6,80 |
| 7. kl | 5,96 | 0,45 | 6,94 |
| 9. kl | 8,06 | 1,10 | 14,1 |
| 11. kl | 9,56 | 1,82 | 18,4 |

Vanusel on erinevate sõnavaranäitajatega alati oluline vastastikmõju (vt tabel 3). Tugevasti sõltuvad kirjutaja east nii üldine (G) kui ka keeruka sõnavara varieeruvus (AG) ja keerukate sõnade osatähtsus (LS). Nõrgemini seostub eaga sisusõnade määr (LD).

Tabel 3. Sõnavaramuutujate sõltuvus kirjutaja vanusest

| Sõnavaramuutuja | Möödik | Vanuse ja muutuja korrelatsioonikordaja r |
|--|--------|---|
| Üldine varieeruvus | G | 0,83*** |
| | U | 0,59*** |
| Keeruka sõnavara varieeruvus ja ulatus | AG | 0,81*** |
| | LS | 0,74*** |
| Sisusõnade osakaal | LD | 0,52*** |

Märkus. Kordaja $r < 0,30$ puhul on korrelatsioon olematu, väga nõrk; $r < 0,70$ puhul keskmise tugevusega; $r > 0,70$ puhul tugev. Olulisuse kahe-suunaline tõenäosus on siin alati maksimaalne: *** $p < 0,001$.

Selgelt erineb muutujate ja ea seose tugevus kahe vanuserühma sees. 5.-7. kl rühmas on näitajate erinevus väike, nende korrelatsioon eaga alati statistiliselt ebaoluline ja väga nõrk. Kahe vanema klassi erinevus on silmnähtav, muutujate korrelatsioon eaga keskmine/tugev ja alati väga oluline. Enim eristavad 9. ja 11. klassi sõnavara variatiivsuse näitajad ($r = 0,69$).

Sõnavara üldvariatiivsust mõõdavad nii Guiraud' kui ka Uberi indeks, kuid viimane seostub eaga nõrgemini. Kuna peamine huviobjekt on õpilase areng, siis viimase tulemusi (vt U tabelis 3) kõigi parameetritega ei seostata.

KIRJUTAJA SUGU ei mõjuta valimis sõnavara üksikmuutujaid praktiliselt üldse. Nõrgalt seostub sooga vaid 9. ja 11. kl sõnavara üldvarieeruvus ($r = 0,27$; $p = 0,015$). Üksikmuutujad sõltuvad kirjutajast, alates 7. klassist mõjutab neid oluliselt iga, mitte sugu. Küll aga muutub sooline pilt tugevasti, kui vaadata sõnavara ealisi näitajaid koos (vt osa 4.6).

4.2. Hinded ja sõnavara

Hinded on nooremas rühmas veidi paremad kui vanemas: 5. ja 7. kl keskmine ja mediaan on puhas 4 ning 9.-11. kl keskmine 4-, mediaan pigem 3+.

Teisalt mõjutab sõnavara hinnet harva oluliselt (vt tabel 4), enamasti juhuslikult (vrd korrelatsioone ja nende suunda klassiti ühe muutuja piires) ja vähe (vt kordajate tähendusi tabeli 3 alt).

Ainus keskmiselt tugev ja oluline seos on hindel 5. kl sõnavara üldvarieeruvusega (G). Väheoluliselt ja vähem mõjutab 5. kl hinnet kõik: hinnet tõstab keeruka leksika varieeruvus (AG), aga langetavad keerukate sõnade ja sisusõnade suurem määr (LS, LD). Arvestatav, kuid väheoluline on keeruka sõnavara osakaalu vastasuunaline seos hindega 11. klassis (LS). Muud seosed puuduvad või on statistiliselt ebaolulised. (Vt tabel 4)

Tabel 4. Hinde ja olulisemate sõnavaramuutujate korrelatsioonid earühmiti

| Sõnavaramuutuja | Möödik | Klass | Hinde ja muutuja korrelatsioonikordaja <i>r</i> |
|--|--------|--------|---|
| Sõnade üldine varieeruvus | G | 5. kl | 0,54*** |
| | | 7. kl | -0,23 |
| | | 9. kl | 0,20 |
| | | 11. kl | 0,19 |
| Keerukama sõnavara varieeruvus ja ulatus | AG | 5. kl | 0,34* |
| | | 7. kl | -0,15 |
| | | 9. kl | -0,03 |
| | | 11. kl | -0,14 |
| | LS | 5. kl | -0,35* |
| | | 7. kl | -0,24 |
| | | 9. kl | -0,18 |
| | | 11. kl | -0,34* |
| Sisusõnade osakaal | LD | 5. kl | -0,32* |
| | | 7. kl | -0,06 |
| | | 9. kl | -0,19 |
| | | 11. kl | -0,16 |

Märkus. Olulisuse kahe-suunaline tõenäosus p on märgitud järgmiselt: * $p < 0,05$, ** $p < 0,01$, *** $p < 0,001$.

Klasse rühmitades⁶ on pilt teistsugune. 5. ja 7. klassil rühmana ei korreleeru hindegaga enam ükski tunnus. Seevastu vanematel, s.o 9. ja 11. kl osas hinnatakse kõrgemini mitmekesisema sõnavalikuga teksti: hinnet mõjutab märgatavalt ja oluliselt ($p < 0,001$) sõnavara üldvarieeruvus G ($r = 0,40$); ebaoluliselt ja väga nõrgalt ka keeruka sõnavara varieeruvus AG ($r = 0,21$; $p = 0,060$). Muud muutujad hindegaga ei seostu.

Helmut Dalleri ja David Phelani (2007) ning Roeland van Houti ja Anne Vermeeri (2007) järgi peaks kirjutise hinnet tõstma sõnavara suurem üldvarieeruvus – seda kinnitasid selgesti 5. kl tulemus ja vanemad klassid rühmana. Samadel viidetel peaks hinnet enam tõstma keeruka sõnavara osakaal, mis 5. ja 11. klassis hinnet väheoluliselt hoopis langetas. Tegurite alamõju õpilaskirjutiste hindele võib põhjustada see, et võrreldavad uurimused põhinevad L2-l, kus sõnavara arengut hoolega jälgitakse (vrd mõjuga 5. klassis). Halvema hinde võivad põhjustada keeruka sõnavaraga kaasnevad vead ja õpetaja hindamisstiil. Piloteerimisel ilmnis nt hinnete oluline seos hindajaga 9. ja 11. klassi osas: $r = 0,54$; $p < 0,001$. Hindetegurid, sh korrektsus vajavad aga eraldi uurimist.

Õpilase emakeelega hinne praktiliselt ei seostu ($r = -0,13$; $p = 0,111$): noorematel puudub sõltuvus üldse, vanema rühma lõikes on see väga nõrk ja ebaoluline ($r = -0,25$; $p = 0,255$), sest seost mõjutab nn kakskeelsete osalejate väike osatähtsus – juttu on 16 (10%) kirjutisest enamuskeelses koolis.

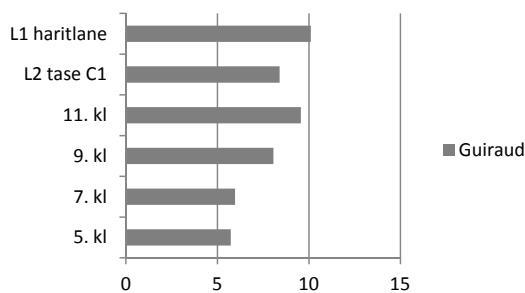
Hinde ja sõnavara nõrku seoseid seletab see, et õpetajal tuleb ka kirjutisi holistiliselt hinnates intuitiivselt arvestada eri tüüpi sisu- ja vormiseiku, kuid kirjutiste hindamise tahud ja rubriigid on praeguseks omaette teadus. Barbara Moskal (2003) toob esile kümneid rühmitatud komponente, mida kirjutamise eeldusena arvesse võtta ning hiljem kirjutiste juures hinnata, ja need on selgunud pärast tuhandete tööde läbilugemist. Oma liigitusega rõhutab ta rubriikhindamise eeliseid holistilise ees. Vicky Spandel ja Rick. J. Stiggins (1990) pakuvad välja seitse efektiivsele kirjutamisele iseloomulikku tunnust: sisu, struktuur, sõnum, sõnavalik, lausestus, kokkulepped (grammatika ja õigekiri, k.a interpunktsioon) ning esitus (käekiri, paigutus, kujundus). Kõiki neid hiljem on arvestatud ja arendatud (nt McMackin, Siegel 2002, Tierney, Marielle 2004 jpt, Uusen 2006a). Kirjutusoskuse arendamise olulisima tahuna emakeeletunnis rõhutab Ellen Krogh (2012) kirjutaja oskust sõnumit oma vaatepunktist organiseerida ja põhjendada (ingl *voice*). Niiugusel taustal on ilmne, et kirjutiste rubriikhindamine vajab eraldi tähelepanu, kusjuures sõnavara puhul on ilmseid eeliseid elektroonilisel hindamisel (vt nt Enright, Quinlan 2010, Cushing Weigle 2013). Viimane vajab aga eraldi käsitlust. (Vt ka osa 4.6.)

Tuleb siiski veel kord rõhutada, et hindamisuurimus on eraldi teadusharu – siinne alustav analüüs ei räägi midagi hindamise kvaliteedist, vaid annab teada kvalitatiivanalüüsi vajadusest ning lubab valida suurest materjalist välja sellised tööd, mida kvalitatiivuurimuses käsitada kindla parameetri või nende kombinatsiooni tasememudelina (vrd nt Tillema jt 2012).

⁶ Rühmitamine on siin küsitav, sest õpetaja kasutab eri eas sama hindedkaalat.

4.3. Üldine leksikaalne varieeruvus

Nagu eespool näidati (vt tabel 2) ja võiski oletada, on sõnavara üldvarieeruvus tugevasti seotud eaga ja need seosed alati väga olulised. Haritud emakeelekõnelejate toimetamata esseedes uurisime sama muutujat võrdluseks mõõdikuga G (mediaan 10,1), sest varem oli kasutatud vaid Uberi indeksit (Pajupuu jt 2009). 11. klassis jõuab G (9,6) haritlastele väga lähedale. Sama L2-näitaja sarnaneb sõltumata täiskasvanu suuremast üldkogemusest veel C1-tasemel 9. klassiga, jõudes 8,4-ni (Alp jt 2013: 280).⁷ Ülevaate leksika üldise varieeruvuse arengust annab joonis 1.



Joonis 1. Sõnavara üldise varieeruvuse loomulik kasv kirjutistes

Guiraud' indeks G näitab seoseid eaga tugevamana kui Uberi indeks U (vt tabel 3). Seejuures ei eristu varieeruvuselt 5. ja 7. kl sõnavara; ainult U näitab nõrka ja väheolulist pöördkorrelatsiooni ($r = -0,26$; $p = 0,025$). Küll erinevad taas keskmiselt tugevasti ja oluliselt ($p < 0,001$) 9. ja 11. klass – nii G järgi ($r = 0,68$) kui ka U järgi ($r = 0,49$). Seega leidis kinnitust, et Uberi indeksil on seda enam eeliseid, mida lühem on tekst (vt Vermeer 2000).

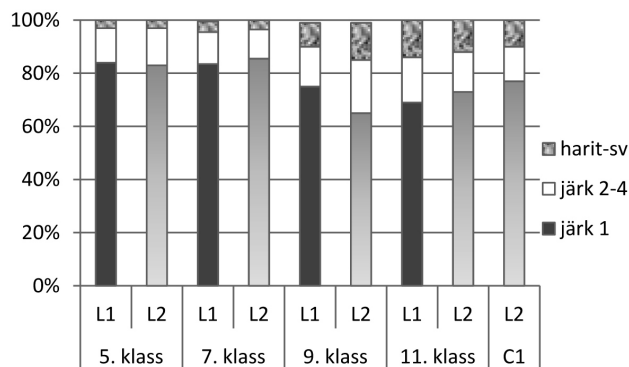
Kummatigi ei ole tulemused küsimärkideta. Näiteks võib leksika enam varieeruda neil, kes kasutavad ülesande täitmisel alusteksti sõnu (Baba 2009), nagu siinses katses 9. ja 11., kuid mitte 5. ja 7. klass. Niisugusel ülesandepüstitusel on aga koolieas selged põhjendused. 1) Täiskasvanu ei kirjuta asjalikku teksti kunagi ilma asjakohaseid varasemaid tekste lugemata ja arvesse võtmata. 2) Noorematel ei saa lugemistekste samas kasutada: eas, kus ümberjutustusega ollakse harjunud, kuid arvamuskirjutisega mitte, oleks niisugune katse tõenäoliselt lõppenud domineeriva ümberjutustamisega ja kirjutiste loomulikku sõnavara olnuks raske mõõta. Ka muutnuks lugemine kirjutuskatse noorematele väsitavaks. Küll viitab Kyoko Baba väide sellele, et loetud alusteksti ja ülesande sõnavarale tuleks kirjutise sõnastiku varieeruvust ja ulatust mõõtes anda selgelt vähem kaalu.

4.4. Leksikaalne profiil ja sõnavara keerukus

Võrdlesime sõnede sagedusjärgude jagunemist eraldi emakeelsete ja kakskeelsete laste individuaaltöodes, ehitades kolmel üldistataval sagedusjärgul põhineva LFP igas vanuserühmas. Profiilide võrdlus näitab hüpet kognitiivses arengus pärast 7. klassi: harituma-keerukama sõnavara osatähtsus sõnedes kasvab 5. kl kõrval põhikooli lõpus umbes kahekordseks ning gümnaasiumis kolmekordseks. (Vt ees

⁷ C1-eksamil ja õpilaskatses on kaalul ka hinne ja pingutus vahest suurem kui L1 haritlastel.

tabel 2 ja all joonis 2.) Joonisel 2 näeb ka 10% kakskeelsete laste eripära võrreldes kirjaliku täiskasvanutasemega C1 (Pajupuu jt 2010).

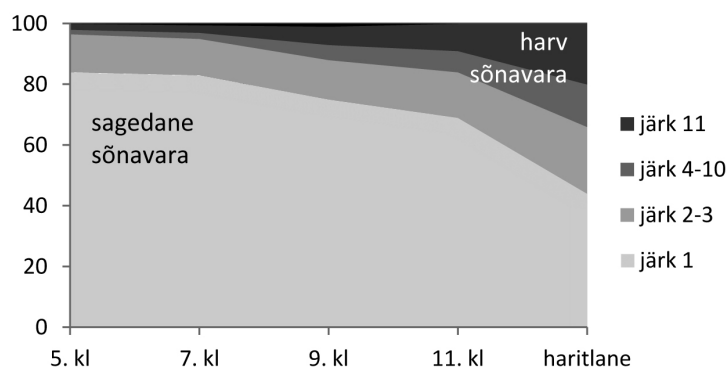


Joonis 2. Leksikaalsete profiilide võrdlus ea ja emakeelte järgi. Lühend *harit-sv* viitab keerukale sõnavarale (sagedusjärg 5–11). L2 (sagedusjärg 1 eristavana helehall) tähistab tinglikult ka kakskeelseid õpilasi

Keerukama sõnavara osatähtsus sõnedes (LS) on emakeelsetel õpilastel eraldi võetuna 5.-7. kl rühmas 6,8% ning 9.-11. kl rühmas 15,4%. Kakskeelsete laste rühmad teistest ei erine: nooremate mediaannäit kattub emakeelsetega (6,8%) ning vanemate rühmas, k.a nõrgema keskhindega venelased, on see teistest ebaoluliselt paremgi (17%).

Uuritud eestikeelse kooli vähesed kakskeelsed lapsed kasutavad keerukat sõnavara poole enam kui täisealised L2 kõnelejad, kelle haritud sõnavara kasvab A2-taseme 6,5%-lt 9,8%-ni C1-tasemel (Alp jt 2013). L2 leksikon on küll alati suhteliselt väike (vt Laufer, Nation 1995: 307), kuid sinne uurimus toob esile kakskeelsuse eelised, olgu siis tegu kakskeelse kodu või emakeelest erineva hariduskeelega. Muidugi vajab tulemus kontrollimist vähegi usaldusväärset valimil.

Kuna varasemad andmed (Pajupuu jt 2009) ei kajasta eesti haritlaste leksikat sagedusjärgudes 1–4 ja 5–11, ehitasime üldistava võrdluse teistsugustel alustel. Joonisel 3 on näha, et õpilased ei jõua 3000 sagedasima sõna osas (järgud 1–3) veel haritlaste kirjaliku arvamusevalduse tasemele: viimastel on nende osatähtsus u 65% sõnedest ja kahe vanimagi earühma õpilastel paarkümmend protsenti suurem.



Joonis 3. Õpilaskirjutiste sõnede sagedusjärgud (%) võrreldes eesti haritud täisealiste mittefiloloogidega (võrdluse allikas Pajupuu jt 2009: 192)

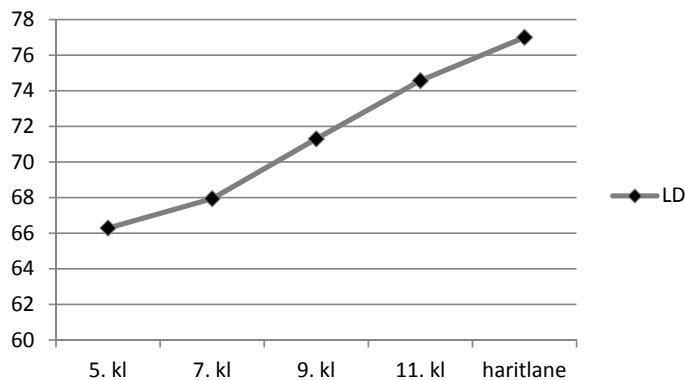
Kuigi Pilvi Alp jt (2013) 1. järku eraldi ei kaalu, vajavad sõnavarauurimustes ilmselt kvalitatiivset tähelepanu kaks äärmist sagedusrühma.

- 1) 1000 sagedasima sõna sõnesid on L1 haritlastel kolmandik vähem kui gümnasistidel (u 44% 11. kl 69% vastu). Põhikooli lõpus on nende määr u 74% sõnedest ning 5. ja 7. klassis võrdselt 83-84%. Ülisagedaste sõnade osatähtsus erineb noorematel (5. ja 7. klassi rühmal) haritlase keskmisest ligi kaks korda ja kahaneb teismeea jooksul u 14%.
- 2) Suur erinevus on sagedussõnastikust väljapoole jäävate sõnade osas – sellest sagedusrühmast, mida suuresti täidab sõnamoodustus, on haritlaskirjutistes sõnesid 2,3 korda enam kui 11. kl õpilastel (haritlaste 20% gümnasistide 9% vastu) ning 8–10 korda enam kui 5. ja 7. klassis (2% ja 2,5%). Ka põhikooli lõpetajad (6%) erinevad siin oluliselt täiskasvanutest ja gümnaasiumist.

Kuna aga haritlasi on siinkohal mõõdetud korpusena, vajab see võrdlus kontrolli individuaalsete kirjutiste alusel.

4.5. Leksikaalne tihedus

Tekstide leksikaalne tihedus (LD) kasvab suhteliselt ühtlaselt (joonis 4) ja kõigub õpilaste mediaanväärtuse 69 lähedal, mis on omane ka teismeea meilidele. 5. klassi tekst on mediaanist u 4% lõdvem, sarnanedes enim täisealiste erameilidele, ja 11. klassil u 6% tihedam (75). (Vt meilide kohta Kerge 2010: 47.) Täiskasvanud L1-esseedes on LD näitaja 77, toimetatud ja toimetamata kirjalike žanrite mediaan 71 (vt Puksand, Kerge 2012: 189).



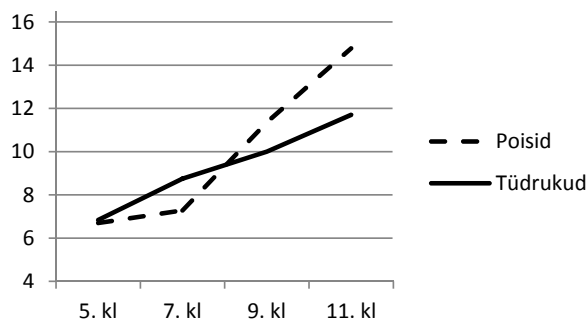
Joonis 4. Kirjutiste leksikaalne tihedus vanuserühmiti (haritlaste andmed: Pajupuu jt 2010)

Teksti tihedus sõltub vanusest üldiselt ($r = 0,52$; $p < 0,001$) ning kahe vanema rühma lõikes ($r = 0,34$; $p < 0,001$), kuid mitte noorematel. Leksikaalsel tihedusel on olulised seosed ($p < 0,001$) variatiivsuse näitajatega G ($r = 0,49$) ja AG ($r = 0,56$) ning keerukama sõnavara osatähtsusega LS ($r = 0,54$).

Lähemalt käsitleme seda teemat mujal kirjutiste süntaksi raames, jälgides sisu- ja funktsioonisõnu liigiti.

4.6. Sõnavara rikkuse kombineeritud indeksid ja hinne

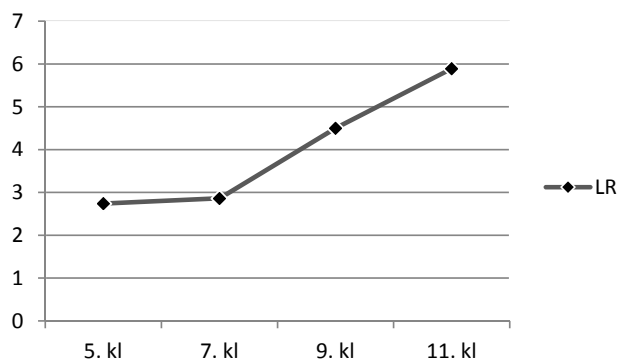
Kuivõrd sõnavara rikkuseks on soovitatud nimetada kõigi uuritud parameetrite kooskõla (Yu 2009), siis arvutasime kirjutiste sõnavara rikkuse tingliku kombineeritud indeksi LR esiteks kõigi 10-le keskmistatud mediaannäitude keskväärtusena klasside ja sugude kaupa. Poistel on sõnavara rikkuse tinglik indeks 10,03 ja tüdrukutel 7% madalam (9,32). Sugude LR-erinevus ei ole statistiliselt oluline 5. ja 7. klassil, kuigi tüdrukud lähevad 7. klassis tuntavalt ette, ega 7. ja 9. klassil, mille vahel poiste LR kasvab järsult. Küll aga on poiste sõnavara tüdrukute omast oluliselt paarkümmend protsenti rikkam 9. ja 11. klassis ($r = 0,64$; $p = 0,001$). Arengukõveraid võrdleb joonis 5.



Joonis 5. Sõnavara rikkuse üldistatud areng sooti

Ootamatuna mõjub poiste sõnavaraandmete taustal see, et poiste hinded on oluliselt nõrgemad nii 5.-7. klassis (hinde ja soo seos: $r = 0,47$; $p = 0,001$) kui ka 9.-11. klassis ($r = 0,30$; $p = 0,005$). Koguvalimis on soo mõju hindele märgatav ja statistiliselt väga oluline ($r = 0,36$; $p < 0,001$).

Teise meetodina kasutasime LR-indeksi ilmestamiseks sõnavara kõigi 1-le keskmistatud mediaannäitajate summeerimist klassiti. Earühmade sõnarikkust võrdleb keskmistatud näitajate summana joonis 6.



Joonis 6. Sõnavara rikkuse areng eanäitajate summaarse indeksi alusel

Hindamise sooaspekt sarnaneb mõneti riigieksamikirjanditega (vt Külli Habichti analüüsi 2007).⁸ Mediaanis oli poiste sisuhinne seal 11,1% võrra tüdrukutest nõrgem (sisuhinde kaal on 45% üldhindest). Keelehinne (kaal 55%), mida kujundanud õigekiri (35%), sõnastus ja stiil (15%) ning vorm ja kompositsioon (5%), jääb osi

⁸ Vt http://www.ekk.edu.ee/vvfiles/0/Kirjandi_analyys_2007.pdf (28.12.2013). Hiljem on riik igakülgse eksamianalüüsi asendanud koolide võrdlusega.

kaaludes kokku tüdrukutele alla juba 17,3% võrra. Kuna sõnavalik seostub otseselt sisuga, siis võib see alateadlikult mõjutada poiste suhteliselt paremat sisuhinnet. Habicht analüüsib vaid passiivset sõnavara (teemade mõistmist), kuid soovib õpetada vältima kõnekeelsust ja üldsõnalisust. Analüüs puudutab aktiivset sõnavara 2-3 näitega tähenduseksimuste ja kõnekeele mõju kontekstis, kus 2/3 näiteid ilmestab üld- ja erislängi.

Praegu on riigieksamikirjutise sisul võrdne 25% kaal õigekirjaga, sõnavara ja stiiliga ning vormi ja ülesehitusega; iga tahu hinne 1–5 korrutatakse teguriga 5 (vt Riigieksam 2013). Ilmne on tarve arvutieksamite järele, et sisu ja kompositsiooni saaks hinnata subjektiivselt, kuid keelekasutust objektiivselt.

5. Kokkuvõte ja järeldused

Lapsed alustavad 5. klassis väga erineva sõnavaraga – individuaalselt erineb üldine varieeruvus pea kaks korda ja keeruka sõnavara osas isegi 10–12 korda.

5. ja 7. kl statistiliselt lähedased mediaannäitajad vajaks arengupsühholoogilist selgitust. Teisalt võis 7. klassi sõnavalikut kaudselt mõjutada ka uurimuse disain: noorimatega ühesugune ülesanne võis kahandada oma arvamuse võimetekohast argumentatsiooni. Katset tuleks pisteliselt korrata raskema ülesandega.

7. klassi järel toimub kirjaliku sõnavara arengus ilmne hüpe, mis enim väljendub keeruka sõnavara lisanemise näol. 9. kl sõnavara erineb 11. kl omast väga oluliselt, kuid tuleb arvestada, et nende kahe klassi vahel on juba toimunud valik gümnaasiumisse, nii et kognitiivse arengu kohta ei ole see võrdlus kõnekas – põhikooli lõpetajatega tuleks võrrelda ka täiskasvanugümnaasiumi ja kutsekooli õpilasi.

Nagu võiski oletada, kasvab teismeeas kirjaliku sõnavara varieeruvus ja ulatus: 5. klassist 11.-ni tõuseb variatiivsusindeks keskmiselt 1,7 korda, kuid üle kolme korra kasvab keerukam sõnavara. Arenevad ja õpivad kooli jooksul kõik – tundub, et nõrgemad kirjutajad enamgi: Guiraud' järgi on kõige üheülbalisema sõnavaraga kirjutis 11. klassis 39% parem kui 5. klassi nõrgimal; enim varieeruva sõnavaraga kirjutised neis klassides erinevad vähem (27%). Summeeritud LR-näitaja järgi kasvab leksikarikkus teismeeas üle kahe korra.

Eritähelepanu vajavad soolise arengu erinevus ja kirjutise hinde kujunemine.

Sõnavara eristab eri vanuses lapsi hästi, kuid seostub hindega vaid osati, suhteliselt juhuslikult ja nõrgalt – ainult 5. klassis tõstab sõnavara üldine varieeruvus hinnet; nõrk pöördmõju on hindetele keerukal sõnavaral 5. ja 11. klassis. See on ka kõik.

On ilmne, et emakeeleomandaja korpusi on vaja arendada pidevalt ning kõiki vanuseid ja eri liiki tekste hõlmates, pluss tingimusi ja kirjutajaid igati (sh hindega) iseloomustades.

Järgmises artiklis tõlgendame õpilaskirjutiste juba uuritud tekstisüntaksi arengut. Siinsete kombineeritud ja üksiknäitajate järgi on võimalik leida ja kvaliteetiliselt analüüsida n-õ mudelkirjutisi. Ajapikku tuleks uurida harvema sõnavara koostist ja arvestada ülesande ja lugemisteksti mõju sellele; suhestada õigekeelsust ja konstruktsioonide sagedust jpm.

Kõik see aitab ajakohastada ka loovkirjutiste eakohase hindamise juhendeid. Pole aga kindel, kas leitud andmed annavad juhenditele selge sisu, kui viimaseid

endistviisi rakendada vaid intuitiivselt. Õigem on teha ja analüüsida õpilase elukäiku mõjutavaid töid (tasemetööd, lõpu- ja riigieksamid, k.a muude ainete esseed) üha enam arvutis ja varustada hindaja niiviisi objektiivsete lisaandmetega. Arvuti ei asenda kirjutiste subjektiivset hindamist, vaid täiendab seda objektiivsete näitajatega, k.a sõnavara omad.

Rakenduslikult aga on, mida siinsetest tulemustest õppida.

Ühelt poolt vajab laps veel 7. klassiski väga individuaalset lähenemist – alles pärast seda saab arvestada individuaalse sõnavara väiksema kõikumisega ja üldise loomuliku edasimineku. Teisalt juhivad tulemused taas tähelepanu asjaolule, et puudub ülevaade eri ainete õpetajate üldisest teadlikkusest kirjaoskuse, kirjutusoskuse ja sõnavara arengu või arendamise alal. Selgeks rääkimist vajab emakeeleõpetaja ja muude õpetajate vastutus kirjutusoskuse kujunemise eest.

Täheldasime, et kakskeelsetel ja õppekeelest erineva emakeelega lastel on sõnavaranäitajad omaealistega suhteliselt sarnased, ületades nt keeruka sõnavara osakaalus ligi kahekordselt C1-taseme näitajat (vt Alp jt 2013). Siinsed tulemused ei luba alahinnata eestikeelset haridust, nagu see gümnaasiumis on juba kehtestatud. Et aga sinne valim on üliväike ning teise keele täiskasvanud omandajal suurem üldine ja emakeelekogemus, tuleks võrrelda eesti- ja venekeelse kooli muukeelsete õpilaste eesti ja emakeele kirjutisi.

Earühmade eristajana tõestasid oma olulisust kõik kasutatud meetodid, kuid tulemuste kombineerimine vajab kokkuleppeid ja võrdlusi. Leksikaalses profiilis vajavad tähelepanu äärmised sagedusjärgud.

Batia Laufer ja Paul Nation (1995) on mõõtnud ka leksikaalset originaalsust ehk õppijale eriomaste sõnade osatähtsust, s.o leksikoni seda osa, mida teised kasutanud ei ole. Tulevikus väärriks see muutuja noorte sõnavara võrdluseks sobivana tähelepanu.

Tänu

Täname kõiki koole, õpetajaid ja õpilasi, kelle najal kogu uurimus teoks sai; Helin Puksandit, kes oma pika õpetajakogemuse toel aitas katseülesandeid koostada ja katseid korraldada, ning veel kord ETF-i (grant 8605).

Viidatud kirjandus

- Alp, Pilvi; Kerge, Krista; Pajupuu, Hille 2013. Measuring lexical proficiency in L2 creative writing. – Jozef Colpaert, Mathea Simons, Ann Aerts, Margret Oberhofer (Eds.). *Language Testing in Europe: Time for a New Framework?* Antwerpen: Linguapolis Universiteit Antwerpen, 274–286.
- Baba, Kyoko 2009. Aspects of lexical proficiency in writing summaries in a foreign language. – *Journal of Second Language Writing*, 18 (3), 191–208. <http://dx.doi.org/10.1016/j.jslw.2009.05.003>
- Berk, Laura E. 2003. *Child Development*. USA: Pearson Education Inc.
- Brügelmann, Hans 1990. Lapset ja kirjoittaminen. Kuinka lapset oppivat vähitellen kirjoittamaan oikein. – *Kielikukko*, 1, 21–25.
- Chenoweth, Ann N.; Hayes, John R. 2001. Fluency in writing: Generating text in L1 and L2. – *Written Communication*, 18 (1), 80–98. <http://dx.doi.org/10.1177/0741088301018001004>

- Corson, David 1995. *Using English Words*. Dordrecht: Kluwer Academic Publishers. <http://dx.doi.org/10.1007/978-94-011-0425-8>
- Cushing Weigle, Sara 2013. English language learners and automated scoring of essays: Critical considerations. – *Assessing Writing*, 18 (1), 85–99.
- Daller, Helmut; van Hout, Roeland; Treffers-Daller, Jeanine 2003. Lexical richness in the spontaneous speech of bilinguals. – *Applied Linguistics*, 24 (2), 197–222. <http://dx.doi.org/10.1093/applin/24.2.197>
- Daller, Helmut; Phelan, David 2007. What is in a teacher's mind? The relation between teacher ratings of EFL essays and different aspects of lexical richness. – Helmut Daller, James Milton, Jeanine Treffers-Daller (Eds.). *Modelling and Assessing Vocabulary Knowledge*. Cambridge Applied Linguistics. Cambridge: CUP, 234–244.
- Eesti keel 2011 = Õppekava. Põhikooli valdkonnaraamat: Eesti keel. Gümnaasiumi valdkonnaraamat: Eesti keel. Tallinn: REKK. <http://www.oppekava.ee> > Abiks õpetajale (25.9.2013).
- Ehala, Martin; Kerge, Krista; Lepajõe, Kersti; Sõrmus, Kadri 2010. Kõrgkoolide üliõpilaste eesti keele oskuse tase. Uuringukokkuvõte. Tartu: Tartu Ülikool.
- Enright, Mary K.; Quinlan, Thomas 2010. Complementing human judgment of essays written by English language learners with e-rater® scoring. – *Language Testing*, 27 (3), 317–334. <http://dx.doi.org/10.1177/0265532210363144>
- Goodman, Judith C.; Dale, Philip S.; Li, Ping 2008. Does frequency count? Parental input and the acquisition of vocabulary. – *Journal of Child Language*, 35 (3), 515–531. <http://dx.doi.org/10.1017/S0305000907008641>
- Guiraud, Pierre 1960. *Problèmes et méthodes de la statistique linguistique*. Dordrecht: D. Reidel.
- Habicht, Külli 2007. 2007. aasta eesti keele riigieksami analüüs. – Eesti keele eksamid. Tallinn: REKK, 1–26. http://www.ekk.edu.ee/vvfiles/o/Kirjandi_analyys_2007.pdf (28.12.2013).
- Hulstijn, Jan H. 2011. Language proficiency in native and nonnative speakers: An agenda for research and suggestions for second-language assessment. – *Language Assessment Quarterly*, 8 (3), 229–249. <http://dx.doi.org/10.1080/15434303.2011.565844>
- Kaalep, Heiki-Jaan; Muischnek, Kadri 2002. *Eesti kirjakeele sagedussõnastik*. Tartu: Tartu Ülikooli Kirjastus.
- Karlep, Karl 1998. *Psühholingvistika ja emakeeleõpetus*. Tartu: Tartu Ülikooli kirjastus.
- Kerge, Krista 2010. Kirjažanrite keeleparameetrid mitme tekstiliigi taustal. – *Emakeele Seltsi aastaraamat*, 55, 32–62.
- Kerge, Krista; Pajupuu, Hille 2010. Text-types in speech technology and language teaching. – Jorge L. Bueno Alonso, Dolores González Álvarez, Úrsula Kirsten Torrado, Ana E. Martínez Insua, etc. (Eds.). *Analizar datos > Describir variación / Analyzing data > Describing variation*. Vigo: Universidade de Vigo (Servizo de Publicacións), 380–390.
- Kern, Richard 2000. *Literacy and Language Teaching*. Oxford: Oxford University Press.
- Krogh, Ellen 2012. Writing in the literacy era: Scandinavian teachers' notions of writing in mother tongue education. – L1 – *Educational Studies in Language and Literature*, 12 (special issue). Iris Pereira, Brenton Doecke (Eds.). *A contribution to the inescapability of language*, 1–28.
- Krull, Edgar 2000. *Pedagoogilise psühholoogia käsiraamat*. Tartu: Tartu Ülikooli kirjastus.
- Laufer, Batia; Nation, Paul 1995. Vocabulary size and use: Lexical richness in L2 written production. – *Applied Linguistics*, 16 (3), 307–322. <http://dx.doi.org/10.1093/applin/16.3.307>
- Leiwo, Matti 1993. *Lapse keeleline areng*. Tallinn: Tallinna Pedagoogikaülikool.
- Lepajõe, Kersti 2012a. Kirjand kui tekstiliik. Riigieksamikirjandite tekstuaalsed, retoorilised ja diskursiivsed omadused. *Dissertationes philologiae estonicae Universitas Tartuensis* 21. Tartu: Tartu Ülikooli Kirjastus.

- Lepajõe, Kersti 2012b. Argumenteerimisvõtted riigieksamikirjandites. – Eesti Rakenduslingvistika Ühingu aastaraamat, 8, 123–138. <http://dx.doi.org/10.5128/ERYa8.08>
- Malvern, David; Richards, Brian; Chipere, Ngoni; Durán, Pilar 2004. *Lexical Diversity and Language Development: Quantification and Assessment*. Hampshire, New York: Palgrave Macmillan. <http://dx.doi.org/10.1057/9780230511804>
- McMackin, Mary C.; Siegel, Barbara S. 2002. *Knowing How. Researching and Writing Non-fictions*, 3–8. Portland: Stenhouse Publishers.
- Meara, Paul 2005. Designing vocabulary tests for English, Spanish and other languages. – C. Butler, M. A. Gómez González, S. Doval Suárez (Eds.). *The Dynamics of Language Use: Functional and Contrastive Perspectives*. Amsterdam: John Benjamins, 271–285.
- Mellor, Andrew 2011. Essay length, lexical diversity and automatic essay scoring. – *Memoirs of the Osaka Institute of Technology, Series B*, 55 (2), 1–14.
- Moskal, Barbara M. 2003. Recommendations for developing classroom performance assessments and scoring rubrics. – *Practical Assessment Research & Evaluation*, 8 (14). PAREonline.net. <http://pareonline.net/getvn.asp?v=8&n=14> (30.10.2013).
- Nippold, Marilyn A. 2004. Research on later language development: International perspectives. – Ruth A. Berman (Ed.). *Language Development Across Childhood and Adolescence: Psycholinguistic and Crosslinguistic Perspectives*. Trends in Language Acquisition. Amsterdam: John Benjamins, 1–8.
- Pajupuu, Hille; Kerge, Krista; Alp, Pilvi 2009. Sõnavara loomulik rikkus haritud keeleteadlaste tekstides. – Eesti Rakenduslingvistika Ühingu aastaraamat, 5, 187–196. <http://dx.doi.org/10.5128/ERYa5.12>
- Pajupuu, Hille; Kerge, Krista; Meister, Lya; Asu, Eva Liina; Alp, Pilvi 2010. Natural speaking and how to assess it. – *Trames: Journal of the Humanities and Social Sciences*, 59 (2), 120–140. <http://dx.doi.org/10.3176/tr.2010.2.02>
- Puksand, Helin; Kerge, Krista 2012. Õpikuteksti analüüs kirjaoskuse omandamise kontekstis. – *Emakeele Seltsi aastaraamat*, 57, 162–217. <http://dx.doi.org/10.3176/esa57.09>
- Riigieksam 2013 = Riigieksamite materjalid 2013. Eesti keele riigieksami eristuskiiri. Tallinn: Innove. http://www.innove.ee/UserFiles/Riigieksamid/2013/eesti_kee_re_eristuskiiri_040413.pdf (20.9.2013).
- Schoon, Ingrid; Parsons, Samantha; Rush, Robert; Law, James 2010. Children's language ability and psychosocial development: A 29-year follow-up study. – *Pediatrics*, 126 (1), 73–80. <http://www.ifs.org.uk/caytpubs/schoon9.pdf> (2.7.2013).
- Spandel, Vicki; Stiggins, Richard. J. 1990. *Creating Writers: Linking Writing Assessment and Instruction*. New York: Longman.
- Sõrmus, Kadri 2008. *Emakeeleõppija korpus. Statistiline analüüs ja veamärgendussüsteem. Magistritöö*. Tartu: TÜ eesti ja üldkeeleteaduse instituut.
- Tierney, Robin; Marielle, Simon 2004. What's still wrong with rubrics: focusing on the consistency of performance criteria across scale levels. – *Practical Assessment, Research & Evaluation*, 9 (2). PAREonline.net. <http://PAREonline.net/getvn.asp?v=9&n=2> (15.8.2013).
- Tillema, Marion; van den Bergh, Huub; Rijlaarsdam, Gert; Sanders, Ted 2012. Quantifying the quality difference between L1 and L2 essays: A rating procedure with bilingual raters and L1 and L2 benchmark essays. – *Language Testing*, 30 (1), 71–97. <http://dx.doi.org/10.1177/0265532212442647>
- Tuokko, Eeva 2007. Mille tasolle perusopetuksen Englannin opiskelussa päästään? Perusopetuksen päästövaiheen kansallisen arvioinnin 1999 Eurooppalaisen viitekehysten taitotasoihin linkitetty tulokset. *Jyväskylä Studies in Humanities* 69. Jyväskylä: University of Jyväskylä.
- Ure, Jean 1971. Lexical density and register differentiation. – G. Perren, J. L. M. Trim (Eds.). *Applications of Linguistics. Selected Papers of the Second International Congress of Applied Linguistics*, Cambridge 1969. London: Cambridge University Press, 443–452.

- Urm, Ada 2012. Individuaalsete erinevuste stabiilsus eesti laste sõnavara arengus. Magistritöö. Tartu: Tartu Ülikooli sotsiaalteaduskond, psühholoogia instituut.
- Uusen, Anne 2006a. Põhikooli I ja II astme õpilaste kirjutamisoskus. Tallinna Ülikooli sotsiaalteaduste dissertatsioonid 19. Tallinn: Tallinna Ülikooli Kirjastus.
- Uusen, Anne 2006b. Kirjutamisest ja kirjutama õpetamisest. Tartu: Atlex.
- van Hout, Roeland; Vermeer, Anne 2007. Comparing measures of lexical richness. – H. Daller, J. Milton, J. Treffers Daller (Eds.). *Modelling and Assessing Vocabulary Knowledge*. Cambridge: CUP, 93–115.
- Vermeer, Anne 2000. Coming to grips with lexical richness in spontaneous speech data. – *Language Testing*, 17 (1), 65–83. <http://dx.doi.org/10.1177/026553220001700103>
- Yu, Guoxing 2009. Lexical diversity in writing and speaking task performances. – *Applied Linguistics*, 31 (2), 236–259. <http://dx.doi.org/10.1093/applin/amp024>

Krista Kerge (Tallinna Ülikool) huviala on eesti kirjakeele allkeelte, žanrite ja individuaalstiili lingvistiline analüüs, eesti keele omandamine esimese ja teise keelena.
Narva mnt 25, 10120 Tallinn, Estonia
krista.kerge@gmail.com

Anne Uuseni (Tallinna Ülikool) huviala on emakeele kirjaoskuse omandamine ja õpetamine algkoolieas.
Narva mnt 25, 10120 Tallinn, Estonia
uusen@tlu.ee

Halliki Põlda (Tallinna Ülikool) huviala on kvalitatiivanalüüsi meetoodika, mõistete ja fenomenide analüüsi ja interpreteerimise võimalused avalikus kommunikatsioonis.
Narva mnt 25, 10120 Tallinn, Estonia
halliki.polda@gmail.com

TEENAGE VOCABULARY AND ITS ASSESSMENT IN CREATIVE WRITING

Krista Kerge, Anne Uusen, Halliki Põlda

Tallinn University

Within a study designed to find parameters of educated adult non-philologist L1-writings which give “ideal” benchmarks for assessing high-stakes L1 and L2 writing tests, students’ written vocabulary was targeted in grades 5, 7, 9, and 11 (age 12 to 18) to describe development towards those benchmarks (i.e., higher language cognition; see Hulstijn 2011). 159 students (90% of them native Estonians) were tasked to write an age-appropriate argumentative text on individual and social values. All writings (see table 1) were holistically assessed by the participants’ mother tongue teachers and then analyzed individually for their LFP (lexical profile), measuring the frequency range of all tokens (Laufer, Nation 1995, Meara 2005); their LS (lexical sophistication), measuring the percentage of advanced tokens (see Laufer, Nation 1995); the Uber, Guiraud, and Advanced Guiraud indices, measuring the diversity of general and advanced vocabulary (van Hout, Vermeer 2007, Guiraud 1960, Daller et al. 2003); and the LD (lexical density), showing the proportion of content words to function words (Ure 1971).

No significant difference was found between the vocabulary of grades 5 and 7, although individual values of parameters of the strongest writings may be up to ten times higher than those of the weakest ones. Differences between grades 9 and 11 were quite large and all statistically significant. From grade 5 to 11, variation of general vocabulary grows by 1.6 times and that of advanced vocabulary two times; the amount of sophisticated lexis grows by three times.

The writings of the few non-native participants (80% of them being bilingual also at home) were not typical of L2. The text length and vocabulary of this group did not differ from those of their age peers using L1. The lexical sophistication of this group was twice as big (17) as that of L2-speakers who had passed a C1-test (9.6; see Alp et al. 2013). There is an obvious need to test non-native writing in Estonian comparatively in two local school systems (with Russian or Estonian as the language of instruction).

As supposed, teachers assessing writing holistically were not able to take into account vocabulary parameters. In grade 5, ratings correlated significantly with Guiraud; in grade 11, there was a weak significant correlation between ratings and advanced vocabulary measures. Despite having better late vocabulary (Figure 5), boys’ ratings were significantly lower. This seems to argue in favour of partial computer-scoring.

Keywords: teenage, cognitive development, vocabulary variation and range, lexical sophistication and density, majority language in teaching