

## Õpikute sõnavara analüüsi koondtulemused

### Kommentaari tabelile 1

#### Tabeli kirjeldus

Tabeli 1 esimeses veerus on kõik sõnad, mida kasutatakse 1.–9. klassi muukeelse kooli eesti keele õpikutes ja ka eesti keele sagedussõnastiku 10 060 sagedasemat sõna. Välja on jäetud õpikutes kasutatud pärisnimed ja mittesõnad, mis on antud eraldi failidena. Sisse on jäetud ekslikult trükitud sõnad, et autoritel oleks kergem neid leida. Samas suurendavad viimased veidi sõnade arvu, eriti harvaesinevate sõnade hulka.

Tabeli teises veerus on esitatud nimisõnade abstraktsuste hinnangud ülaltoodud kolmepallilise skaala järgi. Teistel sõnaliikidel abstraktsust määratud ei ole ja nende puhul seisab selles veerus kriips.

Kolmandas veerus on sõna esinemissagedus eesti keele sagedussõnastiku aluseks olnud ühe miljoni sõna seas. Kõige sagedasemad sõnad eesti keeles on “olema” ja “ja” (vt tabel 1000)\*. Sagedussõnastikus on 10 060 eesti keele sagedasemat sõna, nende minimaalne sagedus on 5. Eesti keeles harvemini esinevate sõnade puhul on selles veerus sageduseks märgitud kriips, mis tähendab, et sagedus on väike ja teadmata.

Tabel 1000

#### Eesti keele sagedasemad sõnad

| <b>Sõna</b> | <b>Sagedus</b> |
|-------------|----------------|
| olema       | 44897          |
| ja          | 27232          |
| tema        | 21836          |
| see         | 18441          |
| ei          | 13813          |
| mina        | 13137          |
| et          | 12318          |
| kui         | 8600           |
| mis         | 8230           |
| ka          | 6194           |
| saama       | 5757           |
| oma         | 5738           |

Sõna sagedusest olulisem on teada, kas sõna on keeles suhteliselt sageli või suhteliselt harva kasutusel — selle iseloomustamiseks kasutasime järgmist tunnust “sagedusgrupp keeles”. Sagedusgruppi saamiseks reastasime sõnad sageduse järgi ja omistasime neile grupi vastavalt tabelile 1001.

\* Kasutame nii suuri numbreid, et ei tekiks segadust lepingu teksti tabelite numeratsiooniga.

Tabel 1001

Seos sõna keeles esinemise sageduse ja keele sagedusgrupi vahel

| Eesti keele sõnad,<br>mis kasutussageduselt on .... sõna seas | Sõna<br>sagedus<br>vähemalt | Sagedusgrupp<br>keeles |
|---|-----------------------------|------------------------|
| 1 - 2004  | 53                          | 1                      |
| 2005 – 3988   | 21                          | 2                      |
| 3989 – 5886   | 12                          | 3                      |
| 5887 – 7639   | 8                           | 4                      |
| 7640 -10060   | 5                           | 5                      |
| 10060 -   |                             | -                      |

Sagedusgrupid püüdsime koostada nii, et neisse kuuluks võrdne arv sõnu. See ei õnnestunud päris täpselt, kuna gruppide piirialadel on hulk ühesuguse sagedusega sõnu, mis tuleb kõik koos lülitada ühte rühma. Muidu juhtuks nii, et sama sagedusega sõnadest kuulub üks osa suuremasse sagedusgruppi, teine osa aga väiksemasse.

Analoogiliselt kahele eelnevale veerule on tabelis 1 moodustatud ka kaks järgnevat tulpa, ainult et aluseks on võetud analüüsitud õpikute sagedussõnastik. Sõnade sagedused õpikutes kokku on antud veerus pealkirjaga “Sagedus õpikuis”. Väärtus 0 selles veerus tähendab, et sõna ei esinenud õpikuis, see esines ainult eesti keele 10 060 sagedasema sõna seas.

Sõna sagedus keeles ja sõna sagedus õpikuis pole omavahel võrreldavad, kuna esimene on arvatud 1 000 000 ja teine 155 800 sõnakasutuse alusel.

Veerus “Sagedusgrupp õpikutes” on näidatud, mitmendasse tuhandesse sõna satub, kui õpikute sõnad reastada esinemissageduse järgi. Selle tunnuse saamiseks omistasime sõnadele grupi vastavalt tabelile 1002.

Tabel 1002

Seos sõna õpikutes esinemise sageduse ja õpikute sagedusgrupi vahel

| Sõnad õpikutes,<br>mis kasutussageduselt on ....<br>sõna seas | Sõna sagedus<br>õpikutes vähemalt | Sagedusgrupp<br>õpikutes |
|---|-----------------------------------|--------------------------|
| 1 - 929   | 23                                | 1                        |
| 930 – 1821  | 10                                | 2                        |
| 1822 – 2735   | 6                                 | 3                        |
| 2736 – 3770   | 4                                 | 4                        |
| 3771 - 4641   | 3                                 | 5                        |
| 4642 - 6208   | 2                                 | 6                        |
| 6209 - 10632  | 1                                 | 7                        |

Sagedusgrupp 7 koosneb sõnadest, mille sagedus on 1; sellesse gruppi kuulub 4423 sõna. Sagedusgrupp 6 koosneb sõnadest, mille sagedus on 2; sellesse rühma kuulub 1564 sõna.

Sagedusgrupp on keele puhul kaks korda mahukam kui õpikute puhul. Selle põhjuseks on soov lihtsamini võrrelda keele ja õpikute sagedusgruppe ja fakt, et keele 10 000 esimesest sõnast esineb õpikutes umbes 5000 sõna. Õpikute sagedusgruppid 6 ja 7 näitavad harvaesinevaid sõnu. Üldkeele sagedussõnastik viimaseid ei sisalda — seetõttu puuduvad sealt sagedusgruppid 6 ja 7.

Tabeli 1 viimastes veergudes on sõnade esinemissagedused üksikutes õpikutes. Numbriga 123 on tähistatud 1.–3. klassi õpikuid koos. Viimastes veergudes toodud arvude summa võrdub veerus “Sagedus õpikuis” toodud numbriga.

### Tabeli kasutamine

Tabel 1 on aluseks kõikidele järgnevatele tabelitele. Lisaks lepingu tekstis märgitud tabelitele 2 – 6 on tabeli 1 põhjal arvatud ka tabel 7: keeles sageli tarvitatavad sõnad, mis puuduvad õpikuis.

Samas saab tabeli 1 põhjal teha soovi korral otsustusi üksikute sõnade kohta. Toome mõned näited sellest tabelist tehtud väljavõtte põhjal (vt tabel 1003).

Tabel 1003

#### Õpikute analüüsi koondtulemused (väljavõtte tabelist 1)

| Sõna               | Abstraksus | Sagedus keeles | Sagedusgrupp keeles | Sagedusgrupp õpikuis | Sagedus õpikuis | Sagedus 123 | Sagedus 4 | Sagedus 5 | Sagedus 6 | Sagedus 7 | Sagedus 8 | Sagedus 9 |
|--------------------|------------|----------------|---------------------|----------------------|-----------------|-------------|-----------|-----------|-----------|-----------|-----------|-----------|
| aabits             | 1          | 16             | 3                   | 2                    | 11              | 9           | 2         | 0         | 0         | 0         | 0         | 0         |
| aade               | 3          | 9              | 4                   | 7                    | 1               | 1           | 0         | 0         | 0         | 0         | 0         | 0         |
| aadress            | 3          | 72             | 1                   | 1                    | 40              | 10          | 13        | 0         | 4         | 7         | 6         | 0         |
| aadress-<br>andmed | 3          |                |                     | 7                    | 1               | 0           | 0         | 0         | 0         | 1         | 0         | 0         |
| aamen              | 3          | 10             | 4                   |                      | 0               | 0           | 0         | 0         | 0         | 0         | 0         | 0         |
| aastaring          | 2          |                |                     | 5                    | 3               | 0           | 1         | 1         | 1         | 0         | 0         | 0         |

Sõna “aabits” on konkreetne sõna, mis esineb keeles ja õpikuis suhteliselt sageli. Sagedusgrupp keeles on 3 ja sagedusgrupp õpikuis 2. Need arvud on peaaegu võrdsed ja see on hea. Sõna “aabits” kasutatakse algklasside õpikuis 9 korda ja neljanda klassi õpikus 2 korda, hiljem ei kasutata. Seegi on hea.

Sõna “aade” on abstraktne sõna, mis esineb keeles keskmise sagedusega, aga õpikuis harva. See esineb vaid ühes algklassi õpikus ühe korra. Kuna sõna on raske, soovitaks selle välja jätta, kuid sellel nõuandel pole mõtet, sest algklasside õpikuid praegu ei muuda.

Sõna “address” on väga sageli tarvitatav sõna nii keeles kui ka õpikuis (sagedusgrupid võrduvad ühega). See esineb mitme klassi õpikus mitu korda. Sõna omandamisprobleeme ilmselt ei ole.

Sõna “addressandmed” ei ole keele 10 000 sagedasema sõna seas ja esineb vaid ühe korra seitsmenda klassi õpikus. Formaalselt võttes võiks soovitada selle sõna välja jätta, kuid hetkel niisugust nõuannet anda ei julge, kuna tegemist on liitsõnaga, mille mõlemad osad on tuntud (sõna “andmed” esineb õpikuis 13 korda) ja seitsmenda klassi õpilased ilmselt oskavad liitsõna tähendust selle osade järgi leida.

Sõna “aamen” on keeles keskmise sagedusega, kuid õpikuis ei esine. Kui on soov õpikute sõnavara rikastada, siis võiks selle sõna sisse tuua mõnes vanema klassi õpikus (abstraktne sõna).

Sõna “aastaring” pole eesti keele 10 000 sagedasema sõna seas, kuid esineb neljanda, viienda ja kuuenda klassi õpikus, igaihes üks kord. Uue sõna õppimise algul oleks otstarbekas seda rohkem korrata või see sõna üldse välja jätta.

Tabelist 1002 ilmneb, et õpikute 10 700 sõna seas on 4423 sõna, mille sagedus on 1 ja veel 1564 sõna, mille sagedus on 2 — seda on üsna palju. Võrdluseks võiks tuua omal ajal populaarsete Reneli ja Sotteri inglise keele õpikute analüüsi tulemused (vt tabel 1004)\*.

Tabel 1004

Inglise keele algõpetuse õpikute analüüsi tulemused

| Karakteristik                                | 4. klass | 5. klass | 6. klass |
|--|----------|----------|----------|
| Nädalatunde                                  | 3        | 2        | 2        |
| Sõnade koguarv õpikus                        | 7132     | 11355    | 8853     |
| Erinevate sõnade arv                         | 302      | 555      | 636      |
| Erinevate uute sõnade arv                    | 302      | 352      | 267      |
| Sõnade arv, mis kordusid õpikus alla 6 korra | 118      | 283      | 420      |

**Comment [JM1]:** •Mikk J., Mikk E., Tirmaste J. Computerised readability analysis of textbook of English. In: J. Mikk (Ed) Problems of textbook effectivity. Tartu: University of Tartu, 1991, pp 112 – 121.

Nendes õpikutes oli kolme klassi kohta 420 sõna, mis kordusid õpikuis vähem kui 6 korda. Analüüsitavais eesti keele õpikuis on üle 6000 sõna, mis korduvad alla 3 korra. Nüüd on õppetunde tunduvalt rohkem, kuid siiski tundub olevat vastuolu kahe seisukoha vahel: ühelt poolt soov kasutada autentseid tekste ja teiselt poolt soovitus korrata uut sõna õpikus umbes 7 korda selle kindlaks omandamiseks.

Lisaks sellisele analüüsile sõnade kaupa arvutasime tabeli 1 põhjal lineaarkorrelatsiooni sõnade sagedus keeles ja sõnade sagedus õpikuis vahel. See oli 0,927, mis tähendab, et keeles sagedamini tarvitatamaid sõnu kasutatakse tihedamini ka õpikuis. See väga tugev seos on õpikuile plussiks.