

# PÕHIKOOLI AINESÕNASTIKUD\*

HEIKI-JAAN KAALEP, JAAN MIKK

## Sissejuhatus

**A**astal 2004 kuulutas Mitte-eestlaste Integratsiooni Sihtasutus välja Eesti vene õppekeelega koolide keskastme ainesõnastike loomise konkursi. See oli seotud asjaoluga, et vene õppekeelega koolid lähevad järk-järgult üle eestikeelsele õppele, mida peab toetama sobiv õppematerjal. Kui õppetöö ei toimu emakeeles, siis on õpitavast arusaamiseks vaja sõnastikke (vähemalt aeg-ajalt, sõltuvalt õpilase keeleoskuse tasemest). Tuleb arvestada, et keskastme õppekavad on küllalt keerulised ja näevad ette suure hulga selliste terminite omandamist, mida üldkeele sõnastikus ei pruugi leiduda.

Ainesõnastike koostamise ülesandeks oli anda vene õppekeelega koolide õpilastele ja õpetajatele praktiline vahend ainete õppimiseks ja õpetamiseks eesti keeles. Ainesõnastikesse tuli valida ainult vastava aine põhiterminid, s.o need, mis on tähtsad aine põhiprintsiipide mõistmisel ja mis on defineeritud õpikus või ainekavas. Ainesõnastike terminite koguarv pidi olema 13 000–15 000.

Mitte-eestlaste Integratsiooni Sihtasutuse tellitud sõnastikud kavandati poolkakskeelsetena, s.t ükskeelsete eesti ja kakskeelsete eesti-vene sõnastike kombinatsioonina: neis pidid olema nii eestikeelsed märksõnad ja tähendusseletused kui ka venekeelsed vasted.

Eesti sõnavara õppimine ei piirdu ainult sõnade algvormide ja tähenduste omandamisega, selgeks tuleb saada ka sõnade käänamine ja pööramine. Lähteülesandes oligi ette nähtud, et iga termini juures on antud ka põhilised käände- ja pöördvormid. Selles osas pidid sõnastikud sarnanema EKI „Eesti-vene sõnaraamatuga” (1997–2006).

Peale eesti-vene tõlkesuuna nägi lähteülesanne ette ka vastassuuna, kuid seda lihtsustatud loendite kujul, andes vene terminitele eesti vasted ilma seletuste, käänamise ja muu infota.

Artikkel annabki sõnastiku tegemise planeerija ja korraldaja vaatepunktist lähtudes ülevaate, kuidas loodi 7.–9. klassi kaheteistkümne aine sõnastikud: ajaloo (Hallik 2005), bioloogia (Toom, Teller 2005), füüsika (Pärtel 2005), geograafia (H.-M. Pedastsaar, T. Pedastsaar 2005), keemia (L. Tamm, T. Tamm, Tuulmets 2005), kehalise kasvatus (Hein 2005), kunstiõpetuse (Parmasto, Laur, Kidron 2005), matemaatika (Abel, Lepmann 2005), muusika-

\* Ainesõnastike koostamist, trükkimist ja vene õppekeelega koolidele tasuta jaotamist toetas EL-i PHARE programm. Sõnastike Interneti-versiooni koostamist toetasid EL-i eContenti projekt „Eurotermbank” ja Eesti riiklik programm „Eesti keele keeletehnoloogiline tugi”.

Artikli autorid on tänulikud sõnastike autoritele, toimetajatele, retsensentidele, arvutispetsialistidele ja teistele töötajatele (kokku 75 spetsialistile), kes projektides osalesid. Samuti oleme tänulikud artikli anonüümsele retsensendile, kelle soovistest oli palju kasu.

õpetuse (Leppoja 2005), käsitöö, kodunduse ja tehnoloogiaõpetuse (Peedisson, Rihvk, Soobik 2005), inimeseõpetuse (Kõiv 2005), ühiskonnaõpetuse (Piir 2005) sõnastik. Kirjeldatakse sõnastike eesmärke ja kirjestruktuuri ning loomisprotsessi etappe, kusjuures pikemalt peatutakse märksõnavalikul ja tähenduse seletuste koostamisel. Eraldi vaadeldakse eesti keele morfoloogiaalase tarkvara kasutamise võimalusi sõnastiku loomise hõlbustamisel. Tegemist on tagantjärele vaatega kaheaastasele perioodile, mille jooksul need 12 oskuskeesõnastikku valmisid.

Sõnastikud on välja andnud Tartu Ülikooli Kirjastus ja nende tiraaž oli suurema nädalatundide arvuga ainete puhul 2000 ja teiste ainete puhul 1000. Eesti vene õppekeele koolid on saanud sõnastikud tasuta. Need on kättesaadavad kooliraamatukogudes ja ka suuremates üldraamatukogudes ning elektrooniliselt Keeleveebis (<http://www.keeleeveeb.ee>).

Sõnastikke teevad inimesed inimeste jaoks. Nii tegijad kui ka kasutajad soovivad, et tulemus oleks võimalikult hea. Sellest ka vajadus analüüsida nii valmis sõnastikke kui ka nende loomise protsessi. Teadmine, kuidas mingi tulemuseni on jõutud, on oluline nii tulemuse enda hindamisel kui ka uute sõnastike tegijatele eesmärkide püstitamisel.

Suunatus just vene keelt emakeelena kõnelejatele ning valdkonnaspetsiifiline sõnavara eristavad neid ainesõnastikke „Eesti-vene õppesõnastikust” (Kaarma, Külmoja, Lagle, Leemets, Linnas, Öim 1984), milles on üldkeele 7000 märksõna, väljendid näitematerjalina ja just vene vastete juures ära toodud ka nende muutevormid.

## 1. Sõnaraamat ja keeleõpe

Üldkeele sõnastike kasutamise ja keeleõppe uuringud on näidanud, et kui võõrkeele õppijal on võimalik valida nii võõrkeelne seletav sõnastik kui ka kakskeelne tõlkesõnastik, siis eelistab õppiija viimast (Piotrowsky 1989). Samal ajal on küllalt levinud seisukoht, et ükskeelne seletav sõnastik pakub adekvaatsemat ja täpsemat informatsiooni (Laufer, Levitzky-Aviad 2006), nii et keeleõppe seisukohalt oleks tulusam kasutada just seda (Koren 1997). Ehkki eeltoodud vaatepunkti põhjendatuse üle võib arutleda (Piotrowsky 1989), on praeguseks välja antud üsna palju õppijatele mõeldud poolkakskeelseid sõnastikke, mis peaksid ühendama ükskeelse sõnastiku tugevad küljed kakskeelse sõnastiku kasutusmugavusega. Sellistes sõnastikes on võõrkeelsed märksõnad ja tähendusseletused, millele on lisatud emakeelsed vasted. Eestis väljaantud poolkakskeelsete sõnastike näiteks võib tuua Kernermani kirjastuse poolkakskeelsete sõnastike sarja alusel loodud inglise-eesti sõnastikud (EELIS 2001; Passport 2000; Password 1995).

Tuleb küll tunnistada, et keele omandamise edukuse ja sõnastike kasutamise seosed ei ole päris selged. Leidub nii uurimusi, mis tõestavad sõnastike kasutamise positiivset mõju (nt Knight 1994), kui ka uurimusi, mis sellist mõju ei ole tuvastanud (nt De Ridder 2002), ning uurimusi, mis kinnitavad poolkakskeelsete sõnastike positiivset mõju (nt Laufer 1995; Laufer, Levitzky-Aviad 2006), kui ka uurimusi, milles järeldatakse, et kõige paremad on just tavalised kakskeelsed sõnastikud (Bogaards 2005). Eri uurimuste tulemused ei pruugi olla iseenesest vastuolulised, kui võtame peale uurimismetoodika

erinevuste arvesse, et keeleõppijad erinevad nii oma keeleoskuse taseme kui ka õppimisstrateegiate poolest, samuti seda, et sõnastikud ei erine mitte ainult tüübi, vaid ka mahu ja kvaliteedi poolest. Samal ajal tähendavad vastuolulised tulemused seda, et praegu on veel ühese vastuseta küsimused, kuidas on võõrkeele õppimise edukus seotud sõnastiku kasutamisega ning milline oleks parim sõnastik.

Poolkakskeelsete sõnastike mall, kus mõiste definitsioon on ainult ühes keeles, terminite vasted aga mitmes keeles, on terminoloogiasõnastike puhul küllalt tavaline. Eesti keele kesksete näidetena võib siin tuua Stockholmis ilmunud viiekeelse „Kunsti leksikoni” (Paris, End 1986) ja Tallinnas ilmunud viiekeelse (eesti-inglise-soome-saksa-vene) „Elektroenergeetikasõnastiku” (2005). Mõlemas sõnastikus on definitsioonid ainult eesti keeles, terminid ise aga mitmes keeles (sõnastike elektroonilised versioonid on kasutatavad ka Keeleveebis, <http://www.keelveeb.ee>).

## 2. Sõnastikuartiklite ülesehitus

Meie loodud ainesõnastikes on nagu tähestikuliselt järjestatud sõnastikes ikka nii põhiartiklid kui ka viiteartiklid. Esimesed sisaldavad esitatavat infot, teised on vajalikud selleks, et info leidmist mugavamaks muuta. Sõnastiku põhiartiklis on termin ise ja tema kohta käiv info esitatud kaheksal väljal järgmises järjekorras ja kujundusega:

**termin** <õigeks käänamiseks või pööramiseks vajalikud vormid> VALDKOND (= **sünonüümne termin**) termini tähenduse seletus ▷ Näited Vt ka *seotud terminit* • *vene vaste*

Näiteks:

**.arv\_m.õõt** <-möödu, -m.õõtu, -m.õõte e -m.õõtusid> ARITMEETIKA (= **arvmõõtkava**) kaardi või plaani \*mõõtkava esitus \*arvude jagatisena, kus \*jagatav on 1. ▷ arvmõõt 1 : 100 tähendab, et kaardil või plaanil on kõik pikkused 100 korda väiksemad kui tegelikkuses. Vt ka *joonmõõt* • *числовой масштаб* (Abel, Lepmann 2005)

Väljad 3 (ainevaldkond), 4 (sünonüümne termin), 6 (näited) ja 7 (viide seotud terminile) võivad põhiartiklist ka puududa.

Mitte igas sõnastikus ei eristata ainevaldkondi ega viidata seotud kirjetele. Põhikooli ainekavas neljas sõnastikus (Leppoja 2005; Hallik 2005; Peedisson, Rihvk, Soobik 2005; Abel, Lepmann 2005) tuleb aga ette ka võõrkeelseid termineid, mille hääldust on vaja kirjeldada. Võõrkeelse termini kirjes asendab hääldus termini muutevorme, näiteks:

**staccato** (loe: stakaato) lühidalt, teravalt, lüh *stacc.* Vt lisa 19 • *стаккато* (Leppoja 2005)

Näidetes on märksõnades hääldusmärke ja seletuses viitemärkideks tärnid.

Eesti keele ortograafia ei ole päris häälduspärane: kirjas ei eristata alati teist ja kolmandat vältet, ei märgita ebaregulaarset rõhku ega liitsõnas sõna-piiri. Põhiküsimuseks on, kas inimesele, kelle emakeel pole eesti keel, on vaja eesti keele hääldamiseks abi? Meie arvates on vaja, et õppijad suudaksid sõnastiku abiga hääldada õigesti selliseid sõnu nagu näiteks *käristi* ja *turisti*, või et nad saaksid aru, mis vahe on näiteks sõna *piste* nimetaval ja omas-taval käändel.

Meie pakutud hääldusalane lisainfo on eesti spetsialiseeritud sõnastikes esimene omataoline ja see on küllalt haruldane ka eesti üldkeele sõnastike hulgas. Vaid õigekeelsussõnastikud on seda liiki infot kajastanud järjepidevalt alates 1933. aastast, mil ilmus Elmar Muugi „Väike õigekeelsus-sõnaraamat”, nagu nähtub Tiiu Erealti ülevaatest (2007). Peale ÕS-ide on vastav info esita-tud ainult „Eesti keele raskete sõnade sõnaraamatus” (2001) ja kaks- ja mit-mekeelsete sõnastike hulgas ainsa erandina „Eesti-vene sõnaraamatus” (1997–2006).

Üks põhjusi, miks sõnade hääldamist hõlbustavat infot õppijatele mõeldud sõnastikesse lisada, on asjaolu, et hääldusele tuginevad aktiivse morfoloogia reeglid. Meie arvates on aktiivse ja passiivse morfoloogia eristamine ja tund-mine keeleõppe seisukohalt oluline. Eesti keele sõnadest kuulub passiivsete-se muuttüüpidesse üsna väike hulk: 2700 sõna 1976. aasta „Õigekeelsussõna-raamatust” (EKG: 139). Seega näeme, et on küllalt vähe sõnu, mille kääna-mine või pööramine pole sõna kuju alusel mõistatav, nii et muutumismall tuleb pähe õppida. Valdav osa eesti keele sõnavarast muutub regulaarsete mallide alusel.

Kui soovime aidata vene koolilastel eesti keele sõnamuutmist omandada, siis tuleks neile anda informatsiooni, mille valgusel reeglid selgemalt esile tulevad, ent mida kirjapilt alati ei paku.

Entsüklopeedias ja sageli ka terminoloogiasõnaraamatutes tähistatakse tähenduse seletustes viitemärgiga (enamasti tärniga) sõnad, mis esinevad sama allika märksõnadena. Sellist tähistust sisaldavad näiteks „Kunsti lek-sikon” (Paris, End 1986) ja „Eesti standard EVS-ISO/IEC 2382. Infotehnoloogia. Sõnastik” (<http://www.riso.ee/et/it-standardimine/eesti-infotehnoloogia-standardid>). Nii tehakse definitsioonist arusaamine kasutajale lihtsamaks ja hoitakse kokku aega, kui mõne sellise sõna tähendust vaatama peab: kasuta-ja teab, kas selle leiab parajasti kasutatavast allikast või peaks otsima kusa-gilt mujalt. Otsustasime samasugused viitemärgid lisada ka ainesõnastikes-se. Sõnastike veebiversioonis on viitemärkide asemel hüperlingid.

Viiteartikkel suunab kasutaja põhitermini juurde, kus on nii selle kui ka sünonüümsete terminite seletus. Viiteartikkel sisaldab kolme välja ja ta kuju on järgmine:

**sõna** <õigeks käänamiseks või pööramiseks vajalikud vormid> Vt **asja-kohane termin**

Näiteks:

**.arv\_m.õõt\_kava** <-kava, -kava, -kavu e -kavasid> Vt **arv mõõt** (Abel, Lep-mann 2005)

### 3. Töökorraldus

Sõnaraamatu koostamine on tavaliselt kollektiivne töö, mis koosneb mitmest etapist, mille jooksul eri inimesed tegelevad sõnaraamatu erinevate aspektidega.

Sõnaraamatut võib käsitleda teataval viisil struktureeritud tekstina, aga ka andmebaasina. Mis tahes andmebaas on teatavasti esitatav temaga sisuliselt ekvivalentse relatsioonilise andmebaasina ehk tabelite süsteemina. Sõnastiku kirje vastab tabeli reale, sõnastikus esitatav infoväli aga tabeli veerule. Sõnastiku loomine võib toimuda ridade kaupa, s.t et autor koostab sõnastikku kirjete haaval. Kuid sõnastikku võib koostada ka veergude kaupa, s.t et kõigi kirjete puhul vaadatakse korraga ainult üht infovälja ja täidetakse seda. Segaduste vältimiseks rõhutame, et sõnastiku andmebaasina käsitlemine iseloomustab koostajate lähenemisviisi, mitte kasutatavat tarkvaraplatvormi. Meie käsitlesime sõnastikke kui andmebaase, mida täideti mõnes etapis rida-, mõnes etapis aga veerghaaval, kusjuures osa sõnastiku sisust genereeris arvuti.

Kõige mugavam on kasutada selliseid pabersõnastikke, milles märksõnad on esitatud tähestikulises järjestuses, kusjuures sõna erinevad tähendused on esitatud samas sõnastikuartiklis:

**r.uut** <ruudu, r.uutu, r.uute e r.uutusid> 1 ARITMEETIKA, ALGEBRA (= **teine aste**) \*avaldis, mis saadakse antud \*avaldise \*korrutamisel iseendaga.  $\Delta a \cdot a = a^2$ , loetakse: a ruudus. • *квадрат; вторая степень* 2 ГЕОМЕТРИА (= **korrapärane nelinurk**) võrdsete \*külgedega \*ristkülik. • *квадрат; правильный четырехугольник* (Abel, Lepmann 2005)

Terminil võib olla mitu tähendust, aga ka sünonüüme. Kui sõnastiku koostaja peaks peale sisu pidevalt silmas pidama vormi, milles sõnastik lõpuks kasutajale esitatakse – sõnaartiklite esitamise järjekorda, artiklisisese info esitamise viisi (näiteks tähenduste numeratsiooni) ja artiklite omavahelisi seoseid (viiteid teistele artiklitele) –, siis muutuks ta ülesanne palju raskemaks. Et koostajate tööd lihtsustada ning seega vältida inimvõimete piiratudusest tulenevaid eksimusi – viiteid terminitele, mida sõnastikus pole, olemasolevatele terminitele viitamata jätmist, ebaühtlast sünonüümide tähistamist –, otsustasime sõnastikusiseste viidete loomise usaldada arvutile ning sõnastike sisu koostamisel järgida põhimõtet, et keskne koht peab olema mõiste tähendusel. Nii sai koostaja keskenduda tähendusele, millega seoti ühelt poolt eestikeelne termin ja selle sünonüümid ning teiselt poolt venekeelne termin koos sünonüümidega.

Selleks kasutati andmebaasiprogrammi „Microsoft Access” jaoks loodud Arvi Tavasti rakendust või tekstiredaktoreid ja tabelarvutusprogramme, kust A. Tavast andmed oma andmebaasi importis.

Ühe sõnaraamatu koostamisel läbiti vähemalt kolm korda järgmised neli etappi:

1. Terminiloendi koostamine ja/või parandamine.
2. Mõistete seletuste ja näidete kirjutamine, terminitele vene vastete leidmine.

### 3. Arvutitöötlus:

- 3.1. viitemärksõnade genereerimine
- 3.2. märksõna põhivormide genereerimine
- 3.3. seletustes kasutatud terminitele viitemärkide lisamine, kui need terminid esinevad sama sõnastiku märksõnaloendis
- 3.4. sõnastiku trükiks vormistamine

### 4. Sõnastiku hindamine.

Sõnastike hindamisel osalesid eri etappidel eri eksperdid, kellelt autorid said vastavalt ka erinevat tagasisidet.

## 4. Sõnastike koostamine

**4.1. Terminivalik.** Põhiterminite loetelusse võtsime kõigepealt vastavas ainekavas toodud terminid. Nende leidmisel olid põhiallikaks eestikeelsed õpikud, mis haridus- ja teadusminister kinnitas koolide kasutada 2004/2005. õppeaastaks. Sõnastike autorid otsisid põhitermineid eelkõige õpetükkide pealkirjadest, õpikus esiletõstetud sõnade seast ja õpiku lõpus olevast sõnastikust, kui see olemas oli. Abiks oli kogu õpiku tekst, õpetaja aineraamat ja sõnaraamatu autori aineõpetamiskogemus. Mõnel juhul, näiteks bioloogias, oli kasutada ka varem koolidele koostatud terminite sõnastik. Mõnes aines aga oli osa ainekavast õpikutega katmata ja siis pöördusid autorid terminiloendi koostamisel ka kolleegide poole.

Terminite valikul püüdsime arvestada ka seda, kui võrd rasked need on vene õppekeelega koolide õpilastele. Terminiloendi koostamisel eelistasime raskemaid termineid.

Sõnastike autoritele oli terminite hulga kohta orienteeruvad arvud ette antud. Need arvud leiti, lähtudes vastava aine tundide arvust põhikooli viimases astmes ja arvestades ka ainekava mahtu.

Sõnastike autoriteks olid kõrgkoolide õppejõud-metoodikud või eesti õppekeelega koolide vastavate ainete õpetajad. Et aga sõnastike kasutajateks on vene õppekeelega koolide õpilased ja õpetajad, siis palusime koostatud loendeid retsenseerima eeskätt vene õppekeelega koolide õpetajaid. Retsensendid soovitasid enamasti sõnastikku termineid lisada ja neid soovitusi arvestati.

**4.2. Käändevormide ja hääldusmärkide lisamine: automaatne morfoloogia praktikas.** Käändevormide ja hääldusmärkide lisamine tähendas meie sõnastikus kõigi kirjete juurde kahe esimese infovälja – lisamärke sisaldava termini ning põhivormide – automaatset loomist.

Eestis on sõnastikukirjete osalist genereerimist arvutile usaldatud varemgi, näiteks „Eesti-vene sõnaraamatu” loomisel EKI-s (Viks 2000). Kasutatud tarkvara kirjeldusest võib välja lugeda, et EKI kirjegeraator võimaldas sellel ajal, kui meie sõnastikeprojekt algas, küll määrata sõna muuttüübi, kuid raskusi oli liitsõnapiiri, kolmandavärtelise silbi ning rõhu automaatsel tuvastamisel. Seetõttu oli ikkagi vaja lingvisti ettevalmistatud sisendit.

Õnneks oli võimalik kasutada meile sobiva funktsionaalsusega lingvistilist tarkvara: eesti keele morfoloogilist analüsaatorit ning süntesaatorit ESTMORF (Kaalap, Vaino 2000). ESTMORF on mõeldud ammendavalt katma üldkeelt ja ta sisaldab ka võimalust kasutada erinevaid heuristilisi võt-

teid käändevormide analüüsiks ja sünteesiks selliste sõnade puhul, mis programmi leksikonist puuduvad.

Märksõnadest käändevormide moodustamine ei olnud siiski nii sujuv, nagu võinuks loota, sest sõnaraamatu märksõna ei pruugi olla algvormis üksiksõna, mida morfoloogilise sünteesi programm just eeldab. Termin võib olla mitmesõnaline, võörkeelne, mitmuse vormis ja tal võib olla kasutamispääsmeid käändevormide osas, kusjuures ühel terminil võib olla korraga mitu sellist kõrvalekallet. Enne põhivormide genereerimist tuli teha märksõnade morfoloogiline analüüs, et määrata kindlaks, mis tüüpi juhtumiga on tegemist (eestikeelne sõna ainsuse nimetavas käändes või mitmuses, võörkeelne sõna või mitmesõnaline termin). Alles seejärel asuti genereerima põhivorme, kasutades iga tüübi jaoks just talle sobivat algoritmi. Väär algoritm annaks valesid vorme.

Sõnaraamatute koostamise ajal ilmnes mõneti üllatavalt, et nii põhivormide esitusviisi kui ka nende valiku puhul oli vastakaid seisukohti. Esiteks oli probleemiks liitsõna osasõnu ja hääldust tähistavate lisamärkide kasutamine. Inimesed, kelle emakeeleks ei ole eesti keel, suhtusid sellesse positiivselt. Siiski oli ka arvamusi, eranditult eesti keelt emakeelena kõnelejal, et lisamärgid muudavad märksõnade lugemise ebamugavaks. Teiseks tekitasid vaidlusi sõnad, millel (keelereeglite järgi) võivad küll esineda teatavad käändevormid, mida aga mõnes ainevaldkonnas kunagi ei kasutata. Näiteks keemias käsitletakse mitmesuguseid aineid, kuid konkreetne aine (näiteks sinihape) esineb ainult ainsuses.

Töö käigus otsustati kõik sellised näited juhtum juhtumi haaval. Otsused ei pruugi seejuures olla väga järjepidevad. Esiteks tegid neid eri inimesed, erinevate teemavaldkondade sõnaraamatute autorid. Teiseks on eri sõnade kasutamisel erinevaid semantilisi piiranguid ja seda peegeldab ka nende käändevormide kasutatavus. Sõnaraamat ei saa siiski liigselt detailidesse laskuda, vaid peab andma süstemaatilist informatsiooni, otsuse langetab kasutaja ise. Väikest arvu selgelt erandlikke sõnu on võimalik eraldi tähistada.

Põhikäändevormide automaatne moodustamine lähtub sõna algvormist. Kuid eesti keele omapäraks on, et ainsuse nimetav kääne võib eri sõnadel olla samasugune, ehkki muud vormid erinevad, näiteks *niit: niidi/niidu*. Algvormist põhikäändevorme genereerides tuleb otsustada, mis neist konkreetsel juhul sobivad. Paraku tugineb see otsus sõna tähendusele, mis on kirjas mõiste definitsioonis või (enamasti, kuid mitte alati) oletatav liitsõna ülejäänud komponentide põhjal, kuid tähenduse mõistmist arvutilt eeldada ei saa. Ülle Viks on homonüümiat pidanud lausa peamiseks takistuseks, mis segab märksõna muutevormide täisautomaatset genereerimist (Viks 2000). Homonüümse algvormiga sõnad esinevad ka liitsõna järelkomponendina küllaltki sageli. Et liitsõnamoodustus on eesti keeles vaba ja produktiivne, võib selliseid liitsõnu olla erialasõnastikus üsna palju, kusjuures osa neist võib muudest sõnastikest hoopis puududa. Seetõttu ei ole ka leksikonile tugineval algoritmil mõnikord andmeid, mille alusel õigeid vorme valida. Probleemi lihtsustab küll asjaolu, et konkreetses valdkonnas on homonüümse algvormiga sõnade tähendustest tavaliselt kasutusel ainult üks. Näiteks geograafias on juttu mitmesugustest niitudest, käsitöös aga niitudest.

Seda probleemi ei olnud põhikäändevormide genereerimisel võimalik lahendada muud moodi, kui et kõik sellised homonüümse algvormiga sõnad pidi igas sõnastikus ikkagi eraldi üle vaatama inimene.

Eesti keeles on mitmuses paralleelvorme nii passiivsesse kui ka aktiivsesse morfoloogiasse kuuluvate muuttüüpide korral. Peale selle on terve hulk aktiivsesse morfoloogiasse kuuluvaid sõnu, mis võivad käänduda eri muuttüüpide järgi (Viks 1992), näiteks *senioreid* (nagu *õpikuid*), *seniore* (nagu *seminare*).

Tegelikus keeletarvitusel mõningaid vorme muidugi eelistatakse ja teised võivad tunduda kasutajatele veidrad või lausa valed. Kui põhivormid on sõnastikus iga märksõna juures välja toodud, hakkavad kasutamisharjumustest tingitud küsitavused selgemalt silma kui ainult muuttüübi numbri või tüüpsõna esitamise puhul. Kuidas peaks käituma koolisõnastiku koostaja? Kas esitama ainult sagedamini kasutatavad vormid või kõik lubatavad? Ainesõnastike sihtgruppi arvestav lihtsuse nõue toetaks justkui esimest valikut. Siiski otsustasime esitada kõik lubatavad vormid. Põhjendus sarnaneb häälde- ja liitsõna osasõna piiride lisamise põhjendusega: aktiivsesse morfoloogiasse kuuluvate sõnade paralleelvormide regulaarne olemasolu peegeldab sõnamuutmissüsteemi sisemisi reegleid, mille tundmine peaks keele omandamist lihtsustama. Kui annaksime õpilastele ainult sagedasemad vormid, siis oleks info aktiivse morfoloogia reeglite mõttes puudulik ja ebasüsteemaatiline ning sõnamuutmissüsteemi omandamine hoopis raskem. Sama lahendus on kasutusel ka näiteks „Eesti-vene sõnaraamatus“ (1997–2006).

**4.3. Mõistete seletuste koostamine.** Sõnastikku valitud mõistete tähenduseletused pidid olema hästi lihtsad ja ka vähese eesti keele oskusega õpilastele arusaadavad, aga samal ajal teaduslikult õiged. Et sõnastikud on mõeldud eeskätt õpikute heaks mõistmiseks, siis on tähenduseletused õpikutes antule võimalikult lähedased. Keerukate mõistete puhul oli autoritel võimalus lisada näiteid. Mõnel juhul on seletus antud joonisena.

Kõige sagedamini on autorid kasutanud sõnastikes reaalse definitsiooni: on esitatud üldmõiste ja defineeritava mõiste liigitunnused, aga on kasutatud ka nominaaldefinitsiooni: seletatavat mõistet on kirjeldatud.

Selleks et tähenduseletused saaksid hästi mõistetavad, soovitati autoreil järgida lihtsa keele reegleid (Mikk 2000), mis jaotati kolme gruppi: 1) tundmatutest sõnadest hoidumine; 2) konkreetsema esituse eelistamine; 3) liiga pikkadest lausetest hoidumine.

Õpilastele tundmatud sõnad näidati autoreile kätte arvutianalüüsis, kus morfoloogilise analüsaatori abiga loodi kõigi tähenduseletustes kasutatud sõnade algvormide loend, mida võrreldi sama sõnaraamatu märksõnade ning 10 000 kõige sagedamini esineva sõna nimekirjaga (Kaalap, Muischnek 2002). Eesmärgiks oli leida sõnad, mis puuduvad mõlemast nimekirjast ja on seega eeldatavasti rasked, mistõttu neid tuleks kas tähenduseletustes vältida või lisada uute märksõnadena ja omakorda selgitada. Sõnaraamatute koostajad said need loendid oma käsikirjade ülevaatamise ajal, kuid ajapuudusel jõudsid neid vähe kasutada.

Tähenduseletustes kasutatud sõnade tähendust aitavad mõista ka lisatud näited.

Kui tähenduseletuses on olnud vaja kasutada sama aine terminit, siis näidatakse viitemärgi abil, et see termin on samas sõnastikus selgitatud.

Eestikeelses tekstis võib sõna esineda peale algvormi ka mõnes muus vormis. Selleks et määrata sõnavormi ja märksõna seost, tuleb morfoloogilise



analüsaatori abil kõigepealt leida sõna algvorm. Kui see on leitud, on arvutil lihtne seda otsida märksõnade loendist ja leidmise korral lisada vastava sõnavormi juurde viitemärk (tärn).

Kirjeldatud algoritm eeldab esiteks, et keele kogu sõnavarast on sõnastikus märksõnadena esitatud ainult piiratud hulk, ja teiseks, et sõnavormi kasutatakse samas tähenduses nagu vastav defineeritud märksõna. Üldkeele sõnastiku puhul ei kehtiks ilmselt kumbki eeldus. Piiratud valdkonna sõnastike puhul aga juhtub harva, et sõna on ühes valdkonnas kasutusel mitmes tähenduses (vt tabeli 2 veergu „mitmetähenduslikud terminid”) või et mõni määr- või kaassõna langeb kokku märksõna vormiga, nagu *vahel* langeb kokku matemaatikatermini *vahe* alalütlevaga.

Esituse konkreetsuse üle otsustasime nimisõnade järgi, jaotades need kolme kategooriasse: 1) konkreetsed nimisõnad, mis tähistavad meeltega vahetult tajutavaid esemeid ja olendeid; 2) nimisõnad, mis tähistavad meeltega vahetult tajutavaid nähtusi; 3) abstraktsed nimisõnad, mis tähistavad meeltega vahetult tajumatuid mõttekonstruktsioone.

Autoritel soovitati kasutada konkreetseid nimisõnu rohkem ja abstraktsed vähem. Esituse konkreetsust aitavad tõsta ka näited ja illustratsioonid.

Liiga pikad laused võivad lugeja mälu üle koormata ja siis ei ole lausest võimalik aru saada. Lühemate lausete kirjutamiseks soovitati autoritel: 1) hoiduda liigsetest sõnadest; 2) hoiduda lauselühendeist; 3) kasutada tegusõnu pöördelistes vormides; 4) paigutada omavahel seotud sõnad üksteise lähedusse; 5) eelistada rindlauset põimlausele.

Autoritele esitati ka orienteeruvad andmed, milline tekst sobib 7.–9. klassi õpilastele (tabel 1).

Tabel 1.

### Põhikooli õpilastele sobiv tähendussetuste keerukus

	7. klass	8. klass	9. klass
Lause pikkus sõnades	11	12	13
Konkreetsete ja abstraktsete nimisõnade suhe	9:1	5:1	3:1

Autorid on arvestanud neid soovitusi oma võimaluste kohaselt ja see on soodustanud arusaadavate selgituste koostamist. Projekti juht ja autorid arutasid läbi esimesed 10–20 tähendussetust, otsides koos võimalikult täpset ja arusaadavat seletust. Tüüpilised eksimused olid algul järgmised: 1) tähendussetuses polnud esitatud mõiste kõik vajalikud liigitunnused või puudusid need üldse; 2) üldmõiste oli esitatud tähendussetuse lõpus, mistõttu defineeritav mõiste oli lauses üldmõistest väga kaugel ja tähendussetusest oli raske aru saada; 3) kasutati raskesti mõistetavaid sõnu; 4) oli liiga palju sõnu ja detaile, mis muutsid tähendussetuse raskesti arusaadavaks.

Alljärgnevas näites on abstraktset mõistet selgitatud nii lihtsalt, nagu oli meie taotlus: on kasutatud tuntud sõnu, lause on lühike.

**südame\_tunnistus** <tunnistuse, -tunnistust, -tunnistusi> inimese sisemine hääl, mis aitab tal head halvast eristada. • советъ (Kõiv 2005)

## 5. Koostatud sõnaraamatud

Vaadeldavad sõnaraamatud sisaldavad kokku 12 373 mõistet, mis on esindatud 14 170 terminiga. Suurim on ajaloo sõnaraamat (Hallik 2005) 1590 mõistega, mida kirjeldab 1897 terminit, ja väikseim inimeseõpetuse sõnaraamat (Kõiv 2005) 361 mõistega, mida kirjeldab 408 terminit (vt tabelit 2). Seega võib neid kõiki liigitada väikesemahuliste oskussõnastike hulka (Erelt 1982: 145).

Tabel 2.

### Mõistete ja terminite hulk sõnaraamatutes

Teema	mõisted	terminid	sh mitmetähenduslikud
Ajalugu	1590	1897	63
Bioloogia	1384	1523	21
Füüsika	959	1133	12
Geograafia	1213	1424	1
Inimeseõpetus	361	408	0
Keemia	1151	1410	4
Kehaline kasvatus	1003	1031	12
Kunst	777	950	3
Käsitöö, kodundus, tööõpetus	1343	1455	17
Matemaatika	1231	1422	6
Muusika	660	728	2
Ühiskonnaõpetus	701	789	1
Kokku	12373	14170	142

Suurim keskmine terminite arv (1,23) on ühe mõiste kohta keemia sõnastikus, kuid kõige rohkem sünonüümseid termineid – seitse – pakub ühele mõistele hoopis ajaloo sõnastik (*teorent, teotöö, teoorjus, teokoormis, teokohustus, mõisatöö, mõisategu*). Väikseim terminite ja mõistete arvu (1,03) suhe on kehalise kasvatus sõnastikus.

Lisaks 142 juhtumile, kus termin tähistab mitut mõistet ühes sõnastikus, on ka termineid, mis esindavad mitme valdkonna mõisteid. Kahes või enamas sõnastikus on selgitatud 1011 mõistet: üht kuues, kaht viies, 13 neljas, 129 kolmes ja 866 mõistet kahes sõnaraamatus. Kõige populaarsem mõiste on *periood*, mis esineb kuues sõnastikus. Mõistel, mis esineb mitmes sõnaraamatus, võib olla täiesti erinev tähendus: näiteks *periood* 'ajavahemik' (Kõiv 2005), 'ajajärk; vahemik, mille järel mingi nähtus kordub' (Parmasto, Laur, Kidron 2005), 'näitab täisvõnke sooritamise kestust' (Pärtel 2005), 'perioodilisustabeli horisontaalne rida, mille moodustavad ühesuguse elektronkihtide arvuga elemendid' (L. Tamm, T. Tamm, Tuulmets 2005), 'väikseim muusikaline lõpetatud vormiosa, mis võib esineda ka iseseisva teosena' (Leppoja 2005), 'korduva te numbrite rühm kümnendmurru murdosas' (Abel, Lepmann 2005).

Tagasiside on olnud positiivne. Maie Soll on küsitlenud sõnastike kasutamise kohta seitsme vene kooli direktorit. Üks ei olnud sõnastike olemasolust teadlik, kuid ülejäänud hindasid neid väga positiivselt. Sõnastikke kasutatakse kõigis või enamikus ainetes, mille jaoks nad on olemas. Õpetajate sõnul peavad õpilased sõnastikke kasulikuks mitmel põhjusel. Õpilased olevat

meeldivalt üllatunud, et leiavad kõik õppeainega seotud terminid ühest allikast. Nad kasutavad sõnastikke nii uute eestikeelsete sõnade tähenduste otsimisel kui ka vene terminitele eesti vastete leidmisel.<sup>1</sup>

Sõnastikke (põhiliselt füüsika ja keemia sõnastikku) kasutavad ka Eestis õppivad hispaaniakeelsed õpilased, kellele pakkuvat eelkõige huvi terminite tähenduse ja sõnade käänamisega seonduv.<sup>2</sup>

Õpetajate sõnul on sõnastike põhiline probleem nende vähesus: kõigile õpilastele ei pruugi neid tunnis jätkuda.

## Kokkuvõte

Artiklis oleme vaadelnud 12 ühetüübilise sõnastiku loomise protsessi: töö etappe, üleskerkinud probleeme ning väljapakutud lahendusi. Tegemist on eesti termineid (kokku üle 14 000) ja tähendussetusi ning terminite vene vasteid sisaldavate ainesõnastikega, mis on mõeldud vene õppekeelega põhikoolile ning mis on vabalt kasutatavad aadressil <http://www.keeleeveeb.ee>. Mitme aine puhul on see sõnastik vastava aine esimene eestikeelne terminoloogiasõnastik. Sõnastike omapäraks on, et iga termini juures on antud ka ta põhikäändevormid.

Kõigi sõnastike loomisel läbiti ühesugused etapid, alates terminite loendi koostamisest ja lõpetades sõnastike trükkimise ning Internetti panemisega. Sõnastike tegemise käigus hinnati tehtu taset mitmel etapil ning autoritele anti tagasisidet.

Sõnastike loomise protsess ise sõltus ka kasutatud tehnilistest vahenditest: mõistepõhiseks koostamiseks mõeldud andmebaasirakendusest ning keeletehnoloogilisest tarkvarast. Ilma nendeta oleks sõnastike loomine nõudnud rohkem aega ja tööd ja lisaks oleks kvaliteet kannatanud.

## Kirjandus

- Abel, E., Lepmann, L. 2005. Matemaatika mõisted 7.–9. klassile. Eesti-veneeesti sõnastik. Tartu: Tartu Ülikooli Kirjastus.
- Bogards, P. 2005. [Retsensioon] Which Dictionary for Whom? Receptive Use of Bilingual, Monolingual and Semi-Bilingual Dictionaries by Polish Learners of English. – *International Journal of Lexicography*, nr 18, lk 114–116.
- Deridder, I. 2002. Visible or Invisible Links. Does the Highlighting of Hyperlinks Affect Incidental Vocabulary Learning, Text Comprehension, and the Reading Process? – *Language Learning & Technology*, kd 6, nr 1, lk 123–146. <http://lt.msu.edu/vol6num1/pdf/deridder.pdf>.
- EELIS = Inglise-eesti seletav sõnaraamat. Tallinn: Festart, 2001.
- Eesti keele raskete sõnade sõnaraamat. Koost R. Mägi, K. Randvere, M. Vaba. Tallinn: TEA Kirjastus, 2001.
- Eesti-vene sõnaraamat 1–4. Tallinn: Eesti Keele Instituut, 1997–2006.
- EKG = M. Ereht, R. Kasik, H. Metslang, H. Rajandi, K. Ross, H. Saari, K. Tael, S. Vare, Eesti keele grammatika. I Morfologia. Sõnamoodustus. Tallinn: Eesti Teaduste Akadeemia Eesti Keele Instituut, 1995.

<sup>1</sup> M. Soll, eravestlus oktoobris 2007.

<sup>2</sup> M. Soll, eravestlus oktoobris 2007.

- Elektroenergeetikasõnastik. Koost M. Meldof. Tallinn: Tallinna Tehnikaülikooli elektroenergeetika instituut, 2005.
- Erelt, T. 1982. Eesti oskuskeel. Tallinn: Valgus.
- Erelt, T. 2007. Õigekeelsussõnaraamatud läbi sajandi. – ÕSi lätted. Õigekeelsussõnaraamatud läbi sajandi. Eesti keele õigekirjutuse-sõnaraamat 1918. Koostanud U. Sutrop. Tallinn: Eesti Keele Sihtasutus, lk 5–34.
- Hallik, T. 2005. Ajaloo mõisted 7.–9. klassile. Eesti-vene-eesti sõnastik. Tartu: Tartu Ülikooli Kirjastus.
- Hein, V. 2005. Kehalise kasvatuse mõisted 7.–9. klassile. Eesti-vene-eesti sõnastik. Tartu: Tartu Ülikooli Kirjastus.
- Kaalap, H.-J., Muischnek, K. 2002. Eesti kirjakeele sagedussõnastik. Tartu: Tartu Ülikooli Kirjastus.
- Kaalap, H.-J., Vaino, T. 2000. Teksti täielik morfoloogiline analüüs lingvisti töövahendite komplektis. – Arvutuslingvistikalt inimesele. Tartu: Tartu Ülikooli Kirjastus, lk 87–99.
- Kaarma, G., Külmoja, I., Lagle, T., Leemets, H., Linna, A., Õim, A. 1984. Eesti-vene õppesõnastik. Tallinn: Valgus.
- Knighth, S. 1994. Dictionary Use while Reading. The Effects on Comprehension and Vocabulary Acquisition for Students of Different Verbal Abilities. – The Modern Language Journal, kd 78, lk 285–299.
- Koren, S. 1997. Quality versus Convenience. Comparison of Modern Dictionaries from the Researcher's, Teacher's and Learner's Points of View. – Teaching English as a Second or Foreign Language, kd 2, nr 3. <http://tesl-ej.org/ej07/a2.html>.
- Kõiv, K. 2005. Inimeseõpetuse mõisted 7.–9. klassile. Eesti-vene-eesti sõnastik. Tartu: Tartu Ülikooli Kirjastus.
- Lauffer, B. 1995. A Case for a Semi-Bilingual Dictionary for Productive Purposes. Kernerman Dictionary News, nr 3. <http://www.kdictionaries.com/newsletter/kdn3-2.html>.
- Lauffer, B., Levitzky-Aviad, T. 2006. Examining the Effectiveness of 'Bilingual Dictionary Plus' – A Dictionary for Production in a Foreign Language. – International Journal of Lexicography, kd 19, lk 135–155.
- Leppoja, K. 2005. Muusikaõpetuse mõisted 7.–9. klassile. Eesti-vene-eesti sõnastik. Tartu: Tartu Ülikooli Kirjastus.
- Mikk, J. 2000. Textbook. Research and Writing. Frankfurt am Main–Berlin–Bern–Bruxelles–New York–Oxford–Wien: Peter Lang.
- Passport. English-Estonian Learner's Dictionary. Tallinn: Festart, 2000.
- Password. Inglise-eesti seletav sõnaraamat. English Dictionary for Speakers of Estonian. Tallinn: TEA Kirjastus, 1995.
- Paris, R., End, E. 1986. Kunsti leksikon. Maali, skulptuuri, graafika, arhitektuuri ja tarbekunsti terminid eesti, inglise, prantsuse, saksa ja rootsi keeles eestikeelse seletusega. Eesti Keele ja Kirjanduse Instituudi toimetised 18. Stockholm: Eesti Keele ja Kirjanduse Instituut.
- Parmasto, A., Laur, K., Kidron, K. 2005. Kunstiõpetuse mõisted 7.–9. klassile. Eesti-vene-eesti sõnastik. Tartu: Tartu Ülikooli Kirjastus.
- Pedastaar, H.-M., Pedastaar, T. 2005. Geograafia mõisted 7.–9. klassile. Eesti-vene-eesti sõnastik. Tartu: Tartu Ülikooli Kirjastus.
- Peedisson, M., Rihvik, E., Soobik, M. 2005. Käsitöö, kodunduse, töö ja teh-

- noloogia-õpetuse mõisted 7.–9. klassile. Eesti-vene-eesti sõnastik. Tartu: Tartu Ülikooli Kirjastus.
- Piir, I. 2005. Ühiskonnaõpetuse mõisted 7.–9. klassile. Eesti-vene-eesti sõnastik. Tartu: Tartu Ülikooli Kirjastus.
- Piotrowsky, T. 1989. Monolingual and Bilingual Dictionaries. Fundamental Differences. – Learners' Dictionaries. State of the Art. Ed. by M. L. Tickoo. Singapore: SEAMEO RELC.
- Pärtel, E. 2005. Füüsika mõisted 7.–9. klassile. Eesti-vene-eesti sõnastik. Tartu: Tartu Ülikooli Kirjastus.
- Tamm, L., Tamm, T., Tuulmets, A. 2005. Keemia mõisted 7.–9. klassile. Eesti-vene-eesti sõnastik. Tartu: Tartu Ülikooli Kirjastus.
- Toom, M., Teller, M. 2005. Bioloogia mõisted 7.–9. klassile. Eesti-vene-eesti sõnastik. Tartu: Tartu Ülikooli Kirjastus.
- Viks, Ü. 1992. Väike vormisõnastik 1. Sissejuhatus & Grammatika; 2. Sõnastik & Lisad. Tallinn: Eesti Teaduste Akadeemia Keele ja Kirjanduse Instituut.
- Viks, Ü. 2000. Kuidas tekib sõnastikukirjese grammatika. – Keel ja Kirjandus, nr 7, lk 486–495. <http://www.eki.ee/teemad/kirjegeneraator.html>.
- Väike õigekeelsus-sõnaraamat. Koostanud E. Muuk. Tartu: Eesti Kirjanduse Seltsi kirjastus, 1933.

### Subject Dictionaries for Grade Schools

Keywords: subject dictionaries, lexicography, language technology, computer morphology

Experience of composing Estonian-Russian-Estonian dictionaries for schools is described. The aim of the project was to create elucidative dictionaries of terms for 12 subjects in the upper grades of compulsory schools. The selection of terms, composing explanations, and different characteristics of the dictionaries are explained. Computers were used in the work for transforming concept-based manuscripts into term-based ones, for creating inflectional forms of the terms, and for marking terms used in the explanations.

*Heiki-Jaan Kaalap (b. 1962), PhD, University of Tartu, Institute of Estonian and General Linguistics, senior researcher, heiki-jaan.kaalap@ut.ee*

*Jaan Mikk (b. 1939), University of Tartu, professor emeritus, jmikk@hotmail.ee*