

Diakroonilised nihked eesti ilukirjanduskeele kasutusmustrites 1890–1990

Jekaterina Trainis

Tallinna Ülikooli lingvistika doktorant
jekaterina.trainis@tlu.ee

Teesid: Artiklis võrdlen 1890. ja 1990. aastate ilukirjanduskeele kasutusmustriteid, toon esile saajandi vältel toimunud statistiliselt olulised kvantitatiivsed nihked ja püstitan nende alusel hüpoteese võimalike kvalitatiivsete diakrooniliste arengute kohta. Analüüsi tulemuste põhjal väidan, et eesti keeles on suurenenud analüütilisus (nt liitsete üksuste kujunemine seoses adverbi sisaldavate mustrite laienemisega) ning et avatud ja suletud sõnaliikide funktsionaalsel võimekusel on tekstilised piirangud (nt konjunktsiooni tekstikasutus laieneb analüütiliste rinnastav-alistavate rühmsidendite arvelt, adpositsiooni funktsioonide kitsenemisega juurdub adpositsioonide hulk). Taustateguritest tulevad arvesse nii tekstiloomelised kui ka keeleuenduslike pürgimuste mõjud kirjakeele kujunemisele.

Märksõnad: diakroonilised nihked, eesti ilukirjanduskeel, korpuslingvistika, lingvistiline klasteranalüüs, morfosüntaktilised mustrid, statistika

1. Teoreetilised lähted ja meetodid

Kirjeldades keelt, tuleb silmas pidada, et keel muutub ajas ebaregulaarselt. Joan Bybee järgi on selle nähtuse esile toomisel ja teoreetilisel mõtestamisel olulised keele üldised kasutusmustrid ning nende esinemise sagedus. Funktsionalistliku kasutuspõhise keeleteaduse primaarne hüpotees on, et keelekasutus kujundab grammatikat (Bybee 2006: 6, 269). Kvantitatiivse analüüsi tulemusel esile tulnud keelemuustrid on olulised mõistmaks, kuidas inimesed keelt mõtestavad, milline roll on selles kordustel ja kordamisel (vt Bybee 2006: 7–9). Need kaks seisukohta haakuvad käesoleva analüüsi eesmärgiga – leida ja võrrelda ajas ilmneva muutusi 1890ndate ja 1990ndate keelekasutuses, keskendudes kvantitatiivsetele ja kvalitatiivsetele morfosüntaktilistele ning leksikaalsemantilistele nihetele, k.a sõnajärg.

1.1. Sünkrooniline ja diakrooniline lähenemine

Keelt võib analüüsida sünkrooniliselt (nt 1890ndate ilukirjanduskeel) või diakrooniliselt (nt 1890ndate ja 1990ndate ilukirjanduskeele võrdlus). Kuigi Ferdinand de Saussure pidas õigeks sünkroonilist ja diakroonilist lähenemist keele “läbilõike” uurimisel lahus hoida, kuna saadud andmed erinevad (vt Saussure 1999: 89–91), ei ole keelelist arengut uurides mõistlik üht lähenemist teisest eraldada, vaid kombineerida (vt McMahon 1994: 9–11; ka Ehala 1996: 375). Vastandina sünkrooniale ja diakrooniale on kasutusel ka *akroonia* mõiste, mis viitab sünkroonia-diakroonia piiride hägususele ja keelestruktuuride piiritlematusele ajaliste raamidega. Nüüdiskeele analüüs ei näita päris täpselt, kuidas keelekasutus on kujunenud, mis on selle taga, missugused arenguetapid on meie tänane keelekasutus aja jooksul läbi teinud ning millega võiks seda seletada, ent tänase keelekasutuse põhjal annab oletada, kust miski võiks pärineda ja miks.

Susan Hunstoni (2002: 160) järgi on diakroonilisel keeleuurimisel kaks raskuspunkti:

- 1) (kaugema) mineviku tekstid pole sama kättesaadavad nagu oleviku tekstid ning seetõttu on raske moodustada võrdluskorpust;
- 2) nii mineviku tekstikorpuse siseselt kui ka kahe (või enama) korpuse võrdluses leidub kirjaipildi varieerumist.

Ka 1890ndate korpuses esinevad täheortograafia variatsioonid (*uueste, kriipisivad, saivad, hää, pääle, naesterahvad, kudas, nüid, piab kuuldamama* jm), võrdluses 1990ndatega ilmnevad sõnajärje- (... , *et mul wißt waja ei ole ennaßt eßiteledagi*) ning rektsoonimallide erinevused (... , *kus ka pere tütar Salme ja Kase Kati arvustamise alla võeti*), mis raskendab automaatset ühestamist, lemmade, vormisageduse ja kollokatsioonide loendamist. Seetõttu asendasin analüüsitavas korpusmaterjalis saksakeelse β -tähe ühe- või kahekordse *s*-ga (nt *isfeeneßelt > iseeneselt, ßamasße > samasse*), *w* asemel sisestasin *v*-tähe (nt *kargasiwad > kargasivad*). 1890ndate ilukirjanduskorpuse põhjal on märgata, et liigutakse eesti kirjakeele ühtsuse poole, kuid nn vanad reeglid on visad kaduma, nt eesti keelele mitteomased fonotaktika reeglid ja nende vaheldumine kirjaipildis: *olid, olivad; hea, hää; peale, pääle; sealsed, säälsed; kergesti, kergeste; sinna, senna; oleksivad jõuudnud, tikkusivad* jms. Muuhulgas on sarnaste juhtumite esinemine 1890ndate tekstides tingitud ka algallikate autorite keelelisest kogemusest (vt ka Bybee 2006: 280).

Niisiis on diakroonial kirjakeele kujunemisloos põhjendav-seletav ning sünkroonial küsimusi-kahtlusi tõstatav roll. Martin Haspelmath (2008b: 18) väidab teiste ja enda varasematele uurimustele toetudes, et diakrooniline muutus on oluline ühenduslüli keelekasutumustrite ja grammatiliste struktuuride vahel.

William Croft (2003: 272) on seisukohal, et järeldotsi diakrooniliste muutuste kohta tuleks teha keele sünkroonilise seisundi põhjal, kuna keeled ei paku tavaliselt diakrooniliste muutuste otsest tõendusmaterjali. Oma uurimuses olen lähtunud kahest konkreetsest eraldiseisvast aastakümnest, et leida eesti keele kasutusmustrite võrdluse alusel jälgi diakroonilistest muutustest ja saada teavet nendega seotud lingvistiliste protsesside kohta.

1.2. Uurimissuund

Siinne uurimus lähtub vormi-funktsiooni loomupärasest ühtsusest tekstis, kus tuleb esile vormide funktsionaalne võimekus. Selles näen seoseid nii Praha koolkonna funktsionalismi kui ka Aleksandr Bondarko vene keele funktsionaalse grammatika põhimõtetega (nt Mathesius 1967; Bondarko 1983), kuna mõlemal juhul keskendutakse sellele, kuidas keelt kasutatakse. Seetõttu analüüsin oma uurimuses keelekasutusele omaseid morfoloogilisi ja morfosüntaktilisi mustreid, mis sisaldavad kindlaid sõnaliigijärgnevusi, vorme ja lekseeme ning millel on tekstis kindlad funktsioonid.

Arenguliste ja võimalike võõrmõjuliste muutuste ning erinevate perioodide keelekasutuses ilmnenud nihete mõtestamisel tuginen loomuliku morfoloogia põhimõtetele, mille Wolfgang Ulrich Dressler, Wolfgang Ullrich Wurzel ja Willi Mayerthaler arendasid välja 1970ndatel loomuliku fonoloogia teooriast. Käsitlus rajaneb Praha lingvistilise koolkonna funktsionalismile, *markeerituse* ja (*morfoloogiliste*) *universaalide* mõistetele, mida saab rakendada ka diakrooniliste muutuste jm lingvistiliste nähtuste kirjeldamisel, sh keelekasutuse loomulikuse hindamiseks ja seletamiseks. (McGregor 2015: 331–333, 339; Habicht 2006: 13–15; Habicht *et al.* 2006: 612–613; McMahan 1994: 97–106.)

Oma uurimuses lähtun korpusainestikust ehk keeleandmetest (ingl *corpus-driven research*, ka *data-driven research*), mis sobib keelekasutusmustrite võrdlev-diakroonilise ja sünkroonilise kirjeldamise eesmärgiga, võimaldades teha avaramaid üldistusi. Nimetatud analüüsimeetodist oluliselt laiem mõiste on *corpus-driven linguistics* ehk korpusest tulenev keeleteadus, mida nt Elena Tognini-Bonelli (2001: 177–179) näeb iseseisva uurimisdistsipliinina.

Korpusest tulenevad lähenemised on uuenduslikumad, sest võimaldavad avada keelelisi nähtusi, mida pole täheldatud traditsiooniliste teooriate raames (Biber 2009: 223), ja lubavad keelekasutust analüüsides liikuda üksikult üldisele ning esitada uurimisküsimusi ja hüpoteese ainestikust tulenevalt. Olemuselt on see kaugemale ja sügavamale ulatuv kui korpuspõhine kirjeldus, sest probleemide leidmise ja lahendamise allikas pole mitte lingvisti keeleteaju, sisetunne või teoreetilised seisukohad, vaid mahukas tekstikogu. Seda on

töestanud varasemadki uurimused (nt Hunston & Francis 2000; Römer 2005; Brunner & Steyer 2007). Ka siinses uurimuses ei ole korpus teooriate tõestamiseks mõeldud tekstikogu, vaid keeleaines, millel on võimalik rakendada erinevaid arvutiprogramme ning analüüsimeetodeid ja saada põhimõtteliselt teistsugust teavet võrreldes traditsioonilise korpuspõhise lingvistilise analüüsiga (vt Francis 1993: 137–156; Esilon & Matsak 2009: 84–85; Feng 2014: 83–86). Olen seda kogenud ka oma eelnevas uurimistöös (Trainis & Allkivi 2014; Trainis 2015). Samas on mõistetav, et kuitahes mahukas korpus ei hõlma kogu keelelist rikkust ja eripära, kuid ühe või teise keelendi puudumine ei tähenda, et seda pole olemas (vt ka Piperski 2015).

1.3. Meetodid

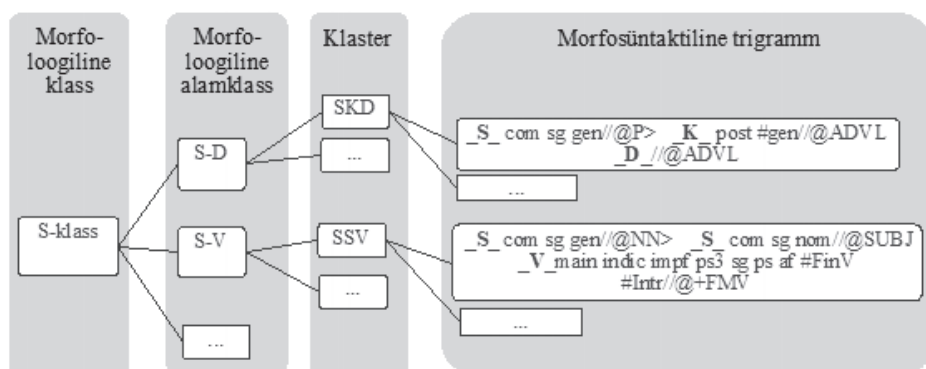
Järgmine küsimus on sobiva meetodi valik uurimiseesmärkidest ja keeleainest lähtudes, mis võimaldaks samadel alustel esile tuua ning võrrelda erinevate perioodide (1890ndad ja 1990ndad) keelekasutusmustreid, nende morfoloogilist struktuuri, morfosüntaktilise ja leksikaalse varieerumise piire, vormi- ja sõnamoodustust, vormide funktsionaalset võimekust jm seoseid. Nagu on näidanud arvukad kasutusgrammatilised uurimused (nt Biber 1991; Bybee & Hopper 2001; Bybee 2006; Haspelmath 2006, 2008a, 2008b; Gries 2007; Gries & Hilpert 2012; Gries & Stefanowitsch 2012; Gries & Ellis 2015), saab nimetatud eesmärkidel kõige objektiivsemat teavet erinevate statistiliste meetoditega. Siinses uurimuses olen rakendanud selleks lingvistilist klasteranalüüsi (vt allpool) ja log-tõepära funktsiooni (edaspidi LL, vt punkt 1.4).

Esmalt peatun *klasteranalüüsi* (ingl *cluster analysis*, ka *clustering*) mõistetel, kuna ma pole rakendanud seda meetodit tavapäraselt, st kui statistilist grupeerimismeetodit, mille puhul on võimalikud kolm lähenemist: hierarhiline klasterdamine (objekte on suhteliselt vähe või klastrid erinevad üksteisest suhteliselt selgelt), k-keskmiste klasterdamine (uuriija määrab klastrite arvu) ja kombineeritud meetodid (andmeid jaotatakse järk-järgult, kuni igas rühmas on ainult üks objekt) (vt Everitt 1997: 466, 468–472; Remm *et al.* 2012: 75–77; Mooi & Sarstedt 2011: 241). Meetodit kasutatakse sotsioloogias, bioloogias, meditsiinis, majanduses ja mujal, et teha kindlaks, milliseid seoseid ja nähtusi on statistika abil võimalik leida (Gore 2000: 318; Everitt 1997: 472; Trainis 2015).

Siinses uurimuses on tegu *lingvistilise klasteranalüüsiga*, milles rakendan eesti keele tarkvaral põhinevat programmi Klastrileidja (Ots 2012), mis töötab andmekaeve põhimõttel, ja Excelit. Klastrileidja sisendiks on parseriga EstCG 1.0 pindsüntaktiliselt eelmärgendatud tekstid (kasutatud kitsenduste grammatika reeglipõhist versiooni). Klastrileidja koondab tekstis sarnase morfoloogilise,

morfosüntaktilise ja süntaktilise märgendusega kahe-, kolme-, nelja- ja viiekomponendilisi järjendeid ehk n-gramme ning fikseerib nende tekstisageduse.

Programm otsib järjendeid libiseval meetodil nii (osa)lause algusest, keskelt, kui ka lõpust. Uurimuse tarvis piirasin Klastrileidja otsingu kolmest komponendist koosnevate üksuste ehk trigrammidega ja seejärel reastasin Excelis sarnase lingvistilise vormistusega kaks ja enam korda esinenud trigrammid tähestikuliselt esikomponendi sõnaliigilise kuuluvuse alusel. Nii moodustuvad morfoloogilised klassid. Iga klassi trigrammide lõpukomponendi sõnaliik eristab klassis rea morfoloogilisi alamklasse ning iga alamklassi trigrammide keskmise komponendi sõnaliik toob välja morfoloogilised klastrid. Iga klaster sisaldab morfosüntaktiliselt ja leksikaalselt varieeruvaid trigramme. Tegu on morfoloogiliste klasside, alamklasside ja klastrite hierarhiaga, siit ka lingvistilise klasteranalüüsi hierarhia mõiste, mida uurimistulemuste süstematiseerimisel ja kirjeldamisel kasutan (vt joonis 1 substantiivialgulise klassi kohta).



Joonis 1. Lingvistilise klasteranalüüsi hierarhia.

Kuigi ma ei kasuta hierarhilise klasteranalüüsi meetodit ega sellele omast viisuaali – dendrogrammi – korpuslingvistikas tavapärasel kujul, sest ei arvesta mõõdetavaid kaugusi, struktureerin andmeid divisiivse hierarhilise klasteranalüüsi loogika järgi (ka *top-down clustering* ehk ülalt alla klasterdamine)¹. Niisiis on Excelis sorteerimine (vt joonis 1) oma olemuselt dendrogrammi tekitamine. Meetodina on lingvistiline klasteranalüüs paindlik, võimaldab sätestada kindlaid klasterdamisreegleid ning toob sageduse alusel esile ühesuguseid morfo- ja süntaksimärgendite lineaarseid järgnevusi.

Teisalt seisab lingvistilise klasteranalüüsi loogika, seotuna grammatilise ja leksikaalse analüüsiga, väga lähedal Ameerika strukturalistlikule distributiivsele analüüsile ning sellega lähedaste põhimõtete rakendamisele nt Vene (arvuti)lingvistikas. Mõlemal juhul käsitletakse keeleüksuste võimet siduda endaga teisi üksuseid kas piiramatult või (range) valiku printsiibil.

Klassikalise distributiivse analüüsi iseloomulikuks jooneks on tähenduse väljajätmine: kogu protseduur tugineb keeleüksuste koosesinemise võimalustele, teisisõnu keeleüksuste omavahelistele formaalsetele kombineerimisreeglitele, mis toovad esile nende vastastikuseid seoseid erinevates kontekstitüüpides ja üldistuvad keelesüsteemile tervikuna. Sellest on välja kasvanud erinevaid distributiivse semantika mudeleid, mis on arvutilingvistikas saanud leksikaalse semantika tööriistaks, võimaldades mh konstrueerida tähenduskirjeldusi sõnade esinemuse põhjal ulatuslikes tekstikorpustes (vt Orav & Parm 2014: 403).

Käesolevas uurimuses rakendatud lingvistilist klasteranalüüsi seob klassikalise distributiivse analüüsiga keele elementide kooskasutusmustrite leidmine formaalsete tunnuste põhjal – Klastrileidja otsib n-gramme sarnaste morfo- ja süntaksimärgendite koosesinemise alusel ning fikseerib mustrite sageduse.

1890ndate ja 1990ndate ilukirjanduse korpuste tekstidest Klastrileidja väljundi põhjal saadud trigrammid olen jaganud morfoloogilistesse klassidesse, alamklassidesse ja klastritesse. See hierarhiline järjestus üldiselt konkreetsemale on süstematiseeriv ja klassifitseeriv ning sobib seetõttu analüüsi tulemuste lingvistilise interpreteerimise ja kirjeldamise aluseks. Klasside tasandil ilmnenuid sõnaliikide kasutustendentsid konkretiseeruvad alamklasside ja klastrite tasandil. Üksikasjalikum kvalitatiivne lingvistiline analüüs trigrammide tasandil aitab välja selgitada seoseid semantika, vormikasutuse, leksikaalse ja funktsionaalse varieerumise piirangute vahel, mis võimaldavad seletada, kuidas inimesed suhtluseesmärgil keele elemente tavaliselt kombineerivad ning mida seejuures oluliseks peavad. Martin Ehala (1996: 453) toob välja, et diakrooniline keeleteadus peaks pöörama tähelepanu eeskätt pisihälvetele, kui on eesmärgiks seletada, kuidas muutused keeles tekivad ja kujunevad. Trigrammide kasutusmustreis esinevate pisihälvete põhjal saab võrrelda ja objektiivsemalt hinnata 1890ndate ja 1990ndate eesti ilukirjanduskeele kasutuses toimunud kvantitatiivseid ja kvalitatiivseid nihkeid, lootes leida sisemistest arengulistest ja võimalikest võõr mõjulistest faktoritest tingitud nähtusi.

1.4. Representatiivsus, valiidsus ja võrreldavus

Mistahes uurimistöös on oluline põhjendada valimi representatiivsust, analüüsitulemuste valiidsust ja võrreldavust. Selleks on erinevaid võimalusi. Oma analüüsis olen lähtunud kahest homogeensest samade printsiipide alusel koostatud

valimist. Nii 1890ndate kui ka 1990ndate ilukirjandustekstikogud on tüübilt suletud korpused ning selles mõttes võrreldavad (vt ka Biber 1993: 243–257). Valimite erinev maht pole takistus, kuna keelekasutusmuustrite võrdlus tugineb klasside, alamklasside, klastrite ja trigrammide osakaalule kummaski valimis. Nii eraldas Klastrileidja 1890ndate ilukirjanduskorpusest 99 687 trigrammi, kus kaks ja enam korda esinenud trigramme oli 9879, ainukordseid 89 808. Ainukordsete ($89\,808 \times 100\% : 99\,687$) ning kaks ja enam korda esinenud trigrammide ($9879 \times 100\% : 99\,687$) suhe on vastavalt 90,1% : 9,9%. Kehtib Zipfi needus, mida Ehala on seoses sageduse ja häälikupikkusega kirjeldanud seaduspärasusena, kus “sageduse ja häälikulise pikkuse vahekord on pöördvõrdeline” (Ehala 1996: 383). Sama saab öelda 1990ndate valimi kohta, kus on 49 211 trigrammi: “suhe kaks ja enam korda esinevate trigrammide ($14\,316 \times 100\% : 49\,211$) ning ainukordsete ($34\,895 \times 100\% : 49\,211$) suhe on 29,09% : 70,9%” (Trainis & Allkivi 2014: 285).

Edasine samm on kindlaks teha võrreldavate valimite optimaalne trigrammide hulk ehk statistiline olulisus, millest alates võib rääkida uurimistulemuste valiidsusest ja kahe valimi võrreldavusest. Selleks olen valinud arvnäitaja k , mis arvutatakse järgmise valemi alusel: optimaalne trigrammide hulk k 1890ndate valimis võrdub ruutjuurega kaks ja enam korda kasutatud trigrammide hulgast – 9879 – jagatud 2-ga, tulemuseks saame 70 ($k \approx \sqrt{9879 \div 2}$). 1990ndate valimis $k \approx 85$. Niisiis, kui $k \geq 70$ või $k \geq 85$, siis on klaster esinemus vastavas valimis statistiliselt oluline.

Järgnevalt võrdlesin valimitevahelist statistiliselt olulist sarnasust-erinevust sümmeetrilise mõõdiku abil. Selleks rakendasin log-tõepära funktsioonile üles ehitatud programmi (loonud Paul Rayson²), mis tuvastab sageduse absoluutarvude põhjal kahe valimi statistiliselt olulist erinevust (vt lisad 1–3). Log-tõepära ehk LL-näitaja toob 1890ndate ja 1990ndate vahel klastrite esinemuses välja samalaadsust ja erinevusi. Ühes või mõlemas valimis optimaalsuskoeffitsiendi põhjal ebaolulise klaster esinemus võib osutada 1890ndate ja 1990ndate korpuste võrdluses oluliselt erinevaks või ühes valimis kõrgema statistilise olulisusega klaster osutada ebaolulisel määral esindatumaks.

Lisades 1–2 on arvutatud kahe sagedaima, substantiivi- ja adverbialgulise klassi klastrite osakaalud alamklassides. Tõin helehalli värviga esile statistiliselt olulised klastrisagedused ning määrasin nende LL-näitaja.

Mida suurem on LL väärtus, seda vähem tõenäoline on tulemuse juhuslikkus. Erinevust kahe korpuse vahel võib lugeda oluliseks, kui $LL \geq 3,84$ ($p < 0,05$). Plussmärk LL-näitaja ees viitab, et trigrammi kasutatakse rohkem 1890ndate kui 1990ndate korpuses. Miinusmärk kajastab vastupidist tendentsi.

Tähis p näitab nullhüpoteesi kehtimise tõenäosust. Kui p väärtuseks on 0,05, siis viiel juhul sajast võib seos juhuslike kokkulangevuste tõttu välja paista

ka siis, kui seda tegelikult ei ole. Sõltuvalt uuringu tüübist valitakse olulisusnivooks lubatud viga 10%, 5% või 1%. Keeleteaduslikes töodes on suhteliselt tavaline piir 5%, millega olen arvestanud ka siinses uurimuses. Et tegu on konstantse suurusega, siis pole ma 1890ndate ja 1990ndate keelekasutusmustrite võrdluses LL-näitajatele seda lisanud. Näiteks kui alamklassisiseselt ei ole osakaalu järgi klastri substantiiv-adverb-adverb (SDD) esinemuses sajandi jooksul märkimisväärseid muutusi toimunud (vastavalt 18,6% ja 18,1%), võib seda erinevust kahe korpuse vahel siiski oluliseks pidada, kuna LL on $-82,28$ ($p < 0,05$). Järelikult on valimi suurus piisav, kui suhteliselt väike protsendiline vahe osutub mõõdetavaks. Ka sellest vaatevinklist saan pidada 1890ndate ja 1990ndate ilukirjandustekstide valimeid representatiivseteks.

2. Analüüsi tulemused ja arutelu

1890ndate ja 1990ndate ilukirjanduskeele lingvistilise klasteranalüüsi tulemusi võrdlen ning kirjeldan morfoloogiliste klasside, alamklasside, klastrite ja trigrammide hierarhia alusel (vt punkt 1.3). Ilukirjanduskeele kasutusmustrite leksikaalsemantilise, morfosüntaktilise ja funktsionaalse varieerumise piire vaatlen substantiivi- ja adverbialgulise alamklassi sagedamate klastrite trigrammide näitel.

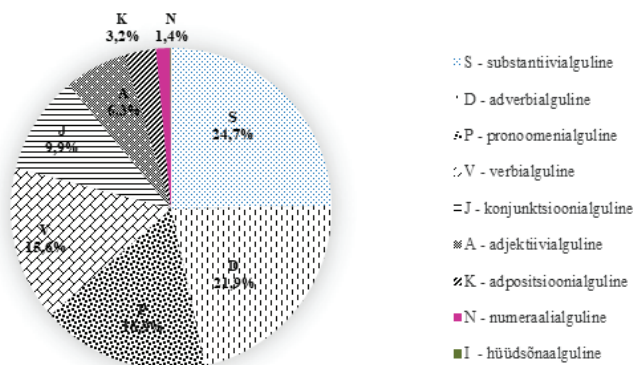
2.1. Morfoloogiliste klasside võrdlusest tulenevad hüpoteesid

Morfoloogilised klassid toovad keelekasutusmustrites esile sõnaliikide kasutuseelistusi. Hierarhia järgmistel tasanditel täpsustuvad järk-järgult lingvistilised tingimused, mis iseloomustavad sõnaliikide kooskasutust muustris. Selle taga on näha 1890ndate ja 1990ndate eesti ilukirjandustekstidele omaseid morfoloogilisi eelistusi ja lingvistilisi protsesse, trigrammide tasandil ka keelekasutuse morfosüntaktilisi, leksikaalseid ja funktsionaalseid piiranguid, mis võimaldavad seletada kasutuselise eesti keeles sajandi vältel toimunud kvantitatiivseid ja kvalitatiivseid muutusi.

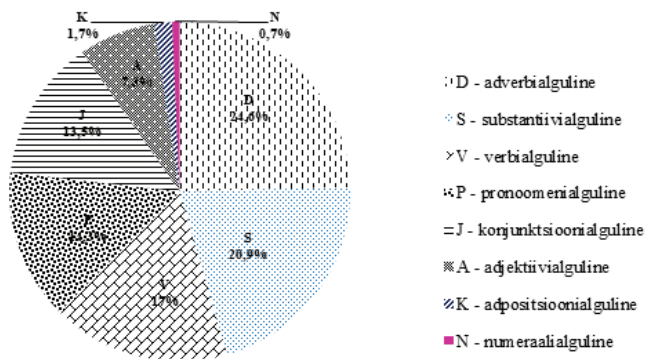
1890ndate aastate ilukirjanduskeele analüüsi tulemuste põhjal (vt joonis 2) on tekstikasutuses suurim osakaal substantiivi- (24,7%) ja adverbialgulisel klassil (21,9%), umbes kolmandiku võrra väiksem pronoomeni- (16,9%) ja verbialgulisel klassil (15,6%). Järgnevad konjunktsiooni- (9,9%), adjektiivi- (6,3%), adpositsiooni- (3,2%) ja numeraalialguline (1,4%) klass. Täiesti marginaalseks jääb hüüdsõnaalguline klass (0,1%), mis pole ringdiagrammil välja toodud.

1990ndatel on toimunud morfoloogiliste klasside jagunemises rida muutusi (vt joonis 3), neist olulisim on substantiivialgulise klassi asendumine adver-

bialgulisega. Kui 1890ndate korpuse sagedaim morfoloogiline klass on substantiivialguline (24,7%) ning adverbialguline klass on teisel kohal (21,9%), siis 1990ndatel on sagedam adverbialguline klass (24,6%) ja järgneb substantiivialguline (20,9%). Seega on substantiivi esinemus ilukirjanduskeele kasutusmustrites 1990ndateks kahanenud 3,8% ja adverbi esinemus kasvanud 2,7%. Saja aasta möödudes on toimunud nähtav nihe.



Joonis 2. Morfoloogiliste klasside osakaal 1890ndate ilukirjanduskeeles.



Joonis 3. Morfoloogiliste klasside osakaal 1990ndate ilukirjanduskeeles (Trainis & Allkivi 2014: 286 andmetel).

Järgmise rühma moodustavad 1990ndate ilukirjanduskeele kasutusmustrites verbi- (17%), pronomeni- (14,3%) ja konjunktsioonialguline (13,5%) klass. Võrreldes 1890ndate andmetega on verbi ja konjunktsiooni esinemus tõusnud – vastavalt 1,4% ja 3,6% – ning pronomeni 2,6% võrra kahanenud.

Samalaadseid kvantitatiivseid nihkeid võib näha ka väikese esinemusega morfoloogiliste klasside puhul: 1990ndate keelekasutuses on 1% võrra suurenenud adjektiivialgulise klassi osakaal, samas kui adpositsiooni- ja numeraalialgulise klassi osakaal on kahanenud vastavalt 1,5% ja 0,7%, st adpositsioonil poolteist korda ja numeraalil poole võrra. Kuigi vahed pole suured, viitavad need siiski nihetele adjektiivi-, adpositsiooni- ja numeraalialgulistest mustrites, mille alusel võib toimuda erinevaid arenguid.

Tõstatub küsimus, millest võiksid olla tingitud eespool kirjeldatud kvantitatiivsed nihked. Võimalike põhjenduste ning küsimuste sõnastamisel lähtun sõnaliikide jaotamisest avatud ja suletud klassideks³. Fred Karlssoni järgi on avatud sõnaliikideks tüüpiliselt substantiivid, adjektiivid, verbid ja adverbid. Neis sisaldub palju mitmetähenduslikke sõnu, mille hulk küünib kümnete tuhandeteni. Suletud sõnaliikide alla koonduvad adpositsioonid, konjunktsioonid, pronoonemid, artiklid jm väiksemad sõnarühmad abisõnade funktsioonis. Suletud sõnaliigid sisaldavad piiratud hulka keelendeid ning tavaliselt neid niisama juurde ei teki. Kui avatud sõnaliike annab võrrelda semantiliste morfeemidega, siis suletud sõnaliikide puhul saab tõmmata paralleeli grammatiliste morfeemidega. (Karlsson 2002: 218.)

Kuna nii substantiivid kui ka adverbid kuuluvad avatud sõnaliikide hulka, siis põhjendab see nende suurt esinemissagedust ja funktsionaalset võimekust. Et adverb on avatud sõnaliikides ainsana muuteparadigmata, siis on loogiline, et sellesse "lisandub eri tüüpi ainek, millest osa on üleminekustaadiumis mitme sõnaliigi vahel" (Habicht *et al.* 2011: 38). Adverbi avara funktsionaalse võimekuse tõttu on kasutusnihked tavalised – esineb nt süntaktiliselt, semantiliselt ja fonoloogiliselt tõmbunud adverbikooslusi, mida vaadeldakse etapina potentsiaalsete liitsete üksuste kujunemisel (vt Eslon 2017: 22 jj). Ka substantiiv kui avatud sõnaliik läbib samalaadseid arenguid. On tavaline, et nt kivilinenud käändevormid leiavad kasutamist nii substantiivile kui ka adverbile omastes funktsioonides, vrd *meel* (ainsuse nominatiiv), *Su meel* (subjekt) *on kurb* ja *meelega* (ainsuse komitatiiv), *Meelega* (viisimäärus) *ta seda küll ei teinud*.

Ühelt poolt viitab adverbi kasutusala laienemine keele toimimise üldisele seaduspärasusele, teisalt aga näitab, et ajavahemikus 1890–1990 on eesti keele kasutuses olulisemaks saanud mustrid, milles adverbi eelistatakse substantiivile. Sel nähtusel võib olla mitmeid seletusi. Täheledatakse lekseemide tähenduse tuhmumist, mistõttu nende funktsioonid avarduvad (nt Vaiss 2004), lisandub liitsete üksusi, millel oma tähendus ja funktsioonid (nt Jürine 2016; Habicht & Penjam 2006). Neid protsesse on kirjeldatud grammatisatsiooni- ja leksikaliiseerumise raames (nt Heine 2003; Metslang 2006; Velsker 2010; Küngas 2014), mis toimivad keelemuutuste suunal. Ehala on rõhutanud, et keele olulisemaid iseorganiseerumise ilminguid on grammatisatsioon:

See on kõigis maailma keeltes toimiv diakrooniline universaal, mille toimel iseseisvad sõnad muutuvad grammatilisteks markeriteks. [---] Arvatakse, et grammatikaliseerumine toimub kahe protsessi – reanalüüsi ja analoogia – koosmõju tulemusel. (Ehala 1996: 381–382.)

Kuni kumbki neist protsessidest pole ületanud n-ö kriitilist piiri, võime rääkida funktsionaalsetest nihetest, mitte muutustest. Muustrite kvantitatiivsete nihetega võivad kaasned kvalitatiivsed lingvistilised nihked, sh morfosüntaktiliselt piiratud või kinnistunud muustrite leksikaalsemantilise varieerumise laienemine ja morfosüntaktiliselt varieeruvate muustrite leksikaalsemantiline kitsenemine ning kinnistumine.

Ajavahemikul 1890–1990 on adverb kahtlemata täienenud uue ainesega, sh kasutatuna liitsete üksuste koosseisus; ilmselgelt on avardunud ka adverbi tekstifunktsioonid. Seda kõike võib seostada laiemate protsessidega eesti keele arengus teel suurema analüütilisuse poole (vt Penjam 2005; Jürine & Habicht 2013). Loodan, et ka minu uurimistulemused toetavad seda seisukohta.

Sajandi vältel on toimunud kvantitatiivne nihe samuti pronoomeni- ja verbialgulise morfoloogilise klassi vahel, st suletud ja avatud sõnaliigi vahel: 1890ndatel on pronoomenialgulise klassi osakaal 16,9% ja verbialgulise klassi osakaal 15,6% (erinevus 1,3%), samas kui 1990ndatel on olukord vastupidine – vastavalt 14,3% ja 17% (erinevus 2,7%). Seda funktsionaalset nihet võib pidada reeglipäraseks. Pronoomenid on loendatavad, nende tekstifunktsioonid jäävad suhteliselt piiratuks (nt stiilikaalutlustel tehtud asendused), nagu ka grammatilised ja semantilised rollid (grammatiline subjekt või objekt, omaja või kogeja). Pronoomenite kasutus on seotud pragmaatiliste tähendustega (nt *keegi*, *miski* tähistavad ebamäärasust, umbisikulisust; *mina* isiksusekesksust). Verb aga täieneb uute analüütiliste üksustega, sel sõnaliigil on avar funktsionaalne võimekus (nt tegevuslaadi ja aspektilisuse täpsustamine tekstis).

Pronoomeni/verbi kui suletud/avatud sõnaliigi tekstikasutuses ilmnenud nihkeid saab seletada ilmselt tekstifunktsioonidega, sest nende kasutus sõltub esituslaadist ja terviktekstide loomise kommunikatiivsetest vajadustest. Kui pronoomeneid kasutatakse nt stiilikaalutlustel ja viiteseose vahendina, siis verbil on tekstiloomes veelgi olulisemaid funktsioone: erinevate grammatiliste ajavormide samalaadne või vahelduv kasutamine ning ajaplaanide kombineerimine võimaldab mitte ainult lauseid loogiliselt siduda, vaid ka koostada eri liiki tekste, mille esituslaad on valdavalt kirjeldav (tegevuste samaaegsus, seisundid), jutustav (ajaplaanide vaheldumine) või arutlev (erinevate ajaplaanide kombineerimine). Ilukirjandustekstides on need kolm esituslaadi omavahel põimunud ja sulandunud tervikuks. Sellest mõttekäigust tuleneb siinse uurimuse teine seisukoht. Väidan, et sajandi jooksul toimunud kvantitatiivset nihet pronoomeni- ja verbialguliste morfoloogiliste klasside vahel saab seletada

tekstiloomes seisukohalt. Muutunud on mitte ainult see, millist sisu ilukirjandus lugejale pakub, vaid ka see, kuidas sisu on esitatud ning kas ja millistel pronoomenitel on seejuures keskne roll (personaalsuse/impersonaalsuse või üldisiku eelistamine).

Juba kirjeldatud kahele olulisemale keelekasutusnihele, kus kaks morfoloogilist klassi vastastikku teineteist asendavad, leidub nihkeid ka sama morfoloogilise klassi esinemuses. Seda on näha konjunktsiooni-, adjektiiv-, adpositsiooni- ja numeraalialgulise klassi puhul. Konjunktsiooni esinemus 1990ndate keelekasutusmuustrites on laienenud 3,6% (1890ndatel 9,9% ja 1990ndatel 13,5%) ja adjektiivialgulise morfoloogilise klassi kasutusala 1% võrra (vastavalt 6,3% ja 7,3%). Seevastu adpositsiooni- ja numeraalialguliste klasside osakaal on 1990ndate ilukirjanduskeele muustrites vähenenud – adpositsioonil poolteist korda ja numeraalil poole võrra (vastavalt 0,7% vs. 1,7% ja 1,4% vs. 3,2%), vt joonised 3 ja 4.

Ka konjunktsiooni, adjektiivi, adpositsiooni ja numeraali kasutuses ilmenud kvantitatiivsetel muutustel on oma seletus, mille algeid vaatlen eesti kirjakeele kujunemisloos kontekstis ja seostan keelelistest taotlustest tulenevate valikutega, millele andis tõuke nooreestlaste pürgimus väljendusrikkama eesti keele poole, et eesti keel ei jääks alla suurtele kultuurkeeltele. Selle jätkuks oli Johannes Aaviku keeleuuenduslik tegevus. Nende sotsiaal-kultuuriliste protsesside mõju avaldub enim ilukirjandustekstides. Siit kolmas seisukoht: väidan, et pürgimus väljendusrikkama eesti keele suunas kajastub ka ilukirjanduskeele muustrites esile tulnud kvantitatiivsetes ja kvalitatiivsetes nihetes ning on kooskõlas eesti keele loomuliku arenguga. (Vt ka Laanekask 2004: 36.)

Seda mõtet toetab adjektiivi kui avatud sõnaliigi laienemine 1990ndate muustrites ja adjektiivse sõnavara täienemine sellele omastes tekstifunktsioonides (nt eestäiend), mis seostub tavaliselt kirjeldava esituslaadi vajadustega. Ka verbi kasutusala laienemist vaatlesin eespool seoses teksti esituslaadiga ning konjunktsiooni kui suletud sõnaliigi suurem esinemus 1990ndate aastate ilukirjanduskeele muustrites võrreldes 1890ndatega võib olla seletatav pürgimusega väljendusriikka kultuurkeele suunas, vajadusega sõnastada keerulisi mõttekäike ja abstraktseid arutlusi. Mõttekäigu keerukus on sünkroonne teksti süntaktiliste seoste keerukusega, mida tõendab ka näiteks rinnastav-alistavate rühmsidendite kui analüütiliste üksuste laienemise 1990ndate ilukirjanduskeeles (vt Eslon 2017). Sellele kaasub küsimus konjunktsioonist kui suletud sõnaliigist, kuhu niisama lihtsalt uusi liitseid üksuseid ei tohiks lisanduda, ent tekstiloomes tuleb neid adverbide, kivilinenud substantiivivormide, pronoomenite ja konjunktsioonide kombineerimisel järjest juurde. Mis tingimustel, seda peaks näitama 1890–1990 keelekasutusmuustrite üksikasjalik analüüs trigrammide tasandil.

Ning lõppeks, ka numeraali kui suletud sõnaliigi kasutusala kitsenemine 1990ndate ilukirjanduskeeles muutrites on arvatavasti esile tulnud keeleuuen-duslikel põhjustel, millest eespool juttu. Ilukirjanduslik esituslaad ei vaja arve ega numbreid, nende esinemus on seotud teema, tekstiliigi ja -pragmaa-tikaga. Näiteks *üks*, päritolult põhiarv, esineb tekstis sageli ülekantud tähenduses, markerides umbisikulisust või ebamäärasust: *oli üks pikk poiss must king jalas* 'keegi pikka kasvu poiss mustades kingades'; saksa keeles *Da war ein Mann, wer ...* 'siin oli keegi mees, kes...'; inglise keeles *One man* 'üks mees, st keegi'. Selles funktsioonis paigutub põhiarv *üks* indefiniitpronoomenite alla. Analoogsed nähtused on keeleülesed (vt Aikhenvald & Dixon 2007: 1–66).

Mis puudutab sajandi jooksul tekkinud nihet adpositsiooni kui suletud sõnaliigi kasutuses, siis selle seletus peitub eesti keele sisemises arengus. Kuigi tegu on saksa keele mõjul tulnud nähtusega, on adpositsioonid eesti keeles juurdunud sõnaliik, nende hulk on loendatav, neid kasutatakse enamasti koha- ja ajamarkeritena kas ainult substantiivi ees, järel või mõlemas positsioonis (vt Palmeos 1982). Adpositsioonide esinemuse vähenemine näitab, et selle suletud ja piiritletud funktsioonidega sõnaliigi kasutus on eesti keeles välja kujunenud ja omaks võetud.

2.2. Substantiivi- ja adverbialgulise klassi alamklassid ja klastrid

Arutelu jätkan substantiivi- ja adverbialgulise morfoloogilise klassi keeleka-sutusmuutrite võrdleva analüüsiga alamklasside ja klastrite tasandil, tuues näiteid sagedamate trigrammidest. Loodan leida seletusi eespool esitatud seisukohtadele ja kirjeldada sajandi jooksul toimunud kvantitatiivsete nihete lingvistilist olemust. Keskendun substantiivi ja adverbi kasutusnihele. Ees-pool väitsin, et adverb kui avatud sõnaliik on vahemikus 1890–1990 täienenud uue ainesega, teisenenud on adverbi tekstifunktsioonid. Seostan seda eesti keele avaramate sisemiste arengutega teel suurema analüütilisuse poole ja otsin vastuseid, millised nihked sõnaliikide kooskasutuses adverbi sisaldavaid mustreid iseloomustavad ning millistes tekstifunktsioonides on neid 1890ndatel ja 1990ndatel eelistatud kasutada.

2.2.1. Substantiivialgulise klassi alamklassid ja sagedamad

klastrid

1890ndate ilukirjanduskeeles on substantiivialgulises klassis kaheksa alam-klassi (vt tabel 1), neist suurima esinemusega substantiiv-adverb (S-D), nt *istme pääle maha* (31,0%).

Tabel 1. Substantiivialguliste alamklasside võrdlus 1890–1990.

	Alam- klass	S-D	S-V	S-S	S-K	S-A	S-J	S-P	S-N
1890ndad	Osakaal	31,0%	23,9%	19,7%	8,6%	5,5%	5,5%	5,1%	0,7%
1990ndad	klassis	33,3%	22,5%	24,0%	5,5%	5,2%	7,1%	2,4%	0,1%

Järgnevad substantiiv-verb (S-V), nt *ratast ära kiskuda* (23,9%), ja substantiiv-substantiiv (S-S), nt *abielu vanade eestlaste* (19,7%). Väiksema osakaaluga alamklassid on substantiiv-adpositsioon (S-K), nt *käe südame vastu* (8,6%), substantiiv-adjektiiv (S-A), nt *taevas nii sinine* (5,5%), substantiiv-konjunktsioon (S-J), nt *silmad maha ja* (5,5%), substantiiv-pronoomen (S-P), nt *sugulaste juurde oma* (5,1%) ja substantiiv-numeraal (S-N), nt *kord nõuuks 12* (0,7%).

1990ndatel on substantiivialgulise klassi osakaal tervikuna kahanenud, laiinenud on kolme alamklassi esinemus: substantiiv-adverb (S-D), nt *mees lõpuks välja*, substantiiv-substantiiv (S-S), nt *jutt või luuletus*, ja substantiiv-konjunktsioon (S-J), nt *käed ära ja*. Nii on alamklassi S-S osakaal suurenenud 4,3% (19,7%-lt 24,0%-ni), S-D osakaal 2,3% (31%-lt 33,3%-ni) ja S-J osakaal 1,6% (5,5%-lt 7,1%-ni). Ülejäänud viie alamklassi (adpositsiooni-, pronoomeni-, verbi-, numeraali- ja adjektiivilõpuline) esinemus näitab kahanemise tendentsi: S-K 3,1%, S-P 2,7%, S-V 1,4%, S-N 0,6% ja S-A 0,3% võrra (vt tabel 1).

Järgnevalt kirjeldan täpsemalt, milliste ilukirjanduskeele kasutusmustritega on nihked alamklasside esinemuses kõige sagedamini seotud.

1) Alamklass substantiiv-adverb (S-D)

1890ndate keelekasutuse analüüs tõi alamklassis S-D esile kaheksa morfoloogilist klastrit (vt lisa 1): SKD (40,5%), SVD (25,9%), SDD (18,6%), SSD (11,2%), SPD (1,9%), SAD (1,1%), SJD (0,5%) ja SND (0,3%). Neist kõige laialdasema kasutusega on **SKD**, mille sagedamad trigrammid algavad substantiiviga ainsuse genitiivis, järgnevad adpositsioon ja adverb, nt *<hüppas> laua pealt üles*, *<laskus> ülema ette põlvili*. Adpositsioonidest esinevad enim kohatähenduslikud *peal*, *ees* ja *sees*, adverbidest on korduvalt kasutatud verbipartikleid *tagasi*, *kinni*, *maha*.

1990ndate keelekasutuses sisaldab alamklass S-D kuut morfoloogilist klastrit (vt lisa 1): SVD (58,1%), SDD (18,1%), SKD (11,0%), SSD (7,3%), SJD (3,9%), SPD (1,6%). SKD esinemus on sajandi vältel märgatavalt kahanenud (40,5%-lt 11%-le, vt tabel 1, lisa 1). See erinevus on statistiliselt oluline (LL +/6,25). Suurima esinemusega on SVD (58,1%), mille keskmine komponent on öeldisverb indikatiivis imperfekti ainsuse 3. pöördes, nt *töö edenes vaevaliselt*.

Eespool seostasin adpositsiooni kui suletud sõnaliigi osakaalu vähenemist selle kasutusmallide kinnistumisega eesti keeles. Vajalik eeltingimus on paralleelsete võimaluste olemasolu sama tähenduse edastamisel – saab kasutada nii adpositsioonifraasi kui ka käändetunnust. Viimase tähendus on palju abstraktsem ja kasutusala seetõttu avaram (EKK: 191; vt ka Kaalep *et al.* 2000: 628–629). Kuigi Riho Grünthal (2003: 111) on väitnud, et eesti keelele on omane käändetunnuse ja tagasõnaga fraaside lõimumine ning soome keelele segregatsioon, näitab ilukirjanduskeeles kasutusmuustrite võrdlus 1890–1990, et sajandi jooksul on toimunud statistiliselt oluline kvantitatiivne nihe adpositsioonide hulga ja kasutusmallide kinnistumises. Niisama lihtsalt neid juurde ei tule, ka ei saa niisama lihtsalt adpositsioonifraasi asendada substantiivi käändetunnusega ning vastupidi (nt *laua taga, maja kohal, arvuti ümber*). Adpositsioonifraasi kasutusala kitseneb ja muutub spetsiifilisemaks.

Alamklassi S-D kõige levinum klaster 1990ndate ilukirjanduskeeles on **SVD**, mida ka 1890ndatel on aktiivselt kasutatud (esinemuselt SKD järel teisel kohal). SVD sagedamad trigrammid algavad substantiiviga ainsuse nominatiivis, verb on indikatiivi imperfekti ainsuse 3. pöördes, adverbil on erinevaid funktsioone, nt *kahin tõusis nüüüd* (ajamäärus), *õue vaadates vaiksel* (viisimäärus). Klasteri substantiivne sõnavara on rikkalik, verbil piiratum. Korduma jääb substantiiv subjekti, *olema*-verbi imperfekt koopula ja adverb öeldistäite funktsioonis, nt *neiu oli vait*. Erinevust SVD esinemuses 1890ndatel ja 1990ndatel võib pidada väga oluliseks (LL -/615,46), klaster esineb 1990ndatel olulisel määral rohkem kui 1890ndatel (vt lisa 3).

Substantiivi, eriti aga adverbi ja verbi kui avatud sõnaliikide funktsionaalne võimekus tuleb esile tekstiloomes (nt verbi ajavormide roll erineva esituslaadiga tekstilistes terviküksustes või adverbide polüfunktsionaalsus ja analoogia grammatiliste morfeemidega). Seda võimekust toetavad eesti kirjakeele arengu aluseks olnud põhimõtted, samuti sajandi jooksul toimunud muutused väärtushinnangutes, mis kajastuvad ühiskondlikus teadvuses ja väljenduvad keele vahendusel. See on paratamatult leidnud kajastust ka ilukirjanduses. Alates 1980ndate perestroikast on ühiskondlikku teadvust enim mõjutanud ajakirjandus, mille keelekasutus on varasemaga võrreldes muutunud avatumaks, isiksusekessemaks ja stiililt värvikamaks. Mahuliselt on vähenenud kirjelduste ning suurenenud konkreetsete isikustatud tegevuste osakaal (vt Yelenevskaya 2008: 117).

Kas just tingituna nendest mittelingvistilistest asjaoludest, kuid ka eesti 1990ndate ilukirjanduskeeles on alamklassi S-D klasterkasutuses esiplaanil adverbi sisaldavad muustrid. Võrreldes 1890ndatega on suurenenud klasteri **SDD** esinemus. Kuigi osakaalu põhjal pole SDD kasutuses sajandi jooksul märkimisväärseid nihkeid (vastavalt 18,6% ja 18,1%), tuleb erinevust pidada siiski oluliseks, kuna LL on -/82,28.

SDD sagedam trigramm algab substantiiviga ainsuse nominatiivis subjekti funktsioonis, sellele järgnevad kaks adverb, nt *äär peaaegu juba*. Kuigi kõige sagedamini on kasutatud rõhuadverbe *koguni, nii ja ikka*, ei laiene see SDD-klastrile tervikuna, kuna kasutatud on erinevat päritolu adverbe, millel on tekstis erinevaid funktsioone, nt verbipartikkel *ära* (<*võtub*> *mees vist ära*), subjektiivmodaalne hinnangusõna (<*ei tule*> *ettepanek vist lihtsalt*), määruslik laiend (<*on*> *elu ju alati*).

Klastris **SSD** esinemus on 1890ndatel ja 1990ndatel oluliselt erinev (LL -/14,52), mis näitab veel kord, et substantiivialgulise morfoloogilise klassi keelekasutusnihetes mängib olulist rolli adverb. Sellele sõnaliigile omane funktsionaalne võimekus avaldub varieerumises, nt *poiste seadus veel* (ajamäärus), *meeste jutt mitte* (eituspartikkel), *laulu sõnad vist* (subjektiivmodaalne tõeväärtushinnang) jm. 1890ndatel esineb SSD puhul raskusi adverb kui sõnaliigi ja selle tekstifunktsioonide määramisel (vt ka Habicht *et al.* 2011), põhjuseks õigekirjutusreeglid, sõnajärje- ning reksioonimuutused jm, mistõttu esineb mitmeti tõlgendatavust.

SJD pole 1890ndate valimis statistiliselt oluline klaster, kuid võrdlus 1990ndatega (LL -/67) viitab, et on tekkinud oluline erinevus (nelja näite asemel 41), nt *koormus ja lihtsalt, kinos ja sealgi, idee või õigemini*. 1990ndatel esineb substantiiv nominatiivis, partitiivis ja inessiivis, sidesõnadest tarvitatakse nii subordineerivaid (*kui, nagu*) kui ka koordineerivaid (*ja, või*), varieeruvad adverbide liigid ja funktsioonid. Seega on 1990ndaks muutunud mitte ainult mustri esinemus, vaid ka selle morfosüntaktiline keerukus. Võib oletada, et see protsess jätkub.

2) Alamklass substantiiv-verb (S-V)

1890ndate keelekasutuse analüüs tõi alamklassis S-V esile üheksa klastrit (vt lisa 1): SSV (27,4%), SVV (27,2%), SKV (26,5%), SDV (9,9%), SPV (3,4%), SAV (2,4%), SJV (2,4%), SNV (0,5%) ja SXV (0,3%; X – verbi juurde kuuluv sõna, millel eraldi sõnaliigi tähistus puudub). Kuigi esimese kolme klastris esinemus on lähedane, kasutatakse kõige laiemalt **SSV**-d. Trigrammides on substantiiv tavaliselt ainsuse genitiivis eestäiendina, järgnevad substantiiv ainsuse nominatiivis subjektina ja öeldisverb indikatiivis imperfekti ainsuse 3. pöördes, nt *mõisa aidamees suri, Tagametsa perepoeg jahmatas*. Võrdlus 1990ndatega (vt ka lisa 1) näitab, et SSV esinemus on tunduvalt kitsenenud (27,4%-lt 13,4%-le, vahe 14,0%) – see vahe on statistiliselt oluline (LL -/2,0). Samas on 1990ndatel tunduvalt kasvanud klastris **SDV** osakaal (9,9%-lt 25,8%-ni, vahe 15,9%; LL -/185,77). Järelikult on SSV tunduv kahanemine ja SDV esilekerkimine 1990ndate ilukirjanduskeele mustrites märgiline. Ühelt poolt on see seotud

adverbi kui sõnaliigi kasutusala laienemisega, teisalt sisaldab SDV 1990ndatel enamasti ühendverbi mineviku liitaja (<oli> šokolaadi ära söönud) või da-infinitiivi vormis (<võid> ukse lahti teha). Niisiis on SDV laienemine eesti keeles seotud otseselt analüütilisuse suurenemisega. Tänapäevases keelekasutuses on SDV kinnistunud analüütiliste ühendverbide kasutus- ja moodustumallina (vt Eslon 2014a: 63).

Analoogselt on kinnistunud 1890ndatel ebaoluline **SJV**, mis 1990ndate ilukirjanduskeeles esineb osalausepiiril, keskmiseks komponendiks eranditult rinnastav sidesõna (enamasti *ja*), nt *kausi ja tõstis* (vt ka Trainis & Allkivi 2014: 290). See erinevus on statistiliselt oluline (LL -/134). Niisiis kinnistuvad 1990ndate keelekasutuses SDV ja SJV, mille osakaal 1890ndatel ei ole statistiliselt oluline. Selle kvantitatiivse nihkega kaasneb mõlema struktuuri morfosüntaktiline kinnistumine.

1890ndate ilukirjanduskeeles teine oluline klaster on **SVV**, milles substantiiv on tavaliselt ainsuse nominatiivis subjekti funktsioonis, järgnevad eitav abiverb ja finiidne intrantsitiivne põhiverb indikatiivi imperfektis, nt *süda ei ole*. Substantiivina varieeruvad üld- ja pärisnimed (nt kohanimed *Greekamaal*, *Prantsusemaal*). 1990ndatega võrreldes tuleb klasteri SVV kasutuses esile oluline erinevus (LL -/89,21; osakaal vastavalt 27,2% ja 29,7%, vahe 2,5%). Eitava kõne laienemine 1990ndate ilukirjanduskeeles muustrites võib olla tingitud isiksusekeskse arutlev-jutustava esituslaadi laienemisest. Joel Sang (1983: 24) seostab seda teksti pragmaatikaga:

eitavaid lausungeid kasutab kõneleja selleks, et kummutada vestluspartneri eelnevat ütlust või oletatavat arvamust. Seega eeldab eitus juba iseenesest (tegelikku või mõttelist) dialoogi.

Kolmas oluline klaster 1890ndate ilukirjanduskeeles muustrites on **SKV**. Selles sisalduvate trigrammide morfosüntaktiline dominant on adpositsioonifraas: substantiiv ainsuse genitiivis + adpositsioon. Ning kuigi verbi morfosüntaks varieerub (nt ainsuse ja mitmuse 3. pööre, finiidsed ja infiniidset vormid), järgneb adpositsioonile sagedamini siiski intransitiivne öeldisverb indikatiivis imperfekti ainsuse 3. pöördes, nt *silma ees seisis*. Saja aastaga on adpositsiooni sisaldava SKV esinemus ilukirjanduskeeles oluliselt kahanenud – vastavalt 26,5% ja 16,9%, vahe 9,6%. See erinevus on statistiliselt oluline (LL -/13,33), mis räägib adpositsiooni kui suletud sõnaliigi ja adpositsioonifraasi kinnistumise kasuks neile omases määruse funktsioonis (sama kirjeldasin eespool SKD kasutuses).

Kokkuvõttes toetavad ka alamklassi S-V klasterite kvantitatiivsed ja kvalitatiivsed nihked samu tendentse ilukirjanduskeeles arengus, millest eespool juttu: adverbi ja verbi osakaal laieneb, adpositsioonifraasi sisaldava mustri osakaal

kitseneb ja muutub spetsiifilisemaks, esile tõusevad mustrid, mis omased isikusekesksele arutlevale-jutustavale esituslaadile. 1990ndate keelekasutuses on vähenenud ka pronoomenit ja adjektiivis sisaldavate mustrite SPV ning SAV osakaal (LL vastavalt +/3,28 ja +/1,02); täielikult on taandunud 1890ndatel marginaalsed klastrid SNV ja SXV.

3) Alamklass substantiiv-substantiiv (S-S)

1890ndate tekstides koosneb alamklass S-S kaheksast klastrist (vt lisa 1): SJS (26,3%), SSS (16,3%), SKS (12,9%), SVS (12,5%), SAS (12,1%), SDS (9,1%), SPS (8,7%) ja SNS (2,1%). Suurima esinemusega on **SJS**, kus rinnastav sidesõna ühendab kahte substantiivi ainsuse nominatiivis. 1990ndateks on selle klasteri osakaal suurenenud 9,1% (26,3%-lt 35,4%-le; LL -/200,69), mis on oluline kvantitatiivne erinevus. Tegu on kinnistunud mustriga, milles substantiivid võivad olla semantiliselt lähedased (nt *kirin ja kärin, mõte ja meel*), vastandatud (nt *must ja valge, mets ja meri*) või kontekstisidusad teemasõnad (nt *kord ja kohus, mees ja hobune*).

Kuigi oluliselt teise klasteri **SSS** esinemus on 1890ndatel 16,3% (SJS-ga võrreldes tunduvalt kitsama levikuga), taandub see 1990ndateks 3%-le (vahe 13,3%). See on statistiliselt oluline kvantitatiivne nihe (LL +/5,14), muster on iseloomulik vaid 1890ndate ilukirjanduskeelele, nt *kindrali mütsiga mehe <kõrval>*. Kolmanda komponendi käänne varieerub: ainsuse nominatiiv (*Parisi linna valitsus*), genitiiv (*mamma hääl kopli <taha>*), partitiiv (*pool taeva-lae äärt*), illatiiv (*sissetulekust riigi kassasse*), mitmuse adesiiv (*Athena linnas pulmadel*), elatiiv (*hääl kopli võsadest*) ja essiiv (*kohuseid riigi kodanikkudena*). 1890ndatel sisaldab klaster SSS mitmeid kohanimesid, nt *Gusti loigu kaldal, Viljandi-maale Karksi mõisa*.

Alamklassi S-S ülejäänud klastrite (SKS, SVS ja SAS) esinemus on 1890ndatel lähedane. 1990ndateks on neist enim laienenud adjektiivis sisaldav muster **SAS**, nt *<ka> abielu vanade eestlaste <juures>* (vastavalt 12,1% ja 21,9%, vahe 9,8%, LL -/157,89), veidi vähem verbi sisaldav **SVS**, nt *<kümme vana> poissi võtavad naesed* (vastavalt 12,5% ja 17,8%, vahe 5,3%, LL -/34,48). Kitsenenud on adpositsioonifraasi sisaldava klasteri **SKS** esinemus (vastavalt 12,9% ja 10,8%, vahe 2,1%, LL -/34,48), nt *ukse taha põrandale*, sest adpositsioonifraasi kasutus on koondunud SKD ja SKV ümber (vt eespool). Seega kinnitavad siinsed andmed samu tendentse avatud ja suletud sõnaliikide kasutuses, mida kirjeldasin morfoloogiliste klasside põhjal.

4) Alamklass substantiiv-adpositsioon (S-K)

Alamklassis S-K on 1890ndate ilukirjanduskeelee andmetel neli klastrit (vt lisa 1): SSK (84,4%), SPK (9,9%), SVK (3,8%), SDK (1,9%). Alamklassi dominantne klaster **SSK** (<pani> *käe südamet vastu*) säilitab staatuse ka 1990ndatel (vastavalt 84,4% ja 86,6%, vahe 2,2%), kuigi erinevus on oluline (LL -/22,02) ja viitab adpositsioonifraasi kasutuse laienemisele. Klastrite SPK (*poisid enestest seast*) ja SVK (*neiud ronivad üle*) esinemus on võrreldavatel perioodidel suhteliselt ühtlane; SDK (*seadus veel pärast*) kaob 1990ndatel kasutusest; uue muustrina tuleb esile lausealguline SNK (*Kella viie ajal, Kella kahe paiku*). Kirjeldatud tendentsid adpositsiooni sisaldavate muustrite kasutuses tõendavad veel kord, et 1990ndateks on adpositsioonifraasi kasutusala kitsenenud ja muutunud spetsiifilisemaks.

2.2.2. Adverbialgulise klassi alamklassid ja sagedamad klastrid

1990ndate ilukirjanduskeeles on adverbialgulise klassi muustrid saanud avarama kasutuse ja substantiivialguline klass on taandunud teisele positsioonile. 1890ndatel on adverbialgulise klassi kõige sagedam alamklass adverb-adverb ehk D-D (25,9%), nt <kandis> *väga tähtsalt ette*, järgnevad adverb-verb ehk D-V (23,9%), nt *nii ilus on*, ja adverb-substantiiv ehk D-S (21,4%), nt <võttis> *ometi naese kaenlasse* (vt tabel 2).

Tabel 2. Adverbialguliste alamklasside võrdlus 1890–1990.

	Alamklass	D-D	D-V	D-S	D-K	D-P	D-J	D-A	D-N
1890ndad	Osakaal	25,9%	23,9%	21,4%	10,9%	7,4%	5,6%	4,4%	0,5%
1990ndad	klassis	19,1%	25,4%	29,4%	4,4%	7,5%	6,5%	7,5%	0,2%

1990ndatel on seis vastupidine – suurema esinemusega alamklass on D-S (29,4%), nt <oli> *nii selgesti meeles*, järgnevad D-V (25,4%), nt *ennevanasti on olnud*, ja D-D (19,1%), nt *seejärel silmitseb mõtlikult*. 1990ndateks on kitsenenud alamklasside D-S (8% võrra) ja D-D (6,8%) esinemus ning laienenud D-V (1,5%) kasutus. Väiksema osakaaluga alamklasside seas on suurim nihe alamklassi D-K (<naeris> *mõnikord tema üle*) esinemuses (kahanenud 6,5%) ning alamklassi D-A (*nagu uus hiilgav*) kasutuses (suurenenud 3,1%). Alamklasside D-P (<tahtis> *veel enam oma*), D-J (<tuli> *siis jälle ja*) ning D-N (*praegu kaks miljonit*) kasutus on jäänud ühtmoodi marginaalseks.

Niisiis on 1990ndatel laienenud adverbialgulistes alamklassides nende muustrite esinemus, mis lõpevad substantiivi ja verbiga, vähemal määral ka adjektiiviga, samas kui adverbi ja adpositsiooniga lõppevate muustrite osakaal

kahaneb. See on üldiselt kooskõlas sõnaliikide üldiste kasutustendentsidega, mis tulid esile morfoloogiliste klasside ja nende alamklasside alusel.

1) Alamklass adverb-adverb (D-D)

Alamklass D-D koosneb mõlemas valimis kuuest klastrist (vt lisa 2), mille esinamus 1890ndate ilukirjanduskeele mustrites on järgmine: DDD – 68,4%, DJD – 8,7%, DPD – 8,6%, DVD – 8,6%, DSD – 5,4% ja DAD – 0,3%. Domineerib eri liiki määrustest koosnev muster **DDD**, nt <tulevad> *siiapoolle nagu ka, <läks> kaugemale eest ära, <tundus> esiotsa küll natukene*. Klaster iseloomustavad adverbide avar funktsionaalne võimekus ja suur leksikaalne varieeruvus.

1990ndate ilukirjanduskeeles on DDD samuti suurima osakaaluga klaster, kuid võrreldes 1890ndatega on selle esinamus siiski tunduvalt kahanenud – lausa 28,4% (68,4%-lt 40%-le). See võib olla seotud muutustega sõnajärjes, kuigi kolm järjestikust adverbi on eesti keele sõnajärjele üldiselt omane (vt Matsak *et al.* 2010). Kõikide ülejäänud klastrite osakaal alamklassis on aga kasvanud: verbi sisaldav DVD 10,3% (8,6%-lt 18,9%-le), substantiivi sisaldav DSD 8,3% (5,4%-lt 13,7%-le), konjunktsiooni sisaldav DJD 4,7% (8,7%-lt 13,4%-le), pronoomenit sisaldav DPD 4,5% (8,6%-lt 13,1%-le) ning adjektiivi sisaldav DAD 0,6% (0,3%-lt 0,9%-le). Log-tõepära funktsiooni alusel pole vahe DDD esinemuses statistiliselt eriti märkimisväärne (LL -/17,42), hoopis olulisemaid nihkeid on toimunud DVD (LL -/110,98), DSD (LL -/90,04), DJD (LL -/56,85) ja DPD (LL -/55,44) puhul, mille kasutusala on 1990ndatel laienenud koos adverbide esinemuse suurenemisega. Muutunud on ka struktuurid, mille koosseisus adverbe eelistatakse. Niisiis on sajandi jooksul lisaks kvantitatiivsetele toimunud ka kvalitatiivseid nihkeid, nt adverbi sisaldavate klastrite alusel tekib uusi vabu ning seotud ühendeid, mis viitab analüütilisuse laienemisele:

tegu on kinnistunud [---] struktuuridega, mille põhjal saab moodustada praktiliselt piiramatul hulgal mitmesõnalisi kooslusi, millest aja möödudes võivad kujuneda idiomaatilised üksused [---] (Eslon 2014b: 20).

DDD trigrammides on näha ühelt poolt suurt leksikaalset ja funktsionaalset varieeruvust, teisalt aga on 1990ndate keelekasutuses rõhuadverbide, verbi- ja eituspartiklite kõrval esile tulnud ka (potentsiaalseid) liitseid adverbilisi üksuseid nagu <tõstis> *kaugemale eest ära, <ei jäänud> sugugi enam rahule, <vaatas> juba eemalt vastu*, mis on märk suurenevast analüütilisusest. Ka see kvalitatiivne nihe viitab analüütilisuse laienemisele. Adverbidega rõhutatakse (*veel nii kaugel*), tuuakse esile (*pealegi veel nii*), seotakse omavahel lauseid (*Kas mitte sellepärast*) ja osalauseid. Kõige enam korduvad *kuidas, nüüd, mitte ja võib-olla*.

DJD kasutuserinevus on kvantitatiivne (1990ndatel esineb 4,7% võrra rohkem). 1890ndatel seob rinnastav sidesõna (tavaliselt *ja*) kahte viisimäärust (*pik-kamisi ja tähendusrikkalt, igatsevalt ja õrnalt*) ning kohamäärustena esinevaid proadverbe (*siin ja seal*). Ka 1990ndatel kasutatakse selles mustris peamiselt rinnastusseost ja viisimäärusi (*tõiselt ja ka, üliviiisakalt ja korralekutsuvalt*).

DPD alguses on 1890ndatel tavaliselt *kus, kust, kas, eks, vist* ning lõpus *küll, nüüd, sinna, siis* (nt *kus ennast nüüd, eks mind siis*). 1990ndatel avaldub selle mustris kasutuses kvalitatiivne nihe, mille tulemuseks on tugiverbiühendid. Sel juhul on vähemalt üks DPD adverbidest partikli funktsioonis ja kuulub tugiverbiühendisse, põhiverb asub vasakus või paremas kontekstis, nt *<läks> täiesti endast välja, <oli> nõus mind ära <viima>*. Ka see on nihe suurema analüütilisuse poole, sest tugiverbiühendid on analüütilised konstruktsioonid.

Tähelepanu köidab samuti kahte adverbi ja verbi sisaldav **DVD** (*siis on hilja, Küll on hiljuti*), mis on sagedamini vormistatud finiiitverbiga indikatiivi oleviku ainsuse 3. pöördes. 1890ndate keelekasutuses algab DVD ajamäärustega *siis, nüüd, homme* ning lõpeb rõhuadverbidega *ka, jälle, veel, ikka* (üksikjuhtumitel viisimäärustega *aralt ja raskesti*). Ka selle klatri kasutuses on sajandi jooksul toimunud kvalitatiivne nihe – 1990ndatel on rõhuadverbid asendunud *-ki (-gi)* liitega, nt *nii tuli ka <tema> ~ nii tuli temagi; varsti algab ju <kevad> ~ kevadki algab varsti*. Samuti kasutatakse DVD alusel analüütilisi ühendverbe, nt *raskelt elas üle, kuipalju kõneleb kaasa*, kus verbipartikkel asub põhisõna järel.

Klattris **DSD** esineb substantiiv nii 1890ndatel kui ka 1990ndatel subjekti (*<tulevad> ehk mõisasaksad siit*), objekti (*<tõmbas> rahulikult sukkepükse üles*) ning adverbialina (*<oli> kauem osakonnas kinni*). Viimasel juhul eristuvad fraasiadverbialid, kus substantiiv on adverbi laiendsõna (*<astus> praegu uksest välja, <said> ilusasti jõest üle*).

2) Alamklass adverb-verb (D-V)

1890ndate alamklass D-V koosneb seitsmest klattrist (vt lisa 2): DDV (33,4%), DVV (28,8%), DJV (18,7%), DSV (8,9%), DAV (5,0%), DPV (4,8%) ja DNV (0,4%). Suurima esinemusega **DDV** algab kahe adverbiga, järgneb infiniitne sihiline verb objekti funktsioonis, nt *täiesti selgesti lugeda*. Adverbide funktsionaalne võimekus ja leksikaalne varieeruvus on suur, neid kasutatakse peamiselt rõhusõna (*ometi ka lugeda*), verbipartikli (*siit ära saab*) ja subjektiivmodaalse hinnangusõnaga (*vist küll ei*); nendega väljendatakse määra (*igati arusaadavalt kõneleda*) ja aega (*Liiga hilja oli*); rõhuadverb laiendab järgnevat adverbi, moodustades sellega fonoloogilise ja semantilise terviku ehk potentsiaalse liitadverbi, mis on analüütilisuse ilming. DDV-s varieeruvad ka verbi vormid ning funktsioonid: kasutatud on finiiitõeldisena imperfekti ainsuse 3. pööret (*nii kaua vingerdas, jälle edasi tormas*), da-infinitiivi (*kaugele eemale pugeda*)

ja *ma*-infinitiivi (*üksinda valmis saama*). Kõneliik on valdavalt jaatav, kuid esineb ka eitavat kõnet (*veel koguni ei <olnud>*). 1990ndatel on DDV esinemus oluliselt suurem kui 1890ndatel, see kvantitatiivne erinevus on statistiliselt väga oluline (LL -/252,72). Esile tulevad analüütilised ühendverbid nagu *nii kokku lepitud, lõpuks läbi peksnud, otsustavalt vahele astuda*.

DVV on kasutusel eitavas kõnes, nt *edasi ei tule*. 1890ndatel on DVV esinemus 4% võrra suurem kui 1990ndatel (vastavalt 28,2% ja 24,8%). See kvantitatiivne erinevus on statistiliselt oluline (LL -/107, 74). Samalaadne nihe tuli eitava kõne puhul välja ka klastriga SVV (vt alamklass S-V). Sidusin 1990ndatel SVV eitava kõne eelistamise isiksusekeskse arutlev-jutustava esituslaadiga. Lisaksin, et eesti keele õppija valib oma kirjalikes töödes tavaliselt eitava kõne ilmselt oskamatusesest või ebakindlusest luua arutlev-jutustavas esituslaadis sidusaid ja loogiliselt üles ehitatud tekste. Kas 1890ndate keelekasutus kajastab midagi analoogset, sellele sinne analüüs otsest vastust ei paku. Ilmselt kujunesid alles 1990ndateks erinevatele esituslaadidele omased keelekasutusstandardid.

DJV algab adverbiga, järgnevad rinnastav sidesõna ja sihiline põhiverb indikatiivi imperfektis ainsuse (harvem mitmuse) 3. pöördes, nt *ärkas> üles ja hüüdis, <istusid> kokku ja sosistasivad*. Selle mustri kasutusala on 1990ndateks ahenenud (18,7%-lt 12,6%-le, vahe 6,1%), ühtegi imperfekti mitmuse vormi ei esine, ainult ainsus. Tegu on olulise kvantitatiivse erinevusega (LL -/36,22), millega kaasneb kvalitatiivne nihe – verbi morfosüntaktiline kinnistumine.

DSV-l (nt *salamahti nõuu pidas, rohkesti lapsi on*) poleks vajadust peatuda, kuna optimaalsuskoeffitsient jääb alla kriitilise piiri, kuid LL -/74,94 viitab samas olulisele kvantitatiivsele erinevusele võrreldes 1990ndatega (analoogia DJV-ga). 1890ndate keelekasutuses on DSV funktsionaalselt ja vormiliselt mitmekesine. Esineb viisi- (*rohkesti, salamahti*), aja- (*täna*), hulgamäärusi (*palju*) jne; substantiiv on partiivis (*palju raamatuid oled*), abessiivis (*nõnda piirita ei*), aditiivis (*siis kosja ei*), inessiivis (*Kusagil maailmas ei*), nominatiivis (*kus õnn on*), aditiivis (*Ka Muhumaal teati*). Verbi kasutatakse jaatava ja eitava kõne (eitav abiverb *ei*) preesensis ning imperfektis. Seevastu 1990ndate keelekasutuses on DSV kinnistunud lause alguses reeglina eitavas kõnes: adverb, isikunimi ainsuse nominatiivis subjekti funktsioonis ja eitus, nt *Võib-olla Aune ei <tulegi>*. Niisiis kaasneb kvantitatiivse nihkega ka kvalitatiivne.

DPV pole samuti 1890ndatel statistiliselt oluline, kuid võrdlus 1990ndatega näitab olulist erinevust (LL -/65,56). Nii 1890ndatel kui ka 1990ndatel on sagedamad need trigrammid, kus adverb asub lause alguses, kuid 1890ndatel järgneb sellele ainsuse 2. isiku ja 1990ndatel ainsuse 3. isiku personaalpronoomen, nt *Miks sa ei; Kas ta ei*. Kvalitatiivne nihe seisneb vahetu *mina-sina* suhtluse asendumises rääkimisega kellestki kolmandast (*temast*).

3) Alamklass adverb-substantiiv (D-S)

Alamklass D-S koosneb üheksast klastrist, mille esinemus 1890ndate keelekasutuses on järgmine: DAS – 24,2%, DDS – 22,7%, DPS – 18,6%, DSS – 13,9%, DVS – 10,6%, DNS – 4,8%, DJS – 3,5%, DKS – 1,3% ning DGS – 0,4% (G – genitiivatribuut, käändumatu eestäänd), vt lisa 2. 1890ndate keelekasutuses on olulisim **DAS**, mille iseloomulikuks tunnuseks avar leksikaalne ja morfosüntaktiline varieerumine. Esile tuleb kaks sagedamat mustrit: a) adverb, adjektiiv genitiivse eestäändina ja tagasõnafraasi kuuluv substantiiv ainsuse genitiivis, nt *teravamalt liikuva ilmapildi <taga>*; b) adverb, adjektiiv ainsuse partitiivis eestäändina, substantiiv ainsuse partitiivis objekti funktsioonis, nt *<on> kaua rasket tööd <teinud>*. Silma hakkab substantiivi käändevalik: genitiivi ja partitiivi kõrval esineb nominatiivi (*juba suur laps*), komitatiivset ja aditiivset substantiivi on kasutatud viisimäärusena (*ka nähtava uhkusega; nii kentsakal toonil*).

1990ndate ilukirjanduskeeles on klatri DAS osakaal kasvanud 9,5% (24,2%-lt 33,7%-le), erinevus on statistiliselt oluline (LL -/343,18). Ilmselt on kvantitatiivse nihke põhjustanud erinevat liiki adverbiaalide rikkus, mis 1990ndate keelekasutuses esile küünib. Adjektiivne eestäänd sulandub käändelise aja-, koha-, viisi-, seisundi- ja hulgamäärusega (*nüüd kibeda irooniaga, kõrgemal seisvate inimeste, monotoonselt kulgev abielu, kikiarvule tõusnud poisike, palju priskemat naist*), laiendades DAS-struktuurist vasakul paiknevat öeldisverbi (*<lausus> nüüd kibeda irooniaga*). Ilmselt ilukirjandusliku esituslaadi keelelise mitmekesistamise tulemusel on 1990ndatel kasvanud semantiliste käänete kasutamine erinevates adverbiaalsetes funktsioonides, kuid see seisukoht vajab tõestamist.

DDS sisaldab 1990ndate keelekasutuses tavaliselt substantiivi ainsuse partitiivis, nt *isegi küllalt valgust* (sh väljendverbid nagu *<andis> veel enam hoogu, <sai> aeg-ajalt isegi aru*). 1890ndate keelekasutuses esineb substantiiv ka nominatiivis (*siis ainult teadus <saab>*) ja genitiivis (*vast veel Ahvrika <kõrbes>*). Sajandi jooksul on substantiivi morfosüntaktiline varieerumine asendunud kinnistunud vormikasutusega (ainsuse partitiiv). Seetõttu on ka DDS-i esinemus 1990ndateks kahanenud 6,2% (22,7%-lt 16,5%-le). Adverbide kasutus on jäänud samaks: esimene tavaliselt ajamäärus (sagedamad *siis, nüüd, praegu, homme*, nt *praegu veel aega*) või rõhusõna *nii (nii palju raamatuid)*, teine viisimäärus (*<tegi> nüüd tihemini haiget*). Statistiliselt olulise kvantitatiivse erinevuse taga (LL -/93,72) peitub vormikasutuse kinnistumisega seotud kvalitatiivne nihe – laieneb DDS-i alusel moodustatud analüütiliste väljendverbide kasutamine (areng suurema analüütilisuse poole).

DPS on struktuur, kus adverbile järgnevad genitiivse eestäiendina ainsuse 3. isiku personaalpronoomen *ta* ja substantiiv ainsuse nominatiivis subjekti funktsioonis, nt *mil ta elupõhi*. Harvem esineb muster, mis sisaldab adverbi järel determinatiivpronoomenit ainsuse genitiivis ning tagasõnafraasi kuuluvat substantiivi ainsuse genitiivis, nt *surnult oma vere <sisse langes>*. Ka seda klastrit esineb 1990ndatel vähem kui 1890ndatel (vahe 2,8%, vastavalt 18,6% ja 15,8%). Erinevus on statistiliselt oluline (LL -/73,63). Kvalitatiivseid nihkeid vormi ja funktsiooni tasandil pole märgata.

DSS-i kasutuses on 1990ndateks toimunud oluline nihe (13,9%-lt 8,2%le, vahe 5,7%, LL -/34,90). 1890ndatel algab klatri sagedam muster ajamäärusega, järgnev substantiiv ainsuse genitiivis on eestäiend ja teine substantiiv ainsuse partitiivis on objekti funktsioonis, nt *siis poisi rinnust <kinni krabas>*. Substantiivi (eriti viimase) kääne ja seega süntaktilised funktsioonid varieeruvad, nt *jälle õnnetusest õnnele, palju tulesi kirikus, jälle seaduse võimu*. 1990ndateks on toimunud kvalitatiivne nihe, sest viimane substantiiv ainsuse nominatiivis on kinnistunud subjekti funktsioonis, nt *vahest armastuse puudus, lausa kulla väärtus, maandatult vägivalla märk*. Tegu on samalaadse kinnistumisprotsessiga, millest eespool korduvalt juttu (nt DDS). Põhjendusena võiksid kõne alla tulla ka verbirektsiooni muutused, kuid see vajab tõestamist.

DVS on 1890ndatel alamklassis D-S esinemuselt viies. Võrreldes 1990ndatega on vahe selle klatri kasutuses statistiliselt oluline (LL -/136,23). 1890ndate ilukirjanduskeeles alustab DVS lauset, morfosüntaktiline varieerumine on piiratud: adverb + verb indikatiivi imperfekti ainsuse 3. pöördes + substantiiv ainsuse genitiivis eestäiendina, nt *Äkitselt kõlas koera <ulg>*. 1990ndatel on toimunud kvalitatiivne nihe, mis seotud substantiivi kinnistumisega ainsuse nominatiivis subjekti funktsioonis, nt *Siis helises uksekell, Arvatavasti oli märtsikuu*. Vastukaaluks morfosüntaktilisele kinnistumisele on rikkalikumaks muutunud DVS-i leksikaalne varieerumine.

DNS-i esinamus on 1890ndate ja 1990ndate ilukirjanduskeeles erinev (LL -33,90). 1890ndatel on see leksikaalselt ja morfosüntaktiliselt varieeruv klaster. Adverbid esinevad peamiselt rõhusõna ja ajamääruse funktsioonis (*alles viis aastat, viimaks kolm kuud, siis kolmas kord*); kvantorina on kasutatud põhi- ja järgarve; substantiiv on genitiivse eestäiendi (*küll ühe advokaadi*), kvantori järellaiendi (*enne mitukümmend korda*), subjekti (*korruga üks sisalik*) ja objekti funktsioonis (*nagu üht sundi*). 1990ndatel taandub mustri avar morfosüntaktiline varieerumine, esile kerkib kvantorfraasi sisaldav muster, milles põhiarv on ainsuse nominatiivis ja substantiiv ainsuse partitiivis (*kokku kakssada kilo*). Ka see on näide morfosüntaktilisest kinnistumisest, mida võib määratleda keeleökonoomiast tingitud kvalitatiivse nihkena, nagu eespool kirjeldatud juhtumid (vt DSS, DDS, DVS).

DJS-i vormi- ja sõnavalik on 1890ndatel piiratud (nt võrdlus *paigal kui soola-sammast*; järgnevad tegevused <läks> ära ja esimees <istus>). 1990ndatel tuleb esile muster, milles subordineerivale sidesõnale järgnev substantiiv ainsuse nominatiivis on subjekti funktsioonis (nt *enne kui tüdruk, ilma et nahk*), harvem käändeline määrus (nt *tuimalt ja kõrvades*). Konjunktsioonidest esineb nii koordineerivaid kui ka subordineerivaid. Tegu on kvantitatiivselt (LL -/43,89), morfosüntaktiliselt ja funktsionaalselt erineva lingvistilise struktuuriga.

DKS (ka DNS ja DJS) esinemus on 1890ndate ilukirjanduskeeles marginaalne. Adverbidest on kasutatud *ikka, ruttu, veel*, eessõnafraasi kuuluv adpositsioon asub substantiivi ees, määrates selle käändekasutust – varieeruvad abessiiv, genitiiv ja partitiiv, nt *ikka ilma pruudita, ruttu ümber telje, veel pärast surma*. 1990ndatel on saanud domineerivaks muster, milles eessõnafraasi kuuluv substantiiv on kinnistunud ainsuse partitiivis, nt *hiljem mööda maad, juba enne keskpäeva*. Kvantitatiivset nihet saab pidada statistiliselt oluliseks (LL -/33,31).

Kokkuvõttes tuleb tõdeda, et struktuuride DSS, DVS, DNS, DJS ja DKS kvantitatiivsed ja kvalitatiivsed nihked näitavad kindlat arenguseaduspära: adverbi sisaldavatest klastritest, mille esinemus 1890ndatel on alla optimaalsuskoeffitsienti ning mis on kasutusmuustrina ebaolulised, on sajandi möödudes saanud kinnistunud morfosüntaksi ja tekstifunktsioonidega mustrid. Selles peitub ka üks põhjusi, miks ületab 1990ndatel adverbi esinemus substantiivi esinemust. Poleks paha veenduda, kas nende muustrite kasutamine on tänaseks muutunud veelgi spetsiifilisemaks.

4) Alamklass adverb-adpositsioon (D-K)

Alamklass koosneb kahest klastrist: DSK ja DPK. Levinum on **DSK** (1890ndate korpuses 78,8%), mis koosneb adverbist, tagasõnafraasi kuuluvast substantiivist ainsuse genitiivis ja adpositsioonist (enamasti kohamäärus), nt *korraga laua peale, tasa kõrva sisse*. 1990ndate valimis on DSK osakaal suurem (82,4%, LL -/7,34). Vahe **DPK** kasutuses pole statistiliselt oluline (LL +/0,21, vastavalt 21,2% ja 14,4%, vahe 6,8%). Erinevus absoluutarvudes näitab, et adpositsioonifraasi sisaldavaid mustreid DSK ja DPK on 1990ndatel vähem kasutatud.

5) Alamklass adverb-pronoomen (D-P)

Alamklass D-P koosneb kuuest klastrist, mille osakaalud 1890ndate valimis on: DVP – 66,5%, DDP – 16,1%, DPP – 6,2%, DJP – 5%, DSP – 5%, DKP – 1,2%. Enim kasutatud **DVP** asub (osa)lause alguses: a) adverb + verb indikatiivi imperfekti ainsuse 3. pöördes finiiitöeldisena + personaalpronoomen ainsuse 3. pöördes subjektina, nt *Murtult lähenes ta*; b) osalause alguses on kinnistunud proadverb *siis* ajamääruse funktsioonis, nt *siis saada teada, siis suri*

see. 1990ndateks on DVP kasutus laienenud (LL -/98,99, vahe on statistiliselt oluline). Adverbide valik on muutunud rikkalikumaks, nende tekstifunktsioonid avardunud: lisaks sidususfunktsioonile (osa)lause alguses esinevad ka aja- (*praegu näisid kõik*), viisi- (*väliselt andis tema, jäägitult pole tema*) ning hulgamäärusena (*rohkem tahtsin ma*). Tegu pole mitte niivõrd kvalitatiivse, kui võrd kvantitatiivse nihkega ja viitab pigem ilukirjanduskeele leksikaal-semantilisele mitmekesisumisele.

DDP (nt *ainult ettevaatlikult oma*) pole statistiliselt oluline, kuid LL -/39,54 näitab, et sajandi jooksul toimunud kvantitatiivne nihe on üsna märkimisväärne. Nagu DVP puhul, nii ka siin ei kaasne sellega kvalitatiivset nihet. Tegu on ilukirjanduskeele muutumisega leksikaalselt ja semantilisele mitmekesisemaks (nt *võimalikult kauaks enda, siis juba mõnda, ju eelkõige iseennast*), mis paistab olema 1990ndatele iseloomulik.

Kokkuvõte

Artiklis kirjeldasin 1890ndate eesti ilukirjanduskeele kasutusmustreid võrrelduna 1990ndatega. Log-tõepära näitaja alusel tulid esile statistiliselt olulised kvantitatiivsed nihked mustrite kasutuses. Tavaliselt kaasnevad nendega ka kvalitatiivsed nihked.

Selgus, et 1890ndate keelekasutuses valdav substantiivialguline klass asendub 1990ndatel adverbialgulise klassiga. Järelikult laieneb adverbi kui avatud sõnaliigi esinemus, suureneb adverbide hulk, mitmekesisuvad nende tekstilised funktsioonid. Sama tendents ilmneb ka substantiivialgulise klassi adverbi sisaldavate klastrite võrdlusest. Põhjuste seletamine – eriti silmas pidades adverbi rolli liitsete analüütiliste üksuste kujunemisel 20. sajandil ja praegu – vajab üksikasjalikku diakroonilist diskursuspõhist analüüsi. Käesoleva analüüsi tulemused näitavad, et see protsess avardub, pidevalt lisandub näiteid eesti keele analüütilisusest. Põhjusi on nii lingvistilisi kui ka tekstiloomelisi ja taotluslikke. Näiteks SVD-mustri laienemist 1990ndate ilukirjanduskeeles võib seletada rõhuasetuste muutumisega – väheneb kirjelduste ning suureneb konkreetsete isikustatud tegevuste osakaal. Oma mõju on sellele avaldanud ka 1980ndatel alanud perestroika, mis tõi ajakirjandusse ja sealt ilukirjandussegi isiksusekesksema keelekasutuse. SVD laienemist 1990ndatel kui olulisimat erinevust võrreldes 1890ndatega kinnitavad log-tõepära andmed (vt lisa 3): oluliselt erineva kasutusega 82 mustri seas on suurim vahe SVD kasutuses (LL -/615,46).

LL-näitaja alusel reastatud kvantitatiivsete nihetega kaasnevaid kvalitatiivseid lingvistilisi nihkeid ei saa siinse analüüsi põhjal käsitleda keelemuu-

tustena. Analüütilisuse suurenemine, muustrite morfosüntaktiline ja funktsionaalne kinnistumine, kaassõnafraasi stabiliseerumine omakeelse muustrina, adverbi jt avatud sõnaliikide leksikaalsemantilise varieerumise avardamine jne pole iseenesest veel diakrooniline muutus, vaid keelesisesed lingvistilised protsessid. Nii kinnistuvad sageli korduvate sõna(vormi)de järjendid, eriti morfosüntaktiliste ja leksikaalsete dominantide olemasolul. Näiteks DVP (*samas leidis ta, samas märkas ta, samas oli ta*), kus aluseks muster *samas* + öeldisverb imperfekti 3. pöördes + ainsuse 3. isiku personaalpronoomen *ta*. Adverbi leksikaalse ja pronoomeni morfosüntaktilise kinnistatusega kaasneb öeldisverbi suurem leksikaalne varieerumine. See on väga iseloomulik näide keelekasutuse lingvistilisest olemusest: ühelt poolt ökonoomsusele tuginev ja ratsionaalne, teisalt rikkalikult varieeruv ning väljendusvõimalusi pakkuv. Tegu on loomuliku protsessiga: muistri komponentide funktsioonide kitsenedes suureneb selle kasutussagedus. Protsess võib olla ka vastupidine (nt SDV funktsioonide laienemine).

Morfosüntaktilise kinnistumise ja leksikaalsemantilise avardamise protsessi näitavad ka DSS, DVS, DNS, DJS ja DKS, mis 1890ndatel kuuluvad statistiliselt ebaooluliste hulka, kuid on sajandiga läbi teinud nii kvantitatiivse kui ka kvalitatiivse arengu – igas klastris on esile kerkinud üks dominantne muster, mistõttu muutub ka mustreid ümbritsev kontekst, k.a esituslaad. Kerkib küsimus, mis edasi. Kas need mustrid on tänaseks sama olulised, on nende areng kulgenud marginaliseerumise suunas või on neist saanud statistiliselt olulised mustrid.

Teine osa kasutusnihetest tuleneb keelekasutaja eelistustest jm konsensuslikest faktoritest ega ole olemuslikult ja tekkelt lingvistilised, vaid lingvistiliste vahendite suhtes valivad (nt noor-eestlaste pürgimused, Aaviku keeleuuendus). See mõjutab ka ilukirjanduse temaatikat, tekstide sisu ja avaldub sõnastustes (nt pronoomeni- ja verbialgulise morfoloogilise klassi kasutussageduse muutumine). Püüe väljendusrikkama eesti keele poole ilmneb ilukirjanduskeeles muustrites esile tulnud nihetes ja on kooskõlas eesti keele loomuliku arenguga.

Ilukirjanduskeeles muustrite diakroonilise analüüsi tulemustel on rakenduslik väljund eesti keele grammatikakirjeldustes ja keeleõpikutes. Ilukirjanduskeel kui kirjakeele ideaalvariant sobib malliks nii emakeelekõnelejale kui ka eesti keele õppijale, et lihvida või omandada tekstiloomes vajalikku vormi- ja sõnakasutust.

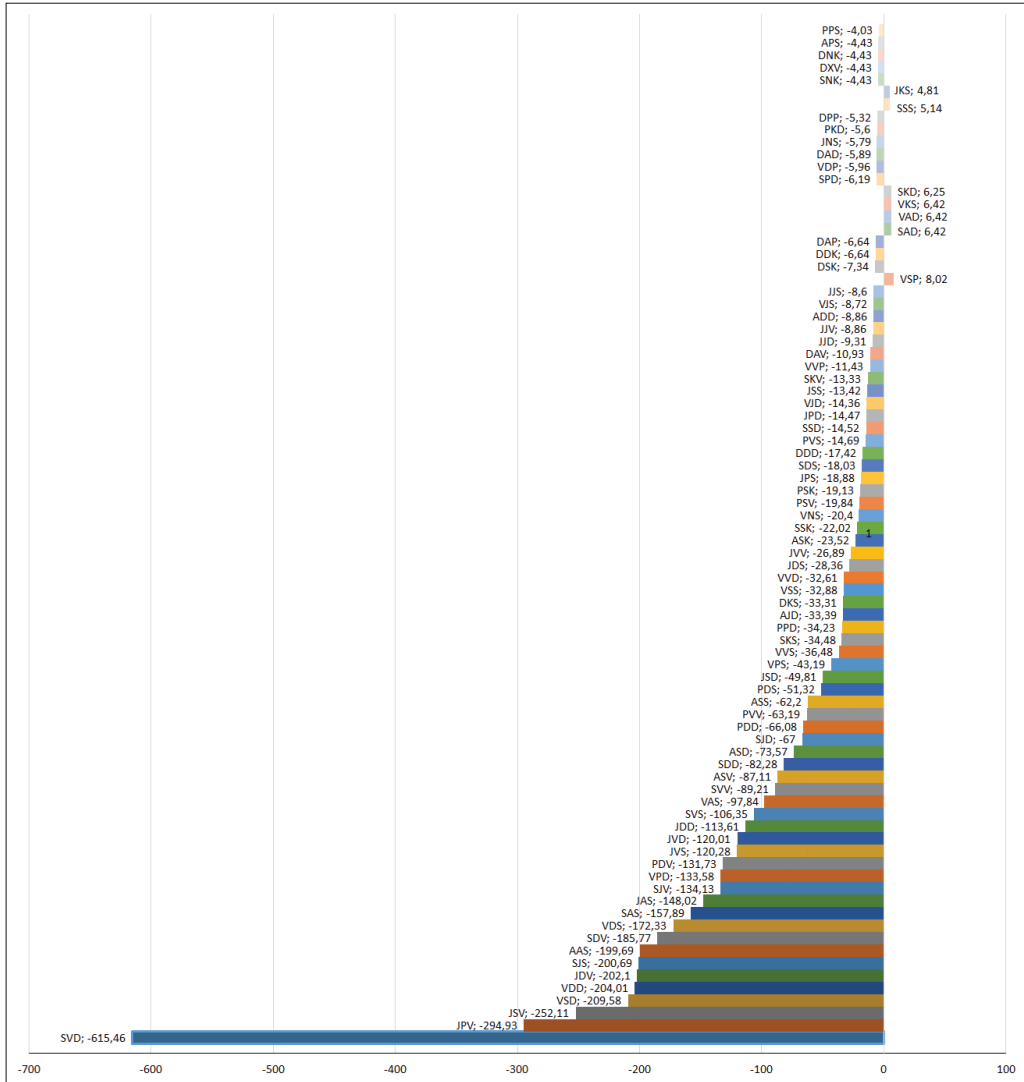
Lisa 1. Substantiivialguline klass

		1890ndad		1990ndad		LL (> 3,84)
Alam- klassid	Klastrid	Klastrisagedus (k ≈ 70)	Klastrite osakaal alamklassis	Klastrisagedus (k ≈ 85)	Klastrite osakaal alamklassis	
S-D	SKD	307	40,5%	116	11,0%	+6,25
	SVD	196	25,9%	611	58,1%	-615,46
	SDD	141	18,6%	190	18,1%	-82,28
	SSD	85	11,2%	77	7,3%	-14,52
	SPD	14	1,9%	17	1,6%	-6,19
	SAD	8	1,1%	-	-	+6,42
	SJD	4	0,5%	41	3,9%	-67
	SND	2	0,3%	-	-	+1,6
Kokku		757	100%	1051	100%	
S-V	SSV	160	27,4%	95	13,4%	-2,0
	SVV	159	27,2%	211	29,7%	-89,21
	SKV	155	26,5%	120	16,9%	-13,33
	SDV	58	9,9%	183	25,8%	-185,77
	SPV	20	3,4%	4	0,6%	+3,28
	SAV	14	2,4%	4	0,6%	+1,02
	SJV	14	2,4%	93	13,1%	-134,13
	SNV	3	0,5%	-	-	+2,41
SXV	2	0,3%	-	-	+1,60	
Kokku		585	100%	710	100%	
S-S	SJS	126	26,3%	268	35,4%	-200,69
	SSS	78	16,3%	23	3%	+5,14
	SKS	62	12,9%	82	10,8%	-34,48
	SVS	60	12,5%	135	17,8%	-106,35
	SAS	58	12,1%	166	21,9%	-157,89
	SDS	44	9,1%	52	6,9%	-18,03
	SPS	42	8,7%	27	3,6%	-1,12
	SNS	10	2,1%	4	0,5%	+0,13
Kokku		480	100%	757	100%	
S-K	SSK	178	84,4%	149	86,6%	-22,02
	SPK	21	9,9%	15	8,7%	-1,16
	SVK	8	3,8%	6	3,5%	-0,58
	SDK	4	1,9%	-	-	+3,21
	SNK	-	-	2	1,2%	-4,43
Kokku		211	100%	172	100%	

Lisa 2. Adverbiaalguline klass

		1890ndad		1990ndad		LL (> 3,84)
Alamklassid	Klastrid	Klastrisagedus (k ≈ 70)	Klastrite osakaal alamklassis	Klastrisagedus (k ≈ 85)	Klastrite osakaal alamklassis	
D–D	DDD	383	68,4%	265	40,0%	-17,42
	DJD	49	8,7%	89	13,4%	-56,85
	DPD	48	8,6%	87	13,1%	-55,44
	DVD	48	8,6%	125	18,9%	-110,98
	DSD	30	5,4%	91	13,7%	-90,04
	DAD	2	0,3%	6	0,9%	-5,89
Kokku		560	100%	663	100%	
D–V	DDV	173	33,4%	352	39,8%	-252,72
	DVV	149	28,8%	219	24,8%	-107,74
	DJV	97	18,7%	111	12,6%	-36,22
	DSV	46	8,9%	99	11,2%	-74,94
	DAV	26	5,0%	31	3,5%	-10,93
	DPV	25	4,8%	70	7,9%	-65,56
	DNV	2	0,4%	-	-	+1,60
	DXV	-	-	2	0,2%	-4,43
Kokku		518	100%	884	100%	
D–S	DAS	112	24,2%	344	33,7%	-343,18
	DDS	105	22,7%	169	16,5%	-93,72
	DPS	86	18,6%	161	15,8%	-73,63
	DSS	64	13,9%	84	8,2%	-34,90
	DVS	49	10,6%	142	13,9%	-136,23
	DNS	22	4,8%	46	4,5%	-33,90
	DJS	16	3,5%	46	4,5%	-43,89
	DKS	6	1,3%	27	2,6%	-33,31
	DGS	2	0,4%	3	0,3%	-1,52
Kokku		462	100%	1022	100%	
D–K	DSK	186	78,8%	126	82,4%	-7,34
	DPK	50	21,2%	22	14,4%	+0,21
	DDK	-	-	3	1,9%	-6,64
	DNK	-	-	2	1,3%	-4,43
Kokku		236	100%	153	100%	
D–P	DVP	107	66,5%	175	66,8%	-98,99
	DDP	26	16,1%	54	20,6%	-39,54
	DPP	10	6,2%	13	5,0%	-5,32
	DJP	8	5,0%	9	3,4%	-2,84
	DSP	8	5,0%	8	3,1%	-1,95
	DKP	2	1,2%	-	-	-1,6
	DAP	-	-	3	1,1%	-6,64
Kokku		161	100%	262	100%	

Lisa 3. Oluliselt erinevad klastrid 1890ndate ja 1990ndate keelekasutuses



Kommentaariid

- ¹ Vt ka <https://nlp.stanford.edu/IR-book/html/htmledition/divisive-clustering-1.html> (6. aprill 2017).
- ² Vt <http://corpora.lancs.ac.uk/clmtp/2-stat.php> (15. september 2017), ka McEnery, Tony & Hardie, Andrew. *Corpus Linguistics: Method, theory and practice. Part 2: Accessing and analysing corpus data*. Cambridge University Press, 2012.
- ³ Eesti keeles on sünonüümselt kasutusel nii sõnaliigi kui ka morfoloogilise klassi mõiste. Oma kirjeldustes jään sõnaliigi mõiste juurde, kuna morfoloogilise klassi mõiste tähistab siinses uurimuses lingvistilise klasteranalüüsi hierarhia kõrgemat tasandit (nt substantiivi-, adverbi-, adjektiivialguline morfoloogiline klass jne).

Allikad

Eesti Kirjakeele Korpuse 1890ndate alamkorpuse (<http://www.cl.ut.ee/korpused/baaskorpus/1890/index.php?lang=et> – 4. oktoober 2017).

Eesti Kirjakeele Korpuse 1990ndate alamkorpuse (http://lepo.it.da.ut.ee/~heli_u/SA – 4. oktoober 2017).

Kirjandus

Aikhenvald, Alexandra Y. & Dixon, Robert M. W. (toim) 2007. *Grammars in Contact: A Cross-Linguistic Typology. Explorations in Linguistic Typology Series 4*. Oxford: Oxford University Press, lk 1–66.

Allkivi, Kais & Eslon, Pille & Trainis, Jekaterina 2017. Kontseptuaalselt sidusa mõistevara kujunemine: klasteranalüüs. *16. rakenduslingvistika kevadkonverents, 20.–21.04.2017*. Teesid. Eesti Rakenduslingvistika Ühing (<https://www.rakenduslingvistika.ee/wp-content/uploads/2016/04/Teesid-2017-3.pdf> – 25. september 2017).

Biber, Douglas 1991. *Variation Across Speech and Writing*. Cambridge: Cambridge University Press.

Biber, Douglas 1993. Representativeness in Corpus Design. *Literary and Linguistic Computing* 8 (4), lk 243–257.

Biber, Douglas 2009. Corpus-Based and Corpus-driven Analyses of Language Variation and Use. Heine, Bernd & Narrog, Heiko (toim). *The Oxford Handbook of Linguistic Analysis*. lk 193–224 (doi: 10.1093/oxfordhb/9780199544004.013.0008).

Bondarko 1983 = Bondarko, Aleksandr 1983. *Printsipy funktsional'noi grammatiki i voprosy aspektologii*. Leningrad: Nauka.

Brunner, Annelen & Steyer, Kathrin 2007. Corpus-driven study of multi-word expressions based on collocations from a very large corpus. *Proceedings of the 4th Corpus Linguistics Conference* (<https://www.deutsche-digitale-bibliothek.de/binary/QVTYID5PT42VIZJMBRLMCK5UIKBK3LBK/full/1.pdf> – 19. august 2017).

- Bybee, Joan L. 2006. *Frequency of Use and the Organization of Language*. New York: Oxford University Press.
- Bybee, Joan L. & Hopper, Paul J. 2001. Frequency and the Emergence of Linguistic Structure. *Typological Studies in Language* 45, lk 1–24 (doi: 10.1075/tsl.45).
- Croft, William 2003. *Typology and Universals*. 2. tr. Cambridge: Cambridge University Press.
- Ehala, Martin 1996. Integreeritud keeleteooria võimalikkusest tänapäeva keeleteaduses (II). *Keel ja Kirjandus* 7, lk 447–460.
- Eslon, Pille 2014a. Morfosüntaktilise ja leksikaalse varieerumise piiridest. *Eesti Rakenduslingvistika Ühingu aastaraamat* 10, lk 55–71 (doi: 10.5128/ERYa10.04).
- Eslon, Pille 2014b. Adverbi sisaldavate struktuuride tekstifunktsioonidest eesti ilukirjandus- ja õppijakeeles. *Lähivõrdlusi. Lähivertailuja* 24, lk 15–46 (doi: 10.5128/LV24.01).
- Eslon, Pille 2017. Keelekasutusmustrid verbist paremal: morfosüntaktiline ja leksi-kaalne varieerumine. *Lähivõrdlusi. Lähivertailuja* 27, lk 17–64 (doi: 10.5128/LV27.01).
- Eslon, Pille & Matsak, Erika 2009. Eesti keele kasutusvariandid: korpusest tulenev käändevormide võrdlev analüüs. *Eesti Rakenduslingvistika Ühingu aastaraamat* 5, lk 79–110 (doi: 10.5128/ERYa5.06).
- Everitt, Brian S. 1997. Cluster Analysis. John P. Keeves (toim). *Educational Research, Methodology, and Measurement: An International Handbook*. New York: Pergamon, lk 466–472.
- Feng, Haoda 2014. *A Corpus-driven Study of Chinese Translators' Use of English Collocations in Commercial Chinese to English Translation*. Auckland University of Technology (<http://aut.researchgateway.ac.nz/bitstream/handle/10292/8004/FengH.pdf?sequence=3&isAllowed=y>–23. märts 2017).
- Francis, Gill 1993. A corpus-driven approach to grammar: Principles, methods and examples. Baker, Mona & Francis, Gill & Tognini-Bonelli, Elena (toim). *Text and Technology: In honour of John Sinclair*. Philadelphia/Amsterdam: John Benjamins Publ. Co., lk 137–156.
- Gore, Paul A. jun. 2000. Cluster analysis. Tinsley, Howard E. A. & Brown, Steven D. (toim). *Handbook of Applied Multivariate Statistics and Mathematical Modeling*. New York: Academic Press, lk 297–321.
- Gries, Stefan Th. & Stefanowitsch, Anatol 2010. Cluster analysis and the identification of collexeme classes. Rice, Sally & Newman, John (toim) *Empirical and experimental methods in cognitive/functional research*. Stanford, CA: CSLI, lk 73–90.
- Gries, Stefan Th. & Hilpert, Martin 2012. Variability-based neighbor clustering: a bottom-up approach to periodization in historical linguistics. Nevalainen, Terttu & Traugott, Elizabeth Closs (toim). *The Oxford Handbook on the History of English*. Oxford: Oxford University Press, lk 134–144.
- Gries, Stefan Th. & Ellis, Nick C. 2015. Statistical measures for usage-based linguistics. *Language Learning* 65 (Supplement 1), lk 1–28.

- Gries, Stefan Th. 2007. Finding groups in chronologically-ordered corpus data: variance-based neighbor clustering. *Paper presented at ICAME 28, Birmingham, 26 May 2007* (<http://www.birmingham.ac.uk/documents/college-artslaw/corpus/conference-archives/2007/9Abstract.pdf> – 22. juuni 2017).
- Grünthal, Riho 2003. *Finnic Adpositions and Cases in Change*. Helsinki: Finno-Ugrian Society.
- Habicht, Külli 2006. Meie muutuv kirjakeel. *Oma Keel* 1, lk 12–19.
- Habicht, Külli & Keevallik, Leelo & Tragel, Ilona 2006. Keele muutumine kasutuskontekstis. *Keel ja Kirjandus* 8, lk 609–625.
- Habicht, Külli & Penjam, Pille 2006. Kaassõna keeleuurija ja -kasutaja käsituses. *Emakeele Seltsi aastaraamat* 52, lk 51–68.
- Habicht, Külli & Penjam, Pille & Prillop, Külli 2011. Sõnaliik kui rakenduslik ja lingvistiline probleem: sõnaliikide märgendamine vana kirjakeele korpus. *Eesti Raken-
duslingvistika Ühingu aastaraamat* 7, lk 19–41 (doi: 10.5128/ERYa7.02).
- Haspelmath, Martin 2006. Against markedness (and what to replace it with). *Journal of Linguistics* 42 (1), lk 25–70 (doi: 10.1017/S0022226705003683).
- Haspelmath, Martin 2008a. Frequency vs. iconicity in explaining grammatical asymmetries. *Cognitive Linguistics* 19 (1), lk 1–33 (doi: 10.1515/COG.2008.001).
- Haspelmath, Martin 2008b. Creating economical morphosyntactic patterns in language change. Good, Jeff (toim). *Language Universals and Language Change*. Oxford: Oxford University Press, lk 185–214.
- Heine, Bernd 2003. Grammaticalization. Joseph, Brian D. & Janda, Richard D. (toim). *The Handbook of Historical Linguistics*. Oxford: Blackwell, 575–601.
- Hunston, Susan 2002. *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.
- Hunston, Susan & Francis, Gill 2000. *Pattern Grammar: A Corpus-driven Approach to the Lexical Grammar of English*. Amsterdam: John Benjamins Publ. Co.
- Jürine, Anni 2016. Täendusmuutuse roll grammatika tekkimise protsessis. *Oma Keel* 33 (2), lk 17–22.
- Jürine, Anni & Habicht, Külli 2013. Kaassõnade tsükliline areng. *Keel ja Kirjandus* 10, lk 736–750.
- Kaalep, Heiki-Jaan & Muischnek, Kadri & Müürisep, Kaili & Rääbis, Angela & Habicht, Külli 2000. Kas tegelik tekst allub eesti keele morfoloogilistele kirjeldustele? Eesti kirjakeele testkorpuse morfosüntaktilise märgendamise kogemusest. *Keel ja Kirjandus* 9, lk 623–633.
- Karlsson, Fred 2002. *Üldkeeleteadus*. Tallinn: Eesti Keele Sihtasutus.
- Küngas, Annika 2014. *Pragmaatiliste markerite kujunemine ja funktsioonid eesti keeles lt-sõnade näitel*. Doktoritöö. Tartu: Tartu Ülikooli Kirjastus.

- Laanekask, Heli. 2004. *Eesti kirjakeele kujunemine ja kujundamine 16.–19. sajandil*. Dissertationes of philologiae estonicae Universitatis Tartuensis 14. Tartu: TÜ Kirjastus (<http://dspace.utlib.ee/dspace/bitstream/handle/10062/1138/Laanekask.pdf?sequence=5> – 9. aprill 2017).
- Mathesius 1967 = Matezius Vilem 1967. O tak nazyvaemom aktual'nom chlenenii predlozheniia. Kondrashov, Nikolai (toim). *Prazhskii lingvisticheskii kruzhek*. Sbornik statei. Moskva: Progress, lk 239–245.
- Matsak, Erika & Eslon, Pille & Kippar, Jaagup 2010. Eesti keele sõnajärje vealeidja prototüübi arendamine. Eslon, Pille & Õim, Katre (toim). *Korpusuuring ja meetodid*. Tallinna Ülikooli eesti keele ja kultuuri instituudi toimetised 12. Tallinn: TLÜ Kirjastus, lk 59–100.
- McGregor, William B. 2015. *Linguistics: An introduction*. 2. tr. London/New York: Bloomsbury Academic.
- McMahon, April M. S. 1994. *Understanding Language Change*. Cambridge: Cambridge University Press.
- Metslang, Helle 2006. Grammatiseerimise teooriaist tänase pilguga. *Teoreetilise keeleteaduse Eestis II*. Tartu Ülikooli üldkeeleteaduse õppetooli toimetised 7. Tartu: Tartu Ülikooli Kirjastus, lk 176–195.
- Mooi, Erik & Sarstedt, Marko 2011. *A Concise Guide to Market Research. The Process, Data, and Methods Using IBM SPSS Statistics*. London/New York: Springer.
- Orav, Heili & Parm, Sirli 2014. Seitsmes “Global Wordnet” konverents. *Keel ja Kirjandus* 5, lk 403–406.
- Ots, Sander 2012. *Statistikapõhise tarkvara loomine morfoloogiliste kollokatsioonide eraldamiseks eesti keele tekstidest*. Bakalaureusetöö. Tallinna Ülikooli informaatika instituut.
- Palmeos, Paula 1982. *Eesti keele grammatika*. II osa. Neljas vihik. Kaassõna. 3. tr. Tartu: Tartu Riiklik Ülikool.
- Piperski, Alexander Ch. 2015. To be or not to be: corpora as indicators of (non-)existence. *Computational Linguistics and Intellectual Technologies*. Papers from the Annual International Conference “Dialogue” (2015) 14 (1), 515–522 (http://www.dialog-21.ru/media/2777/dialogue-2015_voll.pdf – 5. september 2017).
- Remm, Kalle & Remm, Jaanus & Kaasik, Ants 2012. *Ruumiliste loodusandmete statistiline analüüs*. Õpik-käsiraamat. Tartu: Tartu Ülikooli Ökoloogia ja Maateaduste Instituut.
- Römer, Ute 2005. *Progressives, Patterns, Pedagogy: A Corpus-driven Approach to English Progressive Forms, Functions, Contexts and Didactics*. Studies in Corpus Linguistics 18. Amsterdam: John Benjamins.
- Sang, Joel 1983. *Eitus eesti keeles*. Tallinn: Valgus.
- Saussure 1999 = Sossiur, Ferdinand de 1999. *Kurs obshchei lingvistiki*. Ekaterinburg: Ural'skii universitet.
- Tognini-Bonelli, Elena 2001. *Corpus Linguistics at Work*. Amsterdam/Philadelphia: John Benjamins.

Trainis, Jekaterina 2015. Linguistic cluster analysis: a method for describing language units and indicating regularities in language. *Within Language, Beyond Theories. Volume III. Discourse Analysis, Pragmatics and Corpus-based Studies*. United Kingdom: Cambridge Scholars Publishing, lk 229–243.

Trainis, Jekaterina & Allkivi, Kais 2014. Ilukirjanduskeelest uue pilguga. *Eesti Raken-
duslingvistika Ühingu aastaraamat* 10, lk 283–306 (doi: 10.5128/ERYa10.18).

Vaiss, Natalia 2004. *Eesti keele aspekti väljendusvõimalusi vene keele taustal*. Magistri-
töö. Tallinna Ülikool (<http://hdl.handle.net/10062/44143> – 20. juuli 2017).

Velsker, Eva 2010. Leksikaliseerumine keelemuutuses. *Keel ja Kirjandus* 7, lk 521–536.

Yelenevskaya, Maria 2008. Russian: From Socialist Realism to Reality Show. Rosen-
house, Judith & Kowner, Rotem (toim). *Globally Speaking: Motives for Adopting English
Vocabulary in Other Languages*. Clevedon/Buffalo: Multilingual Matters, lk 98–120.

Summary

Diachronic language changes in usage patterns of belletristic Estonian in 1890s–1990s

Jekaterina Trainis

PhD student of linguistics at the Tallinn University
jekaterina.trainis@tlu.ee

Keywords: diachronic language change, Estonian belletristic language, corpus linguistics, linguistic cluster analysis, morphosyntactic patterns, statistics

This paper compares the morphosyntactic usage patterns of Estonian belletristic language in the 1890s and 1990s. A program called Cluster Catcher (developed at the Tallinn University) is applied to find similar n-grams, i.e. sequences of words, on the basis of morphological and syntactic tags, taking into account the frequency of use. The appropriateness of this statistics-based program has been proven previously and the results have opened a new view to the usage-based grammar of Estonian.

In this case, trigrams are grouped into morphological clusters (part of speech sequences) which are, in turn, classified into morphological classes based on the part of speech of their first component, and divided into subclasses based on the first and final component, so that the middle component varies.

Remarkable statistical shifts that have taken place during the period are revealed in this corpus-driven research. Based on the distribution of morphological classes, I put forward hypotheses about possible diachronic changes. After that, step by step I find evidence on the level of subclasses, clusters and trigrams.

It can be suggested that analyticity has been on the rise in Estonian (formation of multi-word units in relation to the widening of patterns consisting of adverbs by the 1990s). Open and closed parts of speech have different functional adeptness and constraints in texts (e.g. two closed parts of speech – conjunction and adposition – have

differences in usage; while the usage of conjunctions is widening thanks to the rise of coordinative-correlative words, the usage of adpositions is declining due to the decreasing number of adpositions and narrowing of functions). Background factors, such as text creation (type of text, coherence, wholeness of text, pragmatics) and also the target of language reform towards more expressive Estonian, can be taken into consideration.