

## ILUKIRJANDUSKEELEST UUE PILGUGA

Jekaterina Trainis, Kais Allkivi

**Ülevaade.** Artiklis kirjeldatakse 20. sajandi teise poole eesti ilukirjanduskeele kasutusmustreid, mis näitavad, kuidas emakeelekõneleja eelistab oma mõtteid kirjalikult väljendada. Uurimisaines on saadud Tartu Ülikooli pindsüntaktiliselt märgendatud ja ühtlustatud ilukirjanduskorpusest. Rakendatud on klasteranalüüsi, mis toob tekstivalimis sageduse alusel esile ühesuguseid morfo- ja süntaksimärgendite lineaarseid järgnevusi. Valitud on trigrammid, mis esinesid korpuses kaks ja enam korda.

Keelekasutuse üldised mustrid on välja toodud trigrammide esikomponendi sõnaliigilise kuuluvuse järgi. Nii tulevad esile morfoloogilised klassid ja alamklassid, mis moodustuvad ühise algus- ja lõpukomponendiga sõnaliigijärjenditest. Iga alamklass hõlmab morfoloogilisi klastreid, kuhu trigrammid koonduvad keskmise komponendi sõnaliigi varieerumise alusel. Morfosüntaktilise trigrammi moodustavad sõnajärjendid, mille komponendid kattuvad nii sõnaliigi ja vormi kui ka süntaktilise funktsiooni poolest. Morfoloogiliste klasside, alamklasside, klastrite ja morfosüntaktiliste trigrammide hierarhia toob esile keelekasutuse üldised tendentsid ja morfosüntaktilise ning leksikaalse varieerumise piirid.

**Võtmesõnad:** ilukirjanduskeel, klasteranalüüs, morfosüntaktiline varieeruvus, leksikaalne varieeruvus, eesti keel

### 1. Meetod ja teoreetilised lähted

Uurimuse eesmärk on kirjeldada eesti ilukirjanduskeele morfoloogilist, morfosüntaktilist ning leksikaalset varieerumist, välja tuua regulaarselt/ebaregulaarselt ilmnevad grammatilised ja leksikaalsed valikud. Ühtlasi on eesmärk näidata lingvistilise klasteranalüüsi kasutuspotentsiaali keelekirjelduses.

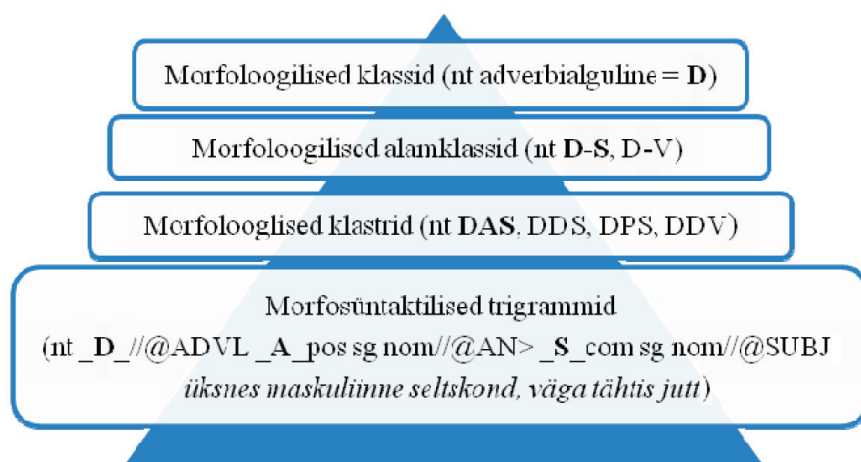
Klasteranalüüs on meetodina objektiivne, kui uurija paneb eelnevalt paika kindlad klasterdamisreeglid: erinevate klasterdamismeetoditega saadud klassifikatsioonid võivad erineda (Remm jt 2012: 75–76). Käesolevas uurimuses on

rakendatud S. Otsa programmeeritud klastrileidjat, mille sisendiks on eesti keele parseri väljund (Ots 2012). Klastrileidja otsib tekstist sarnaseid morfo- ja süntaksimärgendite lineaarseid järjendeid ning fikseerib nende sageduse. Kasutatud on nn libisevat tükeldamist: programm otsib järjendeid põhimõttel sõna1 sõna2 sõna3; sõna2 sõna3 sõna4 jne. Klastrileidja väljundit on töödeldud Excelis.

Varem ei ole eesti ilukirjanduskeelt sellise meetodi abil analüüsitud, ent sama meetodit on kasutatud eesti ilukirjandus- ja õppijakeele morfoloogiliste ja morfosüntaktiliste mustrite võrdlemisel (vt Eslon 2013 ja Eslon 2014 käesolevas kogumikus).

Uurimuse tarvis on klastrileidja otsing piiratud kolmest komponendist koosnevate üksuste ehk trigrammidega. Excelis reastatakse tekstikasutuses kaks ja enam korda esinenud trigrammid tähestikuliselt esikomponendi sõnaliigilise kuuluvuse alusel. Nii moodustuvad morfoloogilised klassid (nt adverbialguline ehk D). Alamklasse moodustavad trigrammid, mis tulevad esile mustritena ühesuguse algus- ja lõpukomponendi alusel, keskmise komponendi sõnaliik varieerub (nt adverb-substantiiv ehk D–S). Morfoloogilised klastrid eristuvad kõigi kolme komponendi morfoloogilise varieerumise alusel alamklassis (nt adverb-adjektiiv-substantiiv ehk DAS). Iga klaster sisaldab morfosüntaktiliselt ja leksikaalselt varieeruvaid trigramme, mis põhinevad samasugustel morfo- ja süntaksimärgendite järgnevustel (nt `_D_//@ADVL_A_pos sg nom//@AN>_S_com sg nom//@SUBJ` ehk adverb adverbialina + adjektiivne eestäiend ainsuse nominatiivis + substantiiv ainsuse nominatiivis subjektina: *pisut piinlik vaikus*).

Morfosüntaktilised trigrammid ja morfoloogilised klastrid, alamklassid ning klassid moodustavad hierarhilise liigituse (vt joonis 1): iga klastri sagedaim trigramm on selle klastri invariant; iga alamklassi sagedaim klaster on selle alamklassi invariant; iga klassi sagedaim alamklass on kindla sõnaliigiga algava klassi invariant. Kõige enam esinenud klass on keelekasutuse invariant.



**Joonis 1.** Lingvistilise klasteranalüüsi hierarhia

Siinses uurimuses on klasterdamist mõistetud kui lingvistiliste objektide rühmitamist morfo- ja süntaksimärgendite regulaarse koosinemise ja varieerumise põhjal. Morfosüntaktiliste kirjelduste alusel pannakse samm-sammult omavahel kokku lingvistiliselt sarnased objektid. Liitmisprotsess lõpeb, kui kõik ühiste tunnustega objektid on ühes klastris. Analüüsi tulemusel ilmnevad tekstikasutuse lingvistilised

mustrid, mille alusel saab kirjeldada grammatika ja leksika omavahelisi seoseid ning keelendite funktsioone.

Uurimuse keeleaines pärineb Tartu Ülikooli eesti kirjakeele süntaktiliselt analüüsitud ja ühestatud ilukirjanduskorpusest. Uurimuse tarvis on sellest leitud kõik trigrammid, mida valimis on esinenud kaks ja enam korda, k.a (osa)lause piiiril.

Kokku oli ilukirjanduskorpuses 49 211 morfosüntaktilist trigrammi, sellest kaks ja enam korda kasutatuid 14 316, mis teeb trigrammide üldmahuks 147 633 sõnet (trigrammide arv  $49\,211 \times$  pikkus 3 sõnet). Kaks ja enam korda kasutatud trigrammide üldmaht on 42 948 sõnet ( $14\,316 \times 3$ ). Suhe kaks ja enam korda esinevate trigrammide ( $14\,316 \times 100\% : 49\,211 = 29,09\%$ ) ning ainukordsete ( $34\,895 \times 100\% : 49\,211 = 70,9\%$ ) vahel on  $29,09\% : 70,9\%$ . Seega kehtib 1929. aastal kirja pandud Zipfi needus (vt ka Tuldava 1977, Tullo, Hurford 2003), mis näitab, et keeles esineb suhteliselt väike rühm väga sageli kasutatud konstruktsioone ning palju keskmise ja väikese sagedusega konstruktsioone (suhe  $1/3$  ja  $2/3$ ).

Optimaalse klastriarvu määramist peetakse klasteranalüüsi kitsaskohaks, kuna see etapp on valdavalt subjektiivne (Burns, Burns 2008: 557). Alamklasside kirjeldamisel on arvestatud optimaalse trigrammide hulgaga, mis leitud valemi  $k \approx \sqrt{n} : 2$  järgi, kus  $n$  on objektide arv (siin: kaks ja enam korda esinud trigrammide arv), mille järgi optimaalne trigrammide hulk  $k$  on 85 (vt lisad, märgitud halli värviga). Trigrammide statistilist olulisust hinnates lähtutakse sellest arvnäitajast. Morfosüntaktiliste trigrammide tasandil analüüsitakse klasside kahe sagedasema alamklassi klastreid, ülejäänud alamklasse kirjeldatakse ülevaatlikumalt.

Kuna klasteranalüüsis ei ole ühte ainuõiget lahendit, rõhutavad mitmed autorid, et kvantitatiivse analüüsi tulemust peab tõlgendama teooria seisukohast (Burns, Burns 2008: 552, Remm jt 2012: 76, vt ka Niglas s. a.). Uurimuse aluseks on kasutuspõhine lähenemine, mis võrsunud konstruktivismi (Croft 2003) ja funktsionaalse grammatika (Halliday 1994) seostest.

M. Barlow ja S. Kemmer (2000) ning J. Bybee (2003) on selgitanud, et keelekasutusmudelid (ingl *models of the language use*) on sageduse põhjal esile toodud keeleliste teadmiste mudelid, nt võistlusmudel (*Competition Model*), leksikaalse tähendusvõrgustiku mudel (*Lexical Network Theory*) jm. P. Norvig (1989) on tutvustanud leksikaalse tähendusvõrgustiku mudelit, mis aitab toime tulla üksiksõnade ja sõnadevahelise tähendusega. A. Veismann (2008) on kaassõnade kirjeldamisprobleemide üle arutledes välja toonud radikaalse tähendusvõrgustiku mudeli, kus kesksed liikmed on sõna tähenduste prototüüpsemad esindajad.

Ka kasutuspõhine grammatikakirjeldus (ingl *usage-based grammar description*) on olemuselt standardiseeritud grammatikakirjeldusest erinev ja kasutuspõhisest keelekäsitlusest (*usage-based language description*) kitsam mõiste. Vastupidiselt generatiivsele lähenemisele, mis toetub seisukohale, et keelekasutaja memoreerib vaid ebareeglipäraseid vorme, püüab kasutuspõhine keelekirjeldus näidata, et esinemissageduse tõttu jätab keelekasutaja meelde ja rakendab valdavalt, enamasti või harva vaid teatud liiki konstruktsioone (Barddal 2003). Morfoloogilisi universaale uurivas loomuliku morfoloogia teoorias tuuakse välja, et teatud nähtused on nii vormi- kui ka sõnamoodustuses ootuspärasemad ja sagedasemad, st keelekasutaja silmis loomulikumad kui teised (Habicht 2006: 13–15).

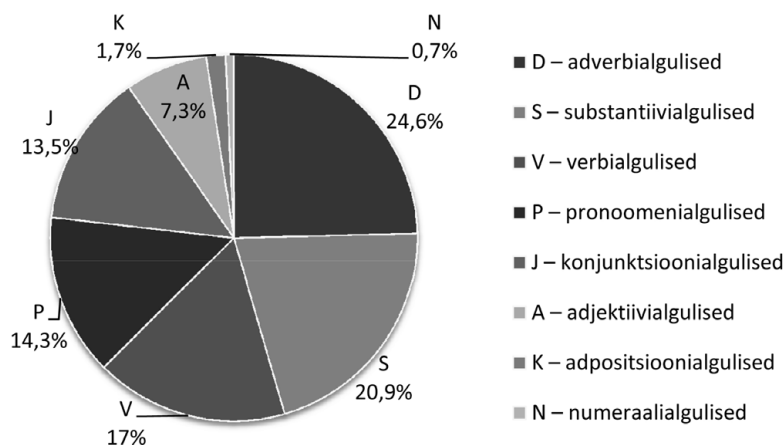
Esinemissageduse olulisust on keelekasutuse uurimisel rõhutanud ka J. Bybee, P. Hopper (2001:1–24) ja M. Haspelmath (2006, 2008). Sagedus dikteerib

keelekasutusmustrit. Seejuures on kõrge esinemissagedusega vormid vähem muutusealtid kui madalama sagedusega vormid, mis on omakorda sõnade ja nende koosinemise suhtelise autonoomia väljendus keeles (Bybee 2010: 50, 75).

Käesolev keelekasutusest tulenev empiiriline uurimus tugineb invariandi, esinemissageduse ja osakaalu mõistetele ning morfosüntaktilise ja leksikaalse varieeruvuse piiridele (vt ka Eslon 2013: 13–38).

## 2. Adverbialguline klass

Trigrammide jagunemine morfoloogilistesse klassidesse esikomponendi sõnaliigi alusel näitab, et ilukirjanduskeeles on suurim osakaal adverbiga algavatel trigrammidel: 24,6% valimi mahust (vt joonis 2, lisad 1–6). Suhteliselt sagedased on ka substantiivi- (20,9%), verbi- (17%), pronoomeni- (14,3%) ja konjunktsioonialguline (13,5%) klass, väiksema kasutussagedusega adjektiivialguline (7,3%). Harva esinenud adpositiooni- (1,7%) ja numeraalialgulist (0,7%) klassi artiklis ei kirjeldata. Järgnevalt antakse ülevaade morfoloogiliste klasside osakaalust valimis.



Joonis 2. Morfoloogiliste klasside osakaal valimis

Adverbialgulise klassi invariantne alamklass on adverb-substantiiv (D–S) *nii selgesti meeles* (vt lisa 1), järgnevad adverb-verb (D–V) *ennevanasti on olnud*, adverb-adverb (D–D) *seejärel silmitseb mõtlikult*. Alamklassid adverb-pronoomen (D–P) *veel kaua-kaua tema*, adverb-adjektiiv (D–A) *hirmus ilusad valged*, adverb-konjunktsioon (D–J) *nägupidi omad ja* ning adverb-adpositioon (D–K) *otse silmade ees* on juba oluliselt vähem esindatud. Alamklass adverb-numeraal (D–N) *möödas juba seitsesada* on teiste klassidega võrreldes marginaalne (vt tabel 1).

Tabel 1. Adverbialguliste alamklasside osakaal valimis

Alamklass	D–S	D–V	D–D	D–P	D–A	D–J	D–K	D–N
Osakaal klassis	29,4%	25,4%	19,1%	7,5%	7,5%	6,5%	4,4%	0,2%

## 2.1. Alamklass D–S

Adverbiga algava klassi sagedaim alamklass D–S koosneb üheksast klastrist: DAS (33,7%), DDS (16,5%), DPS (15,8%), DVS (13,9%), DSS (8,2%), DJS (4,5%), DNS (4,5%), DKS (2,6%) ja DGS (0,3%). Adverbidest korduvad alamklassis D–S enim koha-, seisundi-, rõhu- ja ajamäärsõnad. Klatri trigrammid näitavad, et üks komponent seob teisi, tõmmates oma poole magnetina vaid kindlaid elemente.

Alamklassi D–S invariantne klaster on **DAS**. Selles esineb koha- (*üles, seal, kõrvale*), viisi- ja seisundi- (*rahatult, kapriisselt, kollakalt*), kvantiteedi- (*liiga, palju, üsna*) ja rõhumäärsõnade (*kahjuks, liiga, üha*) leksikaalset varieerumist. Klatri invariantne morfosüntaktiline trigramm on adverb adverbialina + adjektiivne eestäiend ainsuse nominatiivis + substantiiv ainsuse nominatiivis subjektina, nt *mõneti kummaline lugu, tõesti süütu kurat*.

Klatri **DDS** invariantne trigramm on kaks adverbilist adverbialia + substantiiv ainsuse partitiivis objektina, nt *liigagi täpselt aimu, ka mõnikord pahandust*. Adverbide valik ja kombineerumine on väga mitmekesine, tavaliselt esinevad järjestikku aja- ja rõhumäärsõna (*ju ometi juhtnööre, siis ootamatult kaela*).

Klatri **DPS** invariantne trigramm on adverb + possessiivpronoomen ainsuse genitiivis + substantiiv ainsuse partitiivis objektina, nt *ainult oma väitekirja, siin igal õhtul*. Pronoomeni valik on liigiti piiratud: *oma* on kujunemas kõrge esinemissageduse tõttu stampsõnaks.

Klatri **DVS** invariantne trigramm on lausealguline adverb adverbialina + öeldisverb indikatiivis imperfekti ainsuse 3. pöördes + substantiiv ainsuse nominatiivis subjektina, nt *Siis helises uksekell, Vahel ulatus vesi*. Adverbidest korduvad lause alguses *Siis, Pärast, Ees*. Pea pooltel juhtudel on trigrammi moodustavaks verbivormiks *oli*, mis on klatri DVS kasutusele tüüpiline.

## 2.2. Alamklass D–V

Adverbiga algava klassi alamklass D–V koosneb seitsmest klastrist: DDV (39,8%), DVV (24,8%), DJV (12,6%), DSV (11,2%), DPV (7,9%), DAV (3,5%), DXV (0,2%). Alamklassi D–V adverb väljendab põhiliselt subjektiivset modaalsust, rõhutab öeldut või viitab öeldule.

Invariantse klatri **DDV** invariantne trigramm on kaks adverbi adverbialina + eitav abiverb, nt *nüüd küll ei, enam sugugi ei*. Adverbi on kasutatud subjektiivse modaalsuse väljendamise ja millegi rõhutamise eesmärgil (*tegelikult, tingimata*), tagasiviite vahendina lause alguses (*Küllap, Alles, Kui*), harva kohamääрусena.

Klatri **DVV** invariantne trigramm on adverb adverbialina + öeldisverb eitava kõne indikatiivi imperfektis, nt *veel ei liikunud, muidugi ei olnud*. Adverbe iseloomustab leksikaalne rikkus ja funktsioonide mitmekesisus, neid on kasutatud sageli osalause alguses sidendina (*siis ei osanud*), subjektiivse modaalsuse väljendamiseks (*nähtavasti ei pidanud*) või millegi rõhutamise eesmärgil (*tõepoolest ei alustanud*).

Klatri **DJV** invariantne trigramm on adverb + rinnastav sidesõna + transitiivne öeldisverb indikatiivis imperfekti ainsuse 3. pöördes, nt *üles ja küsis, pooleks ja virutas*. Adverbide leksikaalne varieeruvus on suur. Enamjaolt esineb adverb ühendverbide osana (<panen> *kinni ja jätan*, <andis> *tagasi ja läks*).



Klastrid **DSV** invariantne trigramm on adverb adverbialina + substantiiv ainsuse partitiivis objektina + partitsiip öeldisverbi osana, nt *ikka puhtust pidanud, ka märki tabanud*. Eitavas kõne puhul rõhutab adverb öeldut (*Isegi Kraaving ei*) või seob (*siis last ei*).

### 2.3. Alamklass D–D

Alamklass D–D koosneb kuuest klastrist: DDD (40%), DVD (18,9%), DSD (13,7%) ja DJD (13,4%), DPD (13,1%), DAD (0,9%). Alamklassi adverbivalik on rikkalik.

Invariantse klastrid **DDD** lekikaalne varieeruvus on suur: adverbid rõhutavad (*ometi natuke aegsamini*), toovad esile (*pealegi veel nii*) ja seovad (osa)lauseid (*Kas tõesti nii*). Kõige enam korduvad *kuidas, nüüd, mitte, võib-olla*. Klastrid **DVD** on eelistatud *olema*-verbi indikatiivis preesensi ainsuse 3. pöördes. Lausealgulistest adverbidest korduvad *Siin* ja *Nii*. Klastrid **DSD** täidab substantiiv adverbialina, objekti ja subjekti rolli. Klastrid **DJD** esineb rohkesti viisimäärust (*tõiselt ja ka, üli-viisakalt ja korralekutsuvalt*). Adverbide leksikaalne varieeruvus on suur. Klastrid **DPD** on eelistatud pronoomenit ainsuses.

### 2.4. Alamklassid D–P, D–A ja D–K

Alamklassi D–P moodustavad kuus klastrit: DVP (66,8%), DDP (20,6%), DPP (5%), DJP (3,3%), DSP (3,1%) ja DAP (1,2%). Pea kõikide klastrite trigrammid on tüüpiliselt (osa)lause alguses. Adverbil on enamasti sidususfunktsioon, teistest enam korduvad *Nüüd, samas, siis*.

Alamklass D–A koosneb kaheksast klastrist: domineerib invariantne klaster DDA (58,1%), järgnevad DPA (16,5%), DVA (8,9%), DAA (8,4%), DSA (4,2%), DJA (2,3%), DKA (0,8%), DNA (0,8%). Kõikide alamklasside adverbivalik on rikkalik.

Alamklassis D–K on neli klastrit: DSK (82,4%), DPK (14,4%), DDK (2%) ja DNK (1,3%). Alamklassi iseloomustab samuti adverbide rikkalik leksikaalne varieerumine.

## 3. Substantiivialguline klass

Osakaalult teine morfoloogiline klass on substantiivialguline, osakaal 20,9% (vt joonis 2, lisa 2). Substantiivialgulised trigrammid liigituvad kaheksasse alamklassi (vt tabel 2). Tüüpilisim on substantiiv-adverb (S–D) *meeste puhul ühtmoodi*, osakaal valimis 33,3%. Järgnevad substantiiv–substantiiv (S–S) *linna karget värskust* ja substantiiv-verb (S–V) *pilgu mujale keeranud*. Oluliselt vähem on esindatud klassid substantiiv-konjunksioon (S–J) *töö pooleli ja*, substantiiv-adpositsioon (S–K) *ema selja taha*, substantiiv-adjektiiv (S–A) *süli oli pehme*, substantiiv-pronoomen (S–P) *sisetunne ütles mulle*, ja substantiiv-numeraal (S–N) *ema tütar esimesest*.

Alamklasside keskmine komponendi varieerumine on üsna avar. Nii jagunevad S–D ja S–J kuueks ning S–V ja S–A seitsmeks, alamklass S–S kaheksaks klastriks.

**Tabel 2.** Substantiivialguliste alamklasside osakaal valimis

Alamklass	S–D	S–S	S–V	S–J	S–K	S–A	S–P	S–N
Osakaal klassis	33,3%	24,0%	22,5%	7,1%	5,5%	5,2%	2,4%	0,1%

### 3.1. Alamklass S–D

Alamklass S–D sisaldab kuut morfoloogilist klastrit: SVD (58,2%), SDD (18,1%), SKD (11%), SSD (7,3%), SJD (3,9%) ja SPD (1,6%).

**SVD** on dominantne klaster; selle invariantne trigramm koosneb substantiivist ainsuse nominatiivis subjektina + öeldisverbist indikatiivis imperfekti ainsuse 3. pöördes + adverbist adverbialina, nt *helk kumas otsekui, töö edenes vaevaliselt*. Trigrammi leksikaalne varieerumine on avar substantiivil, veidi piiratum adverbil ja kõige piiratum verbil, kus ligi pool näidetest sisaldab *olema*-verbi, nt *isa oli seal, kool oli tookord, meel oli väga*. Korduvalt esinevad ka tuumverbid *hakkama, tulema* ja *liikuma*, mida on samuti eelistatud kasutada imperfektis ja ainsuse 3. pöördes. Adverbidest on enim viisi- ja seisundiadverbe (*häirivalt, kiuslikult, külmakartlikult, mänglevalt, ropult, rõõmsalt*), tuleb ette temporaal- (*ikka, täna, tookord, siis, varsti, möödus*), lokaal- (*kohal, kõrvale, peale, siia, siinsamas*), rõhu- (*ka, küll, lihtsalt, nii*) ning hulgaadverbe (*väga, üleni*).

Klastrit **SDD** invariantne trigramm on substantiiv ainsuse nominatiivis subjektina + kaks adverbilist adverbialia, nt *mees lõpuks välja, käsi nii enesestmõistetavalt*. Kõige sagedamini on kasutatud rõhuadverbi *ka* (*hing ka sisse, teadus siis ka*).

Klastrit **SKD** invariantne trigramm on substantiiv ainsuse genitiivis laiendina + adpositsioon + adverbiline adverbial, nt *juhtunu üle järele, puu najal kummaliselt*. Adpositsioonidest esineb enim *pärast*, adverbidest korduvad samad sõnad, mis sagedased ka eespool kirjeldatud klastrite puhul (*välja, veel* ja *ka*). Korduvalt esineb adverbialijärgendeid *üle järele* (<*mõtles*> *juhtunu üle järele*) ning *alt välja* (<*paistis*> *teki alt välja*).

### 3.2. Alamklass S–S

Alamklassis S–S on kaheksa klastrit: SJS (35,4%), SAS (21,9%), SVS (17,8%), SKS (10,8%), SDS (6,9%), SPS (3,6%), SSS (3,0%) ja SNS (0,5%). Osakaalu poolest tõusevad esile klastrid, mille keskmine komponent on konjunktsioon, adjektiiv ja verb.

Invariantse klastrit **SJS** invariantnes trigrammis seob rinnastav sidesõna kahte subjekti, mõlemad ainsuse nominatiivis, nt *tüli ja pahandus, jutt või luuletus*. Konjunktsioonide varieeruvus on väga väike, ülekaalukalt sagedaim on *ja*, vaid viiel korral esines *või*.

Klastrit **SAS** invariantne trigramm on substantiiv ainsuse genitiivis eestäiendina + adjektiiv ainsuse partitiivis eestäiendina + substantiiv ainsuse partitiivis objektina, nt *autori igapäevast elamist, naise siidpehmet kõhtu*.

Klastrit **SVS** valitseb suur vormivarieeruvus: klastrit 135 sõnajärgendit liigituvad 55 morfosüntaktiliselt erineva trigrammi alla. Selget tüüpilist trigrammi esile ei kerki. Sagedaim trigramm on lausealguline substantiiv ainsuse nominatiivis

subjektina + transitiivne öeldisverb indikatiivis imperfekti ainsuse 3. pöördes + substantiiv ainsuse genitiivis objektina, nt *Jumal lõi maailma, Tüdruk surus nina*.

### 3.3. Alamklassid S–V ja S–K

Alamklassis S–V on seitse klastrit: SVV (29,7%), SDV (25,8%), SKV (16,9%), SSV (13,4%), SJV (13,1%), SAV (0,6%) ja SPV (0,6%). Alamklassi klastreid iseloomustab suur vormivarieeruvus: enamikku trigrammidest on kasutatud vaid kahel-kolmel korral.

Invariantnes klastris **SVV** järgneb substantiivile tavaliselt eitusverb (*isa ei lähe, uhkus ei lubanud*), harvem verbi mineviku liitajavorm (*pilk oli otsinud, tähtedega oli kirjutatud*) või finiitverbi ja infinitiivi ühend (*naine võib sattuda, juuksed läikima pani*). Klaster **SDV** sisaldab enamasti ühendverbi, tavaliselt mineviku liitaja (<oli> *šokolaadi ära söönud*) või da-infinitiivi vormis (<võid> *ukse lahti teha*). Klaster **SKV** täidab adpositsioonifraas enamasti adverbialli rolli, adpositsioonidest korduvad enim *ees, eest, juures, poole, taha* ja *otsa*. Klaster **SSV** esineb sageli lause alguses (*Mehe vaim oli, Riide võilillekiri oli, Vannitoa uks avanes*). Klaster **SJV** hõlmab osalausepiiril esinevaid sõnajärgendeid, mille keskmine komponent on eranditult rinnastav sidesõna, tavaliselt *ja* (*kausi ja tõstis, õhuluugi või paneb*).

Alamklassi S–K nelja klatri seas on selge dominant **SSK**, mille osakaal on 86,6% (*poeg tooli najal, viiulikast käe otsas*). Ülejäänud kolme klatri osakaal pole kuigi märkimisväärne: SPK (8,7%), SVK (3,5%) ja SNK (1,2%).

## 4. Verbialguline klass

Verbiga algava klassi osakaal valimis on 24,6% (vt joonis 2, lisa 3). Invariantne alamklass on verb-adverb (V–D) *mõtleb Leopold nukralt*. Teisel kohal on verb-substantiiv (V–S) *kõlas katkendlik kellatirin*. Pea kolm korda harvem esineb alamklass verb-adjektiiv (V–A) *algab üks tüütu*. Ülejäänud alamklasside osakaal valimis pole kuigi märkimisväärne: verb-verb (V–V) *võib pihta hakata*, verb-konjunktsioon (V–J) *paistsid kaugete otsekui*, verb-pronoomen (V–P) *möödudes vahetasid nad*, verb-adpositsioon (V–K) *kadus vahemüüri taha* ja verb-numeraal (V–N) *astuda need kakssada* (vt tabel 3).

**Tabel 3.** Verbialguliste alamklasside osakaal valimis

Alamklass	V–D	V–S	V–A	V–V	V–J	V–P	V–K	V–N
<b>Osakaal klassis</b>	41,7%	26,3%	9,5%	8,2%	5,2%	4,6%	4,5%	0,2%



#### 4.1. Alamklass V–D

Invariantne alamklass V–D sisaldab viit klastrit, neist neli olulist: VDD (36,6%), VSD (22,0%), VPD (21,1%), VVD (17,9%), VJD (2,4%). Verbivalik on küll rikkalik, kuid eelistatakse siiski *olema*-verbi ainsuse 1. ja 3. pöördes.

Invariantses klastris **VDD** varieeruvad verbi ainsuse ja mitmuse pöörded ning *nud*-partitsiip, ka leksikaalne varieeruvus on avar, kuid sagedaim verb on *olema* (*oli üsna kõrgelt, olin juba päris, on ju lõpuks, oleks nagu veel*). Klastris **VDD** invariantne trigramm on öeldisverb indikatiivis imperfekti ainsuse 3. pöördes + kaks adverbilist adverbialina, nt *oli väga ammu, muutus ajapikku väga*.

Klastris **VSD** invariantne trigramm on transitiivne öeldisverb indikatiivis imperfekti ainsuse 3. pöördes + substantiiv ainsuse nominatiivis objektina + adverb adverbialina, nt *pööras pilgu ära, võttis buketi vastu, pani noa kõrvale*.

Klastris **VPD** invariantne trigramm on *olema*-verb indikatiivis imperfekti ainsuse 3. pöördes abitegusõnana + ainsuse 3. isiku pronoomen nominatiivis + adverb adverbialina, nt *oli ta salaja, oli ta sünkandis, oli ta vahepeal*. Verbivormidest domineerib ainsuse 3. pööre, nagu ka ainsuse 1. pööre (*sain ma liigagi, püüdsin ma küllalt*).

Klastris **VVD** invariantne trigramm on öeldisverb indikatiivi eitava kõne preesens + adverb adverbialina, nt *ei kao kuhugi, ei suuda kauem, ei lähe sinna, ei ole võlgu, ei ole ainult*. Klastris trigrammides tuleb hästi esile verbide leksikaalne rikkus.

#### 4.2. Alamklass V–S

Alamklass V–S koosneb seitsmest klastrist: VDS (27,2%), VAS (24,0%), VPS (17,9%), VSS (14,4%), VVS (10,4%), VJS (3,4%), VNS (2,7%). Verbide leksikaalne varieerumine on selles alamklassis avar, kuid siiski enim korduvad ainsuse 3. pöördes *olema, leidma ja tegema*.

Klastris **VDS** on verb üle poole juhtumitest imperfekti ainsuse 3. pöördes, kuid see pole seotud *olema*-verbiga. Samas klastris invariantne trigramm sisaldab koopulat (*olema* indikatiivi imperfekti ainsuse 3. pöördes) + adverb adverbialina + substantiiv ainsuse nominatiivis predikatiivina, nt *oli seal keskkoolidirektor, oli rohkem hingeart, oli teadagi kaunitar*.

Klastris **VAS** invariantne trigramm on öeldisverb indikatiivis imperfekti ainsuse 3. pöördes + adjektiiv ainsuse nominatiivis eestäiendina + substantiiv ainsuse nominatiivis subjektina, nt *tekkis ähmane lootuskiir, kasvas kõrge vili*. Klastris verbivalik on leksikaalselt piiratud: ligi poole näidetest moodustavad trigrammid *olema*-verbiga, varieerub vaid aja- ja pöördevorm. Seevastu adjektiivide leksikaalne varieeruvus on avar.

Klastris **VPS** korduvad verbid *olema, leidma ja tegema*. Substantiiv esineb nii subjekti (*näisid kõik mõtted*), objekti (*täita mingit ülesannet*), adverbialina (*tuleb meile külla*), predikatiivi (*on minu nimi*) kui ka tagasõna laiendina (*istus ka klaveri <taha>*). Invariantne trigramm on öeldisverb indikatiivis preesens ainsuse 3. pöördes + demonstratiivpronoomen ainsuse nominatiivis eestäiendina + substantiiv ainsuse nominatiivis subjektina, nt *tuleb see nüanss, on too paratamatus*.

Klastris **VSS** esineb suur verbi ja substantiivi leksikaalne varieeruvus. Verbidest on korduvalt kasutatud *olema* ainsuse ja mitmuse 3. pöördes. Invariantne trigramm on koopula (*olema* indikatiivi preesensi ainsuse 3. pöördes) + substantiiv ainsuse genitiivis eestäändina + substantiiv ainsuse nominatiivis predikatiivina, nt *on asjatundja pilk, on inimese tegu*.

#### 4.3. Alamklassid V–A ja V–K

Alamklass V–A koosneb kuuest klastrist: VDA (65,0%), VPA (14,6%), VSA (9,3%), VAA (7,5%), VVA (2,7%), VKA (0,9%). Dominantne klaster **VDA** vastandub osakaalu poolest ülejäänutele. Alamklassile on üldiselt iseloomulik piiratud verbivalik: domineerivad *olema*-verbi erinevad vormid. Adverbidest on korratud lekseeme *vist, nii, pisut, õige* ja *väga*, mis väljendavad valdavalt kõneleja suhtumist või rõhutavad öeldut.

Alamklass V–K koosneb neljast klastrist: VSK (80,2%), VPK (14,2%), VDK (3,8%), VNK (1,9%). Selge dominant on **VSK**, ülejäänud klastrite osakaal on samuti väike, nagu alamklassis V–A. Erinevus viimasest seisneb verbi tunduvalt avaramas leksikaalses varieerumises, kõige enam korduvad *olema, tulema, ronima* ja *minema* erinevad vormid.

### 5. Pronoomenialguline klass

Pronoomeniga algava klassi osakaal on valimi üldmahust 14,3% (vt joonis 2, lisa 4). Invariantne alamklass on pronoomen-adverb (P–D) *ma arvan küll*, järgnevad pronoomen-verb (P–V) *ta lõpuks saavutanud*, pronoomen-substantiiv (P–S) *selle võika olendi*, pronoomen-adpositsioon (P–K) *miski nende vahele*, pronoomen-pronoomen (P–P) *ta seda endale*, pronoomen-adjektiiv (P–A) *enda vastas paljast*, pronoomen-konjunktsioon (P–J) *selline näolapp ja* ning pronoomen-numeraal (P–N) *tal umbes üks* (vt tabel 4). Alamklasside P–D, P–V ja P–S osakaal on pronoomenialgulises klassis valdav, ülejäänud alamklasside osakaal ei ole kuigi oluline.

**Tabel 4.** Pronoomenialguliste alamklasside osakaal valimis

Alamklass	P–D	P–V	P–S	P–K	P–P	P–A	P–J	P–N
Osakaal klassis	39,6%	22,8%	20,4%	7,6%	3,7%	3,6%	2,2%	0,1%

#### 5.1. Alamklass P–D

Alamklass P–D koosneb viiest klastrist: PDD (32,0%), PVD (30,0%), PSD (26,3%), PPD (6,9%) ja PKD (4,8%). Alamklassi pronoomenivalik on piiratud, domineerivad personaalpronoomen *ta* ja demonstratiivpronoomen *see*.

Klastris **PDD** invariantne trigramm on ainsuse 3. isiku personaalpronoomen subjektina + kaks adverbilist adverbialia, nt *ta mõnikord ka, ta ükskord vahele*.

Klastris **PVD** invariantstes trigrammis on lausealguline demonstratiivpronoomen + öeldisverb indikatiivis imperfekti ainsuse 3. pöördes + adverb adverbiaalina, nt *See oli nii See hakkas nüüd*.

Klastris **PSD** korduvad demonstratiivpronoomen *see* ja impersonaalsuse marker *üks*; mõlemat lekseemi on kasutatud erinevates käändevormides. Klastris invariantne trigramm on demonstratiivpronoomen ainsuse nominatiivis substantiivi eestäiendina + substantiiv ainsuse nominatiivis subjektina + adverb adverbiaalina, nt *see külamees liiga see asi siin*.

## 5.2. Alamklass P–V

Alamklass P–V sisaldab kuut klastrit: PVV (33,1%), PDV (31,4%), PSV (29,0%), PPV (3,9%), PKV (2,2%) ja PAV (0,4%). Kolme esimese klastris osakaal valimis on oluline, kolme viimase osakaal aga marginaalne. Alamklassis domineerivad demonstratiivpronoomeni ainsuse 1. ja 3. isiku lühivormid ning demonstratiivpronoomen *see*.

Klastris **PVV** keskmise komponendi leksikaalne varieerumine on piiratud, kasutatakse *olema*-verbi jaatava ja eitava kõne vorme, pronoomenitest eelistatakse ainsuse 1. ja 3. isiku lühivormi (*ma ei esitanud, ta ei kiirustanud*). Invariantne trigramm on lausealguline personaalpronoomen ainsuse 3. isikus + pluskvamperfekti liitaja vorm öeldisena, nt *Ta oli sõitnud Ta oli maganud*.

Klastris **PDV** domineerivad personaalpronoomeni 1. ja 3. isiku lühivormid. Invariantne trigramm on ainsuse 1. isiku pronoomen subjektina + adverb adverbiaalina + verbi eitus, nt *ma lihtsalt ei mina muidugi ei*.

Klastris **PSV** on sage demonstratiivpronoomen *see* (*see töö edenes*), personaalpronoomenite pikk vorm ainsuses ja mitmuses (*meie linnas sündis, minu naisel oli*) ning lühivorm ainsuses (*ta prillid läigatavad, mu grimassid vaheldusid*). Invariantne trigramm on lausealguline demonstratiivpronoomen *see* ainsuse nominatiivis substantiivi eestäiendina + substantiiv ainsuse nominatiivis subjektina + öeldisverb indikatiivis imperfekti ainsuse 3. pöördes, nt *See tunne oli, See mullatükk paindus*.

## 5.3. Alamklassid P–S ja P–K

Alamklass P–S koosneb seitsmest klastrist: PAS (53,5%), PDS (15,0%), PVS (10,7%), PSS (8,7%), PPS (7,8%), PNS (2,4%) ja PKS (1,9%). Alamklassi P–S pronoomenite valik ei ole nii piiratud, kui eespool kirjeldatud alamklassides. Domineeriva klastris **PAS** (pronoomen-adjektiiv-substantiiv) invariantstes trigrammis on kasutatud vaid possessiivpronoomenit *oma* nt *oma vana sõpra oma professionaalset üleolekut*. Esile tuleb substantiivi ning sellega ühilduva adjektiivi käände varieerumine. Kõrvuti grammatiliste käänetega kasutatakse ka adessiivi (*tollel vihmasel detsembripäeval*), elatiivi (*nendest uutest apsudest*) ja inessiivi (*ta koduses kabinetis*). Käändevalik oleneb eelnevas kontekstis kasutatud põhisõna reksioonimallist (nt *nägin keda? millal? teda tollel vihmasel detsembripäeval*).

Alamklassis P–K on kolm klastrit: dominantne **PSK** (94,1%) ning osakaalu poolest väheolulised **PPK** (3,3%) ja **PDK** (2,6%). Pronoomenitest korduvad

peamiselt personaal-, demonstratiiv- ja refleksiivpronoomenid, eriti *see, oma* ja *tema* grammatiliste käänete vormid.

## 6. Konjunktsioonialguline klass

Konjunktsioonialgulise klassi osakaal valimis on 13,5% (vt joonis 2, lisa 5). Klass koosneb seitsmest alamklassist. Tüüpiline alamklass on konjunktsioon-verb (J–V) *aga võib-olla oli*; enam-vähem võrdse osakaaluga esindatud alamklassid on konjunktsioon-adverb (J–D) *ning veel rohkem* ja konjunktsioon-substantiiv (J–S) *ja soojad kindad*; alamklasside konjunktsioon-pronoomen (J–P) *ja vaevalt ta*, konjunktsioon-adpositsioon (J–K) *kuid aja jooksul*, konjunktsioon-adjektiiv (J–A) *kui ta kõrvalseisva*, ning konjunktsioon-konjunktsioon (J–J) *ja selge nagu* osakaal pole märkimisväärne (vt tabel 5). Kõik konjunktsioonialgulised alamklassid on suure klastrivarieeruvusega (igas 5–7 klastrit).

**Tabel 5.** Konjunktsioonialgulistele alamklassidele osakaal valimis

Alamklass	J–V	J–D	J–S	J–P	J–K	J–A	J–J
Osakaal klassis	38,6%	24,9%	22,3%	6,9%	3,7%	3,1%	0,5%

### 6.1. Alamklass J–V

Suurima osakaaluga alamklass J–V kätkeb kuut klastrit: JPV (34,9%), JSV (33,2%), JDV (26,3%), JVV (4,7%), JJV (0,5%) ja JAV (0,3%).

Invariantse klasteris **JPV** sisaldub 84 erinevat morfosüntaktilist trigrammi, mida on kasutatud enam-vähem ühtlaselt. Seetõttu ei tule esile tüüpilist trigrammi, kuid sagedaim on alistav sidesõna + ainsuse 3. isiku personaalpronoomen nominatiivis subjektina + öeldisverb indikatiivis preesensis ainsuse 3. pöördes, nt *et ta seisab, kui tema hakkab*. Selles trigrammis on kasutatud valdavalt alistavat sidesõna *et*, vaid korra *kui* ja *sest*. Verbidest domineerib *olema* indikatiivi preesensi vormis. Personaalpronoomenite puhul on eelistatud ainsuse 3. isiku lühivormi (*ta*).

Vormilise mitmekesisusega paistab silma ka klaster **JSV**, mille all 71 erinevat trigrammi. Invariantne trigramm on alistav sidesõna + substantiiv ainsuse nominatiivis subjektina + öeldisverb indikatiivis preesensis ainsuse 3. pöördes, nt *et trepp on, kuni tee kaob*. Konjunktsioonidest eelistatakse alistavat sidesõna *et*, *olema*-verbi indikatiivi preesensis, harvem on samas vormis kasutatud verbe *hakkama*, *kaduma*, *muigama*, *painduma*, *pöörduma* ja *surisema*.

Klasteris **JDV** sisaldub 47 morfosüntaktilist trigrammi. Invariantne on lausealguline rinnastav sidesõna + adverb adverbialina + öeldisverb indikatiivis imperfektis ainsuse 3. pöördes, nt *Aga siis juhtus, Ja nõnda tuli*. Konjunktsioonidest on ülekaalus *aga*, verbidest *olema*, adverbidest *siis*.

## 6.2. Alamklass J–D

Alamklass J–D jaguneb viieks klastriks: JVD (35,5%), JDD (29,8%) ja JPD (20,2%), JSD (12,8%) ja JJD (1,7%). Konjunktsioonidest on valdav rinnastav sidesõna *ja*.

Klastri **JVD** invariantne trigramm on rinnastav sidesõna + öeldisverb indikatiivis imperfekti ainsuse 3. pöördes + adverb adverbialina, nt *ja tuli uuesti, ja naeratas salapäraselt, vaid lamas kinnisilmi*. Trigramm eristub rikkaliku leksi-kaalse varieerumise poolest, eriti verbide osas. Nii on tuumverbide kõrval võrdselt esindatud täistähenduslikud verbid, nt *rebenes, moondus, räntsatas, paindus ja heldis*. Adverbidest korduvad sageli *siis, jälle, üles ja ometi*, ent samas esineb ka ilukirjanduskeelele iseloomulikke onomatopoeetilisi adverbide, nt *kärinal ja rabinal*.

Klastri **JDD** invariantne trigramm koosneb rinnastavast sidesõnast + kahest adverbilisest adverbialist, nt *ning omamoodi ka, ja isegi nagu, aga siit enam*. Esile tuleb kaks mustrit: 1) *ja* trigrammi alguses ja *nagu* lõpus, nt *ja isegi nagu, ja vist nagu, ja pisut nagu, ja kuskilt nagu* ning 2) rõhusõna *ka* trigrammi keskel või lõpus, nt *ja ka natuke, aga ka päris, ja muidugi ka, ning omamoodi ka*.

## 6.3. Alamklass J–S

Alamklass J–S on konjunktsioonialgulises klassis kõige suurema klastri varieeruvusega: JAS (34,8%), JVS (20,9%) ja JPS (19,3%), JSS (11,8%), JDS (8,5%), JJS (2,4%) ja JNS (2,4%). Invariantse klastri **JAS** invariantne trigramm algab rinnastava sidesõnaga (tavaliselt *ja*), järgneb adjektiiv eestäiendina ja substantiiv ainsuse nominatiivis, nt *ja valge karikakar, aga tüütu karvanäss*. Substantiivi leksikaalne varieeruvus on võrdlemisi suur.

## 7. Adjektiivialguline klass

Adjektiivialgulise klassi osakaal valimis on 7,3% (vt joonis 2, lisa 6). Selle klassi invariantne alamklass on adjektiiv-substantiiv (A–S) *märge läikivat nahkuube*. Järgnevad adjektiiv-verb (A–V) *noored naised meeldivad* ja adjektiiv-adverb (A–D) *vanad trepiastmed rõõmsalt*. Ilukirjanduskeelele vähem iseloomulikud alamklassid on adjektiiv-adpositsioon (A–K) *suletud ukse ees*, adjektiiv-adjektiiv (A–A) *kauge ja salapärane*, adjektiiv-konjunktsioon (A–J) *hea sõber ning*, adjektiiv-pronoomen (A–P) *hilissuvisel päikesel enda* ja adjektiiv-numeraal (A–N) *hõõguvad peopesad tuhandetest* (vt tabel 6).

Kõige suurema osakaaluga alamklassid A–S, A–V ja A–D on samas ka suurima klastri varieeruvusega, adjektiivi kasutatakse enamjaolt substantiivi eestäiendina. Alamklassid A–K, A–J ja A–N koosnevad vaid ühest klastrist, mille keskmine komponent on substantiiv.

Tabel 6. Adjektiivialgulistele alamklassidele osakaal valimis

Alamklass	A–S	A–V	A–D	A–K	A–A	A–J	A–P	A–N
Osakaal klassis	26,7%	18,7%	16,6%	14,5%	11%	10,7%	1,5%	0,2%



## 7.1. Alamklass A–S

Alamklass A–S on ilukirjanduskeelele tüüpiline. Keskmise komponendi varieerumise alusel eristub neli klastrit: AAS (54,3%), ASS (42,0%), AJS (2,9%) ja APS (0,7%).

Dominantse klatri **AAS** vormide varieerumine on avar (kokku 40 morfosüntaktilist trigrammi). Varieeruvad adjektiivide võrdlusaste ning substantiivi ning adjektiivide arvu- ja käändekategooria. Invariantne trigramm on kaks adjektiivset eestäiendit + substantiiv ainsuse (harvem mitmuse) nominatiivis subjektina, nt *mahe kurblik hääl, keskealine tüse naisterahvas; ilusad valged lilled, pikad rohekad juuksed*. Trigrammi substantiivide leksikaalne varieerumine on maksimaalne – enamuse lekseeme on kasutatud ainukordselt. Adjektiivide hulgas esineb kordusi, nt *pikad, suur ja väike*.

Klaster **ASS** on veelgi suurema vormide varieerumisega: invariantne trigramm on adjektiivne ja substantiivne eestäiend ainsuse genitiivis + substantiiv ainsuse nominatiivis subjektina, nt *kauge automootori surin, valge pesuriide pind*.

## 7.2. Alamklass A–V

Alamklass A–V koosneb neljast klastrist: ASV (95,9%) on ülitugev dominant, üksiknäidetega on esindatud klastrid AVV (2,1%), AJV (1,0%) ja ADV (1,0%).

Klaster **ASV** sisaldab 65 morfosüntaktilist trigrammi, grammatiline varieerumine on avar. Vahelduvad verbi aja- ja pöördevormid, finiitsus-infiniitsus, tegevusnime- või kesksõnatunnus. Klastrisse kuuluvad ka eitava kõne trigrammid, nagu *kurnatud inimene ei*, kus eituspartiklile järgneb verbi finiitvorm või moodustub liitpredikaat. Adjektiividel varieerub kääne ja arv, võrdlusastmetest on valdav positiiv, vaid korra esines komparatiiv. Klatri ASV invariantne trigramm on adjektiiv ainsuse nominatiivis eestäiendina + substantiiv ainsuse nominatiivis subjektina + öeldisverb indikatiivis imperfekti 3. pöördes, nt *täielik hullumeelsus oli, paha lugu juhtus, vintske vanamees suri. olema-verbi kasutus on tavaline*.

## 7.3. Alamklassid A–D, A–K, A–J ja A–A

Alamklassi A–D dominantne klaster on **ASD**, osakaal valimis 82,6%. Ülejäänud klastrite osakaal jääb selgelt alla (AJD 14,0%) või on marginaalne (ADD 2,3%) ja AVV (1,2%). Klatri **ASD** täiendab adjektiiv enamasti objekti või adverbiaali rollis olevat substantiivi, harvem on substantiiv subjekti funktsioonis. Iseäranis siis, kui substantiivi on kasutatud objektina, kuulub adverb ühendverbi koosseisu (<tegi> *kohustusliku töö ära*). Ehkki võiks oodata, et adjektiivide leksikaalne varieeruvus on suurem kui adverbidel, leidub korduvaid lekseeme umbes sama palju, nt sagedaimat adjektiivide (*suur*) ja adverbide (*juba*) on kasutatud võrdses matus.

Alamklassis A–K on üks klaster **ASK**, mis hõlmab 23 erisugust trigrammi, nt *kaharate okste all*. Adpositsioonidest kipuvad korduma *eest, juures, alla, käes ja pärast*.

Ka alamklass A–J koosneb ühest klastrist **ASJ**. Stereotüüpselt on kasutatud rinnastavat sidesõna *ja*, nt *suur südamevalu ja*.

Alamklassis adjektiiv-adjektiiv (A–A) on kolm klastrit: dominantne AJA (83,3%), vähese osakaaluga ASA (14,9%) ja marginaalne AAA (1,8%). Klasteris **AJA** kasutatakse alati rinnastavat sidesõna *ja*, nt *tühi ja vaikne*.

## 8. Kokkuvõte

Uurimuse põhjal saab väita, et eesti ilukirjanduskeeles esineb kõige sagedamini adverbi-, substantiivi- ja verbialgulisi trigramme (sagedaimad sõnaliigijärjendid on SVD, VDD, DDV ja DAS). Ajavormidest eelistatakse indikatiivi imperfekti, preesens esineb harvem; ainsus on sagedasem kui mitmus; 3. isik domineerib *mina*-vormi üle; sagedasemad pronoomenid on *see* ja *ta*. Rinnastuse puhul eelistatakse sidesõna *ja*, alistuse puhul *et*, pronoomenipaarides *tema-ta* ja *mina-ma* lühivormi. Leksikaalselt rikkalikult esindatud sõnaliigid on substantiiv, adjektiiv ja adverb. Verbidest domineerib *olema*, väga sageli kasutatakse ka *hakkama*-verbi, harva onomatopoeetilisi verbe (*kõssitama*, *potsatama*, *räntsatama*) ja kestust tähistavaid verbe (*sugenema*, *valendama*).

Ilukirjanduskeele lingvistiline klasteranalüüs tõi esile seose vormide ja leksikavariatsioonide vahel: mida suurem on vormivariatsioon, seda väiksem kipub olema leksikaalne, ja vastupidi. Nt alamklassis verb-adpositsioon on verbi leksikaalne variatsioon väga suur, kuid vormivalik piiratud: VSK (*liikus ukse poole*) on nelja klasteri seas dominant.

Hea näide suurest vormi- ja piiratud leksikaalsest variatsioonist on pronoomeni- ja konjunktsioonialgulised trigrammid. Esimene hõlmab kaheksat alamklassi, teine seitset; mõlemal on suur klasterivariatsioon, kuid pronoomeni ja konjunktsiooni leksikaalne variatsioon on väga piiratud. Klasterites ei kerki esile tüüpilisi ja stereotüüpseid trigramme, nagu adverbi-, substantiivi- ja adjektiivialgulistes klasterites. Selgeimad dominandid joonistuvad välja adjektiivialgulistes klasterites, kusjuures alamklassisisene klasterivariatsioon on sel juhul äärmiselt väike. Samas ei ole välistatud ka avar vormivariatsioon koos leksikaalse rikkusega. Ühes klasteris võib samaaegselt olla hulgaliselt trigramme koos vaheldusrikka leksikaga. Niisugused on substantiivi- ja verbialguliste klasside alamklassid, kus klasterivariatsioon on väga suur ning klasterite invariantseid trigramme ei ole ülejäänud trigrammidest oluliselt sagedamad. Osati käib see ka adverbialgulise klassi kohta.

Niisiis toob eesti ilukirjanduskeele klasteranalüüs esile kindlaid keelekasutusmustrid. Sageli korduvad sõna(vormi)de järjendid kinnistuvad, teatud sõnade ja vormide vahele tekib justkui magnetväli, mis neid üksteise poole tõmbab. Nt klasteri DVP (*samas leidis ta*, *samas märkas ta*, *samas oli ta*) aluseks on muster *samas* + öeldisverb lihtmineviku 3. pöördes + 3. isiku personaalpronoomen. Adverbi ja verbi leksikaalse variatsioon alusel (*siiski lükkas ta*, *siiski jättis ta*) kinnistub see struktuur keelekasutusmustrina. Niisugusest teabest on kasu eesti keele autoomaattöötlusvahendite tarvis, aitab seletada kasutuspõhise keelekäsitluse üld- ja üksikküsimusi, mh võrrelda eesti keele arengutendentse diakroonilises aspektis.

Ilukirjanduskeele klasteranalüüsi tulemuste võrdlus eesti keele grammatika, käsiraamatu, keeleõpikute ja koolile mõeldud programmide sisuga aitab leida vastuseid küsimusele, kuid võrd need allikad kajastavad autentsele keelekasutusele omaseid tendentse. Nähtustele, mis on ilukirjanduskeeles harvad või marginaalsed,

võiks keeleõppes pöörata vähem tähelepanu, samas kui keelekasutusmuustritele ja suurema esinemissagedusega keelestruktuuride kasutamisele võiks panna rohkem rõhku. Ilukirjanduskeel kui kirjaliku keelekasutuse ideaalvariant sobib malliks nii emakeelekõnelejale kui ka eesti keele kui teise keele omandajale.

## Viidatud kirjandus

- Barddal, Johanna 2003. Case in Icelandic. – A Synchronic, Diachronic and Comparative Approach. Lundastudier i Nordisk Språkvetenskap A 57. Department of Scandinavian Languages, Lund University. <http://org.uib.no/iecastp/barddal/chapter2.pdf> (24.6.2013).
- Barlow, Michael; Kemmer, Suzanne (Eds.) 2000. Introduction. – Usage Based Models of Language. Stanford: CSLI Publications. <http://www.press.uchicago.edu/ucp/books/book/distributed/U/bo3624529.html> (10.2.2014).
- Burns, Robert; Burns, Richard 2008. Cluster analysis. – Business Research Methods and Statistics Using SPSS. Additional advanced chapters online: <http://www.uk.sagepub.com/burns/chapters.htm> (5.8.2013). SAGE Publications Ltd.
- Bybee, Joan L. 2003. Phonology and language use. Cambridge: Cambridge University Press.
- Bybee, Joan L. 2010. Language, Usage and Cognition. Cambridge University Press. <http://dx.doi.org/10.1017/CBO9780511750526>
- Bybee, Joan L.; Hopper, Paul J. 2001. Frequency and the Emergence of Linguistic Structure. Typological Studies in Language 45. John Benjamins Publishing Company. <http://dx.doi.org/10.1075/tsl.45>
- Croft, William 2003. Typology and Universals. Second edition. Cambridge Textbooks in Linguistics. Cambridge University Press.
- Eslon, Pille 2013. Kahe keelekasutusvariandi võrdlus: morfoloogilised klassid ja klastrid. [The comparative study of language use: Morphological classes and clusters.] – Lähivõrdlusi. Lähivertailuja, 23, 13–38. <http://dx.doi.org/10.5128/LV23.01>
- Eslon, Pille 2014. Morfosüntaktilise ja leksikaalse varieerumise piiridest. – Eesti Rakenduslingvistika Ühingu aastaraamat, 10, 55–71. <http://dx.doi.org/10.5128/ERYa10.18>
- Habicht, Külli 2006. Meie muutuv kirjakeel. [Our Literary Language in Change.] – Oma Keel, 1, 12–19.
- Halliday, Michael A. K. 1994. An Introduction to Functional Grammar. London: Arnold.
- Haspelmith, Martin 2006. Against markedness (and what to replace it with). – Journal of Linguistics, 42 (1), 25–70. <http://dx.doi.org/10.1017/S0022226705003683>
- Haspelmith, Martin 2008. Frequency vs. iconicity in explaining grammatical asymmetries. – Cognitive Linguistics, 19 (1), 1–33. <http://dx.doi.org/10.1515/COG.2008.001>
- Niglas, Katrin s. a. Klasteranalüüs. [Cluster Analysis.] <http://www.cs.tlu.ee/~katrin/wp/wp-content/uploads/2013/11/klaster.pdf> (24.6.2013).
- Norvig, Peter 1989. Building a large lexicon with lexical network theory. – Proceedings of the IJCAI Workshop on Lexical Acquisition. USA: University of California at Berkeley. <http://krextown.googlecode.com/svn/trunk/papers/lnt.pdf> (26.9.2013).
- Ots, Sander 2012. Statistikapõhise tarkvara loomine morfoloogiliste kollokatsioonide eraldamiseks eesti keele tekstidest. [Software for Morphosyntactic Cluster Extraction from Estonian Texts.] Bakalaureusetöö. Tallinn: Tallinna Ülikooli informaatika instituut.
- Remm, Kalle; Remm, Jaanus; Kaasik, Ants 2012. Ruumiliste loodusandmete statistiline analüüs. [Statistical Analysis of Spatial Data.] Õpik-käsiraamat. Tartu Ülikooli ökoloogia ja maateaduste instituut. [http://kalleremm.ee/RASA/RASA\\_teooria.pdf](http://kalleremm.ee/RASA/RASA_teooria.pdf) (4.8.2013).

- Tuldava, Juhan 1977. Sagedussõnastik leksikostatistilise uurimise objektina. [The frequency dictionary as an object of lexicostatistical investigation.] – Tööd keelestatistika alalt. Keelestatistika II. Tartu: Tartu Riikliku Ülikool, 141–171.
- Tullo, Catriona; Hurford, James R. 2003. Modelling Zipfian distributions in language. – Proceedings of Language Evolution and Computation Workshop / Course at ESSLLI. University of Edinburgh, 62–75. <http://www.ling.ed.ac.uk/~jim/zipfjrh.pdf> (16.2.2013).
- Tartu Ülikooli ühtlustatud ilukirjanduskorpus. [Disambiguated Fiction Corpus of Estonian.] [http://lepo.it.da.ut.ee/~heli\\_u/SA/](http://lepo.it.da.ut.ee/~heli_u/SA/) (24.6.2013).
- Veismann, Ann 2008. Eesti kaassõnade tähenduste kirjeldamise probleemidest. [A cognitive semantic analysis of the Estonian adpositions *üle* 'over' and *peale* 'on, onto': Some questions and queries.] – Keel ja Kirjandus, 5, 335–352. <http://keeljakirjandus.eki.ee/Veismann.pdf> (25.12.2013).

**Jekaterina Trainis** (Tallinna Ülikool) on käsitlenud Johannes Aaviku Rootsi-perioodi sõnavara, doktoritöös uurib semantilisi käändeid diakroonsest aspektist.  
Narva mnt 25, 10120 Tallinn, Eesti  
[jekaterina.trainis@tlu.ee](mailto:jekaterina.trainis@tlu.ee)

**Kais Allkivi** (Tallinna Ülikool) on uurinud alaleütleva käände kasutust, magistritöös võrdleb eesti kirjakeele ja õppijakeele verbikonstruktsioone.  
Narva mnt 25, 10120 Tallinn, Eesti  
[kais.allkivi@tlu.ee](mailto:kais.allkivi@tlu.ee)

# ON BELLETRISTIC LANGUAGE FROM A NEW PERSPECTIVE

**Jekaterina Trainis, Kais Allkivi**

Tallinn University

The article describes the usage patterns of the Estonian belletristic language, which indicate how native speakers prefer to express their thoughts in writing. Research material is taken from the syntactically analyzed and disambiguated corpus of Estonian literary language. Cluster analysis that has been applied brings forth identical linear subsequences of morphological and syntactic tags based on frequency. Morphosyntactic trigrams that appeared in the corpus at least twice have been selected for the analysis.

The patterns of language use have been presented on the basis of the lexical category of the first component of constructions. Every morphological class consists of subclasses that have a common beginning and ending component, and every subclass consists of clusters that are distinguished by the variation of the middle components of trigrams. A cluster contains trigrams that have the same subsequence of morphological and syntactic tags. This hierarchy reveals general tendencies of language use as well as the limits of morphosyntactic and lexical variability.

**Keywords:** belletristic language, cluster analysis, morphosyntactic variability, lexical variability, Estonian



## Lisa 1. Adverbialguline klass

Alam-klassid	Klastrid	Klastrisagedus
<b>D-S</b>	DAS	344
	DDS	169
	DPS	161
	DVS	142
	DSS	84
	DJS	46
	DNS	46
	DKS	27
	DGS	3
Kokku		<b>1022</b>
<b>D-V</b>	DDV	352
	DVV	219
	DJV	111
	DSV	99
	DPV	70
	DAV	31
	DXV	2
Kokku		<b>884</b>
<b>D-D</b>	DDD	265
	DVD	125
	DSD	91
	DJD	89
	DPD	87
	DAD	6
Kokku		<b>663</b>
<b>D-P</b>	DVP	175
	DDP	54
	DPP	13
	DJP	9
	DSP	8
	DAP	3
Kokku		<b>262</b>

Alam-klassid	Klastrid	Klastrisagedus
<b>D-A</b>	DDA	151
	DPA	43
	DVA	23
	DAA	22
	DSA	11
	DJA	6
	DKA	2
	DNA	2
	Kokku	
<b>D-J</b>	DDJ	75
	DVJ	59
	DAJ	51
	DSJ	41
Kokku		<b>226</b>
<b>D-K</b>	DSK	126
	DPK	22
	DDK	3
	DNK	2
Kokku		<b>153</b>
<b>D-N</b>	DNN	4
	DDN	2
Kokku		<b>6</b>
<b>Üldarv kokku</b>		<b>3476</b>

## Lisa 2. Substantiivialguline klass

Alam-klassid	Klastrid	Klastrisagedus
<b>S-D</b>	SVD	611
	SDD	190
	SKD	116
	SSD	77
	SJD	41
	SPD	17
Kokku		<b>1052</b>
<b>S-S</b>	SJS	268
	SAS	166
	SVS	135
	SKS	82
	SDS	52
	SPS	27
	SSS	23
	SNS	4
Kokku		<b>757</b>
<b>S-V</b>	SVV	211
	SDV	183
	SKV	120
	SSV	95
	SJV	93
	SAV	4
	SPV	4
Kokku		<b>710</b>
<b>S-J</b>	SDJ	79
	SVJ	54
	SKJ	44
	SSJ	41
	SAJ	4
	SJJ	2
Kokku		<b>224</b>

Alam-klassid	Klastrid	Klastrisagedus
<b>S-K</b>	SSK	149
	SPK	15
	SVK	6
	SNK	2
Kokku		<b>172</b>
<b>S-A</b>	SVA	80
	SDA	37
	SKA	24
	SJA	14
	SSA	4
	SAA	2
SPA	2	
Kokku		<b>163</b>
<b>S-P</b>	SVP	44
	SJP	16
	SDP	8
	SKP	7
Kokku		<b>75</b>
<b>S-N</b>	SVN	2
	SVN	2
Kokku		<b>4</b>
<b>Üldarv kokku</b>		<b>3157</b>

### Lisa 3. Verbiälguline klass

Alam-klassid	Klastrid	Klastrisagedus
<b>V-D</b>	VDD	362
	VSD	218
	VPD	209
	VVD	177
	VJD	24
Kokku		<b>990</b>
<b>V-S</b>	VDS	170
	VAS	150
	VPS	111
	VSS	90
	VVS	65
	VJS	21
	VNS	17
Kokku		<b>624</b>
<b>V-A</b>	VDA	147
	VPA	33
	VSA	21
	VAA	17
	VVA	6
	VKA	2
	Kokku	
<b>V-V</b>	VDV	75
	VJV	41
	VSV	29
	VPV	24
	VVV	24
	VAV	2
Kokku		<b>195</b>

Alam-klassid	Klastrid	Klastrisagedus
<b>V-J</b>	VSJ	51
	VDJ	32
	VAJ	27
	VVJ	13
Kokku		<b>123</b>
<b>V-P</b>	VVP	64
	VDP	31
	VPP	8
	VJP	6
Kokku		<b>109</b>
<b>V-K</b>	VSK	85
	VPK	15
	VDK	4
	VNK	2
Kokku		<b>106</b>
<b>V-N</b>	VDN	2
	VPN	2
Kokku		4
<b>Üldarv kokku</b>		<b>2377</b>

#### Lisa 4. Pronoomialguline klass

Alam-klassid	Klastrid	Klastrisagedus
<b>P-D</b>	PDD	256
	PVD	240
	PSD	210
	PPD	55
	PKD	39
Kokku		<b>800</b>
<b>P-V</b>	PVV	153
	PDV	145
	PSV	134
	PPV	18
	PKV	10
	PAV	2
Kokku		<b>462</b>
<b>P-S</b>	PAS	221
	PDS	62
	PVS	44
	PSS	36
	PPS	32
	PNS	10
	PKS	8
Kokku		<b>413</b>
<b>P-K</b>	PSK	145
	PPK	5
	PDK	4
Kokku		<b>154</b>

Alam-klassid	Klastrid	Klastrisagedus
<b>P-P</b>	PVP	34
	PDP	15
	PSP	14
	PPP	6
	PJP	4
	PKP	2
Kokku		<b>75</b>
<b>P-A</b>	PDA	35
	PVA	20
	PAA	6
	PSA	6
	PKA	5
Kokku		<b>72</b>
<b>P-J</b>	PSJ	30
	PDJ	6
	PAJ	4
	PPJ	4
Kokku		<b>44</b>
<b>P-N</b>	PDN	2
Kokku		<b>2</b>
<b>Üldarv kokku</b>		<b>2022</b>

## Lisa 5. Konjunktsioonialguline klass

Alam-klassid	Klastrid	Klastrisagedus
<b>J-V</b>	JPV	257
	JSV	245
	JDV	194
	JVV	35
	JJV	4
	JAV	2
Kokku		<b>737</b>
<b>J-D</b>	JVD	169
	JDD	142
	JPD	96
	JSD	61
	JJD	8
Kokku		<b>476</b>
<b>J-S</b>	JAS	148
	JVS	89
	JPS	82
	JSS	50
	JDS	36
	JJS	10
	JNS	10
Kokku		<b>425</b>

Alam-klassid	Klastrid	Klastrisagedus
<b>J-P</b>	JDP	58
	JVP	30
	JJP	20
	JPP	15
	JSP	7
	JKP	2
	Kokku	
<b>J-K</b>	JSK	56
	JPK	15
Kokku		<b>71</b>
<b>J-A</b>	JDA	21
	JPA	14
	JSA	12
	JVA	8
Kokku		<b>60</b>
<b>J-J</b>	JSJ	6
	JAJ	3
Kokku		<b>9</b>
<b>Üldarv kokku</b>		<b>1910</b>



## Lisa 6. Adjektiivialguline klass

Alam- klassid	Klastrid	Klastrisagedus
<b>A-S</b>	AAS	150
	ASS	116
	AJS	8
	APS	2
Kokku		<b>276</b>
<b>A-V</b>	ASV	186
	AVV	4
	AJV	2
	ADV	2
Kokku		<b>194</b>
<b>A-D</b>	ASD	142
	AJD	24
	ADD	4
	AVV	2
Kokku		<b>172</b>
<b>A-K</b>	ASK	150
Kokku		<b>150</b>
<b>A-A</b>	AJA	95
	ASA	17
	AAA	2
Kokku		<b>114</b>
<b>A-J</b>	ASJ	111
Kokku		<b>111</b>
<b>A-P</b>	ASP	14
	AVP	2
Kokku		<b>16</b>
<b>A-N</b>	ASN	2
Kokku		<b>2</b>
<b>Üldarv kokku</b>		<b>1035</b>