

Leksika-grammatika piirimail: publitsistlike arvamused keelekasutusmustrid

PILLE ESLON

Tallinna Ülikool

Ülevaade. Uurimus paigutub kasutuspõhise korpuslingvistilise keelekäsitluse raamesse. Nagu varasemates töodes, mis tehtud eesti ilukirjandus- ja õppijakeele korpustel, rajaneb ka sinne arvamused (Postimees ja Õhtuleht) keeleanalüüs kolmest komponendist koosnevatele sõnaliigijärjenditele ehk keelekasutusmustritele, mis leitud morfosüntaktiliselt märgendatud tekstidest andmekaeve põhimõttel töötava programmiga Klastrileidja. Korduvalt esinevasse sõnaliigijärjendisse on sisse kodeeritud selle komponentide kooskasutuse fonoloogilised, süntaktilised ja kontseptuaalsed reeglid, mis avalduvad semantika-grammatika piirimail ning toovad esile keelekasutuse seaduspärasusi. Seetõttu on aktiivselt kasutatavad mustrid sobiv allikmaterjal eesti keele õppes, eri lähtekeelega õppijate keelekasutuse võrdlemisel jm, k.a õppija kirjutatud tekstide automaatse töötlemise vajadusel (sõnastuse parandamine koos viitega reegli(te)le ja sihtkeelepärasemale sõnastusele).

Võtmesõnad: lingvistiline klasteranalüüs; keelekasutusmustrid; leksikaalgrammatiline varieeruvus; keelekasutuse seaduspärasused; eesti keel

1. Sissejuhatus

Eesmärk on leida ja klassifitseerida publitsistlike arvamusaluste keelekasutusmustrid, kirjeldada nende leksikaalgrammatilise ja funktsionaalse varieeruvuse seaduspärasusi substantiivialguliste muustrite näitel.

Arutlused on valitud seetõttu, et neis sulanduvad kõnekeelne ja belletristlik esituslaad lugeja informeerimise ja mõjutamisega. Seega tuleb publitsistikas esile rohkem üldkeelseid aktiivselt kasutatavaid morfosüntaktilisi struktuure, millele mh rajaneb kirjalike tekstide mõistmine ja produtseerimine. Oskus selliseid struktuure kasutada kiirendab Kathy Conklini ja Norbert Schmitti (2008) andmetel oluliselt tekstide produtseerimisprotsessi. See on tähtis keeleõppe seisukohast, sest nii arusaamine kui ka uute tekstide koostamine rajanevad sõnade kooskasutusele leksikaalgrammatiliselt ja funktsionaalselt varieeruvates või stereotüüpse vormistusega muustrites.

Uurimisobjekt on kolmest järjestikusest sõnast koosnevad morfosüntaktilised muustrid, mis on keelekasutuse automaatse analüüsi puhul tavapärane. Näiteks ingliskeelsete pressiteadete ja äriuudiste informatiivsust ning lugeja mõjutamist käsitlevas uurimuses on Sylvie De Cock ja Sylviane Granger (2021) tuginenud samuti kolmesõnalistele järjenditele. Sama pikkusega tekstilõike on tavaliselt analüüsitud ka teise keele omandamise ja leksikograafia alastes uurimustes.

Allikmaterjal pärineb Eesti vahekeele korpuse K1 alamkorpusest, mis sisaldab kõrgharitud mitte filoloogiharidusega emakeelekõnelejate arutlusi 21. sajandi teise kümnendi aktuaalsetel teemadel. Tekstid on automaatselt märgendatud, ebatäpsused käsitsi parandatud. Valim sisaldab 185 teksti (109 326 tekstisõna).

Keelekasutusmuustrite otsimiseks olen kasutanud andmekaevel põhinevat programmi Klastrileidja (Ots 2012), mis on arendatud Tallinna Ülikoolis Erika Matsaku sõnajärjeleidja prototüübi (Matsak jt 2010; Metslang & Matsak 2010) alusel.

Klastrileidja otsing on objektipõhine. Programm tuvastab morfosüntaktiliste märgendite samalaadse järgnevuse põhjal sarnase keelelise

vormistusega mustreid, jagab need rühmadesse (klastritesse) ja lisab sageduse arvandmed. Tegemist ei ole klassikalise, vaid *lingvistilise* klasterdamisega ehk rühmitamise tehnikaga, mida statistilise analüüsiga integreeritud rakendus võimaldab. Meetod välistab uurija subjektiivsete seisukohtade või teoreetiliste eelistuste mõju keeleandmete otsimisel ja süstematiseerimisel.

Andmekaevega leitud keelekasutusmustrite süsteemne käsitlus võimaldab kirjeldada tüüpilisi ja harva esinevaid sõnaliigijärgnevusi tekstis, esile tuua mustris varjatult toimivaid semantilisi ja leksikaalgrammatilisi seoseid, keelekasutaja sõna- ja vormieelistusi, mustri komponentide leksikaalgrammatilise ja funktsionaalse varieeruvuse piire. Tegu on korpuslingvistilise kasutuspõhise uurimusega (Biber 2009: 281 järgi korpusest tulenev analüüs, ingl *corpus-driven approach*), mis rajaneb automaatse tekstitöötamise käigus leitud sõnaliigijärgnevuste süsteemsel lingvistilisel analüüsil, võimaldades saadud tulemuste põhjal üldistada ja seletada nii sõnade loomuliku kooskasutuse keelelisi seaduspärasusi kui ka võrrelda erineva lähtekeelega õppijate keeleoskuse arengut, tuua esile stiili- ja žanrierinevusi jm.

2. Keelekasutusmuster kui morfosüntaktiline terviküksus

Keelekasutusmustrite kaeve ja süstematiseerimine rajaneb sõnaliigi märgenditel: S – substantiiv, V – verb, A – adjektiiv, P – pronoomen, N – numeraal, G – genitiivatribuut, D – adverb, K – adpositsioon, J – konjunktsioon, Y – abreviaatuur. Näiteks tähistab lühend VAS tüüpilist verbi-adjektiivi-substantiivi järgnevust tekstis (*valendab üksik puri*). Seejärel kogub Klastrileidja VAS-struktuuri alusel leitud keelenäited ühte rühma, tuues eraldi välja kõik selle rühma morfosüntaktilised variandid koos keelenäidete ja statistikaga. Sama struktuuri suurima esinemusega varianti olen nimetanud **invariantseks mustriks**, ülejäänud on sageduse kahanemise järjekorras reastatud **variandid**.

Mustrite hierarhilise klassifikatsiooni koostamine algab **mustriklasse** eristamisest esikomponendi sõnaliigi põhjal. Nt belletristikas

on suurima esinemusega klassid, mille esikomponendina kasutatakse adverbi, substantiivi, verbi, pronoomenit ja konjunktsiooni. Iga mustri-klass jaguneb mustri viimase komponendi sõnaliigi alusel **alamklassideks**. Nt osakaalult kolmanda, verbialgulise klassi produktiivsemad alamklassid lõpevad adverbi ja substantiiviga (V-D, V-S). Iga alamklassi mustri keskmise komponendi sõnaliigi vaheldumine toob esile selle alamklassi **mustrirühmad** koos neis sisalduvate morfosüntaktiliste variantide ehk mallidega. Nt verbialgulise mustri-klassi adverbilõpulisel alamklassis on kirjeldatud viit mustrirühma: kolm esimest on produktiivsed (VDD, VSD, VPD), neljas keskmise esinemusega (VVD) ja viimane väheproduktiivne (VJD), nt VDD – *oli üsna kõrgelt, olin juba päris, on ju lõpuks, oleks nagu veel*; VSD – *pööras pilgu ära, võttis buketi vastu, pani noa kõrvale*; VPD – *oli ta salaja, oli ta siinkandis, sain ma liigagi, püüdsin ma küllalt*. (Vt Trainis & Allkivi 2014: 284, 286, 290–292)

Mustri-klasside, alamklasside ja -rühmade hierarhiline liigendus tuleneb Klastrileidja toimesüsteemist. Ühelt poolt on tegemist statistilise klassifitseerimistehnikaga (vt Law 2007: 332), et rühmitada erineva vormitusega keelekasutusmustreid. Teisalt on tegemist lingvistilise klassifitseerimissüsteemi leidmisega (Gore 2000: 300), mida saab rakendada keele elementide kooskasutuse süsteemsel objektipõhisel käsitlemisel. See on olemuslik klassifikatsioon, mis võimaldab korrastada lingvistilisi andmeid.¹

Mustri komponentide omavahelise sidususe aste on mõõdetav, kasutades mitmemõõtmelise statistilise mudeldamise tehnikaid (nt Allkivi-Metsoja 2021a) või muid uurimisobjekti käsitlemiseks sobivaid funktsioone (nt Uiboaed 2010), nt log-tõepära sõnavormide vastastikuse prognoositavuse ja kompositsionaalsuse astme määramisel (Aedmaa

¹ Seda klassifitseerimistehnikat olen põhjalikumalt seletanud artiklis Eslon & Allkivi-Metsoja 2018 ning rakendanud reas varasemates uurimustes (nt Eslon 2014a, 2014b, 2017a, 2017b, 2021a, 2021b; Eslon & Paeoja 2015). Oma panuse meetodi tutvustamise ja kasutamise on andnud Jekaterina Trainis (2015, 2017; Trainis & Allkivi 2014) ja Kais Allkivi-Metsoja (Allkivi 2016; Allkivi-Metsoja 2021b). Sama meetod on olnud tulemuslik erinevate uurimisülesannete lahendamisel (nt Ševtšenko 2014; Paeoja 2015; Hallik 2015; Tšernõšuk 2016; Voolaid 2018).

2015, 2016); mitmesõnalise üksuse komponentide kollokatsioonilisuse astme mõõtmisel eristamaks liitpredikaate püsiühenditest ja abiverbi-konstruksioonidest (Muischnek & Sahkai 2010); ingliskeelsete pressiteadetes ja äriuudistes aktiivselt kasutatud kolmesõnaliste järjendite reastamisel olulise alusel, kus nt kaks sagedamat teemakeskset mustrit on vastavalt *vice president and, today announced that ning per cent to, Bank of England* (De Cock & Granger 2021).

Siinse uurimuse käigus leitud keelekasutusmuustrite alusel pole komponentide vahelise seose tugevust mõõdetud, kuna puudub otsene vajadus.

Esiteks, igal muustril on oma **püsikindel struktuur**. Näiteks arvamslugude adverbialgulise mustriklassi adverbiga lõppeva alamklassi muustrite sekka kuuluvad DDD (*kahjuks ka palju, Nagu eelpool juba*), DVD (*Paraku on nii, Loodetavasti pole päris*), DJD (*täpsemalt ja range-malt, Nii või teisiti*).

Teiseks, mustri komponentidel on **tüüpiline vormistus ja tüüpiline tekstifunktsioon**. Näiteks mustri DDJ alusel moodustatakse ja kasutatakse vastandavat paarissidendi *ei .. ega (ei rohkem ega vähem, ei seest ega väljast, ei ühiskondlikult ega isiklikult)*. Selle sünonüümid moodustuvad DNJ- ja DPJ-struktuuri põhjal (*ei üks ega teine ~ ei see ega teine*). Mustrid DDJ, JDA ja DSJ on alus paarissidendi *nii .. kui (ka)* moodustumisele ja kasutamisele (*nii suuliselt kui kirjalikult, nii kultuuriliselt kui majanduslikult, nii füüsiliselt kui vaimselt, Nii euroskeptitsism kui rahvuslik identiteet, nii fotograafia kui kujutav kunst, nii jumalale kui iseendale*). Struktuurid JDA, JJD ja JJS kasutatuna (osa)lause alguses on samaaegselt alus paarissidendi moodustumisel ja kasutamisel viiteseose vahendina (*Ning kui rahvusvaheline, Ja kui nüüd, Või siis mitte, et kui kodanik, sest kui telg*) jne. Tegu on püsikindlate struktuuridega, mis seovad lause moodustajad süntaktilis-semantiliseks tervikuks kindlas tekstifunktsioonis.

Kolmandaks, mustri **leksikaalgrammatilisel varieeruvusel on pii-rid**, nt VJD-struktuuris tarvitatakse tavaliselt V indic pres 3 sg koos sidesõnaga *aga (on aga kusagil, on aga tuntavalt, tuleb aga ka, tuleb aga kõigepealt, liigub aga edasi, kerkib aga hoopis, seisab aga hoopis)*, millel

on mustris mööndust toonitava partikli funktsioon. Sagedam verb on *olema*, harvem modaalverb *tulema* ('on vaja', 'peab'), ainukordselt ka erinevad täistähenduslikud verbid. Tegu on reeglipärase grammatilise vormistuse, kuid leksikaalselt varieeruva mustriga.

Eelpool loetletud tunnused lubavad käsitleda keelekasutusmustreid sarnaselt muude mitmesõnaliste kooslustega,

- 1) mis esinevad regulaarsete sarjadena, on produktiivselt moodustatud ja eristuvad statistiliselt, nt adessiivsed sooviavaldused, parameetersõnadena toimivate nimisõnavormidega väljendid, genitiivagendiga fraasid, liitpredikaadid (vt Sahkai 2006, 2008, 2011; Muischnek & Sahkai 2010);
- 2) millel on ütluste ja lauseaheladena kindel pragmaatiline funktsioon, kuuludes kinnistunud sõnakasutusega lauseliste kõnevormelite ja jätkuga lausete alla (vt Öim & Öim 2019: 73–109);
- 3) mida on kirjeldatud kui fraase, neist üle poole moodustavad vormelid või kõrge vormelliku väärtusega (ingl *high formulaic value*) sõnajärjendid: keeletükid (*chunks*), leksikaalsed kimbud (*lexical bundles*), mitmesõnalised väljendid (*multi-word expressions*), mitmesõnalised üksused (*multi-word units*) jm, mis eraldatud tekstist statistiliselt korpuslingvistiliste meetodite ja erinevate tekstitötlusvahendite abil (vt Evert 2005: 15–20; Muischnek 2006; Muischnek & Kaalep 2010; Schmitt 2010; Chen & Baker 2010; Martinez & Schmitt 2012: 303–305; Juknevičienė 2013; Selivan 2018; De Cock & Granger 2021).

Mustri püsistruktuur, komponentide grammatiline ja semantiline terviklikkus, tüüpiline vormistus ja leksikaalgrammatilise varieeruvuse piirid on kompleks tunnuseid, mille alusel mustreid süsteemselt kirjeldada, esile tuua keelekasutusele omaseid nähtuseid ning sõnastada seaduspärasusi. Ja kuigi mustrites sisaldub liitmeid sõnakooslusi, konstruktsioone, fraase, kollokatsioone ja idioome, on nende esinemus alati seotud konkreetse sõnaliigijärjendiga. Näiteks:

- 1) SVD on alus tugiverbikonstruktsioonide moodustamiseks ja kasutamiseks – *olema* moodustab järgneva adverbiga semantilise

terviku, nt *seadused on kaldu, kuradid on kinni, Vastus on kirjas, faas on möödas/läbi, Savisaar on nõus, raha on otsas, kurikavalus on süüdi, piimatootjal on tarvis, meil on vaja, meeled on valla/avali*;

- 2) VDD on ühendverbide kasutamiseks, millel olemas ühesõnaline sünonüümne vaste, nt *käib kaasas ka > kaasneb, tuleb sisse juba > siseneb, tuli lihtsalt lahti > avanes, võtaks visuaalselt kokku > üldistaks*;
- 3) VDD alusel moodustatakse ja kasutatakse ütluste seeriat, millega tähistatakse kas hoogsat liikumist, seisundit või subjektiivset arvamust, nt *sammub/liigume/põrutame kindlalt edasi* 'hoogne minek'; *ei saa enam ilma läbi* 'väljapääsmatus'; *ei ole enam kaugel* 'saabub peagi'; *on nüüd kohal* 'on käes', 'olemas'; *ei saa ju ometi* 'pole võimalik'; *on juba praegu* 'kindel suund'; *on järjest rohkem* 'areng' või 'taandareng', 'laienemine' või 'kitsenemine'; *pole enam vaja*; *võiks ju vähemalt*;
- 4) VDD struktuuri alusel tulevad esile piltlikud väljendid, nt *on kõrgel kohal* 'kõrgel ametikohal', *pole teps mitte* 'kategoorigiline eitus', *on kapitaalselt untsus* (sünonüümid *nässus, metsas*), *ei vea enam välja* 'endistviisi ei toimi' jne.

Eelpool toodud seisukohtade ja neid illustreerivate näidete põhjal saab Ray Jackendoffi (2017: 196–197) järgides väita, et mustri struktuuri kui keelekasutuses korduvalt esinevasse sõnaliigijärjendisse on sisse kodeeritud selle komponentide kooskasutuse fonoloogilised, süntaktilised ja kontseptuaalsed reeglid, mis säilitatakse pikaajalises mälus ning aktiveeruvad keelekasutuses semantika-grammatika piirimail. Igal meist on valmidus reastada sõnu kindlate struktuuride põhjal, mitte suvaliselt. Seda on mõistetud eelkõige keeleõppe (nt sarnasuste ja erinevuste leidmine esimese ja teise keele kasutamise vahel), diskursusanalüüsi ning leksikograafia vallas, kus uurimisobjekt on mitmesõnalised leksikaalsed kooslused.

Järgnevalt kirjeldan sageli esinevaid sõnaliigijärgnevusi kui leksi-kaalgrammatiliselt ja funktsionaalselt terviklikke üksuseid, tuues esile

iga järjendi komponentide kooskasutusele omaseid lingvistilisi tunnusoone. Lähemalt kirjeldan seda substantiivialguliste mustrite näitel.

3. Keelekasutusmustrite süsteemne kirjeldus

3.1. Mustriklässid

Lingvistiline klasteranalüüs tõi arvamusalustele esile kõik mustrid. Edasises analüüsis kasutan mustreid, mille sagedus on ≥ 5 (viis ja rohkem). Piirväärtus on eelnevalt välja arvatud (vt Eslon & Matsak 2009: 83–84), edaspidi on sellest lähtunud kõigis lingvistilise klasteranalüüsi meetodil tehtud uurimustes.

Mustrid jagunevad klasside vahel vastavalt esimese komponendi sõnaliigile, lisatud on arvandmed sageduse ja osakaalu kohta valimis (vt tabel 1). Numeraali-, adpositsiooni- ja abreviatuerialgulised klassid jäävad järgnevast käsitlusest välja vähese osakaalu tõttu.

TABEL 1. Mustriklässide sagedus ja osakaal

Mustrikläss	Sagedus	Osakaal
S	1274	25%
D	1158	23%
J	1012	20%
V	687	13%
A	537	11%
P	205	4%
G	93	2%
N	60	1%
K	53	1%
Y	24	0%

Kõrvale jäetud mustriklässid on olnud ebaolulised ka 1890-ndate ja 1990-ndate ilukirjandustekstidest saadud andmete põhjal (Trainis 2017). Ülejäänud klasside puhul ilmnevad erinevused arvamusalustele

publitsistliku ja 1990-ndate ilukirjandusliku esituslaadi keelelistes eelistustes (vrd Trainis & Allkivi 2014), nt arvamusalgudes on suurima osakaaluga substantiivialguline klass (25%), järgneb adverbialguline (23%), samas kui 1990-ndate ilukirjanduses on järjestus vastupidine (24,6% ja 20,9%); arvamusalgude kolmas mustriklasse on konjunktsiooni- ja neljas verbialguline (vastavalt 20% ja 13%), ilukirjanduses jääb verbialguline mustriklasse kolmandale (17%) ja konjunktsioonialguline viiendale kohale (13,5%) jne.

Niisiis tuleb juba mustriklasse võrdluse tasandil esile erinevuse märke publitsistlikus ja ilukirjanduslikus esituslaadis. Arvamusalgude mustriklassest annab konkreetsema ettekujutuse alamklasse ja neis sisalduvate mustriühmade leksikaalgrammatilise varieeruvuse tasand.

3.2. Alamklassid

Alamklassid eristuvad iga mustriklasse sees sõnaliigijärjendi viimase komponendi alusel (vt tabel 2). Alamklasse osakaal arvutatakse mustriklasse piires, edasisest käsitlusest jäävad välja need, kus näidete arv on ≤ 10 (kümme ja vähem).

TABEL 2. Alamklasse sagedus ja osakaal mustriklasse

Substantiivi- alguline klass (25%)			Alamklasse invariantse mustri näited
S-D	388	30%	<i>Castro on veel, meetodid on üsna, turvalisus pole vaid</i>
S-S	338	27%	<i>liha ja veri, Kirjeldus ja ettepanekud, inimesed või grupid, lapsi ja noori, õe ja venna, eestlaste ja lätlaste, riigist ja rahvusest</i>
S-V	309	24%	<i>kohtuistung algus on, dekonstruktsiooni tagajärjeks on</i>
S-K	95	7%	<i>võrdsuse koorma alla, rubla nõrkuse tõttu, reisimine ida poole</i>
S-A	79	6%	<i>kaup on taskukohane, Kirik on vaba</i>
S-J	49	4%	<i>Kersti Nigesen ja, töökorra seadus ja</i>

Adverbiaalguline klass (23%)			
D-S	363	31%	<i>üha süvenev äng, väga vana tava, Ka uued ideed, Samuti hea valiku, ka vastavate volituste, väga suurt huvi, nii väikeses riigis, kui eelmisel nädalal, praegu täie hooga</i>
D-D	259	22%	<i>ikka nii väga, seejuures mitte kunagi, Arvuliselt palju rohkem</i>
D-V	216	19%	<i>enam ei lähe, täna ei paku, Paraku ei leita, Tegelikult ei saanud, Nii ei tasuks</i>
D-A	121	11%	<i>ilmselt ka uus juhatus, kus isegi lihtsad asjad, ka väga selge ülesande, ilmselt väga kavalat skeemi, väga palju kasutatud vahend, veel ka truuks jääda</i>
D-J	108	9%	<i>nii kaua kui, kindlalt üle ja, kas siin või, Ka mitte siis</i>
D-K	56	5%	<i>teadlikult tõestatu kõrval, Mistahes juhtumi puhul, jätkuvalt naiste õlul</i>
D-P	13	1%	<i>Praegu (Tegelikult, Üldiselt) on see</i>
D-N	12	1%	<i>kaheldamatult 1960-1990, Pisut üle kümne</i>
Konjunktsioonialguline klass (20%)			
J-S	398	39%	<i>ja tolerantsem ühiskond, Ent tänased noored, ja uue laine, ent kirgliku vaidluse, ning roheline erakonna, kui vaene sugulane, Aga ülejääva taristu, või pedagoogilist õigustust, Ent suurel naabril, ehk sihilik solvumine</i>
J-V	280	28%	<i>kui läänemaailm on, et seadused on, Kui kaitseväes on, et kingikotis on, sest tegelikkuses ei, sest ravikindlustatul on</i>
J-D	159	16%	<i>kui ka mujal, aga isegi kõige, Kuid märksa efektselt, Aga nii kaua, Nagu eespool juba, et siit ära, ja rangemalt tarvis, ning sinna vahele</i>
J-K	75	7%	<i>või raja taga, ja mandri vahel, et vajaduse korral, ning jooniste näol, Kui tervishoiuteenuste eest, Nagu kriiside puhul, et aastate pärast</i>
J-A	55	5%	<i>Kuna ka potentsiaalse kandidaadi, et väga kõrge taseme, siis nii uhkete lahenduste, Ning kui rahvusvaheline seltskond, kui nüüd truud alamad, et nii ebameeldivat olukorda</i>
J-J	26	3%	<i>Ja isegi kui, et võib-olla ehk, et millal ja, siis rohkem kui; et poliitika ja, ning usk ja, või naine ja, kui riigi ja, et heaolu ja, ehk tarbimise ja</i>
Verbialguline klass (13%)			
V-D	399	58%	<i>on täiesti omaette, Tundub ju igati, võiks ju vähemalt, keriks kohe esile, oli ju ikka, oli toona väga populaarne, ei vea enam välja, ei saa ju ometi, ei söö mitte kunagi</i>

V-S	122	18%	<i>puudub sisuline vajadus, on olulised erisused, on sisemise kasvamise, on inimsusvastaste tegude</i>
V-A	113	16%	<i>on veel selgusetu, on täiesti normaalsed, ei ole samuti võimalik, pole ka tähtsusetu, oli väga tark, oli just lahkunud, on veel pikk tee minna, on üha sarnasem, oleks liiga sinisilmne, oleks erakordselt vale</i>
V-J	24	3%	<i>on vaja ja, on möödas ja, on piisavalt ning, On rohkem kui kindel, et</i>
V-V	18	3%	<i>pole kunagi olnud, pole kuhugi kadunud, pole väga levinud, on ilma jäänud, on varem olnud, on plahvatuslikult kasvanud</i>
Adjektiivialguline klass (11%)			
A-S	151	28%	<i>praegune kaitseväe luurekeskus, tänavune nafta hinna tõus, kehtiva juhtimisõiguse tõestamiseks, populaarse raadiojaama eetrist, võrdse kohtlemise põhimõtet, on sisemise kasvamise protsess, noorte sõdurite väljaõpe</i>
A-K	135	25%	<i>viimase aasta jooksul, eriliste teenete eest, kinniste uste taga, uute tehnoloogiate vastu</i>
A-V	82	15%	<i>kõrgeim otsustuskogu on, Alternatiivsed valikud on, maa-giline puudutus tuleb, tolerantne ühiskond tõmbab, kõhetu rinnakorv murdub, korduv skandaal viitab, Laias laastus on, Tehnilises mõttes on, väikeses poes on</i>
A-D	68	13%	<i>Jube lugu küll, tõsised tagajärjed ka, suurel naabril ikkagi, viimastel aastatel oluliselt, kohalikus omavalitsuses vähemalt, praeguses valitsuses lausa</i>
A-A	49	9%	<i>kalli ja keeruka, Puhas ja lihtne, küünilised ja räiged, tundlikuks ja pirtsakaks</i>
A-J	47	9%	<i>uus nimi ja, juhtiv büroo ning, üksikud komisjonid ning, keerulise olemuse ja, Poliitilise solvamise ja, hüppelise kasvu ning</i>
Pronoomialguline klass (4%)			
P-D	78	38%	<i>See on muidugi, see toimub mujal, see kõlab ehk, kes on küll, mis on siiani</i>
P-V	53	26%	<i>Ma ei tea, Ma ei ole, me ei saa, mis ei toimi, seda ei saa</i>
P-K	33	16%	<i>sellise info põhjal, selle raha eest, nende küünte vahel, oma naabri kohta, sama töö eest</i>
P-A	26	13%	<i>See on sümptomaatiline, mis on võimeline</i>
P-S	15	7%	<i>selle suure riigi, see tohutu meelemuutus</i>

Genitiivtribuudialguline klass (2%)			
G-S	54	51%	Euroopa Liidu laienemine, Eesti teadlaste panus, Berliini Riigiooperi peadirigent, California ülikooli bioloogi
G-K	17	21%	Balti jaama poole, Eesti põhiseaduse järgi, Euroopa Komisjoni poolt
G-V	14	18%	Poola ekspreident on, Eesti Energia on, ukraina rahvas on, Saksamaa tööstus on

3.3. Keelekasutusmustrite süsteemse kirjelduse näide: substantiivialguline klass

Substantiivialgulised mustrid (kokku 16) jagunevad kuue alamklassi vahel (vt tabel 3).

TABEL 3. Keelekasutusmustrite esinemus substantiivialgulise klassi alamklassides

Alamklass	Mustri- rühm	Sagedus	Osakaal	Koht substantiivi- alguliste mustrite pingereas
S-D	SVD	194	16%	1.
	SDD	94	8%	5.
	SKD	65	5%	8.
	SJD	24	2%	13.
S-S	SJS	171	14%	2.
	SSS	91	7%	6.
	SAS	47	4%	11.
	SVS	23	2%	14.
S-V	SSV	154	12%	3.
	SVV	87	6%	7.
	SKV	59	5%	9.
S-K	SSK	95	8%	4.
S-A	SVA	53	4%	10.
	SJA	16	1%	16.
S-J	SSJ	26	2%	12.
	SKJ	18	1%	15.

Järgnevalt kirjeldan alamklasside muustrirühmi neis sisalduvate mustrite kaupa. Käsitlusest jäävad välja mustrid, milles on ≤ 10 (kümme ja vähem) näidet. Sama skeemi olen kasutanud kõikide mustriklasse alamklassides esinevate mustrite valikul.

3.3.1. Adverbilõpuline alamklass

Invariantne muster on SVD (substantiiv-verb-adverb) – 194 näidet, osakaal 16%, esimene koht alamklassi mustrite sageduse pingereas. Tegu on morfosüntaktiliselt keeruka sõnaliigijärgendiga, mille kasutuses tuleb esile kolm malli:

- a) S nom sg/pl subjektina + V *olema* indic pres 3 sg/pl abiverbina + adverb (144 näidet 194-st) on arvamusaluste keelekasutuses produktiivne, nt *Castro on veel, meetodid on üsna*;
- b) S tr sg, S in sg, S ad sg käändelise määrusena + V *olema* indic pres 3 sg + adverb (25 näidet 194-st) on väheproduktiivne mall, nt *sihtmärgiks on ilmselt, kommünikees on vaid, ravikindlustatud on juba*;
- c) eitavas kõnes S nom sg subjektina + eitusverb *pole* + adverb (12 näidet 194-st) on mitteproduktiivne, nt *taristu pole seevastu, betoon pole ka, turvalisus pole vaid*.

Kahes jaatava kõne mallis varieeruvad substantiivi kääne, arv, funktsioon ja verbi grammatiline vorm. Sagedaima malli S nom sg/pl subjektina + V indic pres 3 sg/pl tüüpilisust kinnistab leksikaalne dominant V *olema*, nt *jagamine on ikka, Kohver on siiani, Savisaar on nõus; sõdurid on juba, tööandjad on valmis, erakonnad on samuti*. Leksikaalset varieerumist täistähenduslike verbidega esineb harva (15 näidet 194-st), nt *Uuring sedastab ka, Poliitika professionaliseerub järjest, erakond mureteb niivõrd, riik korraldab ka, Viga tuleb sisse, menetlus jääb pooleli; seadused kehtivad küll, Müüjad vahetuvad pidevalt, Majandusseadused kehtivad ka*. Veelgi marginaalsem ja pigem erandina võetav on *olema*-verbi ajavormi vaheldumine V impf 3 sg (9 näidet), nt *Vanarahvas oli väga, kõrgharidus oli vaid, sõnum oli toona*.

Väheproduktiivses käändelise määruse mallis esineb substantiiv kolmes vormis, millest sagedaim on S tr sg, nt *Põhjuseks on üha, Kannatajaks on reaalselt, tunnuseks on ka*. Selles funktsioonis soovitatakse translatiivi vältida, ent arvamislugude analüüs näitab vastupidist. Küsimus on muutusest normi ja uususe vahekorras, mis tuleb muustrina esile tänases publitsistlikus keelekasutuses.

Inessiivi ja adessiivi kasutamine käändelise määrusena on tava-pärane, nt S in sg – *kommünikees on vaid, Lõppkokkuvõttes on siiski, poes on vähe, täitemenetluses on üle*; S ad sg – *ravikindlustatul on juba, piima-tootjal on tarvis, presidendil on veel*.

Mustri SVD komponentide kasutuses ilmneb muudki huvipakkuvat.

Esiteks, substantiivide leksikaalne valik on rikkalik, ent samas teemapõhine. See on ka mõistetav, kuna arutluste keskmesse jäävad *inimene ja riik*. Kirjutatakse muutustest ja muudatustest, seadustest ja seadusandjatest, erakondadest, ideoloogiast, Eestist ja maailma probleemidest. Teemasõnad jagunevad rühmadesse: majandus (nt *naftavarud, naftatulud, piimatootja, piimatooted*), omavalitsused (*linn, vald, taristu*), riigikord (*riigikogu, president, poliitika, poliitikavaatlejad, oravapartei, otsustajad, liberaalid, liitlased, juhid*), juriidika (*seadusandja, seadused, täitemenetlus, menetlus, piirid, ravikindlustus*), töö ja raha (*tööandjad, tööpõhimõtted, tööstus, laenukoormus, hinnad, raha, rubla*) jne. Arutelud ühiskondlik-poliitilistel ja majanduslikel teemadel ei jäta ruumi idiomatikale: 194 näite seas leidis vaid üks väljend (*mekk jäi manu*).

Teiseks, silma hakkab rikkalik valik adverbe, nende kasutamine pole juhuslik, vaid seaduspärane (vt tabel 4), sest

- a) sõnakorduste arvu suurenedes väheneb korduvalt kasutatud adverbide hulk;
- b) sõnakorduste arvu suurenedes kitseneb adverbi kasutamine määrusena ja laieneb adverbi kasutamine rõhusõnana.

TABEL 4. Adverbikordused mustris SVD: täistähenduslikest sõnadest abisõnade suunas

Adverbikorduste arv						
2	3	4	5	6	7	13
eelkõige, ehk, ette, ilmselt, järjest, kirjas, lihtsalt, mitte, nõus, nüüd, pigem, piisavalt, rohkem, samuti, seal, seejuures, siiani, siiski, sisse, toona, valmis, välja, üksnes, üldiselt, üle	näiteks praegu suuresti tegelikult vaja veidi üha	hoopis nii pidevalt siin vaid	ju kindlasti veel	ikka juba väga	küll	ka

Tabelis 4 tuleb kujukalt esile sõnasageduse seos vormilt lihtsate, lühikeste ja abstraktsete adverbidega: mida suurem sõnakorduste arv, seda lühemaid ja abstraktsemaid adverbe keelekasutaja eelistab. Ning mida abstraktsemad on adverbid, seda rohkem kasutatakse neid abisõnana. See seletab ka mustri SVD alusel tugiverbikonstruktsioonide ja ühendverbide moodustumist ning kasutamist:

- a) levinud on tugiverbikonstruktsioon, nt *seadused on kaldu, kura did on kinni, Vastus on kirjas, faas on möödas/läbi, Savisaar on nõus, raha on otsas, kurikavalus on süüdi, piimatootjal on tarvis, meil on vaja, meeled on valla/avali*; üksiknäiteid leidub ka ahelverbide kohta (*menetlus jääb pooleli, Sõda käib edasi*);
- b) harvem esineb ühendverbe, kus mustris SVD partikkel järgneb põhiverbile, mille staatusele kandideerivad 15 täistähenduslikku verbi, nt *ülikoolid valmistavad ette, Feministid pressivad peale*; teine võimalus – partikkel eelneb põhiverbile, mis paikneb mustrist paremale hargnevas kontekstis, nt *Tegijad on välja kuulutatud, hinnad on üle kontrollitud, majanduskriis võib kaasa tuua*.

Variant SDD (substantiiv-adverb-adverb) – 94 näidet, osakaal 8%, viies koht sageduse pingereas. Substantiivil on erinevaid funktsioone, neist levinuim käändeline määrus:

- a) S in sg – kontekstis väga põhjalikult, läänes nii kaua, kaoses juba varem, riigikorras kuidagi rahul, ühiskonnas kindlasti mitte, Tartus juba praegu, Ukrainas jõudsalt edasi;
- b) S all sg – Maailmale ka lihtsalt, reakodanikule juba täiesti, ettevõtluskeskkonnale tublilt väljapoole, püüdlusele üsna selgelt;
- c) S el sg – mõtlemisest välja üsna, Rahast ehk isegi, olukorrast väga selgesti;
- d) S kom sg – teemaga lihtsalt sümboolselt, tööjõuga kaugeltki mitte. Järgmise rühma moodustavad näited, kus
- a) S nom sg (pl) kasutatakse subjektina – teadis komisjon väga hästi, kiitis vanematekogu eile heaks, risk kapitaalselt kokku vajuda, läks lapselaps kergelt mossi, ütles tüdruk nii siiralt; lepivad vanemad omavahel kokku, teevad juhid nii liiga;
- b) S nom sg/pl totaalobjektina – saada vastutus hiljem ka juhatajale, vaja töötada diskursus läbi juba homseks, jätke näitamine esialgu ära; kasutatakse isikud üsna ühemõtteliselt ära, sõnastatakse ideed ilmselt ümber, jagati ülesanded selgelt ära;
- c) S part pl partsiaalobjektina – sünnib lapsi üha rohkem, kutsutakse inimesi üles skeptiliselt suhtuma, kutsutakse välismaalasi kohe palju;
- d) S gen pl adpositsioonifraasis – juhtivtöötajate hulgas üle, täiskasvanute seas läbi, häälte nimel ka, reaalteaduste poole kaldu, saadikute kaudu rohkem, lahenduste osas juba, läbirääkimiste poole pealt;
- e) S nom sg predikatiivina – on vesi juba lõpukorral, oli märk mitte just.

Substantiivi rikkalik käändevarieeruvus ja funktsionaalne mitmekesisus sõltub verbi reksioonistruktuurist ja grammatilisest vormist lauses, millesse muster SDD kuulub, nt imperatiiv + totaalobjekt S nom sg (jätke näitamine esialgu ära). Substantiivi kasutust mõjutab ümbritsev kontekst, mistõttu pole kerge tuvastada ega sõnastada pealtnäha lihtsa mustri kasutusreegleid. Võib eeldada, et tegu on struktuuri, vormi ja funktsiooni tasakaalustatusega, mis toimib muutumatuid sõnu (adverb,

adpositsioon, konjunktsioon) sisaldavate muustrite põhjal vaikimisi. Seega ei saa keelekasutusmustreid liigitada lihtsateks ega keerulisteks, vaid standardse (dominantse, tüüpilise, stereotüüpse) ja varieeruva vormistusega muustriteks.

SDD puhul on tegemist grammatiliselt varieeruva polüfunktsionaalse muustriga, mille kontekst mõjutab adverbi kasutamist. Tihtilugu raamib muustrit ühendverb ning põhiverb paikneb muustrile eelnevas kontekstis, nt jagati *ülesanded selgelt ära*. Samas võib põhiverb jääda ka muustrile järgnevasse konteksti, nt *risk kapitaalselt kokku vajuda*. Mõlemad sõnajärjemallid on võimalikud nagu ka SVD puhul. Vaja vaid täpsustada, mis tingimustel. Siinses analüüsis piirdun seega adverbi sageduse, korduvkasutuse, funktsiooni ja positsiooni vahel:

- a) muustri lõpus korduvalt kasutatud adverbid on seaduspäraselt verbipartiklid (nt *välja, ära, läbi, kokku, üle*);
- b) olenemata positsioonist on ainukordsed adverbid erinevat liiki määrused: viisi- ja seisundimäärused (nt *põhjalikult, keskmiselt, emotsionaalselt, rahul*), aja- ja kohamäärused (nt *eile, varem, kaua; siit, hulgas, mujal*), määra ja hulka rõhutavad sõnad (nt *üsna, kõige, hästi, vaid, liiga*) jm.

Variant SKD (substantiiv-adpositsioon-adverb) – 65 näidet, osakaal 5 %, kaheksas koht sageduse pingereas. Muster sisaldab adpositsioonifraasi, kus kasutatakse reeglina S gen sg (pl) – *info põhjal emotsionaalselt, võimu juurest eemale, sammu võrra ees, valija jaoks lahti, kibrulehe alt välja, lahendamise asemel mahukalt; spetsialistide jaoks hiljemalt*. Korduvad järgmised adpositsioonid: *sees, ajal* (4); *kõrvalt, kohta, jaoks, asemel* (3); *vältel, vastu, suhtes, poolt, korral, järgi, juures, juurde, eest, ees* (2 korda). Kivinenud käändevormidest adpositsioonid moodustavad leksikaalsemantilisi rühmi nagu paiknemine-suundumine: *all~alt; ees~ette~eest; juures~juurde~juurest; kõrval~kõrvalt; taga~tagant*. Kui kõrvutada omavahel ainukordsete adpositsioonide ja adverbide arvu muustris, siis on adverbe rohkem – *välja* (4), *mitte* (3) ja *hoopis, isegi, juba, ka, kas, näiteks, siiski, umbes, vastu* (2 korda). Nende kasutamisest

annab selgema pildi mustri SKD paremale ja vasakule hargnev kontekst, mida siinkohal ei käsitleta.

Variant SJD (substantiiv-konjunktsioon-adverb) – 24 näidet, osakaal 2%, kolmeteistkümnes koht sageduse pingereas. Analoogselt mustri SDD on ka siin substantiivil üsna avar käände ja funktsiooni varieeruvus:

- a) S nom sg subjektina – *värbamistöo ja edasi-tagasi, tugevus ja ka, sõnavabadus ja ehk, väikeriik ning mitte*;
- b) S part sg partsiaalobjektina – *nii reguleeritust kui ka, palvet ja pisut, ei suhkrut ega üldse, likvideerimist ja vaid, väljakuulutamist või vähemalt*;
- c) S gen sg eestäiendina – *nii maksureformi kui ka, organisatsiooni või ka*;
- d) S in sg käändelise määrusena – *olukorras või lihtsalt, müügistrateegias ning laiemalt*. Kasutatakse rinnastavat seost, sh paarisidendit *nii ... kui ka*, milles adverb kuulub sidendi koosseisu.

3.3.2. Substantiivilõpuline alamklass

Invariantne muster on SJS (substantiiv-konjunktsioon-substantiiv) – 171 näidet, osakaal 14%, teine koht sageduse pingereas. Rinnastava sidesõnaga (stereotüüpselt *ja*, üksikjuhul *ning, või*) pealtnäha lihtsa mustri kasutamise muudab keerukaks substantiivi käändevalik, käänete omavahelised kombinatsioonid ja mitmekesised funktsioonid. Selles on sarnasust juba kirjeldatud adverbilõpulise alamklassi mustritega.

- a) S nom sg (pl) on tavaliselt subjekti funktsioonis – *liha ja veri, Sisekontroll ja järelevalve; poliitikud ja metsandusettevõtjad, väljamõtled ja propageerijad*. Harva esineb juhtumeid, kus mustri esimene komponent on S nom sg ja kolmas S nom pl – *Kirjeldus ja ettepanekud, meepott ja sussid, Keelekasutus ja kujundid*. Erandina kuuluvad siia näited, kus kontekstist olenevalt võib substantiiv olla nii subjekti, eestäiendi (*kaitsejõud ja kaitseministeerium, palk või pension, riik ja rahvas*) kui ka

predikatiivi funktsioonis (*demokraatia ja kaasamine, haridus ja teadus, vallajuht ning linnapea*).

- b) Harvem kasutatakse substantiivi objektina:
- S nom pl totaalobjekt – *inimesed või grupid, sõnavõtud ja kommentaarid, lätlased ja leedulased*;
 - S part sg/pl partsiaalobjekt – *vaesust ja üksikvanemat, suveräänsust ja iseseisvust; kogemusi ja oskusi, lapsi ja noori, maksusoodustusi ja toetusi*;
 - S gen sg võib omada rööpselt nii totaalobjekti kui ka eestäiendi funktsiooni – *lääne ja ida, hea ja kurja, õe ja venna, kiriku ja riigi*.
- c) S gen sg/pl on ees-, harvem järeltäiend – *Kodeerimisprotsessi ja dekonstruktsiooni, õnne ja tervise, päästeameti või politsei; sõjalennukite ja väeüksuste, äratundmiste ja lootuste, muutuste ja uuenduste, eestlaste ja lätlaste*.
- d) S gen sg adpositsioonifraasi koosseisus, adpositsioon jääb mustrist paremale hargnevasse konteksti – *solvamise ja solvumise tõttu, valitsuse ja riigikogu ees, suurendamise ja osalusdemokraatia poolt, osavuse ja tarkuse pärast, riigi ja rahva vastu, sõja ja rahu eest*.
- e) S el sg käändelise määruse ja järeltäiendina – *retoorikast või esteetikast, ravist ja ravimisest, riigist ja rahvusest, suurusest ja ülevusest, küberjulgeolekust ning küberkaitsest*.

Niisiis muudab ka selle väliselt lihtsa mustri keerukaks substantiivide rikkalik käändevaheldus, polüfunktsionaalsus ja erinevad käändekombinatsioonid. Mustri kasutamine sõltub lause kontekstist.

Variant SSS (substantiiv-substantiiv-substantiiv) – 91 näidet, osakaal 7%, kuues koht sageduse pingereas. Muster rajaneb järjestikuste substantiivide kooskasutusele eestäiendi funktsioonis. Selleks on kaks malli, milles esineb erinevaid käändekombinatsioone.

- a) Kaks kolmandikku näidetest sisaldab eestäiendina kahte esimest substantiivi, kolmas on reeglina S nom sg subjekti funktsioonis. Eestäiendi vormistamisel on kasutatud erinevaid

käändekombinatsioon: S gen sg + S gen sg (*Elatisraha tagastamise fond, Viimsi valla laenukoormus*), S gen pl + S gen sg (*maksuliikide kasvuvaenulikkuse järjestus, nõuete täitmise kohustuslikkus*), S nom sg + S nom sg (*Carl Robert Jakobson, teadusajakirjanik Tiit Kändler*), S nom sg + S gen sg (*Eero Epneri tekst, Andrei Portnovi vanaisa*).

- b) Üks kolmandik näidetest sisaldab erinevaid käändekombinatioone eestäiendi funktsioonis: S gen sg + S gen sg + S gen sg (*nafta hinna vähenemise*), S gen sg + S nom sg + S gen sg (*juhi Andrus Ansipi*), S gen sg + S nom sg + S nom sg (*lastekaitse peaspetsialist Kati*).

Tegu on jällegi lause kontekstist sõltuva muustriga, kus lisaks kolme substantiivi rikkalikule vormivarieeruvusele esineb erinevaid käändekombinatioone.

Variant SAS (substantiiv-adjektiiv-substantiiv) – 47 näidet, osakaal 4%, üheteistkümnes koht sageduse pingereas. Mustri esimene ja teine komponent on eestäiendi funktsioonis, kolmandat komponenti kasutatakse peamiselt subjekti, harvem nimisõna eestäiendi ja käändelise määrusena. Vastavalt varieerub ka substantiivi ja adjektiivi kääne. Mustril SAS on kolm morfosüntaktilist malli:

- a) S gen sg + A nom sg + S noms sg subjektina (*inimese heitlik mälu, Asja sisuline külg*);
- b) S gen sg + A gen sg + S gen sg eestäiendina (*raamatu postmodernse poeetika, Kunstikooli uue õppehoone*);
- c) S gen sg + A tr sg + S tr sg käändelise määrusena (*majanduse peamiseks mootoriks, Ajateenistuse nõrgaks kohaks*).

Substantiivide ja adjektiivide kooskasutuses tuleb esile translatiiv käändelise määruse funktsioonis, mida tuleks vältida, asendades translatiivi nominatiiviga: *ajateenistuse nõrgaks kohaks on .. ~ ajateenistuse nõrk koht on ..* (vt eespool adverbilõpulise alamklassi mustri SVD sama-laadne kasutamine).

Variant SVS (substantiiv-verb-substantiiv) – 23 näidet, osakaal 2%, neljateistkümnes koht sageduse pingereas. Sel suhteliselt harva

esineval muustril on oma leksikaalgrammatiline dominant – abiverb *olema* indic pres 3 sg. Mõlema substantiivi kääne ja funktsioon vahelduvad. Selle põhjal eristub kaks morfosüntaktilist malli:

- a) S nom sg subjektina – S nom sg predikatiivina / S gen sg nimi-sõna eestäiendina (*vastavalt vabadus on hüve, valuuta on stabiilsus, Sotsiaalliberalism on ideoloogia; fond on riigi omandus, teema on keskkonnaministeeriumi valdkond*);
- b) S ad sg käändelise määrusena – S nom sg predikatiivina (*Lõpuseringel on uurimisprojekt, riigil on õigus, vanemal on kohustus*). Kee-rukaks teeb mustri kasutamise substantiivide käändevarieeruvus ja erinevates funktsioonides kasutatud käänete omavaheline kombineerumine (vt analoogsed juhtumid eespool).

3.3.3. Verbilõpuline alamklass

Invariantne muster on SSV (substantiiv-substantiiv-verb) – 154 näidet, osakaal 12%, kolmas koht sageduse pingereas. Mustri leksikaalgrammatiline dominant on abiverb *olema* indic pres 3 sg, kuid paaris-kolmes näites kasutatakse modaalseid abiverbe *võima* ja *pidama*, kahekümnes näites 154-st ka erinevaid täistähenduslikke verbe (*tasub, käib, maksab, teeb, kinnitab, manipuleerib, manifesteerib, sedastab, kahekordistab, ütleb, suurendab*), nt *parlamendi põhitöö käib, Juhiloa vahetus maksab, Kaitseliidu liige saab, Raha jagamine teeb, Tehnikaülikooli kogemus kinnitab*.

Substantiivide käände vahelduse ja funktsioonide varieerumise alusel tuleb esile kaks tüüpilist vormistust:

- a) S gen sg (pl) eestäiendina + S nom sg (pl) subjektina – *üksikvanema sissetulek on, toornafta hind on, kohtuistung algus on, Ärimaailma reaalsus on, Fortumi käive on, Artemi Troitski on; Sanktsioonide mõju on, Maksude laekumine on, Valmiskampaaniate loogika on; Välismaa palganumbrid on, täiskogu istungid on*;
- b) S tr sg käändelise määrusena (7 korda) – *Semiootika objektiks on, dekonstruktsiooni tagajärjeks on, seinapannoo teemaks on*. Seda

malli on soovitatud vältida, asendades translatiivi nominatiiviga (*semiootika objektiks on ~ semiootika objekt on*). Tänapäevases publitsistlikus keelekasutuses tuleb aga translatiivne mall esile erinevates mustrites (SVD, SAS, SSV) ning on muutumas järjest produktiivsemaks.

Variant SVV (substantiiv-verb-verb) – 87 näidet, osakaal 6%, seitsmes koht sageduse pingereas. Seda mustrit kasutatakse nii eitavas kui ka jaatavas kõnes, S nom sg/pl on tüüpiliselt subjekti funktsioonis (*nimi ei ole, oligarhid ei ole, lahendamine ei jää, vool ei lakka, riik ei saa, kirjanik ei pea, riik ei piira, oskused ei kao, taimetoitlased ei kelgi, tuhanded ei keerle*). Mustri morfosüntaktiline ja leksikaalne varieerumine seondub verbiga:

- a) sagedaim on eitava kõne preesens (*ei ole 21, ei saa 9, ei pea 5, ei tee 2 korda, ülejäänud 25 verbi on kasutatud ainukordselt*);
- b) kolmandik näidetest sisaldab perfekt (i) (*ekspresident on tunnistanud, ravikindlustus on taganud, kirikukogu on langetanud, Riik on teinud, läänemaailm on tegutsenud, hirm on olnud, Missioon on kulgenud*);
- c) harva kasutatakse ka ahelverbi *võib olla* (*Tagajärg võib olla, sõidukiirus võib olla, tunne võib olla*).

Niisiis tugineb mustri kasutamine ühelt poolt leksikaalgrammatilisele dominantile S nom sg/pl subjekti funktsioonis ja teisalt verbivormi varieeruvusele jaatavas ja eitavas kõnes.

Variant SKV (substantiiv-adpositsioon-verb) – 59 näidet, osakaal 5%, üheksas koht sageduse pingereas. Mustri morfosüntaktiline dominant on adpositsioonifraas. S gen sg (pl) kasutatakse tavaliselt koos V indic pres 3 sg, v.a seitse eitava kõne näidet (nt *ukse vahele ei*). Mustri leksikaalne varieerumine on piiratud:

- a) kasutatakse vaid seitset täistähenduslikku verbi (*kõlab, tuleb, paistab, jääb, keerdub, libiseb, kehtib – pereliikme kohta jääb, juhiloa puhul kehtib*), ülejäänud näited on *olema*-verbiga (*majanduse kõrval on, Perekonnaseaduse järgi on, Aastate jooksul on, laste jaoks on*);

- b) pooled adpositsioonid on kasutuses korduvalt – *järgi, kohta* (6), *puhul* (5), *vahel(e), jaoks* (4), *pärast, taga* (3), *ette, eest, kõrval, suhtes, tõttu* (2 korda), nt *Rusikareegli järgi on, elaniku kohta on, Video puhul on, laste jaoks on, soolestiku vahel on, pöördumise taga on, minuti pärast on, juhi eest on, majanduse kõrval on, töö suhtes on, soovi tõttu ei*.

3.3.4. Adpositsioonilõpuline alamklass

Selle alamklassi ainus ja **invariantne muster on SSK (substantiiv-substantiiv-adpositsioon)** – 95 näidet, osakaal 8%, neljas koht sageduse pingereas. Olenevalt esimese substantiivi funktsioonist jaguneb muster kahe morfosüntaktilise malli vahel:

- a) S gen sg/pl eestäiendina adpositsioonifraasi kuuluva S gen sg ees, nt *infosõja udu vastu, Vabariigi põhiseaduse järgi, võrdsuse koorma alla, kanali loomise asemel, rubla nõrkuse tõttu, üürimajade ehitamise vastu*;
- b) S nom sg subjektina koos adpositsioonifraasi kuuluva S gen pl, nt *mees kuulajate seas, reisimine ida poole*. Adpositsioonidel võib näha kinnistumise tendentsi: ainukordseid on 11, korduskasutatuid 21 (sh kohakäänete kinnistunud vormid nagu *kõrvale-kõrvalt, peal-peale-pealt, all-alla, taga-taha, vahel-vahela, ees-ees*). Näiteks: *kohta* (10), *ees(t)* (8), *vastu, poole, osas, all(a), kaudu* (5), *kõrval (kõrvale, kõrvalt), suhtes* (4), *tõttu, peal (peale, pealt), taga (taha), pärast* (3), *üle, vahel(e), korral, jaoks, asemel, arvelt, ajal* (2 korda).

3.3.5. Adjektiivilõpuline alamklass

Invariantne muster on SVA (substantiiv-verb-adjektiiv) – 53 näidet, osakaal 4%, kümnes koht sageduse pingereas. Tegu on stereotüüpse mustriga, kus S nom sg/pl on alati subjekt, V *olema* indic pres 3 sg abi-verbis funktsioonis ja A nom sg/pl tavaliselt predikatiiv, harva eestäiend

(*kaup on taskukohane, reaktsioon on mõistetav, Vastus on lihtne, eliitkoolid on efektiivsed, Küsimused on reaalsed, kunstisündmused on harvad; võte on demagoogiline, toetus on märgiline, liit on kindel, Kirik on vaba*). Substantiivide ja adjektiivide leksikaalne varieeruvus on rikkalik.

Variant SJA (substantiiv-konjunktsioon-adjektiiv) – 16 näidet, osakaal 1%, kuueteistkümnes koht sageduse pingereas. Varieerub substantiivi kääne ja funktsioon:

- a) S nom sg subjektina – *Taavet ning rohmakas, haridus ja erialane, tööelu ja suur;*
- b) harvem S gen sg eestäiendina – *sidususe ja sotsiaalse, kursi ja ülikiire, energialiidu ning digitaalse.*

Kasutatakse rinnastavat seost (tavaliselt *ja*, harva *ning*). Muster seob omavahel osalauseid.

3.3.6. Konjunktsioonilõpuline alamklass

Invariantne muster on SSJ (substantiiv-substantiiv-konjunktsioon) – 26 näidet, osakaal 2%, kaheteistkümnes koht sageduse pingereas. Mustri esimest komponenti S gen sg (harva S nom sg) kasutatakse eestäiendi funktsioonis (*maailmavaate kriisi ja, Kersti Nigesen ja*) ning teist S gen sg/nom sg samuti eestäiendina (*mõistuslikkuse printsiibi ning, Terminali rajamise ja, töökorra seadus ja, omaniku raha ja*). Olenevalt kontekstist võib S nom sg olla ka subjekti funktsioonis, nt *omaniku raha ja* töötajate professionaalsus tagavad ettevõtte edukuse. Muster lõpeb tavaliselt rinnastava sidesõnaga *ja* (harva *ning, või*).

Variant SKJ (substantiiv-adpositsioon-konjunktsioon) – 18 näidet, osakaal 1%, viieteistkümnes koht sageduse pingereas. Tegu on lihtsa adpositsioonifraasiga, nt *lisandväärtuse suunas ning, linna piires ega, summa eest kui, maagaasi suhtes ning, maa peal ning, lapse juures ja, ukse ees ja*. Reeglina lõpeb muster rinnastava seosega (*ja, ning, ega, kui, või, justkui*).

3.4. Keelekasutuse seaduspärasustest substantiivialgulise klassi näitel

Substantiivialguliste mustrite kvalitatiivne lingvistiline analüüs tõi esile kaheksa juhtumit: ühed on seotud mustri vormistuse, teised keelekasutuse üldküsimumustega.

3.4.1. Vormistus

Lisaks püsistruktuurile eristavad mustreid nende morfosüntaktilise ja leksikaalse varieeruvuse piirid, mida kirjeldades saab tugineda stereotüüpse, tüüpilisuse ja dominantsuse mõistetele.

Stereotüüpse vormistusega sõnaliigijärjendeid, nagu SVA, kasutatakse ühe morfosüntaktilise malli alusel: S nom sg/pl on alati subjekt + V *olema* indic pres 3 sg/pl abiverb + A nom sg/pl predikatiiv (*kaup on taskukohane, reaktsioon on mõistetav*). See muster kuulub näiteks eesti õppijakeele viieteistkümneme mustri hulka, mida kasutatakse läbivalt kõikidel keeleoskustasemetel (vt Hallik 2015), kusjuures mustri osakaal kahaneb sujuvalt A2-tasemelt C1-taseme suunas. Analoogne muster on ka SKV: adpositsioonifraasile (S gen sg + K) järgneb reeglina V indic pres 3 sg (*elaniku kohta on, Aastate jooksul on*). Mõlema mustri stereotüüpset vormistust kinnistab leksikaalgrammatiline dominant V *olema* indic pres 3 sg.

Dominantse vormistusega sõnaliigijärjendeid, nagu SSV, kasutatakse tavaliselt ühe, selgelt sagedama ja produktiivsema morfosüntaktilise malli alusel kolmest võimalikust: S gen sg eestäiendina + S nom sg subjektina + V *olema* indic pres 3 sg (*kohtuistung algus on*). See on mustri SSV domineeriva vormistusega mall, millel lisaks leksikaalgrammatiline dominant V *olema* indic pres 3 sg. Sama võib täheldada mustri SVD puhul, mille sagedam morfosüntaktiline mall on S nom sg/pl subjekt + V *olema* indic pres 3 sg/pl abiverbi funktsioonis + adverb (*Saissaar on nõus, sõdurid on juba*). Keerukama morfosüntaktilise dominandi näide tuleb esile mustri SVV kasutuses, milles lisaks vaheldub kõneliik

ja verbi grammatiline ajavorm: S nom sg/pl subjekti funktsioonis + V *olema* indic neg pres / indic perf (*nimi ei ole, ekspresident on tunnistanud*). Nii SSV kui ka SVD kuuluvad eesti õppijakeele viieteistkümne läbivalt kasutatud mustri hulka, kusjuures mõlema osakaal kasvab B1- ja B2-tasemel ning langeb C1-tasemel.

Hakkab silma, et analoogselt stereotüüpsete morfosüntaktiliste mallidega kasutatakse ka siin leksikaalgrammatilist dominantide *V olema* indic pres 3 sg. Samas pole see omane üksnes substantiivialgulisel klassil verbi sisaldavatele sõnaliigijärjenditele, vaid iseloomustab eesti keele kasutust tervikuna. Nii on tasakaalus korpuse lemma-sõnavormide sageduse koondloendi tipus sõna *olema* (639 802 kordust) ja selle sagedaim vorm indic pres 3 sg (315 507 kordust)².

Kahe tüüpilise vormistusega sõnaliigijärjendeid, nagu SJA, kasutatakse morfosüntaktiliste mallide alusel, mille sagedus oluliselt ei erine: S nom sg subjekti funktsioonis + rinnastav sidesõna (tavaliselt *ja*) + A nom sg eestäiendina (*haridus ja erialane, tööelu ja suur*). Peaaegu samaväärselt esineb teinegi mall S gen sg eestäiendina + rinnastav sidesõna (tavaliselt *ja*) + A gen sg eestäiendina (*sidususe ja sotsiaalse, kursi ja ülikiire*). Niisiis saab järjendit SJA pidada kahe tüüpilise vormistusega mustriks, millel oma grammatiline dominant (rinnastav sidesõna *ja*) ning kahte malli eristab vaid substantiivi kääne. Samamoodi on mustriks SSK kaks tüüpilist vormistust: S gen sg/pl eestäiendina + S gen sg + K (*infosõja udu vastu, Vabariigi põhiseaduse järgi*), pisut harvem S nom sg subjektina + S gen pl (*mees kuulajate seas, reisimine ida poole*). Mõlemas mallis kasutatakse stereotüüpset adpositsioonifraasi, mille eristab mustri esimese substantiivi kääne ja funktsioon.

Varieeruva vormistusega sõnaliigijärjendid sisaldavad keerukat vormivaheldust. Substantiivialgulisel klassil mustrites tuleb ühelt poolt esile substantiivi avar käändevarieeruvus ja polüfunktsionaalsus, teisalt aga erinevad käändekombinatsioonid, mis võivad omakorda varieeruda erinevate mustrite või ka sama mustri kõikide morfosüntaktiliste

² Vt https://cl.ut.ee/ressursid/sagedused1/failid/lemma_sonavorm_kahanevas.txt (30.4.2022).

mallide piires. Üks lihtsamaid näiteid on translatiivse käändelise mää-
ruse kasutusmallide vaheldumine kolme mustri alusel: SVD (S tr sg +
V *olema* indic pres 3 sg + adverb), SAS (S gen sg + A tr sg + S tr sg) ja
SSV (S gen sg + S tr sg + V *olema* indic pres 3 sg), nt *tunnuseks on ka*,
Ajateenistuse nõrgaks kohaks on, seinapannoo teemaks on. Teistlaadne
varieeruvuse juhtum ilmneb mustris SJS, kus substantiivi kasutatakse
subjekti (nom sg), totaalobjekti (nom pl, gen sg) või partsiaalobjekti
(part sg/pl), ees- või järeltäiendi (gen sg/pl) ning käändelise mää-
ruse (el sg) funktsioonis, nt *Keelekasutus ja kujundid, lätlased ja leedulased*,
kogemusi ja oskusi, lääne ja ida, eestlaste ja lätlaste, osavuse ja tarkuse
pärast, ravist ja ravimisest. Muster SSS on aga markantne näide rik-
kalikest käändekombinatsioonidest eestäiendi funktsioonis. Järjendit
rakendatakse kahe morfosüntaktilise malli alusel, esimeses neli ja tei-
ses kolm kombinatsiooni (vt eespool substantiivilõpuliste mustrite
alamklass).

3.4.2. Keelekasutuse üldiseid seaduspärasusi substantiivialgulise klassi näitel

Üks osa seaduspärasusi seletab sõnade kooskasutust, st mustrite funk-
tsionaalset võimekust tekstis, teine osa hõlmab keelekasutust tervikuna,
kolmas mustrite struktuuri alusel kulgevaid keelelisi arenguid. Nende
kolme aspekti käsitlemine üksteisest lahus tuleneb objekti kirjeldamise
vajadustest, kuid olemuslikult on nad omavahel põimunud.

**Sõnade kooskasutuse seaduspärasused (mustri funktsionaalne
võimekus).** Sõnaliigijärjendi SVD põhjal tuli esile vähemalt kaks adverb-
kasutamisele omast tendentsi:

- a) sõnakorduste arvu suurenedes väheneb korduvalt kasutatud
adverbide hulk;
- b) adverbikorduste kasvades väheneb nende kasutamine määru-
sena (nt *ilmselt, lihtsalt, piisavalt, seejuures, seal, toona*) ning laie-
neb piiratud arvu abstraktsete adverbide kasutamine rõhusõna
funktsioonis (nt *ka, küll, ikka, väga, ju*).

Need seaduspärasused kerkisid esile adverbide sageduse põhjal (vt tabel 4): mida suurem on ühe või teise adverbi korduste arv, seda lähemaid ja abstraktsemaid adverbe kasutatakse, mis omakorda soodustab nende funktsioneerimist abisõnana (kas rõhusõna või partiklina). See seletab analüütiliste verbide ja analüütiliste konstruktsioonide esinemust mustri SVD alusel, kus enim levinud on küll tugiverbikonstruktsioon (*seadused on kaldu, kuradid on kinni*), kuid kasutatakse ka ühendverbe, milles verbipartikkel järgneb põhiverbile (*ülikoolid valmistavad ette*), kuuludes mustri SVD koosseisu. Liitajavormi puhul jääb põhiverb mustrist paremale hargnevasse konteksti (*Tegijad on välja kuulutatud*). Niisiis moodustatakse ja kasutatakse järjendi SVD morfosüntaktilise malli S nom pl + V *olema* indic pres 3 pl + adverb alusel analüütilisi sõnu, analüütilisi vorme ning analüütilisi konstruktsioone. Erinevus puudutab mustri lähikonteksti sõnaliikide järjestust (vt eespool).

Järgmine sõnade kooskasutuse seaduspärasus õnnestus leida mustri SDD keelenäidete põhjal. Avanes seos adverbi funktsiooni ja adverbi korduvkasutuse ning positsiooni vahel:

- a) mustri lõpus korduvalt esinenud adverb on reeglina verbipartikkel (*välja, ära, läbi, kokku, üle*), sellele eelnev adverb aga tüüpiliselt viisi- või seisundimäärus;
- b) kui mustrit raamib ühendverb (*jagati ülesanded selgelt ära*), siis jääb põhiverb mustrist vasakule hargnevasse konteksti; kui kasutatakse ühendverbi liitajavorme, siis jääb põhiverb mustrist paremale hargnevasse konteksti (*risk kapitaalselt kokku vajuda, on ülesanded selgelt ära jaotatud*);
- c) ainukordsed adverbid on positsioonist sõltumata erinevat liiki määrused: viisi- ja seisundimäärused (nt *põhjalikult, keskmiselt; emotsionaalselt, rahul*), aja- ja kohamäärused (nt *eile, varem, kaua; siit, hulgas, mujal*), määra ja hulka rõhutavad sõnad (nt *üsna, kõige, hästi, vaid, liiga*) jm.

Niisiis sõltub adverbi funktsioon mustris positsioonist ja sagedusest. Ühendverbide kasutus erinevates ajavormides on piiratud mustrist vasakule ja paremale hargneva konteksti sõnajärjega.

Keelekasutuse üldised seaduspärasused ilmnevad mõnikord üsnagi ootamatutes seostes. Analüüsi tulemustest nähtub, et verbita substantiivialgulised mustrid (eriti koos muutumatute sõnadega) on võrreldes verbi sisaldavate muustritega kontekstitundlikumad. Nii iseloomustab pealtnäha lihtsa mustri SDD kasutust rikkalik käändevarieeruvus. Seletasin seda eelpool verbi grammatika ja rektsioonistruktuuriga lauses, millel on substantiivi käändevaheldusega otsene seos. Analoogne nähtus tuleb esile teistegi verbita muustritega, nagu SJS (tüüpiliselt rinnastava sidesõnaga *ja*) ning SJD, mis samuti on avara morfosüntaktilise ja funktsionaalse varieeruvusega. Ka mustri SSS substantiividel esineb eestäiendi funktsioonis väga keerulisi käändekombinatsioone (kahe morfosüntaktilise malli kohta kokku seitse).

Tekib küsimus, kuidas selliseid juhtumeid seletada? Üks võimalusi on näha siin keelekasutuse *tasakaalustatuse põhimõtte* rakendumist: muutumatuid sõnu või sama sõnaliiki sisaldavate muustrite vormilist lihtsust tasakaalustab substantiivi rikkalik käändevaheldus. See on osa inimese keelekasutust reguleerivast üldisest tasakaalust keelelise varieeruvuse ning keelekasutaja ratsionaalsete valikute vahel.

Põhimõtteliselt sama kinnitavad ka ilukirjanduskeele alusel tehtud verbi ümbruse muustrite uurimused (nt Eslon 2014b; Trainis 2017; Eslon 2017b). Nt verbist vasakule hargneva konteksti sagedamal ja lihtsana tunduval mustril DDV avaneb rikkalik morfosüntaktiline ja leksikaalne varieerumine. Vahelduvad aktiivi/passiivi mineviku liitajavormid (abisõna jääb muustrist vasakule hargnevasse konteksti, nt on *põhjalikult järele päritud*, olid *seal varemgi olnud*); *da*-infinitiiv on infiniitse öeldise (nt *kuhugi ära peita*), objekti (nt tahab *veelgi üle lugeda*) või subjekti funktsioonis (nt vaja *ajutiselt sisse seada*); *ma*-infinitiivi kasutatakse infiniitse öeldisena (nt *veel täna minema*); finiiitverbi ajavormid piirnevad V indic impf/pres 3 sg (nt *Ju siis puudus, üha enam kaugeneb*). DDV struktuuri alusel moodustatakse ja kasutatakse ühendverbe, nende leksikaalne varieerumine on rikkalik.

Analoogne tasakaal valitseb ka loominguilise ja standardse keelekasutuse vahel. Sel on oma tähendus muustrite klassifitseerimisel: mustrid

liigituvad standardse (tüüpilise, stereotüüpse) ja varieeruva vormistusega järjenditeks.

Mustrite alusel toimuvate keeleliste arengute näiteid on substantiivialgulises klassis mitmeid. Rääkides eelpool adverbide kasutamise seaduspärasustest, ei saa mööda minna sõnaliigijärjenditest, mille püsistruktuuri alusel moodustatakse ja kasutatakse erinevaid liitseid üksusi (nt SVD põhjal analüütilisi ühendverbe, tugiverbikonstruktsioone). Mustris SJD tuleb rööpselt rinnastavate sidesõnadega esile liitne paarisidend *nii ... kui ka* (adverb kuulub sidendi koosseisu). Translatiivi kasutus käändelise määrusena laieneb kolme erineva struktuuri alusel: SVD (*sihtmärgiks on ilmselt ~ sihtmärk on ilmselt*); SAS (*ajateenistuse nõrgaks kohaks on .. ~ ajateenistuse nõrk koht on ..*) ja SSV (*semiootika objektiks on ~ semiootika objekt on*). Emakeelekõneleja kasutab erinevate mustrite translatiivset malli automaatselt ega näe siin midagi normivastast.

Toodud näited kuuluvad eri laadi keeleliste arengute alla:

- a) liitsete keeleüksuste puhul on tegemist keele sisemise arengutendentsiga kord-korralt suurema analüütilisuse suunas;
- b) translatiivi laienemine käändelise määruse funktsioonis osutab muutusele või nihkele normi-uususe vahekorras.

4. Kokkuvõte

Keelekasutusmustreid saab käsitleda samas reas muud liiki mitmesõnaliste kooslustega, olgu need leitud mistahes meetodil või automaatse analüüsi vahendi abil. Muster rajaneb sõnaliikide loomulikule järgnevusele tekstis, moodustades püsikindla morfosüntaktiliselt tervikliku struktuuri, mis sulandub automaatselt lause konteksti. Tänu kognitiivsele võimele teeb emakeelekõneleja loomulikke otsuseid sõnavara ning mustri vasak- ja paremkonteksti valikul. Selles on abi ka teadmisest (aktsepteeritud sõnastused, grammatika- ja õigekirjareeglid). Keeleõppijale aga on mustri struktuur koos morfosüntaktiliste mallidega terviklik keeleüksus, mille alusel omandada sihtkeele aktiivset grammatikat, rikastada sõnavara, seostada erinevaid mustreid nii lause kui

ka pikema tekstilõigu tasandil. Selleks pakuvad andmekaave põhimõttel eraldatud keelekasutusmustrid uudset ja süstematiseeritud teavet, mille esitus rajaneb klasside – alamklasside – muustrirühmade – mustrite ja nende morfosüntaktiliste mallide hierarhiale. Niisugune klassifitseerimistehnika võimaldab analüüsida keelekasutust objektipõhiselt erinevatel lingvistilistel tasanditel, võrrelda mustreid ja siduda need erinevate kontekstidega mustrist vasakul ja paremal.

Iga mustri püsistruktuur võib olla vormistatud kas ühe või mitme morfosüntaktilise malli põhjal. Esimesel juhul on tegemist stereotüüpse vormistusega. Sihtkeele õppes on mõttekas alustada just seda liiki mustritega, sest nende alusel sõnu varieerides rikastub õppija sõnavara ning kinnistub üks standardne võimalus sõnaliike ja vorme kombineerida. Teisel juhul (mitu morfosüntaktilist malli) on erinevaid vormistusi:

- a) üks mall on ülejäänud mallidest selgelt sagedam ehk domineeriv – keeleõppe otstarbel omandatakse esmalt just selle malli kasutamine;
- b) kahte enam-vähem võrdset esinemust omava malliga ehk tüüpilise vormistusega mustrid võiksid tulla kasutusse edasijõudnud keeleõppija tasemel;
- c) mustrid, mille morfosüntaktiliste mallide varieeruvusega kaasnevad erinevad ekspressiivsed ja pragmaatilised tähendusvarjundid, oleksid mõeldud pigem iseseisvale keelekasutajale.

Lisaks leidub mustrites eelpool nimetatud juhtumitel leksikaalseid ja grammatilisi dominante, mis võivad olla keelekasutusele üldiselt omased (nt V indic pres 3 sg) või siis kinnistunud kindla sõnaliigijärjendiga (nt erinevad sõnavormide kombinatsioonid käändelise määruse funktsioonis). Vormistuse stereotüüpsus, tüüpilisus, dominantsus ja varieeruvus on mõisted, mis lubavad kirjeldada mustrite leksikaalgrammatilist ja funktsionaalset mitmekesisust, avastada nii mõnegi keelekasutuse seaduspärasuse ning järjestada keeleõppe eesmärgil aktiivselt kasutatavad mustrid nende lingvistilist keerukust arvestades.

Keerukamad on kahtlemata mustrid järjestikuste adverbidega (adverbidel on erinevaid funktsioone), aga ka mustrid substantiivide ja

substantiivi-adjektiivi rikkalike käändekombinatsioonidega määrusliku täiendi funktsioonis. Substantiivialguliste mustrite näitel, mis sisaldavad adverbikooslusi ja adverbide kooskasutust teiste sõnaliikidega, ei saa rääkida otseselt reeglitest, sest kõrvalekallete osa on ilmne. Samas tulevad üsna selgelt välja mõningad adverbide kasutamist ilmestavad tendentsid. Siit ka vajadus täpsemalt teada, kui suur on adverbide kooskasutuse tõenäosus mustris, sh erinevates positsioonides ja funktsioonides. Selleks on vaja mahukat andmestikku kõikvõimalikest adverbide kooskasutuse leksikaalsetest mallidest. Analoogse ressursi võiks koostada ka substantiivide ja adjektiivi-substantiivi käände kombinatsioonidest määrusliku täiendi funktsioonis. Need on spetsiaalsed keeleressursid, mille põhjal saab täpsustada, kui suure tõenäosusega iga kooskasutuse variant esineb. Sel analüüsi etapil oleks abi mõõtmistest, mille tulemused loodetavasti aitavad täpsustada nt kahe järjestikuse adverbide funktsioone, piiritleda adverbide sõnaliigina ja adverbide seoseid teiste sõnaliikidega, määrata adverbide sisaldavate kahe- ja kolmesõnaliste koosluste kollokatsioonilisuse astet. Substantiivialguliste mustrite analüüsi tulemused näitavad, et iga adverbide sisaldava mustri alusel ei pruugi kollokatsioonid kujuneda, kuid keeleteadlase jaoks oleksid niisugused leksikaalsed loendid tänuväärne õppematerjal.

Mustri püsistruktuuri kui keelekasutuses korduvalt esinevasse sõnaliigijärjendisse on sisse kodeeritud komponentide kooskasutuse fonoloogilised, süntaktilised ja kontseptuaalsed reeglid, mis toimivad leksika-grammatika piirimaal. Lingvistika seisukohalt sisaldab see uudset teavet, mille olulisus avaldub keeleteaduses ning automaatsete keeletöötlevahendite arendamises. Emakeelekõneleja talletab need reeglid pikaajalises mälus ja aktiveerib automaatselt tekstiloome käigus. Nt mustri SVD (*seadused kehtivad küll*) kasutamisel tuleb esile morfosüntaktiline mall, millest paremale hargnev kontekst jätkub kõrvallausega, selle alguses paiknevad tavaliselt kõrvutatav-vastandavad või mööndust väljendavad sidendid, nagu *kuid*, *ometi*, *ainult et*, *ent*, *ehkki*, *kuigi*: *seadused kehtivad küll*, ometi tuleb tõdeda, et .: *tööandja on valmis* + mida tegema?

(allkirjastama, teatama, tõdema, arutama jne), ainult et ..; *Kohver on siiani + kus?* (Venemaal, vanglas, vangis, kinni, trellide taga jne), kuigi ...

Niisiis kasutatakse sama sõnaliigijärjendit erinevas süntaktilises, semantilises ja leksikaalses kontekstis. Millist varianti nende seast emakeelekõneleja kirjutatud tekstides eelistatakse, seda näitab ühe või teise malli sagedus ja osakaal. Keeleõppe mõttes on niisugune info väärtuslik, sest võimaldab aktiivselt kasutatavad morfosüntaktilised mallid ja tüüpilised kontekstid eristada passiivsetest ning aktiivse grammatika ja sõnavara passiivsest. Keeleõppe edukust silmas pidades on põhimõtteline vahe, kas õppematerjali leidmise ning sõnavara- ja grammatikamiinimumi koostamisel lähtutakse mahukast tekstikorpusest ning rakendatakse andmekaevet või tehakse valikuid õpetamiskogemuse ja seni kasutuses õpikutekstide vmt alusel. Viimasel juhul ei tule kunagi välja nt sama mustris võrdsed võimalused esineda erinevates sõnajärje- ja lausemallides, mille taga on tegelikult peidus raskesti omandatavad juhtumid, nagu ühendverbide tekstikasutus: SDD-s võib põhiverb paikneda mustrile eelnevas (jagati *ülesanded selgelt ära*) või järgnevas kontekstis (kellel-millel on *risk kapitaalselt kokku vajuda*); SVD-s kuulub ühendverb üldjuhul mustris koosseisu (*üliskoolid valmistavad ette*), kuid mineviku liitajavormide ja modaalpredikaadi kasutamisel jääb põhiverb mustrist paremale hargnevasse konteksti (*hinnad on üle kontrollitud*) jne. Siin ka seletus, miks nt eesti õppijakeele viieteistkümne läbivalt kasutatud sagedama mustris seas puudub SDD. Pealtnäha on tegemist tekstipõhise sõnajärje küsimusega, kuid selle all peitub võimalus valida sama struktuuri kasutamisel erinevate, ent samas võrdväärsete kontekstide vahel. Sobivad kõik variandid, mida emakeelekõneleja peab võrdselt loomulikuks, kuid keeleõppija vajab süstematiseeritud teavet.

Samamoodi on andmekaeveta peaaegu et võimatu leida keelestruktuure, mille põhjal eesti keeles moodustuvad ja kasutatakse analüütilisi sõnu ja konstruktsioone, millega õppijad näevad vaeva. Konkreetseid reegleid ei ole, kuid substantiivialgulistes mustrites, mida näitena analüüsisin, leidub morfosüntaktilisi malle, mille alusel tuleb esile nii liitseid sõnu, sidendeid kui ka konstruktsioone. Neid malle saab edukalt

kasutada harjutustes, et analoogiat rakendades tuletada ühendverbe, tugi- ja ahelverbikonstruktsioone, sidendeid ja muid liitseid keeleüksusi, mis aja jooksul on keelekasutuses kinnistunud. Grammatikatest, sõnas- tikest või lühendatud õpikutekstidest ammutatud õppematerjalid niisuguseid võimalusi ei paku.

Tänuõnad

Uurimus sai teoks tänu riikliku programmi Eesti keeletehnoloogia projektile TKA21200 “Eestikeelse teksti automaatkorrekatuur” ja Tallinna Ülikooli uuringufondi projektile TF2621 “ELLE: eesti keele õpet toetav keeletehnoloogiline ressurss”.

Lühendid

3	kolmanda pöörde vorm	kom	komitatiiv
A	adjektiiv	N	numeraal
ad	adessiiv	nom	nominatiiv
all	allatiiv	P	pronomeni
D	adverb	part	partitiiv
el	elatiiv	perf	perfekt
G	genitiivatribuut	pl	plural
gen	genitiiv	pres	preesens
impf	imperfekt	S	substantiiv
in	inessiiv	sg	singular
indic	indikatiiv	tr	translatiiv
J	konjunktsioon	V	verb
K	adpositsioon	Y	abreviaatuur

Kirjandus

- Aedmaa, Eleri 2015. Statistilised meetodid ühendverbide tuvastamisel tekstikorpusest [‘Statistical methods for Estonian particle verb extraction from text corpora’]. – Eesti Rakenduslingvistika Ühingu aastaraamat 11, 37–54. <https://doi.org/10.5128/ERYa11.03>
- Aedmaa, Eleri 2016. Eesti keele ühendverbide kompositsionaalsuse määramine [‘Detecting the compositionality of Estonian particle verbs’]. – Eesti

- Rakenduslingvistika Ühingu aastaraamat 12, 5–23. <https://doi.org/10.5128/ERYa12.01>
- Allkivi, Kais 2016. C1-tasemega eesti keele õppijate ja emakeelekõnelejate kirjaliku keelekasutuse võrdlus verbialguliste tetragrammide näitel [‘Written language use of C1 learners of Estonian and native speakers in comparison: analysis of verb-initial fourgrams’]. – Lähivõrdlusi. Lähivertailuja 26, 54–83. <https://doi.org/10.5128/LV26.02>
- Allkivi-Metsoja, Kais 2021a. Eesti keele A2–C1-taseme kirjalike tekstide võrdlev automaatanalüüs [‘Written Estonian at the levels A2–C1: Comparative automated analysis’]. – Lähivõrdlusi. Lähivertailuja 31, 13–59. <https://doi.org/10.5128/LV31.01>
- Allkivi-Metsoja, Kais 2021b. C1-tasemel eesti keele õppija kirjalik keelekasutus võrdluses emakeelekõnelejaga [‘Written language use of C1 learners of Estonian and native speakers in comparison’]. – Eesti keele oskuse arendamine ja arendamine. Kirjalik õppijakeel. Tallinn: EKSA, 205–231.
- Biber, Douglas 2009. A corpus-driven approach to formulaic language in English. – International Journal of Corpus Linguistics 14 (3), 275–311. <https://doi.org/10.1075/ijcl.14.3.08bib>
- Chen, Yu-Hua, Paul Baker 2010. Lexical bundles in L1 and L2 academic writing. – Language Learning & Technology 14 (2), 30–49.
- Conklin, Kathy, Norbert Schmitt 2008. Formulaic sequences: Are they processed more quickly than non-formulaic language by native and nonnative speakers? – Applied Linguistics 29 (1), 72–89. <https://doi.org/10.1093/applin/amm022>
- De Cock, Sylvie, Sylviane Granger 2021. Stance in press releases versus business news: A lexical bundle approach. – Text and Talk 41 (5–6), 691–713. <https://doi.org/10.1515/text-2020-0040>
- Eslon, Pille 2014a. Morfosüntaktilise ja leksikaalse varieerumise piiridest: ilukirjandus- ja õppijakeele kasutusmustrite võrdlus [‘Constraints on morphosyntactic and lexical variability’]. – Eesti Rakenduslingvistika Ühingu aastaraamat 10, 55–71. <https://doi.org/10.5128/ERYa10.04>
- Eslon, Pille 2014b. Adverbi sisaldavate struktuuride tekstifunktsioonidest eesti ilukirjandus- ja õppijakeeles [‘On the textual functions of adverbial structures in literary Estonian and in Estonian learner language’]. – Lähivõrdlusi. Lähivertailuja 24, 15–46. <https://doi.org/10.5128/LV24.01>
- Eslon, Pille 2017a. Keelekasutusmustrid verbist paremal: morfosüntaktiline ja leksikaalsemantiline varieerumine [‘Patterns of language use found on

- the right periphery of the verb: Morphosyntactic and lexico-semantic variability']. – *Lähivõrdlusi. Lähivertailuja* 27, 17–64. <https://doi.org/10.5128/LV27.01>
- Eslon, Pille 2017b. Kasutuspõhise keelekäsitluse pedagoogiline perspektiiv ['Usage-based language description: Linguistic cluster analysis and its perspectives for pedagogical purposes']. – *Mäetagused* 69, 217–242. <https://doi.org/10.7592/MT2017.69.eslon>
- Eslon, Pille 2021a. Soome- ja venekeelsete õppijate A2- ja B1-taseme keelekasutusmustritest: varieeruvus vs. stereotüüpsus ['On the language use patterns of Finnish- and Russian-speaking learners of Estonian at the levels A2 and B1: Variability vs. stereotypicality']. – *Lähivõrdlusi. Lähivertailuja* 31, 90–113. <https://doi.org/10.5128/LV31.03>
- Eslon, Pille 2021b. Eesti keele kasutamine A2- ja B1-taseme tekstides soome- ja venekeelsete õppijate näitel ['Language use of Finnish- and Russian-speaking learners of Estonian at the levels A2 and B1']. – *Eesti keele oskuse arenemine ja arendamine. Kirjalik õppijakeel*. Tallinn: EKSA, 117–204.
- Eslon, Pille, Kais Allkivi-Metsoja 2018. Teksti keelekasutusmustrid ja lingvistiline klasteranalüüs ['Patterns of language use and linguistic cluster analysis']. – *Lähivõrdlusi. Lähivertailuja* 28, 21–46. <https://doi.org/10.5128/LV28.01>
- Eslon, Pille, Erika Matsak 2009. Eesti keele kasutusvariandid: korpusest tulenev käändevormide võrdlev analüüs ['Corpus-driven comparative analysis of variants of Estonian']. – *Eesti Rakenduslingvistika Ühingu aastaraamat* 5, 79–110. <https://doi.org/10.5128/ERYa5.06>
- Eslon, Pille, Heleriin Paeoja 2015. Samatähenduslike sünteetiliste ja analüütiliste verbide kasutamine ['Use of the synonymous synthetic and analytical verbs']. – *Lähivõrdlusi. Lähivertailuja* 25, 63–104. <https://doi.org/10.5128/LV25.04>
- Evert, Stefan 2005. *The Statistics of Word Co-occurrences, Word Pairs and Collocations*. PhD dissertation. Institut für maschinelle Sprachverarbeitung Universität Stuttgart.
- Gore, Paul A. Jr. 2000. Cluster analysis. – Howard E. A. Tinsley, Steven D. Brown (eds.). *Handbook of Applied Multivariate Statistics and Mathematical Modeling*. New York: Academic Press, 297–321. <https://doi.org/10.1016/B978-012691360-6/50012-4>
- Hallik, Virgo 2015. Eesti vahekeele korpuse klasteranalüüsi vahendite kasutamine teksti keeletaseme prognoosimisel ['Using Estonian interlanguage corpus cluster analysis tools to predict language level of text']. *Bakalaureusetöö*. Tallinna Ülikool, digitehnoloogiate instituut.

- Jackendoff, Ray 2017. In defense of theory. – *Cognitive Science* 41 (Suppl. 2), 185–212. <https://doi.org/10.1111/cogs.12324>
- Juknevičienė, Rita 2013. Recurrent word sequences in written learner English. – Inesa Šeškauskienė, Jonė Grigaliūnienė (eds.). *Anglistics in Lithuania: Cross-Linguistic and Cross-Cultural Aspects of Study*. Cambridge Scholar Publishing, 178–197.
- Law, Nancy 2007. Comparing pedagogical innovations. – Mark Bray, Bob Adamson, Mark Mason (eds.). *Comparative Education Research: Approaches and Methods*. Hong Kong: The University of Hong Kong, 333–364.
- Martinez, Ron, Norbert Schmitt 2010. A phrasal expressions list. – *Applied Linguistics* 33 (3), 299–320. <https://doi.org/10.1093/applin/ams010>
- Matsak, Erika, Pille Eslon, Jaagup Kippar 2010. Eesti keele sõnajärje vealeidja prototüübi arendamine [‘The development of the prototype for an automatic word order error detector for the Estonian language’]. – *Korpusuuring ja meetodid*. Tallinna Ülikooli eesti keele ja kultuuri instituudi toimetised 12. Tallinn: TLÜ EKKI, 59–100.
- Metslang, Helena, Erika Matsak 2010. Kesksete lausekomponentide järjestus õppijakeeles: arvutianalüüsi katse [‘Automatic word order analysis of Estonian as a second language: The nuclear sentence’]. – *Eesti Rakenduslingvistika Ühingu aastaraamat* 6, 175–193. <https://doi.org/10.5128/ERYa6.11>
- Muischnek, Kadri 2006. Verbi ja noomeni püsiühendid eesti keeles [‘Fixed Expressions Consisting of Verbs and Nouns in Estonian’]. *Dissertationes philologiae estonicae Universitatis Tartuensis* 17. Tartu: Tartu Ülikooli Kirjastus.
- Muischnek, Kadri, Heiki-Jaan Kaalep 2010. The variability of multi-word verbal expressions in Estonian. – *Language Resources and evaluation* 44 (1–2), 115–135. <https://doi.org/10.1007/s10579-009-9096-x>
- Muischnek, Kadri, Heete Sahkai 2010. Liitpredikaadid leksikoni-grammatika kontiinumil: konstruktsioonide produktiivsusest verbiga minema moodustatud liitpredikaatide näitel [‘Complex predicates on the lexicon-grammar continuum’]. – *ESUKA/JEFUL* 1 (2), 295–316. <https://doi.org/10.12697/jeful.2010.1.2.16>
- Ots, Sander 2012. Statistikapõhise tarkvara loomine morfoloogiliste kollokatsioonide eraldamiseks eesti keele tekstidest [‘Software for morphosyntactic cluster extraction from Estonian texts’]. *Bakalaureustöö*. Tallinna Ülikool, informaatika instituut.
- Paeoja, Heleriin 2015. Analüütiliste/sünteesiliste verbipaaride kasutusmustrid 1990ndate aastate eesti ilukirjanduskeeles [‘Usage patterns of analytic and

- synthetic verbal pairs based on 1990s Estonian literature texts']. Magistritöö. Tallinna Ülikool, eesti keele ja kultuuri instituut.
- Sahkai, Heete 2006. Konstruksioonipõhise keelekirjelduse võimalustest adessiivse viisi- ja põhjusemääruse näitel ['Construction-based description of adessive adverbials of manner and reason']. – Keel ja Kirjandus 10, 816–831.
- Sahkai, Heete 2008. Konstruksioonipõhine keelemudel ja sõnaraamatumudel ['Some lexicographic implications of a construction-based model of language']. – Eesti Rakenduslingvistika Ühingu aastaraamat 4, 177–186. <https://doi.org/10.5128/ERYa4.11>
- Sahkai, Heete 2011. Eesti keele genitiivse agendifraasi süntaks ['The syntax of the Estonian genitive agent phrase']. – Keel ja Kirjandus 1, 12–30.
- Schmitt, Norbert 2010. Researching Vocabulary: A Vocabulary Research Manual. Research and Practice in Applied Linguistics. Basingstoke: Palgrave Macmillan.
- Selivan, Leo 2018. Lexical Grammar: Activities for Teaching Chunks and Exploring Patterns. Cambridge Handbooks for Language Teachers. Cambridge: Cambridge University Press. <https://doi.org/10.1017/9781009024624>
- Ševtšenko, Marina 2014. Eesti keele kui teise keele 8. klassi õpiku temaatiline sõnavara ja grammatika ['Vocabulary and grammar of the Estonian as a second language textbook (8th class)']. Magistritöö. Tallinna Ülikool, eesti keele ja kultuuri instituut.
- Trainis, Jekaterina 2015. Linguistic cluster analysis: A method for describing language units and indicating regularities in language. – Wojciech Malec, Marietta Rusinek (eds.). Within Language, Beyond Theories. Vol. III: Discourse Analysis, Pragmatics and Corpus-based Studies. Cambridge Scholars Publishing, 229–243.
- Trainis, Jekaterina 2017. Diakroonilised nihked eesti ilukirjanduskeele kasutusmustrites 1890–1990 ['Diachronic language changes in usage patterns of belletristic Estonian in 1890s–1990s']. – Mäetagused 69, 181–216. <https://doi.org/10.7592/MT2017.69.trainis>
- Trainis, Jekaterina, Kais Allkivi 2014. Ilukirjanduskeelest uue pilguga ['On belletristic language from a new perspective']. – Eesti Rakenduslingvistika Ühingu aastaraamat 10, 283–306. <https://doi.org/10.5128/ERYa10.18>
- Tšernõšuk, Anna 2016. Eesti keele kui teise keele 9. klassi õpiku temaatiline sõnavara ja aktiivne leksikon ['Vocabulary of the Estonian as a second language textbook (9th class)']. Bakalaureusetöö. Tallinna Ülikool, humanitaarteaduste instituut.

- Uiboaed, Kristel 2010. Statistilised meetodid murdekorpuse ühendverbide tuvastamisel [‘Statistical methods for phrasal verb detection in Estonian dialects’]. – Eesti Rakenduslingvistika Ühingu aastaraamat 6, 307–326. <https://doi.org/10.5128/ERYa6.19>
- Voolaid, Katrin 2018. Vene ja soome lähtekeele õppijate eesti keele kasutusmustrid (B1-tase) [‘Estonian language usage patterns among Russian and Finnish students (B1 language proficiency level)’]. Magistritöö. Tallinn: Tallinna Ülikool.
- Õim, Asta, Katre Õim 2019. Lähtekohti eesti fraseoloogia käsitlemiseks [‘On the Starting-Point of Estonian Phraseology Description’]. Tallinn: EKSA.

At the border of lexis and grammar: Language use patterns of journalistic opinion articles

PILLE ESLON

Tallinn University

The study draws on the usage-based corpus linguistic approaches. It presents the linguistic analysis of opinion articles (published in the Estonian daily newspapers *Postimees* and *Õhtuleht*), relying on part-of-speech trigrams. These language use patterns have been extracted from morpho-syntactically tagged texts using the data mining software *Cluster Catcher* (*Klastrileidja*) developed at the Tallinn University. Earlier, similar analysis has been done based on the corpora of Estonian fiction and learner language.

The recurring part-of-speech sequences have the encoded phonological, syntactic, and conceptual rules of their components' co-occurrence that manifest at the border of semantics and grammar, bringing out the regularities of language use. Therefore, the actively used language patterns can be considered a suitable source material for teaching and learning Estonian, comparing the language production of learners speaking different native languages and also for automated processing of learner writings (e.g., proposing corrections in wording with references to rules and examples of the target language use).

Keywords: linguistic cluster analysis; language use patterns; lexical-grammatical variation; regularities of language use; Estonian

Pille Eslon

Tallinna Ülikooli digitehnoloogiate instituut

Narva mnt 29, 10120 Tallinn, Estonia

peslon@tlu.ee