

## SÕNARAAMATU KOHTULINGVISTILINE ANALÜÜS: EESTI PRETSEDENT

Margit Langemets, Piret Voll

**Ülevaade.** 2003. aastal viidi Eestis esimest korda läbi kohtulingvistiline leksikograafiaekspertiis, mille käigus hinnati mitme sõnaraamatu sisu, lähtudes kriminaalsüüdistusest intellektuaalse omandi varguses. Artiklis kirjeldatakse Eesti Keele Instituudis läbi viidud leksikograafiaekspertiisi ja selle jaoks välja töötatud keeletehnoloogilist meetodit. Ekspertiisi tulemuste põhjal on välja pakutud piirmäär kakskeelsete sõnaraamatute originaalsuse üle otsustamiseks. Käsitletav juhtum on Eestis ainulaadne, kujutades säärasena kohtulingvistilist pretседenti.

**Võtmesõnad:** kohtulingvistika, leksikograafiaekspertiis, sõnaraamatud, inglise keel, eesti keel

### 1. Sissejuhatus: kohtulingvistikast ja sõnaraamatuplagiaadist

Eestis on kohtulingvistika pea tundmatu ala. Keeleekspertiisi on vajatud enamasti solvangut või laimu käsitlevate kohtuasjade juures. Kohtulingvistika tegelik töömaa on palju avaram, hõlmates nii õiguskeelt kui praktiliselt kõike, mis seondub keele ja kohtusüsteemi ning seadusandlusega (McMenamin 2002: 67). Enamasti uurivad kohtulingvistika eksperdid keelekasutust nii foneetiliselt kui ka stilistiliselt seoses laimujuhtumitega, nii on pea võimatuks osutunud leida leksikograafiapretседenti. Autoriõigus ja plagiaatlus on kohtulingvistikas suhteliselt uued teemad üldse, mitte ainult Eestis, vaid ka mujal. Leksikograafiaalase seadusandluse puudumise üle on kurtanud näiteks Suurbritannia leksikograaf John Williams (1992: 566).

Iseenesest on sõnaraamatute pärast kohtuteed käidud juba 18. sajandil. Londonist on teada juhtum, milles plagiaatluses süüdistati Ann Fisheri sõnaraamatut “An Accurate New Spelling Dictionary and Expositor of the English Language”. Leiti, et teos võib olla kopeeritud John Enticki “The New Spelling Dictionary” (1765) pealt. Võrdlus näitas, et üks-ühele oli üle võetud 43,5%. Kaks aastat kestnud kohtuvaidluste tulemusel lubati sõnaraamat siiski 1773. aastal välja anda (Rodriguez-Alvarez

2006). Veel varasem, mitte küll kohtusse jõudnud näide on pärit 17. sajandist, kui Londonis ilmus Guy Miège'i "The Great French Dictionary. In Two Parts" (1688). Abel Boyer andis 1699. aastal välja "The Royal Dictionary", nimetades hea tava kohaselt ka kasutatud allikaid. Millegipärast "unustas" ta mainida G. Miège'i, kelle sõnaraamatust, nagu hiljem selgus, võeti kas otse või väikeste muudatustega üle umbes 70% märksõnadest koos vastetega (Cormier 2005).

Ka Eestis pole sõnaraamatuplagiaadi kahtlus tundmatu asi. Arnold Kask (1955) on 18. sajandi sõnaraamatutest ülevaadet tehes leidnud, et nii Anton Thor Helle kui ka August Wilhelm Hupel on kasutanud suurel määral Salomo Heinrich Vestringi (umbes 1720-ndad) sõnaraamatu ainekku, sealhulgas A. T. Helle (1732) käsiraamatus avaldatud sõnastik "toetub täiel määral" S. H. Vestringi käsikirjalisele sõnaraamatule (Kask 1955: 190). Käsiraamatu eessõnas on märgitud, et materjal on kogutud "ühe kristliku sõbra" ja läbi töötatud A. T. Helle poolt. "Kristliku sõbra" üle on arutlenud ka Arvo Krikmann (1986) ja Kristiina Ross (EEVA). A. Krikmani arvates kinnitab A. Kase arvamust kõige tugevamini S. H. Vestringi alfabeet, mis on järjestatud väga veidral viisil ning mida sellisena kordab ka A. T. Helle "Anweisung" (Krikmann 1986: 84). K. Ross ei pea samuti A. T. Helle sõnastikku originaalmaterjaliks, sest see "tugineb S. H. Vestringi (märksa mahukamale) käsikirjalisele sõnastikule", ent kuna kohati on seal S. H. Vestringistki rikkalikumat materjali, siis "on oletatud, et Thor Helle kas lisas neile oma kogutud materjali või kasutas lisaks Vestringi käsikirjale veel kellegi materjale." (EEVA)

Järgnevas kirjeldatakse 21. sajandi algul Eestis läbi viidud leksikograafiaekspertiisi käiku. Alustatakse kohtuasjast ja autoriõiguse vastavatest kirjakohtadest, kirjeldatakse ekspertiisi lähteandmeid (sõnaraamatuid), ülesandeid ja Eesti Keele Instituudis välja töötatud meetodit. Seejärel analüüsitakse vaidlusalust materjali ette antud ekspertiisiülesannete kaupa. Lõpuks antakse ülevaade ekspertiisitulemuste rakendamisest Eesti õigussüsteemis, käsitletakse plagiaadi piiritlemise probleemi ja pakutakse välja piirmäär kakskeelsete sõnaraamatute originaalsuse üle otsustamiseks.

## 2. Kohtuasi: sõnaraamatuvarguse süüdistus

2000. aastal kaebas kirjastus TEA kohtusse firma Festart, süüdistades viimast sõnaraamatuvarguses. Juhtumit menetleti kriminaalajana seitse aastat, kuni 2007. aasta kevadeni. Festarti ja selle juhti Sergei Gluhhovit süüdistati selles, et nad aastatel 1999–2004 andsid CD-ROM-il välja inglise-eesti sõnaraamatu versioone, kasutades nende koostamisel TEA sõnaraamatute elektroonilisi põhjasid, omamata selleks kirjastuse luba. Vaidlustatud sõnaraamat pretendeerib originaalteose nimetusele, sest ei pakendil ega lisatekstides leidu ühtki viidet kasutatud allikatele. Muul juhtumil oleks tulnud ära näidata algse teose autori nimi, teose pealkiri (nimetus) ja allikas, kus teos on avaldatud, nagu sätestab autoriõiguse seadus (AÕS § 35 lg 3). Kohtuasja lahendamiseks sai Eesti Keele Instituut 2002. aastal Tallinna Politseiprefektuuri kriminaalosalakonna majanduskuritegude talituselt ülesande läbi viia leksikograafiaekspertiis.<sup>1</sup>

Olgu vahemärkusena öeldud, et kuni 2007. aasta lõpuni töötas Eestis kaks riiklikku ekspertiisiasutust,<sup>2</sup> mis aga kumbki nii spetsiifilist ekspertiisi ei tei-

<sup>1</sup> Ekspertiisi määras Tallinna Politseiprefektuuri kriminaalosalakonna majanduskuritegude talituse uurija, politseivaneminspektor Carmen Kuus. Ekspertiisimäärus on koostatud Tallinnas 11.11.2002. Ekspertiisi teostaja valiti riigihanke korras.

<sup>2</sup> Kaks asutust – Eesti Kohtuarstlik Ekspertiisibüroo ja Kohtuekspertiisi ja Kriminialistika Keskus – liideti alates 1. jaanuarist 2008 üheks asutuseks nimega Eesti Kohtuekspertiisi Instituut (Riigi Teataja 24.08.2007).

nud.<sup>3</sup> Lähimaks, st intellektuaalse sfääriga lõikuvaks võis ehk pidada dokumendi-, infotehnoloogia-, ka kujutise uuringut, millega eksperdid toona lähemalt tutvust ei teinud.

## 2.1. Autoriõigusest

Teoste hulgas, millele tekib autoriõigus, on autoriõiguse seaduses nimetatud väga mitmesuguseid teoseid: kirjalikke (ilukirjandus, publitsistika, erisugused teadustöö tulemused jms) ja suulisi teoseid (kõned, jutlused jms), muusika-, tantsu- ja kujutava kunsti teoseid, arvutiprogramme, standardeid, andmebaase jpm. Autoriõigus tekib ka tuletatud teosele, milleks on “teose tõlge, algse teose kohandus (adaptsioon), töötlus (arranžeerimine) ja teose muu töötlus” (AÕS § 4 lg 22). Põhiline kriteerium on, et teos oleks “autori enda intellektuaalse loomingu tulemus” (AÕS § 4 lg 2).

Ajal, mil arvutiajastu soodustab hõlpsat kopeerimist, diskuteeritakse ühtelugu autori- ja omanikuõiguse kontrolli ja vabakasutuse üle eri elualadel.<sup>4</sup> Seaduses on sätestatud mõned intellektuaalse tegevuse tulemused, millele autoriõigus (seni) ei laiene, nagu rahvaloominguteosed, päevauudised, faktid, andmed, arvutiprogrammi elemendid jms (AÕS § 5). Sõnu *sõnastik* või *sõnaraamat* ei esine autoriõiguse seaduses kordagi.

Sõnaraamatule võiks autoriõigus tekkida kui originaalsele andmebaasile (“süsteemiliselt või meetodiliselt korrastatud iseseisvate teoste, andmete või muu materjali kogu”) või kui tuletatud teosele. Mille alusel aga eristada autoriõigusega kaitstavat (tuletatud) teost plagiaadist – selle kohta puuduvad autoriõiguses kaudsedki juhtnöörid. Kui arvutiprogramm on tervikuna kaitstud kui teos, aga selle elemendid mitte, siis kuidas on lood sõnaraamatuga? Süütuse presumptsiooni põhimõttel näeb seadus ette, et kui asja õigusküsimuses kaheldakse, siis lasub tõendamiskohustus sellel, kes on vaidlustanud teose autoriõigusega kaitstuse (AÕS § 4), siinses kohtuasjas niisiis kirjastusel TEA.

## 2.2. Ekspertiisi lähteandmed ja ülesanded

Ekspertiisiakti<sup>5</sup> sissejuhatavas osas on ekspertiisi lähteandmetena üles loetud kõik asjasse puutuvad sõnaraamatud, mis jagunevad kahte leeri. Ühel pool on vaidlustatud materjal ise ehk ekspertiisi objekt, milleks on kirjastuse Festart üllitatud inglise-eesti sõnaraamat CD-ROM-il ning seda toetav võrdlusmaterjal. Teisel pool on kannatanu võrdlusmaterjal (kakskeelsed inglise-eesti või eesti-inglise sõnaraamatud). Tabelis 1 on kirjas kõik kohtuasjas osalenud sõnaraamatud koos märkega nende staatuse kohta. (Edaspidi on artiklis kasutatud poolele osutavaid lühinimesid, nt FESTART-sõnaraamat, TEA teaduse ja tehnika sõnaraamat.)

<sup>3</sup> 2008. a loodud asutuse põhikirjas (Riigi Teataja 20.12.2007) on üles loetud asutuses tehtavad uuringud, nt alkoholi-, DNA-, dokumendi-, hääle-, infotehnoloogia-, jälje-, kujutise-, käekirja-, narkootilise aine jpm uuringud (kokku 30 liiki), samuti ekspertiisiliigid: 1) kohtuarstlik, 2) kohtubioloogia-, 3) kohtukeemia-, 4) kohtupsühhiaatria- ja 5) kriminalistikaekspertiis.

<sup>4</sup> Vt nt Evi Arujärve (2008) kirjutist “Kultuur – missioon ja äri”.

<sup>5</sup> Ekspertiisiakt kriminaalasjas nr 00231000138 “Festart Dictionary English-Estonian. 70 000 Entries. CD-ROM Version 3.10”. Käsikiri Eesti Keele Instituudis. Leksikograafiaekspertiisi peaekspert: Piret Voll.

**Tabel 1.** Ekspertiisi lähteandmed (kohtuasjasse puutuvad sõnaraamatud)

Nr	Täiskirje ekspertiisiaktis	Lühinimi siinses artiklis	Materjali staatus
1	"FESTART DICTIONARY ENGLISH-ESTONIAN. 70 000 ENTRIES. CD-ROM Version 3.10" ISBN 9985-9063-4-9 <sup>6</sup>	FESTART-sõnaraamat	vaidlustatud materjal
2	English-Latvian Dictionary / Angļu-latviešu vārdnīca. 3rd improved edition. Rīga: Jāņa Sēta Publishers & Printers Ltd, 1997. 1328 lk	FESTART inglise-lāti	võrdlusmaterjal
3	Teaduse ja tehnika seletav sõnaraamat I-II. Tallinn: TEA Kirjastus, 1997. 1667 lk	TEA teadus ja tehnika	võrdlusmaterjal
4	Teaduse ja tehnika seletav sõnaraamat. Eesti-inglise. Lisaköide. Tallinn: TEA Kirjastus, 1998. 544 lk	TEA teadus ja tehnika lisaköide	võrdlusmaterjal
5	PASSWORD. Inglise-eesti seletav sõnaraamat. Tallinn: TEA Kirjastus, 1995. 856 lk	TEA Password	võrdlusmaterjal
6	Eesti-inglise majandussõnastik. 2. trükk. Tallinn: TEA, 1998. 496 lk	TEA eesti-inglise majandus	võrdlusmaterjal
7	Inglise-eesti majandussõnaraamat. Tallinn: Keelefirma TEA, 1995. 496 lk	TEA inglise-eesti majandus	võrdlusmaterjal
8	Inglise-eesti ja eesti-inglise ärisõnastik. Tallinn: Keelefirma TEA, 1993. 560 lk	TEA inglise-eesti äri	võrdlusmaterjal
9	Festart; English-Estonian Dictionary of Economics 2.47 [elektrooniline teavik]	FESTART inglise-eesti majandus	täiendav võrdlusmaterjal
10	Silvet, Johannes 1989. Inglise-eesti sõnaraamat I-II. Kolmas, parandatud ja täiendatud trükk. Tallinn: Valgus	Silvet 1989	täiendav võrdlusmaterjal
11	Inglise-eesti sõnastik. Eesti Keele Instituudi vabavara Internetis <sup>7</sup>	EKI sõnastik	täiendav võrdlusmaterjal
12	Festart. Keyword, English-Estonian Dictionary 2.44	FESTART Keyword	täiendav võrdlusmaterjal

Ekspertiisimääruses oli sõnastatud neli ülesannet (küsimust), millele ekspertiis pidi vastama.

1. Missugused tunnused viitavad võrdlusmaterjalina esitatud sõnastike ja/või nende elektrooniliste põhjade kasutamisele FESTART-sõnaraamatu koostamisel?
2. Kui palju on vaidlustatud materjalil (FESTART-sõnaraamatul) märksõnu või muid tunnuseid, mis kattuvad võrdlusmaterjalina esitatud sõnastike märksõnade või muude tunnustega?
3. Kuidas võis võrdlusmaterjalina esitatud sõnastike ja/või nende elektrooniliste põhjade alusel tehniliselt toimuda FESTART-sõnaraamatu märksõnade, vastete ja indekse moodustamine?

<sup>6</sup> Täpne märksõnade arv on 71 154.

<sup>7</sup> Vt [www.eki.ee/dict/inglise/](http://www.eki.ee/dict/inglise/) (20.11.2007).

4. Kas FESTART-sõnaraamat on originaalteos või on see tuletatud teos, mis ulatuses see teos kattub võrdlusmaterjalina esitatud teostega?

Järgnevas on neid küsimusi ühekaupa analüüsitud. Aga enne seda meetodist.

### 3. Meetod

Nagu juba eespool mainitud, ei olnud leksikograafiaekspertiisi tegemiseks otsest eeskjuju leida. Peale selle tuli Tallinna Politseiprefektuuri kriminaalosakonna ettekirjutusena ekspertiis läbi viia absoluutses ulatuses, st toetumist statistilisele valimile nad ei aktsepteerinud. Seega oli mõistlik, nagu ka nüüdisajale kohane, kasutada keeletehnoloogia abi, seda enam, et kogu materjal – nii vaidlustatud teos kui ka võrdlusmaterjal – loovutati ekspertide käsutusse elektroonilisel kujul.

Ekspertiisianalüüsi peamine meetod oli sõnaraamatute elektrooniliste versioonide läbiv võrdlus arvutiprogrammide abil. Aga küsimus oli: mida, millist üksust mõõta? Kui element üksi on liiga vähe, et sellele toetuda, siis elementide kombinatsioon võiks teataval määral peegeldada autori loomingut. Kakskeelses sõnaraamatus võiks vähimaks oluliste infoüksuste kombinatsiooniks pidada märksõna koos oma vaste(te)ga: ei saa olla kakskeelset sõnaraamatut, kus märksõnade juures pole tõlkevasteid (viiteartiklis asendab vastet viide) – need üksused on obligatoorsed igas tõlkesõnaraamatus. Vasteüksus võib koosneda ühest vastest (ühest sõnast või mitmesõnalisest väljendist, vt näide 1a). Sagedamini on antud mitu vastet, mis omavahel on eraldatud komaga (1b). Ei ole kuigi tavaline, et märksõnal ongi ainult üks vaste, eriti üldkeele sõnaraamatus, sest pole kaht keelt, mis üks-ühele kattuks.<sup>8</sup> Väga tavaline on vastete loendis ka semikoolon, mis harilikult eraldab tähenduselt veidi kaugemaid vasteid (1b, 2).

- (1a) **post-paid** .. saatekuludeta [FESTART-sõnaraamat]  
w pikkusühiku koormuse sümbol [FESTART-sõnaraamat]
- (1b) **voidable** .. annulleeritav, tühistatav; vaidlustatav [FESTART-sõnaraamat]
- (2) **micro-** väike, miljondik-; pisi-; miljondik, 0,0000001 m [FESTART-sõnaraamat]

Ekspertidid määratlesid analüüsi põhiüksuseks märksõna koos oma vaste(te)ga, lähtudes vasteloendis esinevast semikoolonist. Koma eraldab lähemaid vasteid ühekaupa ega sobi seega nii hästi rühma eraldajaks (märksõna-vaste-grupis oleks alati ainult üks vaste). Peale selle võib koma esineda vaste osana pikemates väljendites, murdarvudes jm (vt näide 2). Seega eraldab võrreldava märksõna-vaste-grupi (edaspidi MV-grupp) alati semikoolon, kusjuures märksõna korratakse iga grupi juures (vt tabel 2).

FESTART-sõnaraamatus, mis sisaldab 71 154 märksõna, on kokku 122 915 niisugust MV-gruppi. Ühe või teise TEA sõnastiku osakaalu FESTART-sõnaraamatus on iseloomustatud protsentides MV-gruppide kaupa. Välja toodud erinevuste või sarnasuste rõhutamiseks on (ekspertiisis esitatud) näidetes vastavad kohad alla joonitud, tabelitesse on lisatud selgituste tulp.

<sup>8</sup> Spetsiaalses, nn kakskeelestatud sõnaraamatus võib olla põhimõtte esitada ainult üks vaste (seda tüüpi sõnaraamat on näiteks TEA Password), aga sealgi "segab mängu" leksikaalne mitmetähenduslikkus: üks vaste vahendab üht tähendust.

**Tabel 2.** Märksõna-vaste-gruppide (MV-grupid) näiteid

Sõnaartikkel	MV-grupid	Selgitus
<b>micro-</b> väike, miljondik; pisi-; miljondik, 0,0000001 m	kokku 3 MV-gruppi: 1) <b>micro-</b> väike, miljondik-; 2) <b>micro-</b> pisi-; 3) <b>micro-</b> miljondik, 0,0000001 m	1. ja 3. MV-grupis on kaks vastet (eraldatud komaga), 2. MV-grupis on üks vaste
<b>realise</b> [verb] aru saama; täide viima; realiseerima	kokku 3 MV-gruppi: 1) <b>realise</b> aru saama; 2) <b>realise</b> täide viima; 3) <b>realise</b> realiseerima	kõik grupid koosnevad ühest (mitmesõnalisest) vastest

## 4. Esimene ülesanne: reetlikud tunnused

Tunnusena, mis võiks viidata sõnaraamatu ja/või nende elektrooniliste põhjade kasutamisele teise sõnaraamatu koostamisel, on ekspertiisis välja toodud järgmised elemendid ja/või seosed: a) korduvad märksõnad, b) märksõnavaliku kattuvus, c) MV-gruppide kattuvus, d) trükivead (õigekirjavead, küljenduselementide jäänu-sed), e) haruldased vasted. Muud tehnilised detailid (nt muutunud tähestikjärjestus, lihtsustatud diakriitikutega tähed, eesti-inglise pöördindeks) on sõnastikuprog-rammi eripärad ja siduvat informatsiooni ei kannu.

Märksõnavaliku või muude tunnuste kattuvust (seosed b ja c) on kirjeldatud järgmises, teist ülesannet käsitlevas 6. osas. Siin on lähemalt vaadeldud ülejäänuid reetlikke tunnuseid.

### 4.1. Korduvad märksõnad

FESTART-sõnaraamatu 71 154 märksõnast<sup>9</sup> on korduvaid märksõnu 4585 (6,4%), st ligi 5000 juhtumil väljastatakse päringu tegemisel vastuseks mitu sama märk-sõnaga artiklit. See tõsiasi ei külvaks kahtlust, kui tegemist oleks seda märksõna *sisaldavate* väljenditega, mitte mitme identse märksõnaga. Näiteks “Eesti keele seletussõnaraamatus” tehtud märksõnapäring *must* toob vastuseks nii lihtsõnast märksõna enda (*must*) kui ka seda osist sisaldavad liitsõnad (*kärbsemust*, *mustjuur*, *mustsada*, *trükimust* jt) ja väljendid (*must lammas*, *must valgel*, *ei musta ega valget* jt). Sõnavarabaaside üks põhinõudeid on märgistada kirjeldusobjekt unikaalseks. Seega kui märksõna on kujult sama (nt homonüüm), on tal ikkagi üldjuhul mingi eristav tähis juures.

FESTART-sõnaraamatus on asi teisiti: lõviosa (u 90%) mainitud 4585-st artiklist on sellised, kus märksõna kordub muutumatul kujul. Väljastatud n-õ teise artikli vastena esineb siis enamasti majandustermin (vt tabel 3). See viib vägisi mõtted elektroonilisele kopeerimisele (majandussõnastikust?), kuivõrd tavalisel moel koostatud sõnaraamatus on sellised juhtumid enamasti ühe märksõna all, kas siis eri tähendusena (eraldatud tähendusnumbriga) või tähendusrühmana (semikooloniga).

<sup>9</sup> Edaspidi on jälgimise hõlbustamiseks kasutatud ümardatud andmeid (70 000 märksõna), mis on kirjas ka toote ümbrisel.

**Tabel 3.** Korduvate märksõnadega artikleid FESTART-sõnaraamatus

Otsisõna	Väljastatav vastus	Selgitus
<b>post-paid</b>	<b>post-paid</b> [adjective] saatekuludeta <b>post-paid</b> [economics] postikulud tasutud, saatekuludeta, franko	märksõna kordub muutumatul kujul, kusjuures teise märksõna järel esineb majandustermini märgend <b>economics</b>
<b>yuppie</b>	<b>yuppie</b> .. [noun] jappi <b>yuppie, yuppy</b> [economics] juppi, jappi	väljastatakse kaks väga sarnast (inimese poolt toimetamata?) sõnaartiklit; märksõna kordub muutumatul kujul, kusjuures teise märksõna järel esineb majandustermini märgend <b>economics</b>

Ülejäänud juhtudel (10%) esineb korduv osis fraasi osana, mis on ka pabersõnastikes tavaline, kuigi seal on sellised juhtumid enamasti kasutusnäite või allmärksõna positsioonis (3), samas kui FESTART-sõnaraamatus on need tõstetud märksõna staatusse (4). Paljudes pabersõnastikes kuulub sõnaartiklisse märksõna tuletisi, mitmesõnalisi püsiühendeid ja väljendeid, samas kui arvutisõnastikes paigutatakse need kõik tavaliselt eraldi (st märksõna positsiooni). Kõrvalekaldena on siin paslik juhtida tähelepanu märksõnade arvule: sel moel eraldi esitatud märksõnad paisutavad sõnaraamatu mahtu, mida on muidugi soodus kasutada (arvuti)toodet reklaamides. Tegelikult võiks FESTART-sõnaraamatu “suurt” mahtu (70 000 märksõna) pidada võrdväärseks umbes keskmise suurusega (st 40–50 000 märksõnaga) pabersõnaraamatuga.

- (3) **voidable** [economics] annulleeritav, tühistatav; vaidlustatav • **voidable act** vaieldav toiming; **voidable title** vaieldav omandiõigus [võimalik artikkel pabersõnaraamatus]
- (4a) **voidable** [economics] annulleeritav, tühistatav; vaidlustatav [FESTART-sõnaraamat]
- (4b) **voidable act** [economics] vaieldav toiming [FESTART-sõnaraamat]
- (4c) **voidable title** [economics] vaieldav omandiõigus [FESTART-sõnaraamat]

#### 4.2. Trükivead: õigekirja- ja poolitusvead, küljenduselementide jäänused

Leksikograafias (nagu muudiski kirjalikes tekstides) on samade trükivigade esinemine üks kindlmaid tõendeid selle kohta, et üks teos kopeerib teist: ilmselget viga ei ole lihtsalt võimalik teistmoodi paljundada, kui maha kirjutades: kas siis täiesti asjatundmatu inimesena või, mis veelgi hõlpsam, elektrooniliselt, kõike mehaaniliselt üle võttes. Ekspertiisis on välja toodud nii korduvaid õigekirja- ja poolitusvigu (vt tabel 4) kui ka küljenduselementide jäänuseid (vt näide 5). Näiteks 69-st ekspertiisi käigus leitud TEA Passwordi poolitusveast kordub 68 (98,5%) FESTART-sõnaraamatus.

**Tabel 4.** Kattuvad trükivead: õigekirja- ja poolitusvead

TEA inglise-eesti majandus	FESTART-sõnaraamat	Selgitus
BOOM buum, kõrgkonjunktuur; ..	<b>boom</b> [economics] buum, kõrgkonjunktuur; ..	õigekirjaviga: p.o <i>kõrgkonjunktuur</i>
quantitative CHANGE	<b>quantitative change</b> [economics] ..	õigekirjaviga: p.o <i>quantitative change</i>
SECOND bill of (exchange) sekundaveksel, ..	<b>second bill of</b> [economics] sekundaveksel	viga märksõnas: programmiga on välja jäetud kõik sulgudes osad
<b>courtly</b> [adjective] .. • peene-kombeline	<b>courtly</b> [adjective] õukondlik; peene-kombeline	poolitusviga

Küljendusvead on reeglina tekkinud uue materjali lisamisel ja/või millegi ümber-  
tõstmisel – see seletab ülearuseid tühikuid (5a), topelt-kirjavahemärke (5b), rea  
lõppu “unustatud” kirjavahemärke (5c) ja muid kirjavahemärkidega seonduvaid  
vigu (5d) (kõik näited FESTART-sõnaraamatust).

- (5a) **carriage free** [economics] franko, tasuta
- (5b) **letter of credit** [economics] akreditiiv, krediitkiri, rambuuri, maksekä-  
sund
- (5c) **compensate** [economics] korvama, hüvitama, tasuma; tasakaalustama,  
balansseerima,
- (5d) **appreciate** [economics] hindama; arvesse võtma; hinnas tõusma, hinda  
tõstma, väärtuses kasvama

Samalaadsele tehnilisele argumentile – veidrale tähestikjärjestusele – on osutatud  
juba 18. sajandil Helle-Vestringi sõnaraamatukriitikas, millest oli juttu siinse artikli  
sissejuhatuses.

### 4.3. Haruldased vasted

Kui suure kasutussagedusega sõnade vastete kattumine on mõnevõrra ootuspärane,<sup>10</sup>  
siis haruldaste märksõnade ja kummaliste vastete kordumine mitmes sõnaraamatus  
seda kindlasti ei ole. Probleem on muidugi, kuidas kõik need haruldased vasted üles  
leida. Kuna see toona ei olnud keeletehnoloogiliselt lahendatav (põhimõtteliselt  
saaks mõõta iga eesti sõna sagedust tekstikorpuses – aga põhjendatud tulemuse  
saamiseks läheks vaja vähemalt 100 mln sõne suurust üldkeele korpust, mida Eestis  
pole veel praeguse ni olemas), siis töid eksperdid n-ö käsitsi välja TEA sõnaraama-  
tuses silma hakanud mitmes mõttes erilisi vasteid.

Leksikograafide arusaamatul põhjusel korduvad need FESTART-sõnaraamatus,  
kuna tavaliselt selliseid asju uustrüki jaoks täiendatakse või parandatakse. Kui  
TEA teaduse ja tehnika sõnaraamatus, mille inglise aluspõhja (Chambers Ltd) on  
viimati uuendatud 1988. a, on esitatud (vananenud) märksõna ja vaste (6a), siis

<sup>10</sup> Võib täiesti mõista Interneti-suhtlejate nõrdimust 2007. a kevadel (kui kohtuprotsess lõppes): “.. tegin ka eks-  
pertiisi: TEA sõnastik: *ajaleht* = *newspaper*; FESTART: *ajaleht* = *newspaper*. Oskad sa ette kujutada, ongi täpselt  
sama sõna pandud Festardil, järelikult maha tehtud! Piraat mis piraat! Kole ju ..” (Eesti Päevaleht Online, aprill  
2007).

sama kordub FESTART-sõnaraamatus (6b), kus toimetajad ei ole kiiresti vananevaid infoteaduse jm termineid uuemate vastu välja vahetanud. Kui TEA teaduse ja tehnika sõnaraamatus on ära toodud spetsiifilisi termineid, mis üldkeele või väiksemates erialasõnastikes puuduvad (7a, 8a), siis on need olemas ka FESTART-sõnaraamatus (7b, 8b). Näidete 7 ja 8 puhul võib tähelepanu juhtida ka identses sõnastuses vastetele. Ebaharilik on ka allpool toodud vaste *tähtpäevuma* (10b). Võrdluseks: ekspertide kogemus sõnaraamatute koostamisel kinnitab, et isegi juhul, kui märksõna tähendusele on lubatud panna ainult üks vaste, esitavad eri tegijad selle vaste pisut isemoodi.

- (6a) **XNS** [Comp] Abbreviation for *Xerox network system*. .. • XNS [TEA teaduse ja tehnika]
- (6b) **XNS** [computers] XNS [FESTART-sõnaraamat]
- (7a) **VOS** [Image Tech] Abbreviation for *video on sound*. See **background video**. • helipealne video, taustvideo [TEA teaduse ja tehnika]
- (7b) **VOS** [photography] helipealne video, taustvideo [FESTART-sõnaraamat]
- (8a) **taungya** [Ecol] .. • agrometsanduse vorm troopilistel aladel [TEA teaduse ja tehnika]
- (8b) **taungya** agrometsanduse vorm troopilistel aladel [FESTART-sõnaraamat]

## 5. Teine ülesanne: kattuvate märksõnade või muude tunnuste määr

FESTART-sõnaraamatu materjali võrreldi ekspertiisis sisuliste alarühmade, valdkondade ja vastavate sõnaraamatute kaupa: 1) majandusterminid (võrdlusaluseks TEA inglise-estl majandus), 2) tehnikaterminid (TEA teadus ja tehnika) ja 3) üldsõnavara (TEA Password ja Silvet 1989). Pärast iga alarühma analüüsimist eemaldati vastav osa (MV-grupid) algsest FESTART-sõnaraamatust. Iga järgnevat alarühma analüüsi eelmise osa võrra vähendatud algmaterjalis.

### 5.1. Majandusterminid

Majandussõnavara osatähtsus on FESTART-sõnaraamatus (70 000 märksõna, 122 915 MV-gruppi) märkimisväärselt suurem kui muudes kakskeelsetes üldkeele sõnastikes: majandussõnad (mürgendiga [*economics*] 'majandus') moodustavad FESTART-sõnaraamatu kogumahust ligi 40% (kokku 27 387 märksõna).

Majandusterminite osa eristub muudest sõnavarakihtidest mitme iseärasuse poolest: märksõnadel puudub häälendus (9-10); kogu FESTART-sõnaraamatus esinevad ainult selles rühmas mitmest sõnast koosnevad märksõnad (10); esineb trükivigu: veatuübid nagu "tühik koma ees" (9a), "koma järel puudub tühik" (9c), "kirjavahemärk rea lõpus" (9b) või "kaks kirjavahemärki üksteise järel" (10a) on selles rühmas ülekaalukas enamuses.

- (9a) **billion** [economics] biljon , miljard
- (9b) **bunco** , **bunko** [economics] pettus,
- (9c) **discuss** [economics] arvamusi vahetama, diskuteerima, polemiseerima, väitlema, vaidlema, arutama, arutlema
- (10a) **letter of credit** [economics] akreditiiv, krediitkiri, rambuuri, maksekä-sund
- (10b) **become due** [economics] saabuma, tähtpäevuma, aeguma; maksmisele kuuluma

Majandustermineid analüüsiti kahes etapis: esmalt võrreldi FESTART-sõnaraamatut teise sama kirjastuse sõnaraamatuga (FESTART inglise-eesti majandus, 27 500 märksõna), seejärel kontrolliti viimase võimalikku seotust TEA inglise-eesti majandussõnaraamatuga.

Kahe Festardi sõnaraamatu vigaste kohtade võrdlemisel ilmnes mitu asja. Esiteks, MV-gruppide kokkulangevus on sama hea kui täielik: võib väita, et FESTART-sõnaraamatu 27 891 MV-gruppi (23%) pärinevad sama kirjastuse majandussõnaraamatust (mis samuti pole originaalne, aga sellest allpool). Ülevõtmisel on automaatselt välja jäetud sulgudes olnud kommentaarid, täpsustused, keeleviited jms, seejärel tehtud korrektuur ja lisatud ka mõningaid uusi märksõnu. Tõenäoliselt on märgend [*economics*] lisatud lõpuks programmi abil ja ilma järgneva korrektuuri või toimetamiseta. Seda illustreerib märksõna *yeepy*, kus märgendi õige koht peaks olema eesti vaste *aktiivsed vanurid* ees, mitte ingliskeelse tähenduse seletuse keskel (11).

(11) **yeepy**, youthful [economics] energetic elderly people involved in everything aktiivsed vanurid [FESTART]

Teiseks tulevad esile tüüpiliste vigade tekkekohad, st kohad, kus midagi on muudetud, kas uue materjali elektroonilisel lisamisel, väljajätu kohal, sulgudes esitatud materjali ärajätmisel, küljenduse poolitusvihjete puhul jne (vt tabel 5).

**Tabel 5.** Festardi sõnaraamatute võrdlus: tüüpilisi veakohti

FESTART inglise-eesti majandus	FESTART-sõnaraamat	Vea tekke selgitus
APPRECIATE hindama; tänulik olema; arvesse võtma; hinnas tõusma, hinda tõstma	<b>appreciate</b> [economics] hindama; arvesse võtma; hinnas tõusma, hinda tõstma, väärtuses kasvama	vahela lisatud vaste ees
IRRETRIEVABLE korvamat; lootusetu; tagasisaadamatu; taastamatu; heastamatu	<b>irretrievable</b> [economics] korvamat, heastamatu; lootusetu; tagasisaadamatu; taastamatu	vahela lisatud vaste ees
REALITY reaalsus, tegelikkus	<b>reality</b> [economics] tegelikkus, reaalsus	kohavahetuse kohal
BILLION biljon (UK, Saksamaa), miljard (USA, Prantsusmaa)	<b>billion</b> [economics] biljon , miljard	sulgudes teksti väljajätu kohal
GROSS .. sum-maarne	<b>gross</b> [economics] .. sum-maarne ..	poolitusvihje kohal

Kolmandaks, kõigile FESTART inglise-eesti majanduse sõnaraamatust pärit märksõnadele on lisatud FESTART-sõnaraamatus erialamärgend [*economics*] (vt tabel 5, 2. veerg). Nii on tekkinud FESTART-sõnaraamatusse hulk “imelikke” artikleid, kus märksõnad korduvad (sellest oli ka eespool juttu), kusjuures ka erialamärgend on esitatud mõlemas artiklis (12a, 12b): sest kuskilt võetuna on materjal is juba olemas hääldusega ja märgendiga varustatud artikkel. Leidub artikleid, kus majanduse osa (nagu ka teised erialaterminid) on sattunud suisa üldsõnavaraosa keskele (13).

- (12a) **account** [õkaunt] [verb] lugema, arvama; punkte saavutama; tapma; [noun] arvestus; [*economics*] arve; seletus; jutustus [FESTART-sõnaraamat]
- (12b) **account** [*economics*] raamatupidamiskonto; konteerima, kirjendama; finantsaruanne; register; alaline klient; riigieelarve täitmise aruanne; börsitehingu arvutus; tasumata võlakohustus; aru/selgitusi andma; aruandlus, äri~~ra~~amatud; vastutama, vastutust kandma [FESTART-sõnaraamat]
- (13) **principal** [ˈprinsipl] [adjective] peamine; [noun] printsipaal; tööandja; direktor; [jurisprudence] peasiüdlane; [building] peatala; [*economics*] kapital; esinäitleja, esilaulja [FESTART-sõnaraamat]

Ekspertiisi käigus selgus, et FESTART inglise-eesti majandussõnaraamatus on võrdluses TEA inglise-eesti majandussõnaraamatuga omakorda täpseid MV-gruppide vastavusi 26 698 (97%). Elektroonilise küljenduspõhja kasutamisest annavad taas tunnistust poolitusvihjete täpne kordamine, kokkulangevad õigekirjavead jm. Näiteks leidub äärmiselt reetlik kirjaviga vastes äri~~ra~~amatud, mis pärineb juba TEA inglise-eesti majanduse pabersõnastikust, esinedes nõnda igas kolmes allikas: esimesena TEA-s, seejärel kahes Festardi sõnaraamatus.

## 5.2. Tehnikaterminid

Pärast majandussõnavara (27 387 märksõna koos oma MV-gruppidega) eraldamist on vaadeldud allesjäänud osa (42 613 märksõna), milles omakorda eristub suur rühm – märksõnad ilma häälduseta, erialamärgenditega [*zoology*], [*chemistry*], [*physics*], [*building*], [*geology*] jne. Selliste erialamärgenditega sõnu oletasid eksperdid pärinevat TEA teaduse ja tehnika sõnaraamatust. Ilma häälduseta märksõnu, mis pole majandusterminid, on kogu FESTART-sõnaraamatus 25% (kokku 18 024 märksõna). Ühelt poolt kuuluvad siia hulka märksõnad, mille puhul sõnastikes hääldust sageli ei esitatagi (akronüümid, keemilised elemendid jms) ja märksõnad ilma igasuguse erialamärgendita. Teiselt poolt kattub osa erialatermineist tavakasutuses olevate sõnadega – siis on üldjuhul ka hääldus olemas (tõenäoliselt üle võetud TEA Passwordist, sellest on juttu allpool). TEA teaduse ja tehnika sõnaraamatuga toimiti tõenäoliselt järgneval viisil.

TEA sõnaraamatu elektrooniline versioon koosnes küljendusprogrammi Ventura failidest, millest mittevajaliku info ja kujunduselementide kustutamine võtab kogenud programmeerijal aega vaid mõne tunni (ekspertidel oli kasutada oma programmeerija, kes selle läbi proovis). Mittevajaliku info hulka kuulusid näiteks

inglisekeelsed seletused ja tähendusnumbrid. Sel viisil “puhastatud” sõnastikku jäid vaid inglise märksõnad, erialamärgendid ja eesti vasted, kokku 51 692 märksõna. Materjali võrdlus näitab, et sellest on otse üle võetud ligi 40% seal leiduvatest märksõnadest (kokku üle 20 000). Kogu FESTART-sõnaraamatust moodustab TEA teaduse ja tehnika sõnaraamatust pärit osa 16%. Ülevõtmine on võinud toimuda järgmiste üldreeglite alusel:

- 1) välja on jäetud mitmesõnalised inglise märksõnad (st märksõnad, mis sisaldavad tühikut);
- 2) kustutatud on kõik, mis oli sulgudes (meetod on tuttav juba varasemast, nii tekib näiteks vastete kordumine);
- 3) teisenduse käigus on kaotsi läinud kursiiv, fondivahetus jm, mis põhjustab sisulisi vigu: väärraid märksõnu ja seletusi; palju jälgi on endast maha jätnud numbrite töötl(emat)us;
- 4) kaotatud on tähendusliigendus: kustutatud on tähendusnumbrid, semi-koolon on asendatud komaga;
- 5) mitmest erialamärgendist on alles jäetud ainult esimene;
- 6) erialamärgendid on asendatud (osalt välja kirjutatud, osalt muudetud) (vt tabel 6).

Kõik loetletud asjaolud kinnitavad arvutiprogrammi, mitte inimesest koostaja kasutamist.

**Tabel 6.** Võrdlus TEA teaduse ja tehnika sõnaraamatuga: tüüpilisi veakohti

TEA teadus ja tehnika	FESTART-sõnaraamat	Selgitus
<b>ageing</b> [Eng] .. • .. <b>2.</b> vananemine (plastide korral); <b>3.</b> vananemine (ferromagneetikute korral)	<b>ageing</b> [physics] vananemine, vanandamine .. [physics] vananemine; vanandamine, vanandus	kustutatud on sulgudes osa: tulemuseks korduv erialamärgend, korduvad vasted
<b>α-particle</b>	<b>a-particle</b>	teisenduse käigus on kaotsi läinud erisümbolid: väär märksõna
<b>sthene</b> jõu ühik meeter-tonn-sekund-süsteemis, ekvivalentne $10^3$ N-ga	<b>sthene</b> jõu ühik meeter-tonn-sekund-süsteemis, ekvivalentne $10^3$ N-ga	teisenduse käigus on kaotsi läinud ülaindeksid: väär seletus
<b>albedo</b> <b>1.</b> albedo, peegeldusvõime; <b>2.</b> neutronite peegeldumise võime pinnalt	<b>albedo</b> [physics] albedo, peegeldusvõime, neutronite peegeldumise võime pinnalt	kaotatud on tähendusliigendus, semi-koolon on asendatud komaga
<b>mucilaginous</b> [Bot, Zool] lima-	<b>mucilaginous</b> [botany] lima-	mitmest erialamärgendist on alles jäetud ainult esimene
[Aero] [Bot] [Behav] [ImageTech] [Space]	[aviation] [botany] (puudub) [photography] [space]	erialamärgendid on asendatud

<b>folly</b> .. • dekoratiivne pargirajatis <u>18. s.</u>	<b>folly</b> .. [architecture] dekoratiivne pargirajatis <u>1s.</u>	programm on (punktiga) numbrit pidanud tähendusnumbriks (8.) ja eemaldanud
<b>grit</b> .. • pulber kõvadest osakestest suurusega üle <u>76<math>\mu</math>m</u>	<b>grit</b> .. [technics] pulber kõvadest osakestest suurusega üle <u>76m</u>	esimene sümbol (kreeka täht) on numbril järel kustutatud
<b>singles</b> .. • kivimitükk: katusekilt 12x8 tolli, süsi 1- <u>0,5 tolli</u> <b>X-band</b> .. • X-laineala: <b>1.</b> sagedusriba 8...12 GHz; <b>2.</b> radarite tööriba 8,8...10,68 MHz	<b>singles</b> .. [mineralogy] süsi 1- <u>0</u> <b>X-band</b> [technics] radarite tööriba <u>8</u>	programm on pidanud numbris esinevat koma vastete eraldajaks, mida nimetoimetaja kunagi ei teeks

Sellele, et FESTART-sõnaraamatu tegemisel võib olla kasutatud TEA teaduse ja tehnika sõnaraamatu elektroonilist põhja, osutavad taas TEA teksti vigased kohad (nt *madalatemperatuurne*, *pinuase nõlv* jt), mis parandati enne TEA teaduse ja tehnika lisaköite ilmumist (*madalatemperatuuriline*, *pinnasenõlv* jt). Samad vead on aga FESTART-sõnaraamatus täiesti olemas. Ekspertiisi käigus leiti ka selliseid kattuvaid kirjavigu, mida lisaköite tegemisel veel polnud märgatud (*harkelundieemaldus pro harkelundieemaldus* jt).

### 5.3. Üldsõnavara

Pärast kahest eelnevast allikast (TEA inglise-eesti majandus, TEA teadus ja tehnika pärinevate MV-gruppide eemaldamist jäi alles umbes pool esialgsest mahust (35 298 märksõna, 48 000 MV-gruppi). Järgnevalt võrreldi seda materjali üldkeelt esindava TEA Passwordiga (23 629 märksõna, sh allmärksõnad – tuletised ja väljendid), kus iga tähenduse puhul on vasteid esitatud harilikult üks, harvemini kaks-kolm. Kuna eksperdid kahtlustasid TEA Passwordi ülevõtmist juba varem ilmunud kirjastuse Festart elektroonilisse sõnastikku (FESTART Keyword), siis lubati viimast kasutada täiendava võrdlusmaterjalina.

FESTART Keyword koosneb Festardi sõnastikele iseloomulikult hästi eristuvast majanduse- ja üldsõnavara osast. Lihtne analüüs annab alust oletada, et vähemalt kolmveerand FESTART Keywordi mittemajandussõnadest on üle võetud TEA Passwordi elektroonilisest põhjast. Kopeeritud on tervikartiklid, st peale märksõnade ja vastete ka hääldused, seletused ja näitelauseid. Eespool juba mainitud 69 TEA Passwordi poolitusveast 68 “esindatus” FESTART-sõnaraamatus kordub ka siin, ajaliselt varasemas Keywordis. Võrdluse käigus erinevaks tunnistatud artiklites on erinevused põhjustatud hääldusmärkide erinevast kodeeringust jt vormistuslikest erinevustest (näiteks punkt iga artikli ja iga tähenduse lõpus, vt tabel 7), mida kuigi tõsiselt ei saa võtta. Ekspertid ei pidanud täpsemat kvantitatiivset analüüsi tarvilikuks, vaid pidasid tõendusmaterjaliks sisulisi (kvalitatiivseid) andmeid. Kuna TEA Passwordi korrektuur ja küljendus on olnud korrektne, siis leidub ka väga vähe trükivigu, millele toetuda.

**Tabel 7.** Festart Keywordi võrdlus TEA Passwordiga

TEA Password	FESTART Keyword	Selgitus
<b>keep away</b> to (cause to) remain at a distance: <i>Keep away -- it's dangerous!</i> • eemale hoidma	<b>keep away</b> to (cause to) remain at a distance: <i>Keep away - it's dangerous!</i> • eemale hoidma.	lisandus: punkt artikli lõpus
<b>keep back 1</b> not to (allow to) move forward: <i>She kept the child back on the edge of the crowd;</i> <i>Every body keep back from the door!</i> • eemal hoidma <b>2</b> not to tell or make known: <i>I feel he's keeping the real story back for some reason</i> • enda teada hoidma	<b>keep back 1</b> not to (allow to) move forward: <i>She kept the child back on the edge of the crowd;</i> <i>Every body keep back from the door!</i> • eemal hoidma. <b>2</b> not to tell or make known: <i>I feel he's keeping the real story back for some reason.</i> • enda teada hoidma.	lisandus: punkt iga tähenduse ja artikli lõpus; viga: p.o everybody (viga pärineb TEA Passwordist, kus rea lõpust puudub poolitusmärk, mida asendab tühik)
<b>aggressively</b> <b>appreciatively</b> <b>comprehensively</b>	<b>aggressively</b> <b>appreciatively</b> <b>comprehensively</b>	ilma häälduseta
<b>mechanic</b> .. • mehhaanik <b>mechanical</b> .. • mehhaaniline <b>mechanically</b> .. • mehhaaniliselt	<b>mechanic</b> mehaanik <b>mechanical</b> mehaaniline <b>mechanically</b> mehhaaniliselt	ebajärjekindlus h/hh tarvituses

Sellegipoolest on võimalik analüüsida (ja hinnata) kõigile sõnastikele omaseid parameetreid nagu sõnavalik, hääldused, vastete kokkulangevused, lisamärgendite kasutamine, artikli struktuur jm.

TEA Passwordis on lihtmärksõnu (vastandina mitmeosalistele) 18 647, neist 17 912 (96%) leidub FESTART-sõnaraamatus. Sõnastike sarnast tüüpi ja võrreldavat mahtu arvestades on see ootuspärane tulemus. MV-gruppe on TEA Passwordis kokku 31 391, millest 21 632-le (69%) leidub täpne paariline FESTART-sõnaraamatus. Samas puudub FESTART-sõnaraamatust umbes 8000 TEA Passwordi MV-gruppi (25% lihtmärksõnadest, nt *keenness* 'ind', *kidnap* 'inimest röövima' jt). Ebatäielikule kattumusele vaatamata võib FESTART-sõnaraamatust leida kokkulangevusi, mis endiselt seovad teda TEA Passwordi sisuga (haruldased vasted), nagu ka elektroonilise põhjaga (küljendusest pärit poolituskriipsud).

Üht osa TEA Passwordi märksõnadest – mis a) on ilma häälduse ja sõnaliigita, ja b) pole pärit TEA teaduse ja tehnika sõnaraamatust (kokku 1141) – võrreldi täiendavalt EKI sõnastikuga. Kui MV-gruppide kokkulangevus TEA ja FESTARDI vahel oli 97%, siis TEA ja EKI vahel ainult 53% (!). Kui TEA sõnaliigimärgendit sisaldavad MV-grupid kattusid FESTART-sõnaraamatuga 92% ulatuses, siis EKI-ga üksnes 24%. Need võrdlevad andmed võiksid olla kõnekas argument sõnaraamatu originaalsuse tõestamisel.

Kuna ekspertidele torkas pidevalt silma FESTART-sõnaraamatu heterogeensus, mis andis alust arvata, et sellegi allesjäänud materjali saamiseks on mehhaaniliselt ühendatud mitu allikat, siis võrreldi materjali ka Silvet 1989 sõnaraamatuga. Kahtlus, et ülejäänud materjal FESTART-sõnaraamatus pärineb Silvet 1989 sõnaraamatust, leidis kinnitust. Sellele viitab mitu asjaolu. Esiteks, peaaegu kõik eesti vasted järelejäänud FESTART-sõnaraamatu materjalis esinevad täpselt samas järjestuses Silvet 1989 vastete hulgas. Teiseks, leidub vasteid, mille tarvituse on ebaharilik, nüüdiskeeles vananenud ja mis ei esine ei tavakeeles ega üheski muus

sõnaraamatus, nt *adobe* – plonnohoone, *brief* – kohtule esitatav kaitse-kokkuvõte, *tight* – pinguline, *bridle* – pahandudes pead üles ning lõuga tagasi tõmbama jpt. Kolmandaks, sellise silmale märkamatu elemendi nagu (elektroonilises versioonis) tühik põhimärksõnaartikli lõpus leiame täpselt nendest kohtadest, kus Silvet 1989 on esitanud sulgudes kommentaare. Tühik – mis elektroonilisel kujul muutub “nähtavaks” – jäi artikli lõppu tõenäoliselt pärast seda, kui sulgudes olnud tekst kustutati (17). Sisuline viga tekkis samal põhjusel (sulgudes osad jäeti mehaaniliselt kõrvale): nõnda sügenes artiklisse kaks identset vastet (18b).

(17a) **adulterer** abielurikkuja (mees) [Silvet 1989]

(17b) **adulterer** abielurikkuja\_ [FESTART-sõnaraamat]

(18a) **current** vool, voolus, (mere)hoovus, (elektri)vool [Silvet 1989]

(18b) **current** vool, voolus, hoovus, vool [FESTART-sõnaraamat]

Ekspertide hinnangul on Silvet 1989 vastete suurest hulgast (umbes 150 000 MV-paari) üle võetud umbes kolmandik. Vastete ülevõtmise kopeerimise teel välistab asjaolu, et 1999. aastal polnud sõnastiku elektroonilist varianti veel olemas: nii on osa kirjavigu tekkinud teksti sisestamisel (nt *silmamoondusrikke* pro *silmamoondusrikke*). Et FESTART on ka omalt poolt vasteid lisanud, sellele viitavad leitud vigased märksõnad või vasted, mis sellisena mujal ei esine (nt märksõna *thunderstruck* pro *thunderstruck*, vaste *äärmuste abinõude* pro *äärmuslike abinõude*).

#### 5.4. Teise ülesande vastus

FESTART-sõnaraamatu majandussõnavara pärineb tõenäoliselt täielikult TEA inglise-eesti majandussõnaraamatust, olles peitva vaheastmena üle võetud varasemast FESTART inglise-eesti majandussõnaraamatu arvutiversioonist. Majandussõnavara osakaal kogu sõnaraamatus on 23% (vt tabel 9).

FESTART-sõnaraamatu tehnikasõnavara pärineb tõenäoliselt TEA teaduse ja tehnika sõnaraamatus, kust on üle võetud ligi 20 000 MV-gruppi. Tehnikasõnavara osakaal kogu sõnaraamatus on 16% (vt tabel 9).

FESTART-sõnaraamatu üldsõnavara pärineb kahest allikast: TEA Passwordist (4000 MV-gruppi) ja Silvet 1989-st (hinnanguliselt 69 000 MV-gruppi). Üldsõnavara osakaal kogu sõnaraamatus on 59% (3% + 56%) (vt tabel 9).

FESTART-sõnaraamatu originaalloomingu suuruseks on hinnanguliselt 2% (vt tabel 9).

**Tabel 9.** Eri sõnastike osakaal FESTART-sõnaraamatus

Sõnavararühm	Võrdlusobjekt	FESTART-sõnaraamatuga kattuvate MV-gruppide arv	Osakaal (%) FESTART-sõnaraamatus
Majandus-sõnavara	TEA inglise-eesti majandus	ligi 28 000	23%
Tehnika-sõnavara	TEA teadus ja tehnika	üle 20 000	16%
Üldsõnavara	TEA Password	4 000	3%
	Silvet 1989	(umbkaudne hinnang:) 69 000	56%
	Kirjastuse FESTART omalooming	(umbkaudne hinnang:) pisut üle 2 000	2%
FESTART-sõnaraamatus kokku		122 915	100%

## 6. Kolmas ülesanne: võimalik koostamisskeem

Eksperdid, kes lähtusid kirjastuse Festart väitest, et vaidlustatud FESTART-sõnaraamatu koostamine algas FESTART inglise-läti sõnastiku põhjal, pakkusid välja järgmise skeemi:

- 1) aluseks võeti märksõnastik ja hääldused FESTART inglise-läti sõnastikust, jättes välja mitmest sõnast koosnevad märksõnad;
- 2) jaotus sõnaliikideks võeti tõenäoliselt koos vastetega üle Silvet 1989 sõnaraamatust;
- 3) koostajad said tõenäoliselt ülesande lisada Silvet 1989 põhjal nii palju vasteid, kui nad tarvilikuks peavad;
- 4) nii valmis sõnastik, mille maht oli vähemalt 32 000 märksõna (koos häälduste, sõnaliikide ja erialamärgenditega). Sellele tehti pealiskaudne korrektuur;
- 5) saadud sõnastikku täiendati puuduvate märksõnade ja vastetega TEA Passwordist: programmiga nopiti välja ühesõnalised märksõnad ja nende vasted. Tekkisid esimesed ebahütlused: osa märksõnu varustati nii häälduse kui ka sõnaliigiga, osa ainult sõnaliigiga, osa üksnes eesti vastega. Märksõnade arv kasvas 3000 võrra (kokku 35 000);
- 6) seejärel täiendati sõnastikku tõenäoliselt TEA teaduse ja tehnika sõnaraamatust pärit märksõnade ja vastetega (välja jäeti mitmest sõnast koosnevad märksõnad). Märksõnade arv kasvas üle 13 000 võrra (kokku üle 48 000). Erialamärgendite ühtlustamisel programmidega tekkis läbivaid vigu;
- 7) saadud sõnastikule lisati tõenäoliselt kogu FESTART inglise-eesti majandussõnaraamat (27 500 märksõna – mis omakorda oli üle võetud TEA inglise-eesti majandussõnaraamatu küljendus põhjal). Kõik need lisatud märksõnad moodustasid eraldi üksuse (seetõttu esineb korduvaid märksõnu);
- 8) sõnastik sai oma lõpliku kuju: valminud FESTART-sõnaraamatus on 71 154 märksõna.

## 7. Neljas ülesanne: kas tegemist on originaal- või tuletatud teosega?

Plagiaatlus selle sõna laiemas tähenduses on sõnastike koostamisel vältimatu. Ühtki sõnastikku ei koostata nullist ja mis tahes uue sõnastiku autor saab materjali olemasolevatest sõnastikest. Kui selle tegevuse käigus a) viidatakse algallikatele, b) omandatakse litsentsiluba, ja c) luuakse loominguiliselt iseseisev ja algsest teosest sõltumatu teos – siis on (autoriõiguse sõnastuses) tegemist tuletatud teosega, mida plagiaatluseks ei nimetata.

FESTART-sõnaraamatut ei pea eksperdid originaalsõnastikuks, sest eksperiti tulemused näitavad, et selle materjal on mehaaniliselt kokku kuhjatud neljast sõnaraamatust. Kirjastuse Festart omapoolne, st originaalne panus on ainult 2%. Sellise mehaanilise ülevõtmise puhul ei tohiks autoriõiguste rikkumise üle otsustamisel saada määravaks asjaolud, et alusmaterjali ei võetud üle tervikuna (Silvet 1989 vastetest on kirjutatud ümber kolm viiendikku, TEA Passwordist viiendik, TEA teaduse ja tehnika sõnaraamatust umbes kolmandik) või et saadud sõnastik ei koosne enam ühest, vaid mitmest eri allikast. Juba algsõnastikele viitamata jätmine on käsitletav autoriõiguse rikkumisena, seda ka juhul, kui on leping, mis lubab nimetatud teoseid kasutada (AÕS § 35 lg 3).

Autorikaitse objektiks saab olla terviklik teos/andmekogu, sõnastike puhul on võimalik asuda seisukohale, et autorikaitse alla peaks kuuluma autori lisatud loominguiline panus, mis võib väljenduda märksõnavalikus, sõnastiku ülesehituses, uute vastete, definitsioonide, kasutusnäidete, kommentaaride, illustratsioonide jm lisamises vm viisil. Vaidlustatud sõnaraamat ei vasta ühelegi neist kriteeriumeist. Koostajate tegevus on mõnevõrra märgatav vaid majandussõnavara puhul.

Kõige selle põhjal on eksperdid FESTART-sõnaraamatu nimetanud tuletatud, mitte originaalteoseks. Autoriõiguse seaduse järgi on tuletatud teos “.. teose tõlge, algse teose kohandus (adaptatsioon), töötlus (arranjeering) ja teose muu töötlus.” (AÕS § 4 lg 21)

## 8. Lõpetuseks: oportuuniteedist

Kuigi eksperitiis tõestas üheselt vaidlustatud sõnaraamatu mehaanilise kompileerimise käigu, ilma selleks luba omamata, lõpetati kriminaalasi, mida Eesti õigussüsteemis oli menetletud seitse aastat, 2007. aastal oportuuniteedi alusel (Truu 2007). Mida selle all mõista? Aga seda, et riigi toimimises, nagu iga inimese elus, arvatakse mõned sündmused, teod või probleemid vähem oluliseks kui teised: oportuuniteediprintsiip on põhimõte, mis seadusena võimaldab otsustada riigi asjade otstarbekuse üle.

Eesti Vabariigi kriminaalkodeksisse lisandus oportuuniteedipunkt (KrMS § 168 täiendusena) uuesti 2004. aastal.<sup>11</sup> Oportuuniteedi põhiline mõte ongi olnud eelkõige vähendada politseiuurijate, prokuröride ja kohtute koormust, st lubada otsustada juba alamal astmel. Kriminaalmenetluse võib oportuuniteedile toetudes lõpetada, kui puudub avalik menetlushuvi ja kui süü ei ole suur (KrMS § 168, 202) või kui karistus on ebaotstarbekas (KrMS § 168, 203). Prokurör Endla Ülviste on

<sup>11</sup> Nõukogude koodeksis oli see samuti olemas, ja toimis väga hästi, nagu on kinnitanud Ain Seppik, kelle sõnusti “seal tegi seda uurija”, mitte prokurör või kohus (väljavõtte Riigikogu õiguskomisjoni istungi protokollist (23.10.2003)).

küll väljendanud oma kergendust: “Te ei kujuta ette, kui palju on kohtutes neidsamu pasteetide, õllepudelite ja saiakeste lugusid” (Püüa 2004). Samas jääb kummitama võimalus “teise astme kuritegu niiöelda lämmatada”.<sup>12</sup> Ja kuidas ikkagi mõõta avalikkuse huvi? Kas lugeda kokku ajakirjanduses avaldatud kirjutised? Või asjasse segatud inimeste hulk? Või veel muud moodi? 2007. aastal avaldatud kokkuvõttest selgub, et lõviosa kriminaalmenetlustest ongi lõpetatud avaliku huvi puudumise tõttu (92% ehk 5239 kriminaalasja), vähem on toetunud ebaotstarbeka karistuse sättele (8% ehk 465 juhtu).<sup>13</sup>

Ehmatava ja kurvastava tõsiasi lisandus 2007. a sellesse pasteedi-õllepudeli-saiakeste loendisse ka loometöö vili – sõnaraamat –, ja samuti avaliku huvi puudumise tõttu. Päris karistusega vaidlustatud sõnaraamatu tegijad siiski ei jäänud: Harju maakohus otsustas, et S. Gluhhov peab riigi tuludesse tasuma 125 000 krooni ning tegema 100 tundi üldkasulikku tööd, firma Festart peab maksma riigi tuludesse 250 000 krooni, tasuma kriminaalmenetluse kulud (413 682 krooni) ja kannatanu TEA Kirjastuse kulud (216 889 krooni).

## 9. Ettepanek tuleviku jaoks

Eesti Keele Instituudis välja töötatud sõnastike kokkulangevust hindava märksõna-vaste-gruppide mõõtmise meetodi ja selle rakendamise põhjal kakskeelsete sõnaraamatute ekspertiisil tuleks kehtestada piirmäär, millest edaspidi lähtuda sõnaraamatu originaalsuse üle otsustamisel.

Tuletame veel meelde siinsest artiklist läbi käinud protsente ja fakte: 18. saj inglise Fisheri õigekirjasõnaraamat, kus üks-ühele oli teisest üle võetud 43,5%, lubati kohtuvaidluste järel siiski välja anda; teise, 17. saj inglise juhtumi puhul olevat üle võetud koguni 70% märksõnadest koos vastetega, aga too asi kohtuni ei jõudnud. FESTART inglise-eesti majandussõnaraamatu MV-gruppidest langes TEA majandussõnaraamatu materjaliga kokku 97%. Ühe kontrollrühma (1141 märksõna) osas oli FESTART-sõnaraamatu kokkulangevus TEA Passwordiga samuti 97%, samas EKI sõnastikuga 53%. Kui samas kontrollrühmas võeti arvesse kolmas üksus (sõnaliigimärgend), siis kattusid MV-grupid TEA-ga 92%, aga EKI sõnastikuga üksnes 24% (!) ulatuses. Korduva märksõna-osisega artiklitest (ligi 5000) on FESTART-sõnaraamatus 90% selliseid, kus märksõna kordub muutu-matul kujul. Ekspertiisi käigus leitud 69-st TEA Passwordi poolitusveast kordub 68 (98,5%) FESTART-sõnaraamatus.

Autorite ettepanek on kehtestada kahes sõnaraamatus kokkulangevate märksõna-vaste-gruppide piirmääraks umbes 60–75% (pöörates muidugi tähelepanu ka muudele “arvutinippidele”). Üle selle on tegemist plagiaadiga, alla selle (tuletatud) originaalteosega.

<sup>12</sup> Märt Raski sõnad Riigikogu õiguskomisjoni istungil (väljavõte protokollist 23.10.2003).

<sup>13</sup> Justiitsministeeriumi kriminaalpoliitika osakonna ülevaade, ptk “Kriminaalmenetluse oponentidega lõpetamine (1.07.2004–31.03.2007)”, vt [www.just.ee](http://www.just.ee) (6.10.2007).

## Kirjandus

- Arujärv, Evi 2008. Kultuur – missioon ja äri. – Eesti Päevaleht, 15.02.2008.
- AÕS = Autoriõiguse seadus. Elektrooniline Riigi Teataja, <https://www.riigiteataja.ee/ert/act.jsp?id=12791124> (20.02.2008).
- Cormier, Monique C.; Fernandez, Herberto 2005. From the Great French Dictionary to the Royal Dictionary. – *International Journal of Lexicography*, Vol. 18, No. 4, 479–507.
- EEVA = Ross, Kristiina. Anton Thor Helle (1683?–1748 ). Piiblitõlkija ja eesti kirja-keele arendaja, Jüri pastor. – Eesti vanema kirjanduse digitaalne tekstikogu. [www.utlib.ee/ekollekt/eeva/](http://www.utlib.ee/ekollekt/eeva/) (20.11.2007).
- Helle, Anton Thor 1732. Kurtzgefazzte Anweisung Zur Ehstnischen Sprache, in welcher mitgetheilet werden I. Eine GRAMMATICCA. II. Ein VOCABULARIUM. III. PROVERBIA. IV. AENIGMATA. V. COLLOQUIA. Tallinn: Eestimaa Konsistooriumi kirjastuskassa.
- Hupel, August Wilhelm 1780. Ehstnische Sprachlehre für beide Hauptdialekte, den revalschen und dörrptschen, nebst einem vollständigen Wörterbuch. Riga-Leipzig: Hartknoch.
- Kask, Arnold 1955. S. H. Vestringi eesti-saksa sõnaraamat ja selle osa eesti leksikograafias. – Ajaloo-keeleteaduskonna töid. Tartu Riikliku Ülikooli toimetised 38. Tartu: Tartu Riiklik Ülikool, 181–192.
- Krikmann, Arvo 1986. Fraseoloogiline aines eesti vanemais grammatikates ja sõnastikes. Tallinn: Valgus.
- KrMs = Kriminaalmenetluse seadustik. Elektrooniline Riigi Teataja, <https://www.riigiteataja.ee/ert/act.jsp?id=12897094> (20.02.2008)
- Landau, Sidney I. 2001. *Dictionaries: The Art and Craft of Lexicography*. 2nd. ed. Cambridge University Press.
- McMenamin, Gerald R. 2002. *Forensic Linguistics. Advances in Forensic Linguistics*. CRC Press.
- Püüa, Marko 2004. Pasteedivaraste karistused leevenesid. – *Postimees*, 25.03.2004. Vt ka <http://www.postimees.ee/260304/esileht/siseuudised/129949.php> (5.03.2008).
- Rodriguez-Alvarez, Alicia; Rodriguez-Gil, Maria 2006. John Entick's and Ann Fisher's Dictionaries: An Eighteenth-century Case of (Cons)Piracy? – *International Journal of Lexicography*, Vol. 19, No. 3, 287–319.
- Truu, Aet 2007. Kohus lõpetas piraatkoopiaste valmistamises süüdistatava kriminaalasja oportuniteedi alusel. – *Äripäev*, 04.04.2007. Vt ka <http://www.ap3.ee/Default2.aspx?ArticleID=5aa5daae-8403-4fc0-a6b2-17848c1adfc3> (5.03.2008).
- Vestring, Salomo Heinrich 1998 [käsikiri ?1720-ndad]. *Lexicon Esthonicum Germanicum*. Trükki toimetanud Ellen Kaldjärv. Tartu: Eesti Kirjandusmuuseum.
- Williams, John 1992. The question of plagiarism and breach of copyright in the dictionary-making process (with particular reference to the UK). – *EURALEX '92 Proceedings*. Tampere: Tampereen Yliopisto, 561–570.

**Margit Langemets** (Eesti Keele Instituut) on lõpetanud Tartu Ülikooli eesti keele ja kirjanduse erialal. Osales TEA-Festardi kohtuprotsessis tunnistajana. Uurimisvaldkonnad: leksikoloogia ja leksikograafia üldiselt, arvutileksikoloogia, leksikaalne semantika.  
[margit@eki.ee](mailto:margit@eki.ee)

**Piret Voll** (Eesti Keele Instituut) on lõpetanud Tartu Ülikooli eesti keele ja kirjanduse erialal. Osales TEA-Festardi kohtuprotsessis peaekspertina. Uurimisvaldkonnad: leksikoloogia ja leksikograafia.  
[piret.voll@eki.ee](mailto:piret.voll@eki.ee)

## **LINGUISTIC FORENSIC ANALYSIS OF A DICTIONARY: AN ESTONIAN PRECEDENT**

**Margit Langemets, Piret Voll**

Institute of the Estonian Language

In 2003 the first linguistic forensic assessment of lexicographic objects took place in Estonia. The case required expert assessment of the contents of several dictionaries involved in criminal accusation of theft of intellectual property. The article describes the process of lexicographic expert assessment carried out at the Institute of the Estonian Language, and the ad hoc language-technological method applied. Relying on the results, a critical measure has been elaborated to judge over the degree of originality of bilingual dictionaries. Being the first and only case of its kind in Estonia, it offers a linguistic-forensic precedent.

**Keywords:** forensic linguistics, lexicographic expert assessment, dictionaries, English, Estonian