

# MORFOSÜNTAKTILISE JA LEKSIKAALSE VARIEERUMISE PIIRIDEST: ILUKIRJANDUS- JA ÕPPIJAKEELE KASUTUSMUSTRITE VÕRDLU

Pille Eslon

**Ülevaade.** Artiklis avaldatud klasteranalüüsi tulemused on saadud eesti ilukirjandustekstide ja õppija kirjutatud loomingulist laadi esseede keelekasutusmustrite võrdlemisel, aluseks formaalgrammatilise ja funktsionaalse lähenemise sidusus kasutuspõhises keelekäsitluses. Ühelt poolt eristatakse keelt kui hierarhilist süsteemi ja teisalt keelekasutust kui sõnade ja vormide koosesinemist tekstis, kus keeleüksused on seotud kindlat liiki leksikaalsete ning morfosüntaktiliste piirangute ja varieerumise piiridega. Tekstide automaatne töötlus toob esile keelekasutusmustrid, võimaldades teha mh järeldusi nii emakeelekõneleja kui ka õppija sõna- ja vormieelistuste ratsionaalsuse kohta.\*

**Võtmesõnad:** kasutuspõhine keelekäsitlus, teksti automaatne töötlus, lingvistiline klasteranalüüs, morfosüntaks, õppijakeel, eesti keel

## 1. Uurimuse lähted

Uurimus tugineb ühelt poolt Saussure'i dihhotoomiade dialektilisele loogikale ja teisalt formaalgrammatilise ning funktsionaalse lähenemise ühtsusele kasutuspõhises keelekäsitluses (nt Barlow, Kemmer 2000: 7–28, Bybee 2007, Bybee, Eddington 2006, Bybee, Hopper 2001, Verhagen 2009: 119–152, Croft 2003, Goldberg 2006). Ka eesti keele uurimustes on lähtutud konstruktsioonigrammatikast ja kasutuspõhisest keelekäsitlusest (Sahkai 2008, Penjam 2008, Veismann 2009, Klavan 2012, Uiboaed 2013 jt).

Käesoleva uurimuse olulised üldmõisted on invariant, sagedus ja keele funktsionaalne potentsiaal, mis kerkivad esile vajadusest eristada või integreerida grammatikat ja keelekasutust.

Grammatika alla kuulub morfoloogiliste paradigmade ja kategooriate süsteemne esitus. Kaasaegne muutemorfoloogia mudel on mitmemõõtmeline: tugineb

\* Tööd on toetanud riikliku programmi "Eesti keel ja kultuurimälu (2009–2013)" projekt "REKKi käsikirjaliste materjalide digiteerimine, Eesti vahekeele korpuse alamkorpuste loomine ja korpuse kasutusvõimaluste populariseerimine (2009–2013)"; ETF-i grant 8222 "Ülekantud tähenduses fraasid eesti keele korpustes (2010–2013)" ja ETF-i grant 8240 "Lähtekeele mõju ja teise keele omandamine: korpuspõhine uurimus (2010–2013)".

erinevate tunnuste ristumisele (nt Roman Jakobsoni universaalne verbikategooriate kirjeldus), binaarsetele tunnustele (nt Nikolai Trubetzkoy fonoloogiasüsteem), algoritmilisele esitusele, kus iga järgnev tunnustepaar tuleneb eelneva(te)st, võimaldades kirjeldada grammatikaseoseid ja vormide funktsionaalset potentsiaali integreeritult ning tuua esile oluliste/mitteoluliste tähenduste eristamisele tuginev semantiline võrgustik, mille alusel üldistuvad invariantseid tähendused (nt Mihhail Šeljakini vene verbi integreeritud kirjeldus).

Grammatikas on vormi invariantseks tähenduseks üldine tunnus, mis seda vormi kõigist teistest vormidest eristab ja samas nendega ühendab. Bondarko väidab, et Jakobson on avanud invariantse mõiste just selles aspektis ehk grammatikasüsteemi keskselt, samas aga üldistuvad invariantseid tähendused tekstikasutuse põhjal, vormi funktsionaalne potentsiaal avaneb. (Bondarko 2002: 159–163 jj) Grammatikas ning tekstikasutuses regulaarselt ilmnevate kategoriaalsete seoste ja semantika vahel on kindlat liiki piirangud, mis eksplitseeruvad erinevates kombinatsioonides ja erineva tekstisagedusega. See, mis on invariant ühtedes seostes, võib teistes olla variant või fakultatiivne nähtus – piir keelesüsteemi ning keelekasutuse vahel on olemuselt hajus.

Kui keele kategoriaalseid seoseid ja vormistikku saab esitada erinevate hierarhiline tunnuste ristumisele ning loogilisele vastandatusele rajaneva paradigmaatilise mudelina, seletada grammatikareeglitega, siis vormide funktsionaalne potentsiaal avaneb keeleüksuste lineaarse kooskasutuse tingimustes, kus grammatikareeglid üksüheselt ei rakendu. Nt ülekantud tähenduses vormikasutust ja ekspressiivsete värvingute lisandumist ei saa seletada morfoloogiareeglitega, küll aga on abi tekstikasutuse morfosüntaktiliste, semantiliste ja pragmaatiliste piirangute tundmisest. Need näitavad elementide kombineerumisvõimalusi ja regulaarsust tekstis (dominantne, sage, harv, kasutamata). On põhimõtteline erinevus, kas lingvistiline kirjeldus rajaneb semantiliste tunnuste loogilisel hierarhial ja kategoriaalsetel seostel (grammatikakirjeldus) või lingvistilistel tervikobjektidel, millel on kindlad leksikaalse ja morfosüntaktilise varieerumise piirid ning sagedus (kasutuspõhine keelekirjeldus).

Sageduse puhul eristatakse sõnastiku- ja tekstisagedust, mõlemad on kesksel kohal semantika ja grammatika integreerimisel, omades seetõttu teoreetilist väärtust (Greenberg 1990: 219–220). Käesolevas uurimuses on vaadeldud keeleüksuste tekstisagedust. See on samasugune üldmõiste kui invariant grammatikas: mõlemat saab võrdväärselt rakendada nii vormide kui ka nende funktsionaalse potentsiaali avanemisel. Tekstikasutuses sagedased struktuurid on seotud kindlate tekstiliste funktsioonidega. Selge dominandi ehk invarianti puudumisel võivad selle staatusele pretendeerida ka nt kaks ja enam keelekasutusmustrit.

Haspelmath (1999, 2007) on märkinud, et keelekasutusest lähtuvad uurimused vajavad teooriat, mis seletaks, mille põhjal sõnu ja vorme valitakse ning reastatakse; mida tekstikasutuses eelistatakse kombineerida ja kuidas seda tehakse; missugused leksikaalsed ja morfosüntaktilised piirangud või varieerumise võimalused seejuures avanevad. Croft (1999: 206) on pidanud oluliseks süntagmaatilise kombineerimise printsiipi, mis tema arvates võimaldavad Haspelmathil integreerida formaalgrammatilise ja funktsionaalse keelekäsitluse. Küsimus on selles, mis tingib keelekasutaja valikuid, missugust kooskasutust peetakse loomulikuks. Ellise (2006) arvates seostub loomulikkus valikute ratsionaalsuse ja optimaalsusega, mh ka keele valdamise tasemega.

Siinses uurimuses on tuginetud eespool sõnastatud arusaamale üldmõistete osast keele elementide funktsionaalse potentsiaali avanemisel ja tekstikasutuses oluliste tervikobjektide kirjeldamisel.

## 2. Uurimuse eesmärk, objekt ja meetod

Avaram eesmärk on seotud keele elementide funktsionaalse potentsiaaliga eesti kirjakeeles, mis väljendub morfoloogilistes, morfosüntaktilistes ja leksikaalsetes piirangutes ning varieerumise piirides. Kitsam eesmärk on keelekasutusmuustrite automaatne tuvastamine, et välja tuua emakeelekõneleja ja õppija tekstiloome morfoloogilised, morfosüntaktilised ja leksikaalsed eelistused ning neid omavahel võrrelda. Rakenduslik olulisus seisneb võimaluses näidata, kuivõrd adekvaatne või nihkes on õppija tekstiloome võrreldes emakeelekõnelejale omaste leksikaalsete ja morfosüntaktiliste valikutega.

Vaadeldud ei ole kõiki tekstikasutusi: uurimisobjektiks on verbist vasakule jääv kontekst (nt *metsa vari ulatus, veel ei liikunud, ise pidid tegema, kord mul oli*). Ilukirjandustekstid (umbes 200 000 sõnet)<sup>1</sup> ja õppijakeele esseed (umbes 500 000 sõnet)<sup>2</sup> sisestati Sander Otsa programmeeritud klastrileidjasse, mis töötab andmekaeve põhimõttel, otsides valimitest sageduse alusel sarnaseid morfo- ja süntaksimärgendite lineaarseid järjendeid ehk n-gramme. Klastrileidja koostamisel on olnud eeskujuks WordSmith Tools, kfNgram ja Xaira (vt Ots 2012: 6, 16). Käesoleva uurimuse tarvis on valitud trigrammid, mis korduvad tekstis viis ja enam korda. Samalaadselt vormistatud trigrammide kogumeid on autor veidi tinglikult nimetanud lingvistilisteks klastriteks ja meetodit lingvistiliseks klasteranalüüsiks.<sup>3</sup> Kuigi võrreldavate valimite maht erineb, ei ole vormide esinemus korrelatsioonis sõnede arvuga (vt Biber jt 1998: 33) – erineva mahuga valimid on võrreldavad vormide osakaalu põhjal (vt tabel 1).

Klasteranalüüsi tulemusena leiti ilukirjandustekstide valimist kokku 904 trigrammi, mis jagunevad 18 klastri ja 6 klassi vahel; õppijakeele valimist 13941 trigrammi, mis jagunevad 38 klastri ja 8 klassi vahel (vt tabel 1). Tegu on hierarhilise jaotusega, mille aluseks trigrammi esimese ja viimase komponendi sõnaliigiline samalaadsus/erinemine ning keskmise komponendi varieerumine. Nt substantiiviga (S) algavate ja verbiga (V) lõppevate trigrammide keskmise komponendi sõnaliigilise varieerumise põhjal eristuvad klastrid SKV ehk substantiiv + adpositsioon + verb (*maja juures on*), SDV ehk substantiiv + adverb + verb (*süda natuke kripeldas*), SSV ehk substantiiv + substantiiv + verb (*metsa vari ulatus*) jne.<sup>4</sup> Need klastrid reastuvad sageduse alusel substantiivialgulisse morfoloogilisse klassi S–V. Tabelis 1 on näha, et klassis S–V on ilukirjanduskeeles viis ja õppijakeeles kuus klastrit. Trigrammid, mis ei alga substantiiviga, moodustavad keskmise komponendi sõnaliigilise varieerumise põhjal uusi klastreid, mis jagunevad omakorda uute morfoloogiliste klasside vahel. Seega on klass ülemmõiste, mis ühendab morfoloogilised klastrid nende sageduse ja osakaalu alusel, ning klaster ülemmõiste, mis ühendab trigrammid nende morfosüntaktilise sarnasuse alusel.

<sup>1</sup> Eesti keele süntaktiliselt analüüsitud ja ühestatud ilukirjanduskorpus: [http://lepo.it.da.ut.ee/~heli\\_u/SA/](http://lepo.it.da.ut.ee/~heli_u/SA/) (15.8.2013).

<sup>2</sup> Eesti vahekeele korpus: [http://evkk.tlu.ee/Search/search\\_results.html](http://evkk.tlu.ee/Search/search_results.html) (15.8.2013).

<sup>3</sup> Sama programmi on eesti ilukirjanduskeele keelekasutusmuustrite uurimisel rakendanud J. Trainis ja K. Allkivi (2014, vt käesolev kogumik).

<sup>4</sup> Juhised eesti keele morfo- ja süntaksimärgendite kohta vt Tartu Ülikooli arvutilingvistika töörühma kodulehel: [math.ut.ee/~kaili/Korpus/pindmine/labels.pdf](http://math.ut.ee/~kaili/Korpus/pindmine/labels.pdf) (4.1.2013).

Artiklis võrreldakse ilukirjandus- (ILU) ja õppijakeele (Õ) morfoloogiliste klasside invariantsete klastrite invariantseid trigramme, kuna trigrammide tasandil avalduvad keekekasutuse leksikaalsed ja morfosüntaktilised piirangud, mis võimaldab siduda invariandi, sageduse ja osakaalu mõisted (invariant tuleb esile kõige sagedamini, selle osakaal valimis on kõige suurem, vt tabel 1).

**Tabel 1.** Morfoloogiliste klasside klastrivariatsioon, sagedus ja osakaal. S – substantiiv, V – verb, D – adverb, P – pronoomen, A – adjektiiv, J – konjunktsioon, G – genitiivatribuut, K – adpositsioon

Ilukirjanduskeel				Õppijakeel			
Klassid	Klastrid	Sagedus	Osakaal	Klassid	Klastrid	Sagedus	Osakaal
<b>D-V</b> <b>412</b>	DDV	206	50%	<b>J-V</b> <b>3590</b>	JPV	1558	44%
	DVV	116	28%		JSV	977	27%
	DJV	69	17%		JVV	582	16%
	DSV	11	3%		JDV	473	13%
	DAV	10	2%				
<b>J-V</b> <b>230</b>	JDV	114	50%	<b>P-V</b> <b>2803</b>	PVV	1881	67%
	JPV	67	29%		PSV	558	20%
	JSV	49	21%		PDV	177	6%
					PKV	67	3%
			PPV		64	2%	
<b>S-V</b> <b>150</b>	SDV	45	30%	<b>S-V</b> <b>2430</b>	SVV	1329	55%
	SVV	43	29%		SSV	412	17%
	SKV	29	19%		SKV	273	11%
	SSV	20	13%		SJV	204	8%
	SJV	13	9%		SPV	116	5%
<b>P-V</b> <b>53</b>	PDV	28	53%	<b>V-V</b> <b>2402</b>	SDV	96	4%
	PVV	15	28%		VVV	1025	42%
	PSV	10	19%		VDV	576	24%
					VJV	305	13%
			VPV		264	11%	
<b>A-V</b> <b>44</b>	ASV	44	100%	<b>D-V</b> <b>2169</b>	VSV	211	9%
					VAV	21	1%
					DVV	832	38%
					DPV	590	27%
					DDV	468	22%
					DSV	224	10%
<b>V-V</b> <b>15</b>	VDV	15	100%	<b>A-V</b> <b>347</b>	ASV	308	89%
					AVV	17	5%
					AJV	16	4%
					ADV	6	2%
				<b>K-V</b> <b>124</b>	KSV	50	40%
					KVV	46	37%
					KJV	18	15%
					KDV	5	4%
<b>G-V</b> 76	KPV	5	4%				
	GSV	76	100%				

Morfoloogiliste klasside ja klastrite võrdlus on siinses artiklis ülevaateline, tugineb klasside-klastrite osakaalu läheduse-erinemise-vastandlikkuse skaalale. Võrdlusaluseks on ilukirjanduskeel.

### 3. Analüüsi tulemused

#### 3.1. Ülevaade ilukirjandus- ja õppijakeele morfoloogilistest klassidest

Morfoloogiliste klasside võrdlus rajaneb nende osakaalule ILU ja Õ valimites (vt tabel 2). Emakeelekõneleja tekstiloomes piisab suhtlusvajadusteks kolmest morfoloogilisest klassist: tavapäraselt kasutatakse adverbialgulist (D–V); harvem konjunktsiooni- (J–V) ja substantiivalgulist (S–V) klassi. Pronoomeni- (P–V), adjektiivivi- (A–V) ja verbialguliste (V–V) klasside osakaal on väike, adpositsiooni- (K–V) ja genitiivtribuudialgulist (G–V) klasside osakaal sisuliselt olematu.

**Tabel 2.** Morfoloogiliste klasside osakaal ilukirjandus- (ILU) ja õppijakeele (Õ) valimites

Korpus	Morfoloogilised klassid							
	D–V	J–V	S–V	P–V	A–V	V–V	K–V	G–V
ILU	45%	25%	17%	6%	5%	2%	0%	0%
Õ	16%	26%	17%	20%	2%	17%	1%	1%

Õppija tekstiloomes on morfoloogiliste klasside varieerumine avaram, erinevate klasside osakaal ühtlasem. Eelistatakse konjunktsiooni- (J–V) ja pronoomeni- algulist (P–V) klassi; järgnevad substantiivi- (S–V), verbi- (V–V) ja adverbialguline (D–V). Seega on emakeelekõneleja ja õppija morfoloogilised valikud erinevad, millest annab täpsema ettekujutuse morfoloogiliste klasside klastrivarieeruvuse võrdlus.

#### 3.2. Ülevaade ilukirjandus- ja õppijakeele klastrivarieeruvusest

Klasside morfoloogilise varieerumise piirid tulevad esile klastrite osakaalu läheduse-erinemise-vastandlikkuse skaalal (täpsem kirjeldus Eslon 2013). ILU ja Õ võrdluse tulemused näitavad nii sarnasust kui ka olulisi erinevusi. Nt paigutub substantiivi- ja pronoomenialgulist klasside S–V ning P–V klastrikasutus läheduse-erinemise skaalale; konjunktsioonialgulist klassi J–V klastrikasutus on erinev; adjektiivialgulist klassi A–V puhul vastandlik; verbialgulist klassi V–V kasutus on täiesti vastandlik; adverbialgulist klassi D–V klastrikasutus aga oluliselt vastandlik. Tuleb rõhutada, et need erinevused ei tähista mittestandardset, vaid nihestunud keelekasutust, mis ei sega suhtlust, kuid on emakeelekõneleja seisukohast ebaratsionaalne.

ILU morfoloogiliste klastrite kasutust iseloomustab rangepiirilisus. Õ klastrivarieeruvus on avaram, morfoloogiliselt mitmekesisem, samas ei saa seda tõlgendada morfoloogilise keerukusena. Pigem on see märk õppija vähesest oskusest kasutada keelt sama ratsionaalselt kui emakeelekõneleja.

### 3.3. Invariantsete trigrammide morfosüntaktiline ja leksikaalne varieerumine

Invariantsete trigrammide võrdlus tugineb viie invariantse klatri vastandusele (v.a ASV): DDV vs. DVV, JDV vs. JPV, SDV vs. SVV, PDV vs. PVV, VDV vs. VVV. Nagu klasside ja klastrite ülevaates, nii on ka invariantsete trigrammide funktsionaalse potentsiaali kirjeldamisel aluseks ilukirjanduskeel (vt tabelid 3 ja 4). Neis tulevad selgelt esile emakeelekõnelejale ja õppijale omased valikud (ILU adverbialgulised või -kesksed morfoloogilised struktuurid; Õ verbi- ja pronoomenikesksed struktuurid), nt DDV vs. DVV (*väga tublisti kosunud vs. sageli ei vali*); JDV vs. JPV (*ja kindlasti oli vs. et see on*); SDV vs. SVV (*süda natuke kripeldas, vee peale valanud vs. riik ei hakka*); PDV vs. PVV (*ma lihtsalt ei vs. ma ei saa*); VDV vs. VVV (*oli seni olnud vs. ei saa elada*).

**Tabel 3.** Eesti ilukirjanduskeele invariantsete klastrid ja trigrammid

Klass	Invariantne klaster	Invariantne trigramm	Näited
D-V	DDV	_D_ // @ADVL _D_ // @ADVL _V_ main partic past ps #Intr // @-FMV	<i>kuskile ära sõitnud, väga tublisti kosunud</i>
J-V	JDV	_J_crd // **CLB @J _D_ // @ADVL _V_ main indic impf ps3 sg ps af #FinV #Intr // @+FMV	<i>ja kindlasti oli, ning siis kadus, kuid äkki tuli</i>
S-V	SDV	_S_com sg nom // @SUBJ _D_ // @ADVL _V_ main indic impf ps3 sg ps af #FinV #Intr // @+FMV	<i>süda natuke kripeldas, põlv ebameeldivalt tuikas</i>
		_S_com sg gen // @OBJ _D_ // @ADVL _V_ main partic past ps #NGP-P // @-FMV	<i>vee peale valanud, mehe üle võtnud</i>
P-V	PDV	_P_pers ps 1 sg nom // @SUBJ _D_ // @ADVL _V_ aux neg // @NEG	<i>ma lihtsalt ei, ma veel ei, ma ju ei</i>
A-V	ASV	_A_pos sg nom // @AN> _S_com sg nom // @SUBJ _V_ main indic impf ps3 sg ps af #FinV #Intr // @+FMV	<i>täielik hullumeelsus oli, vintske vanamees suri</i>
V-V	VDV	_V_ aux indic impf ps3 sg ps af #FinV #Intr // @+FCV _D_ // @ADVL _V_ main partic past ps #Intr // @-FMV	<i>oli kokku tulnud, oli seni olnud</i>

Viie klattrivastanduse ja ASV invariantsete trigrammide võrdlusest saab ettekujutuse, kuidas funktsionaalne potentsiaal erinevates keelekasutusvariantides avaneb, missuguseid morfosüntaktilisi ja kombinatoorseid piiranguid tekstikasutuses esineb ning sõnavara ja semantiliste kitsendustega seostub.

**Tabel 4.** Eesti õppijakeele invariantseid klastrid ja trigrammid

Klass	Invariantne klaster	Invariantne trigramm	Näited
J-V	JPV	_J_sub // **CLB @J _P_dem sg nom // @SUBJ _V_main indic pres ps3 sg ps af #FinV #Intr // @+FMV	<i>et see on, sest see toob</i>
P-V	PVV	_P_pers ps1 sg nom // @SUBJ _V_aux neg // @NEG _V_main indic pres ps neg #FinV #NGP-P // @+FMV	<i>ma ei saa, ma ei taha</i>
S-V	SVV	_S_com sg nom // @SUBJ _V_aux neg // @ NEG _V_main indic pres ps neg #FinV // @+FMV	<i>tunne ei ole, riik ei hakka</i>
V-V	VVV	_V_aux neg // @NEG _V_main indic pres ps neg #FinV #NGP-P // @+FMV _V_main inf // @-FMV	<i>ei saa elada (peita, unustada, teha, töötada)</i>
D-V	DVV	_D_ // @ADVL _V_aux neg // @NEG _V_main indic pres ps neg #FinV #NGP-P // @+FMV	<i>sageli ei vali, kohe ei saa</i>
A-V	ASV	_A_pos sg nom // @AN> _S_com sg nom // @SUBJ _V_main indic pres ps3 sg ps af #FinV #Intr // @+FMV	<i>täpne koht on, tänane põlvkond on</i>

Invariantsete trigrammide võrdlus toob esile emakeelekõneleja ja õppija keelekasutusele omaseid valikuid. Kirjelduse keskmes on morfosüntaktilise ja leksikaalse varieerumise piirid.

### 3.3.1. Adverbiaalgulised klastrid DDV vs. DVV

Adverbiaalgulise klatri kasutus on vastandlik: adverbikesksele muustrile (ILU) vastandub verbikeskne (Õ), keerulisele morfosüntaktilisele ja rikkalikule leksikaalsele varieerumisele (ILU) piiratud vormi- ja sõnakasutus (Õ).

- 1) Emakeelekõneleja eelistab jaatavat kõnet: kaks leksikaalselt varieeruvat adverbi + mineviku liitaja vormi kuuluv *nud*-partitsiip (*seal varengi olnud, väga tublisti kosunud*). Õppija seevastu eelistab eitavat kõnet: adverb + eituspartikkel + verbi finiitvorm (*sageli ei vali, kohe ei saa*);
- 2) ILU adverbide leksikaalne varieerumine on avaram, kui õppijakeeles. Samas on mõningad adverbid, mida eelistatakse korrata: ILU – trigrammi alguses *enam, nüüd, siis, mitte* ja keskel *põhjalikult, täpselt, õigeaegselt* (enamasti modaalpartikli funktsioonis); Õ – trigrammi alguses (*üldse, üldiselt, ka, kunagi, kindlasti, veel, isegi, lihtsalt*). Esile tuleb ka õppija tüüpviga *\*juba <enam>*, kui adverbide *juba* ja *enam* valikul pole mõistetud vaatepunktide erinevust;
- 3) ILU verbi leksikaalne varieerumine on rikkalik, eriti analüütiliste verbide puhul. Enamus neist on ainukordsed ühendverbid (nt *ette tulema (teatama), edasi minema (tormama), järele kuulama (hõikama), kinni mähkima*). Õppija on aga eelistanud piiratud hulka *sageli* korduvaid sünteetilisi verbe (*saama, ütlema, teadma, unustama, mõtlema, meeldima, ilmuma, esinema*);

- 4) ILU verbivormid on keeruka morfosüntaktilise varieerumisega: kõige enam kasutatakse *nud-* ja *tud-*partitsiipe (valdavalt perfekti koosseisus); *da-*infinitiiv on tavaline kas infiniitse öeldise või objekti funktsioonis; finiiit-verbivormivalik on piiratud ainsuse 3. pöördega imperfektis ja preesensis; *ma-*infinitiivi on kasutatud infiniitse öeldisena.

Õ verbi ajavormide valik on leksikaalselt piiratud: eksistentsiaalverb *olema* ja modaalverb *saama* eitava kõne indikatiivi preesensis (*ei ole, ei saa, ei ütle, ei tea*), harvem imperfektis (*ei olnud*). Tegu on tüüpilise trigrammiga (*üldiselt, ka, kunagi, kindlasti, veel*) *ei ole / ei saa ~ ei olnud*.

### 3.3.2. Konjunktsioonialgulisel klastrid JDV vs. JPV

Konjunktsioonialgulistel klastrite invariantsetel trigrammide võrdlus toob esile ühelt poolt vastanduse adverbikesksus (ILU) – pronoomenikesksus (Õ) ja teisalt trigrammide leksikaalse samalaadsuse (*olema*-verbi eelistamine).

- 1) ILU klaster JDV invariantsetes trigrammis on koordineerivat sidesõna kasutatud võrdselt nii lause kui ka osalause alguses, järgnevad adverb ja finiiitverb öeldise funktsioonis (*ja kindlasti oli, ning siis kadus, Kuid kahtlemata oli, Aga ilmselt jäi*). Õ klaster JPV invariantne trigramm on osalause alguses: subordineeriv sidesõna + demonstratiivpronoomen *see* subjekti funktsioonis + finiiitverbi öeldisena (*et see on, et see viib, sest see annab*). Seega eelistab emakeelekõneleja koordinatsiooniseost lause ja osalause alguses, õppija subordinatsiooniseost osalause alguses;
- 2) ILU invariantse trigrammi leksikaalne varieerumine on piiratud. Sagedaimad on koordineerivad sidesõnad *ja, ning, aga, ent, või, kuid* ja adverb *siis*, ent samas leidis rohkesti ka ainukordseid adverbe (nt *ilmselt, kahtlemata, kindlasti*). Õ invariantsetes trigrammis on kasutatud subordinatsiooni (valdavalt *et*). Verbidest on mõlemas keelekasutusvariandis sagedaim *olema*, ülejäänud väike hulk verbe on ainukordsed (nt ILU: *tuli, meenus, astus, ilmus*; Õ: *aitab, võib, jääb, annab, võimaldab, teeb*);
- 3) ILU sagedaim verbivorm on imperfekti ainsuse 3. pöörde, järgnevad preesensi ainsuse 3. pöörde ja imperfekti mitmuse 3. pöörde (*ja siis oli/on*). Vähesel määral esines ka eitava kõne näiteid (*aga järele ei, ent tookord ei*). Õ sagedaim verbivorm on preesensi ainsuse 3. pöörde, harva imperfekti mitmuse 3. pöörde (*et see on/oli*). Seega emakeelekõneleja eelistab *olema*-verbi imperfektis ja õppija preesensis;
- 4) ILU invariantne trigramm on morfosüntaktiline tervikobjekt, mida struktuurianaloogia põhjal on lihtne produtseerida, eriti eitavas kõnes, kus trigrammi kõik kolm komponenti on muutumatud sõnad (*ent tookord ei*). Tegu on pealtnäha lihtsa keelestruktuuriga, millel on oma osa tekstilise terviku loomisel. Ent kui arvestada, et eituspartikkel kuulub siin aktiivi/passiivi mineviku liitaja koosseisu, siis pole see morfosüntaktiline struktuur sugugi lihtne (vrd *ent tookord ei tulnud* mida teha?; *ent tookord ei tulnud* + adverb). Sellele vastupidiselt on Õ invariantse trigrammi näol tegu stereotüübiga *et see on (oli)* ja *et (ja) see ei*, millele järgnevas kontekstis on öeldistäide (tavaliselt lekseem *hea*).



### 3.3.3. Substantiivialgulised klastrid SDV vs. SVV

Substantiivialgulist klastrit on kasutatud erinevalt: ILU jätkuvalt adverbikeskne, Õ verbikeskne ja eitavat kõnet eelistav.

- 1) ILU klatri SDV invariantse trigrammi staatust jagavad kaks morfosüntakti- list struktuuri: subjektialguline (*süda natuke kripeldas, põlv ebameeldivalt tuikas*) ja objektialguline (*šokolaadi ära söönud, mehe üle võtnud*). Õ klatri SVV invariantne trigramm on eitavas kõnes (*tunne ei ole, riik ei hakka*);
- 2) ILU objektiga algava trigrammi leksikaalne varieerumine on avar: nii substantiive kui ka analüütilisi verbe on kasutatud ainukordsetena, v.a üksikjuhtumid (nt *suu(dki) lahti teha, ukse lahti teha, tervis tagasi tuli*). Seevastu Õ klatri SVV invariantstes trigrammis on tüüpsõnavara: substantiivid *inimene, tulemus, baar, mobiiltelefon, probleem, eestlane, keel, töö, raha* ja verbid *olema, saama, tohtima, sobima, unustama, jooma*. Ainukordsete ja kaks korda kasutatud lekseemide osakaal on kesine;
- 3) ILU invariantstes trigrammis varieerub substantiivi kääne ja süntaktiline funktsioon: totaalobjekt ainsuse genitiivis (*inimese ära tapnud, hobuse siit leidnud*), harva partsiaalobjekt ainsuse partitiivis (*meest ärkvele raputada, teekonnajuttu peale suruda*); subjekt ainsuse nominatiivis (*mõte ainult vilksatas, talumees muudkui muheles*). Totaalobjektile järgneva analüütilise verbi vorm varieerub morfosüntaktiliselt: mineviku liitaja koosseisu kuuluv *nud*-partitsiip, *da*-infinitiiv. Subjektiga algavates konstruktsioonides on finiitverb imperfekti ainsuse 3. pöördes, variandid on *ma*-infinitiiv öeldise funktsioonis (<*tundus*> *aeg möödab olema*, <*hakkas*> *sisemus isemoodi sügelema*) ja eitav kõne (*jutt enam ei <sobinud>*, *ema kergelt ei <andesta>*). Substantiiv ainsuse komitatiivis on käändsõnaline viisi- või kaasnevusmäärus (*kiiruga alla jooksnud, vestlusega kaasa läinud*).

ILU klatri SDV invariantsete trigrammide keeruline morfosüntaktiline varieerumine avab seoseid, mida eesti keele funktsionaalne potentsiaal võimaldab: substantiiv objekti, subjekti ja käändelise määruse funktsioonis, millest sõltub öeldisverbi valik. Ilmneb seaduspärasus: totaalobjekt + analüütilise verbi vorm; subjekt + sünteetilise verbi vorm. Keerukust lisab sõnajärg, kuna adverbi kasutatakse subjekti/objekti ja predikaadi vahel. Verbi ja substantiivi avar leksikaalne varieerumine ning idiomatiliselt laadi analüütiliste verbide kasutamine viitab sõnavara rikkusele. Struktuur SDV on eesti keeles morfosüntaktiliselt keerukam ja leksikaalselt varieeruvam kui DDV.

Õ klatri SVV invariantse trigrammi kasutuses ilmnevad piirangud substantiivi funktsionaalse potentsiaali kasutamisel: subjekt on tavaliselt ainsuse (*tunne ei ole, riik ei hakka*), harvem mitmuse nominatiivis (*inimesed ei saa, noored ei tea*), semantilised käänded väljendavad määruslikke tähendusi (ainsuse inessiiv – kohamäärus, ainsuse adessiiv – valdajamäärus; ajamäärus esines harvem). Verbi kasutatakse eitava kõne preesensis (*sõber ei joonud, kirik ei olnud*), mitmusliku subjektiga ka imperfektis. Trigrammi vähene morfosüntaktiline varieerumine ja tüüpsõnavara toovad esile stereotüübi *inimene ei ole/ei saa*. Seega ILU trigrammide morfosüntaktilisele keerukusele ja leksikaalsele rikkusele vastandub Õ stereotüüpne keelekasutusmuster.

### 3.3.4. Pronoomenialgulised klastrid PDV vs. PVV

Pronoomenialguline klaster näitab ILU adverbi- ja Õ verbikesksust. Ühisjooned on eitava kõne eelistamine, personaalpronoomen *ma* subjekti funktsioonis ja kitsa leksikaalse varieeruvusega morfosüntaktiline struktuur.

- 1) ILU klatri PDV invariantne trigramm on ainsuse 1. isiku personaalpronoomeni lühivorm subjekti funktsioonis + adverb + eituspartikkel (*ma lihtsalt ei, ma veel ei, ma ju ei*); Õ klatri PVV invariantne trigramm on ainsuse 1. isiku personaalpronoomeni lühivorm subjekti funktsioonis + eituspartikkel + liitpredikaati kuuluv modaalne abiverb (*ma ei saa, ma ei taha*);
- 2) pealtnäha on ILU invariantne trigramm lihtne, kuid eituspartiklile järgnev kontekst näitab, et tavaliselt on öeldiseks eitava kõne imperfekt või preesens (*ma lihtsalt ei <saanud, saa> + infinitiiv või ma veel ei <teadnud, tea>*). Trigrammi keskmise komponendi leksikaalne varieerumine on piiratud, lekseemide kasutussagedus suur. Järelikult peab trigramm esile tooma adverbid funktsionaalse potentsiaali. Neid kasutatakse kas rõhu-, modaalpartikli, viisi-, aja- või hulgamäärusena (nt ajamäärus *parajasti*, modaalpartikkel *tõepoolest*, rõhupartikkel *ju*). Trigrammi esimese ja viimase komponendi muutumatu kasutus + piiratud hulk sageli kasutatud adverbe moodustavad morfosüntaktilise tervikobjekti, mida on lihtne struktuurianalooogia põhjal produtseerida ja kasutada tagasiviite vahendina.

Õ klatri PVV invariantse trigrammi alguses on samuti personaalpronoomen *ma*, kuid oluline erinevus ILU trigrammist lisaks verbikesksusele seisneb selleski, et subjekti funktsioonis tuleb laialdaselt esile erinevaid pronoomeni liike, eriti sageli demonstratiivpronoomen *see*. Järgneb eitava kõne preesens + infinitiiv. Verbide leksikaalne varieerumine on kitsas (*saama, tahtma, olema, harvem pidama, nägema*).

- 3) ILU invariantse trigrammi morfosüntaktiline varieerumine on seotud pronoomeni ja eitava/jaatava kõneliigi valikuga, jaatavas kõnes ka verbivormi varieerumisega. See on hea näide pealtnäha lihtsa konstruktsiooni järkjärguliselt keerukamaks muutumisest. Nii võib eitavas kõnes subjektina esineda lisaks ainsuse 1. isikule ka ainsuse 3. isik (*tema küll ei, ta siis ei*); jaatavas kõnes kasutatakse subjektina enamasti ainsuse 3. isikut ja finiitverbi imperfekti ainsuse 3. pöördes (*ta äkitselt kadus*). Veelgi keerulisem varieerumise juhtum on indefiniitne pronoomen *miski* partsiaalobjektina koos analüütilise verbi *da*-infinitiiviga, kui trigrammile eelnevas kontekstis seisab modaalne abiverb (<*tuleb*> *midagi ette võtta*, <*vaja*> *midagi välja pressida*).

Õ klatri PVV invariantse trigrammi morfosüntaktiline varieerumine on dikteeritud kõneliigist: a) eitava kõne vormivalik toob esile seose modaalverbide tähenduse ja finiitverbi kõneviisi-ajavormi valiku vahel: *saama, teadma* + indikatiivi preesens (*ma ei saa, ma ei tea*); *saama* + konditsionaali preesens (*ma ei saaks*) ja indikatiivi imperfekt (*ma ei saanud*); b) jaatavas kõnes domineerib konstruktsioon *tahtma* + *da*-infinitiiv totaalobjekti funktsioonis (*ma tahan õppida*), harva esines *ma*-infinitiiviga liitpredikaat (*ma pean ostma*). Esile on tulnud ka üks jaatava kõne veamuster, mis tingitud finiitverbi semantikast tuleneva valikuga *da*- ja *ma*-infinitiivi

vahel (nt *ma pean \*parandada* (\**õppida*, \**minna*, \**lugeda*, \**vaadata*), verbide häälduspärasest kirjaviisist (\**suitsida*, \**soitsida* <*suitsetada*>, \**õpida* <*õppida*>) või isikupärasest sõnakasutusest (\**praktida* <*praktiseerida*>).

### 3.3.5. Adjektiivialguline klaster ASV

Adjektiivialgulise klasteri invariantseid trigramme on kasutatud vastandlikult, kuid nii ILU kui ka Õ invariantne trigramm on substantiivikeskne.

- 1) ILU klaster ASV koosneb vaid ühest trigrammist, mida on kasutatud nii lause sees kui ka alguses (*vintske vanamees suri*, *Liikumatu muie püsis*); Õ klaster sisaldab rohkem trigramme, kuid invariantne struktuur on sama (*täpne koht on*, *suur rühm on*, *suur palk sõltub*). Erinevused ilmnevad trigrammi komponentide leksikaalses ja morfosüntaktilises varieerumises;
- 2) kõik ILU adjektiivid ja substantiivid on ainukordsed, kuid verbikasutuses tuleb esile selge dominant, kuna pool näidetest sisaldab *olema*-verbi. Sama tendents ilmneb ka õppijakeeles, kus *olema*-verb esineb kahes kolmandikus näidetest. Seega ILU tüüpilisele leksikaalsele piirangule vastab Õ leksikaalne dominant;
- 3) ILU trigrammis järgneb adjektiivsele eestäiendile tavaliselt substantiiv ainsuse või mitmuse nominatiivis (subjekt) ja *olema*-verb imperfekti ainsuse või mitmuse 3. pöördes (*täielik hullumeelsus oli*, *Kuldpruunid palkseivad olid*). Variant: substantiiv ainsuse partitiivis (partsiaalobjekt) + *da*-infinitiiv öeldisena (*verbaalset vastet leida*, *õiget naist saada*). Õ trigrammis on substantiiv enamasti ainsuse nominatiivis ja finiiitverb preesensi ainsuse 3. pöördes (*täpne koht on*, *suur palk sõltub*). Esineb ka mitmeid teisi arvu- ja ajavormide kombinatsioone, kuid need on harvad. Kuigi emakeelekõneleja ja õppija kasutavad sama morfoloogilist struktuuri ja eelistavad *olema*-verbi, erinevad trigrammide leksikaalse varieerumise piirid ja morfosüntaktiline lihtsus/keerukus. ILU tekstikasutuses on tüüpiline *olema* imperfekti ainsuse 3. pöördes, Õ tekstikasutuse stereotüüp *olema* preesensi ainsuse 3. pöördes. Eakeelekõneleja jaoks on tavaline ka substantiiv partsiaalobjektina + infiniitne öeldis, mis õppija keelekasutuses puudub.

### 3.3.6. Verbialgulised klastrid VDV vs. VVV

Verbialguliste klastrite kasutamine on täiesti vastandlik: ILU adverbi- ja Õ verbi-keskne, emakeelekõneleja eelistab jaatavat ja õppija eitavat kõnet.

- 1) ILU klasteri VDV trigramm koosneb verbi pluskvamperfekti liitaja vormist, mille “lõhub” adverb (*oli kokku tulnud*, *oli seni olnud*). Tegu on eesti keelele omase sõnajärgiga, mis võib tekitada probleeme nii keeleõppijale kui ka emakeelekõnelejale. Õ klasteri VVV invariantne trigramm langeb kokku eitava kõne liitpredikaadiga, kus modaalsele abiverbile preesensis (harvem imperfektis, konditsionaali preesensis, väga harva impersonaalis) järgneb *da*-infinitiiv infiniitse öeldise funktsioonis, harvem totaalobjektina (*ei saa elada*, *ei tohi suitsetada*);

- 2) ILU trigrammis on kõiki adverbe ja verbe kasutatud ainukordselt. Õ trigrammis vastandub sellele tüüpsõnavara: preesensis on korratud seitset verbi (*saama, tohtima, harvem tahtma, oskama*, paaril korral *viitsima, suutma, armastama*), viimane komponent on *da*-infinitiiv, mis on küll avarama leksikaalse varieerumisega, kuid ka siin tuleb esile selge erinevus sageli (*teha, elada, öelda/ütelda, suitsetada*) ja vaid paar korda esinenud verbide vahel (*loobuda, istuda, olla*). Niisiis on invariantsete trigrammi-de leksikaalne varieerumine ILU ja Õ tekstides täiesti erinev: leksikaalne rikkus *vs.* tüüpsõnavara.

## 4. Arutelu

Eesti ilukirjandus- ja õppijakeele võrdleva analüüsi põhjal selgus, et emakeelekõneleja eelistab piiratud hulka sagedasi morfoloogilisi klastreid ning morfosüntaktilisi trigramme. Nende seas tulevad esile adverbialgulised ja -kesksed mustrid, kus adverbil on erinevaid tekstilisi funktsioone, mis viitab adverbi avarale funktsionaalsele potentsiaalile. Määruse funktsiooni taandumine tekstiliste funktsioonide ees on keeleomane nähtus, mida on seotud diakroonilise arengu üldisemate seaduspärasustega (vt Valdmets, Habicht 2013, Veismann 2009). Seetõttu on ka siinse arutelu keskmes kõik adverbiga seonduv.

Adverbi sisaldavad mustrid on üsna keerulised. Nii eesti kirjakeele<sup>5</sup> kui ka siinses artiklis vaadeldud verbist vasakule jääva konteksti klasteranalüüs toob esile adverbiliste struktuuride tekstilised funktsioonid ja nende seose sõnajärgjega: lause alguses tagasiviite ja eelneva lause sisu konkretiseerimise või hinnangu vahend; lause sees modaali- või rõhupartikkel. Sõltuvalt funktsioonist võib adverb oma kohta lauses vabalt varieerida, paigutudes subjekti/objekti ja liitpredikaadi või verbi liitajavormi vahele. Seetõttu püüab õppija neid struktuure vältida, kasutades adverbi enamasti aja- ja kohamäärusena.

Konjunktsioonialguline adverbikeskne trigramm esineb emakeelekõneleja tekstides regulaarselt lause, harvem osalause alguses, õppijal aga osalause alguses. ILU trigrammis kasutatakse kuut koordineerivat sidesõna (sagedaim *ja*); adverbide valik on leksikaalselt piiratud (sagedaim *siis*), verbidest on sagedaim *olema* imperfekti ainsuse 3. pöördes (*ja siis oli*). Seda struktuuri on lihtne produtseerida, verbi saab leksikaalselt vabalt varieerida, ent emakeelekõneleja pole seda võimalust eriti kasutanud. Ka õppijal on konjunktsioonialguline struktuur sage, kuid koordinatsiooniseost välditakse, adverbi asendab verb, *olema* on preesensi ainsuse 3. pöördes (*et see on*). Erinevused on tingitud sellest, mis otstarbel kirjeldatud trigramme kasutatakse. Emakeelekõnelejal on see tekstilise terviku loomise vahend, kus järgnev lause täiendab informatiivselt eelneva sisu, õppijal aga vahend kõrvallause sidumiseks pealausega (põhjus-tagajärg seos). See seletab kummagi keelekasutusvariandi eelistusi (koordinatsioon *vs.* subordinatsioon ja tagasiviitevahendina sünonüümne asendus demonstratiivpronoomeniga *see*). Siin peitubki nihke olemus emakeelekõnelejaga võrreldes. Ning kui me tahame, et õppija oskaks kasutada adverbistruktuure sama oskuslikult nagu emakeelekõneleja ning neid mitte vältida, siis tuleb õpetada, kuidas kõige ratsionaalsemal viisil sidusaid tekste

produtseerida, subjektiivseid hinnanguid ja muid informatsiooni aktualiseerimise vahendeid kasutada.

Siinkohal paar üldisemat märkust. Esiteks on adverbilised mustrid hea näide grammatika, semantika ja pragmaatika sidususest tekstis. Selle alusel kujunevad keeleüksuste lineaarse kooskasutuse piirangud ehk reeglid, samas kui eesti keele õpe tugineb jätkuvalt grammatika reeglitele. Sagedaste lingvistiliste tervikobjektide asemel õpetatakse vormidest ja sõnadest vaba valiku alusel moodustatud grammatilisi konstruktsioone. Uurimused on näidanud, et olenemata keelest moodustavad kahest-kolmest elemendist koosnevad sagedased terviküksused, nt vormelid (ingl *formulaic sequences*) keskmiselt pool teksti mahust, mistõttu piiratud sõna- ja vormivalikuga tekstilistel struktuuridel on kindel koht keele omandamisprotsessis. Need jäävad hästi meelde, neid loetakse ja produtseeritakse kiiremini. (Conklin, Schmitt 2008: 72–77) Kuid eesti keele õpikutes leidub niisuguseid struktuure vähe (Puhtvend 2009). Ei osata arvestada, et idiomaatilise printsiipi domineerib keeleüksuste vaba varieerumise üle, et idiomaatilist tähendust võivad omada “ka produktiivselt moodustatud väljendid” (Sahkai 2008: 171–172). Keeles on “süntaktilisi püsimalle, [---] [mida] on võimalik täita kümnete, sadade või tuhandete erinevate figuratiivsete vm leksikaalsete elementidega” (Krikmann 2004: 116). See, mis on invariant grammatikas ning mille alusel reegleid tuletatakse, ei ole tekstikasutuse invariant.

Teine üldisemat laadi märkus tuleneb eelnevast ja seostub keelekasutuse ratsionaalsusega. Siinse klasteranalüüsi tulemused näitavad, et emakeelekõneleja eelistab lihtsa struktuuriga morfosüntaktilisi terviküksusi, milles saab lekseeme vabalt varieerida ning mida produtseeritakse struktuurianaloogia alusel. Just niisugune on adverbialguline invariantne klaster DDV, mille alusel saab genereerida ühendverbe. Järelikult on struktuuri DDV kasutusele mõttekas rajada ka eesti ühendverbide produtseerimis- ja kasutamissooetus. Nähtus on universaalsemat laadi, nt itaalia keeles käsitletakse iseseisvat tähendust omavate mitmesõnaliste predikatiivsete üksuste arengut semantika-süntaksi koostoimena (Masini 2005). Sama kehtib saksa keele partikkelverbide, vene ja inglise perifrastiliste ja fraasiverbide kohta. Analoogselt on võimalik kasutada ka substantiivseid trigramme, et genereerida väljendverbe. Mõistetavalt polnud verbist vasakule jäävas kontekstis nende kasutus sage, kuid verbist paremale jäävas kontekstis, prototüüpelt Ø-markeeritusega objektifraasis, ilmnevad need trigrammid regulaarselt (Eslon 2012: 23–26). Tuumverbid *tegema*, *andma*, *saama* ja *võtma* moodustavad koos elutu partitiivse objektiga nii idiomaatilisi terviküksuseid (*nalja tegema*, *nõu andma*, *asja saama*, *sõna võtma*) kui ka vabu ühendeid (*leiba tegema*, *eratunde andma*, *tööd saama*, *ravimeid võtma*), mida õppija on genereerinud normikohaselt. Sageli kasutatuna võivad ka vaba valiku alusel moodustatud ühenditest aja jooksul kujuneda idiomaatilised terviküksused, mis täiendavad eesti väljendverbide loendit. Ka see nähtus pole eesti keele keskne. Nt hispaania keeles genereeritakse *saama*-täheendusega verbe kooskasutuses adjektiivi ja elusa subjektiga. Seda struktuuri rakendatakse võrdväärselt nii muutust tähistavate idiomaatiliste kui ka mitteidiomaatiliste terviküksuste produtseerimiseks (Bybee, Eddington 2006). Kirjeldatud morfosüntaktiliste struktuuride funktsionaalne potentsiaal avaneb tänu struktuurianaloogiale, mis on kõige loomulikum ja ratsionaalsem viis analüütiliste verbi(konstruktsiooni)de produtseerimiseks ning omandamiseks. Keele valdamiseks oluline teave ei pärine grammatikast, vaid piirangutest keeleüksuste lineaarses kooskasutuses.

Kuna teadmised eesti keele kasutusmustrite kohta ei ole piisavad, pole ka õpetamis- ja õppimisprotsess kuigi optimaalne, vaid pigem intuitiivne. See tingibki leksikaalsete ja morfosüntaktiliste stereotüüpide teket ning mittevajalikku variatiivsust. Nt õppijakeeles on ebaratsionaalselt avar pronoomenite varieerumine. Konjunktioonialgulise klatri JPV invariantne trigramm sisaldab stereotüüpselt demonstratiivpronoomenit *see*, lisaks veel 43 sagedat trigrammi personaalpronoomeniga. Valdab *mina*-kesksus, õppija tekstikasutuses vastanduvad *ma-nad*, *ta-nad*. Eesti ilukirjanduskeeles ei ole *see* esile tulnud. Pronoomenialguline morfoloogiline klass on küll olemas, kuid selle osakaal jääb adverbialgulisele klassile tunduvalt alla (vt tabel 2). Samas eristub *mina*-kesksus nt vene- ja ingliskeelses tekstiloomes. Kuna eesti vahekeele korpuse loomingu laadi esseede autorid on valdavalt vene ja vähesel määral inglise emakeelega eesti keele õppijad, siis saab pronoomenikesksust seletada keelemeelelisusega. Struktuuris JPV piirab pronoomen ka verbi leksikaalset ning morfosüntaktilist varieerumist: eitava kõne korral kasutatakse modaalverbide *saama* ja *tahtma* ning jaatavas kõnes *tahtma* finiitvormi. Koos ainsuse 1. isiku personaalpronoomeniga moodustub stereotüüp *ma ei saa*. Seega on õppija keelekasutuses pronoomenikesksusel otsene seos modaalsete hinnangute leksikaalse väljendamisega. Isegi ajakirjanduskeeles, kus hinnangulisus on informeerimise järel teine oluline pragmaatiline funktsioon, ei tule modaalkonstruksioonid niisuguse regulaarsusega esile kui õppijakeeles (Eslon 2009: 46). Samas on emakeelekõneleja hinnangute edastamisel stereotüüpsele trigrammile eelistanud leksikaalset ja morfosüntaktilist varieeruvust. Õppija lihtsalt ei valda seda keelelist mitmekesisust, olgugi et mõningate variatsioonide ja eripäraga on see olemas ka tema emakeeles.

Eesti keeles kõige keerulisema morfosüntaktilise varieerumisega klaster on SDV ja selle kaks invariantset trigrammi. Substantiivi on kasutatud nii objekti, subjekti kui ka käändelise määrusena, substantiivi funktsioonist oleneb verbi valik: totaalobjekt + analüütilise verbi vorm; subjekt + sünteetilise verbi vorm. Tegu on ratsionaalse kasutusreegliga, mida ei teata – osaliselt seetõttu kasutab õppija stereotüüpset eitava kõne konstruktsiooni *inimene ei ole/ei saa*.

Eespool kirjeldatu on otseselt seotud vormikasutuse keerukuse/ratsionaalsuse dialektikaga. Morfoloogiliselt ja morfosüntaktiliselt varieerub ilukirjanduskeel vähem kui õppijakeel, eristudes viimasest vormikeerukuse ja sõnajärjeraskuse poolest. Emakeelekõneleja sõnastus on kahtlemata varieeruvam ja keerulisem, kuid tehtud valikute poolest ratsionaalsem. Dominantsete trigrammide kasutus klattrites ASV ja VDV toob esile korrelatsiooni leksikaalse ja morfosüntaktilise varieerumise vahel: maksimaalse leksikaalse varieerumisega kaasneb vähene morfosüntaktiline varieeruvus. Õppijakeeles pole korrelatsioon esile tulnud, vormide varieerumine võib küll olla avaram, kuid sõnavara ikkagi piiratud. Nihkeid ja nende tekkepõhjust on erinevaid, kuid tekstide produtseerimisoskuse kujundamisel on oluline tugineda tekstikasutuses sagedate mustrite leksikaalgrammatilistele piirangutele, mitte grammatikale ja sõnastikule.

## Viidatud kirjandus

- Barlow, Michael; Kemmer, Suzanne (Eds.) 2000. Usage-Based Models of Language. Stanford: CSLI.
- Biber, Douglas; Conrad, Susan; Reppen, Randi 1998. Corpus Linguistics: Investigating Language Structure and Use. Cambridge University Press. <http://dx.doi.org/10.1017/CBO9780511804489>
- Bondarko, Aleksandr 2002. Teorija znachenija v sisteme funkcional'noj grammatiki: Na materiale russkogo jazyka. [The Meaning Theory in System of the Functional Grammar: On the Russian Language.] Moskva: Jazyki slav'anskoj kul'tury, 159–356.
- Bybee, Joan; Hopper, Paul (Eds.) 2001. Frequency and the Emergence of Linguistic Structure. Amsterdam: John Benjamins. <http://dx.doi.org/10.1075/tsl.45>
- Bybee, Joan; Eddington, David 2006. A usage-based approach to Spanish verbs of 'becoming'. – Language, 82 (2), 323–355. <http://dx.doi.org/10.1353/lan.2006.0081>
- Bybee, Joan 2007. Frequency of Use and the Organization of Language. Oxford: Oxford University Press. <http://dx.doi.org/10.1093/acprof:oso/9780195301571.001.0001>
- Conklin, Kathy; Schmitt, Norbert 2008. Formulaic sequences: Are they processed more quickly than nonformulaic language by native and nonnative speakers? – Applied Linguistics, 29 (1), 72–89. <http://dx.doi.org/10.1093/applin/amm022>
- Croft, William 1999. Adaptation, optimality and diachrony. – Zeitschrift für Sprachwissenschaft, 18 (2), 206–208. <http://dx.doi.org/10.1515/zfsw.1999.18.2.206>
- Croft, William 2003. Typology and Universals. Second edition. Cambridge: Cambridge University Press.
- Ellis, Nick C. 2006. Language acquisition as rational contingency learning. – Applied Linguistics, 27 (1), 1–24. <http://dx.doi.org/10.1093/applin/ami038>
- Eslon, Pille 2009. Eestikeelses tekstiloomes eelistatud konstruktsioonid ja käändevormid. [Contextual preferences the use of case forms in text production in Estonian.] – Korpus-uuringute metodoloogia ja märgendamise probleemid. Tallinna Ülikooli eesti keele ja kultuuri instituudi toimetised 11. Tallinn: TLÜ Kirjastus, 30–53.
- Eslon, Pille 2012. Objekti ja tegevuse leksikaalne markeeritus eesti õppijakeeles. [The lexical markedness of object and action in Estonian learner language.] – Lähivõrdlusi. Lähivertailuja, 22, 15–42. <http://dx.doi.org/10.5128/LV22.01>
- 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>
- Goldberg, Adele E. 2006. Constructions at Work. The Nature of Generalization in Language. Oxford: Oxford University Press.
- Greenberg, Joseph 1990. The relation of frequency to semantic feature in a case language (Russian). – Keith Denning, Suzanne Kemmer (Eds.). On Language. Selected writings of Joseph H. Greenberg. Stanford: Stanford University Press, 207–226.
- Haspelmath, Martin 1999. Optimality and diachronic adaptation. – Zeitschrift für Sprachwissenschaft, 18 (2), 180–205. <http://dx.doi.org/10.1515/zfsw.1999.18.2.180>
- Haspelmath, Martin 2007. Creating economical morphosyntactic patterns in language change. – Jeff Good (Ed.). Language Universals and Language Change. Oxford: Oxford University Press. <http://email.eva.mpg.de/~haspelmt/CreatingEconomy.pdf> (3.2.2013).
- Klavan, Jane 2012. Evidence in Linguistics: Corpus-Linguistic and Experimental Methods for Studying Grammatical Synonymy. Dissertationes Linguisticae Universitatis Tartuensis 15. Tartu: University of Tartu.
- Krikmann, Arvo 2004. “Sai hea obaduse vastu obadust”: löömist ja peksmist märkivad väljendid eesti keeles. [Multi-word expressions denoting hitting in Estonian]. Reetor 3. Tartu: Eesti Kirjandusmuuseum Folkloristika osakond, Eesti Kultuuriloo ja Folkloristika Keskus.

- Masini, Francesca 2005. Multi-word expressions between syntax and the lexicon: the case of Italian verb-particle constructions. – *SKY Journal of Linguistics*, 18, 145–173.
- Ots, Sander 2012. Statistikapõhise tarkvara loomine morfoloogiliste kollokatsioonide eraldamiseks eesti keele tekstidest. [Software for morphosyntactic cluster extraction from Estonian texts.] Bakalaureusetöö. Tallinn: TLÜ informaatikainstituut.
- Penjam, Pille 2008. Eesti kirjakeele *da-* ja *ma-*infinitiiviga konstruktsioonid. [The constructions of *da-* and *ma-*infinitives in written Estonian.] *Dissertationes Philologiae Estonicae Universitatis Tartuensis* 23. Tartu: Tartu Ülikooli Kirjastus.
- Puhtvend, Kädi 2009. I–IV klassi laste kujundlik keeekasutus. Õpetamine ja oskused. [The use of figurative language in elementary school (1st–4th class): Teaching and skills.] Bakalaureusetöö. Käsikiri Tallinna Ülikooli eesti keele ja kultuuri instituudis.
- Sahkai, Heete 2008. Konstruktsioonipõhine keelemudel ja sõnaraamatumudel. [Some lexicographic implications of a construction-based model of language.] – *Eesti Rakenduslingvistika Ühingu aastaraamat*, 4, 177–186. <http://dx.doi.org/10.5128/ERYa4.11>
- Trainis, Jekaterina; Allkivi, Kais 2014. Ilukirjanduskeelest uue pilguga. [On belletristic language from a new perspective.] – *Eesti Rakenduslingvistika Ühingu aastaraamat*, 10, 283–306. <http://dx.doi.org/10.5128/ERYa10.18>
- Uiboaed, Kristel 2013. Kollostruktsioonilised meetodid ja konstruktsioonilise varieerumise tuvastamine. [Collostructional methods and verb constructions in Estonian dialects.] – *Eesti ja soomeugri keeleteaduse ajakiri / The Journal of Estonian and Finno-Ugric Linguistics (ESUKA–JEFUL)*, 4 (1), 185–204.
- Valdmets, Annika 2011. Kahe eesti kirjakeele modaalpartikli arengust viimase sadakonna aasta jooksul. [On the development of two modal particles in written Estonian over the last century.] – *Keel ja Kirjandus*, 10, 764–776.
- Valdmets, Annika; Habicht, Külli 2013. Episteemilistest modaalpartiklitest eesti kirjakeeles. [About epistemic modal particles in Estonian.] – *Eesti ja soomeugri keeleteaduse ajakiri / The Journal of Estonian and Finno-Ugric Linguistics (ESUKA–JEFUL)*, 4 (1), 205–222.
- Veismann, Ann 200. Eesti keele kaas- ja määrsõnade semantika võimalusi. [Semantics of Estonian adpositions and adverbs.] *Dissertationes Linguisticae Universitatis Tartuensis* 11. Tartu: Tartu Ülikooli Kirjastus.
- Verhagen, Arie 2009. The conception of constructions as complex signs. Emergence of structure and reduction to usage. – *Constructions and Frames*, 1 (1), 119–152. <http://dx.doi.org/10.1075/cf.1.1.06ver>

**Pille Esloni** (Tallinna Ülikool) peamine uurimisvaldkond on morfosüntaks, eesti kirjakeele ja õppijakeele morfosüntaktiliste kasutusmustrite võrdlus.  
Narva mnt 25, 10120 Tallinn, Estonia  
[peslon@tlu.ee](mailto:peslon@tlu.ee)



# CONSTRAINTS ON MORPHOSYNTACTIC AND LEXICAL VARIABILITY

**Pille Eslon**

Tallinn University

This study uses cluster analysis in order to describe the functional mobility of a language system. It compares Estonian fiction and learner language texts with regard to the use of language preceding the verb (e.g. *ema lähemale tuli* 'mother nearer came', *kõhn keha vabises* 'scrawny body trembled', *elu ei ole* 'life not is', *ja kindlasti oli* 'and certainly was', *sest see aitab* 'because it (that, this) helps'). The aim of this study is to show the constraints on morphosyntactic and lexical variability of the invariant n-grams. It is presumed that the functional potential of a language system is used more rationally by native speakers than by learners. Results of the analysis indicate that native speakers make optimal choices, which entail morphosyntactic simplicity but at the same time complexity and richness of vocabulary. Learners' language usage shows on the one hand a broader range of structural variability than in the language of fiction, but on the other hand, it reveals oversimplification based on stereotypical language structures, overuse and underuse of morphological classes and clusters and poor vocabulary.

**Keywords:** usage-based language description, parsing, cluster analysis, syntax-morphology interaction, constraints and variability on language use, learner's language, Estonian