

Kahe keelekasutusvariandi võrdlus: morfoloogilised klassid ja klastrid

PILLE ESLON

Tallinna Ülikool

Ülevaade. Kasutuspõhise keelekirjelduse väärtus grammatika jaoks seisneb võimaluses avada keelesüsteemi funktsionaalne potentsiaal, kirjeldades näiteks keelestruktuuride hierarhiat, võrreldes keelekasutusvariantide morfosüntaktilisi mustreid, keelesüsteemist ja keelekasutaja valikutest tulenevaid semantilisi ja grammatilisi kitsendusi, keelekasutuse leksikaalse ja grammatilise varieeruvuse piire. Nende esiletoomine vajab meetodeid, mis annaksid võimalikult usaldusväärseid tulemusi. Siinses uurimuses on kasutatud klasteranalüüsi. Võrreldud on eesti ilukirjandus- ja õppijakeele tekstide verbist vasakule jääva konteksti keelekasutust. Leitud on kõik kolmest üksusest koosnevad konstruktsioonid, rühmitatud need morfosüntaktilise samalaadsuse ja sageduse alusel ning kirjeldatud morfoloogiliste klastrite ja klasside hierarhiana. Klassid näitavad keelekasutuse üldiseid morfoloogilisi mustreid, klastrite alusel saab kirjeldada klasside morfoloogilise varieeruvuse piire. Konkreetset konstruktsioonid toovad esile piiranguid klastrite leksikaalgrammatilisele varieeruvusele. Eesti ilukirjandustekstide ja keeleõppija loomingulist laadi esseede võrdlusest peaks selguma, kuidas kasutavad keelesüsteemi funktsionaalset potentsiaali emakeelekõneleja ja õppija. Analüüsi tulemused näitavad, et emakeelekõnelejad teevad optimaalseid valikuid, mis hõlmavad nii morfosüntaktilist lihtsust kui ka keerukust, nii sõnavara rikkust kui ka kinnistunud korduvkasutust. Selles osas on õppija keelekasutus nihkes, emakeelekõneleja keeletunnetuse põhjal kuidagi ebaharilik. Samas ei tähenda see, et tegu oleks

vigadega. Ühelt poolt näitavad õppija kasutatud keelestruktuurid ilukirjanduskeelest avaramat varieeruvust. Teisalt tuleb selgelt esile stereotüüpsetele keelestruktuuridele tuginev lihtsustamine, morfoloogiliste klasside ja klastrite üle- ning alakasutus, sõnavara kesisus.

Võtmesõnad: klasteranalüüs, grammatilise varieeruvuse piirid, eesti ilukirjandus- ja õppijakeel, kasutuspõhine keelekäsitlus

1. Eesmärk, keeleaines, meetod

Uurimuse üldisem eesmärk on võrrelda eesti ilukirjandus- ja õppijakeele morfosüntaktilist ning leksikaalset varieerumist, välja tuua erinevatele kasutusvariantidele iseloomulikke grammatilisi ja leksikaalseid piiranguid. Artiklis võrreldakse ja kirjeldatakse vaid morfoloogilisi klasse ning klastreid. Selle aluseks on keeleüksuste sageduse, osakaalu ja varieeruvuse mõisted.

Keeleaines pärineb kahest allikast: Tartu Ülikooli eesti kirjakeele süntaktiliselt analüüsitud ja ühestatud ilukirjanduskorpuse osast (maht ~ 200 000 sõnet, kasutatud võrdluskorpuseks) ning Tallinna Ülikooli eesti vahekeele korpusest (~ 500 000 sõnet). Kuigi võrreldakse eri suurusega valimeid, pole vormide esinemissagedus teadaolevalt sünkroonne sõnede arvuga valimis (vt Biber jt 1998: 33). Valimid on võrreldavad vormide osakaalu alusel.

Rakendatud on korpusanalüüsi (vt Eslon jt 2010: 20–22), eesti keele tarkvara ja Sander Otsa programmeeritud klastrileidjat, mille sisendiks on eesti keele parseri väljund. Klastrileidja töötab andmekaeve põhimõttel: otsib tekstist sarnaseid morfo- ning süntaksimärgendite lineaarseid järjendeid ja fikseerib nende kasutussageduse. Klastrileidja väljundit saab eksportida Excelis. Programmi loomisel on eeskujuks olnud statistikapõhine tarkvara nagu WordSmith Tools, kfNgram ja Xaira. (Vt Ots 2012: 6, 16.)

Siinse uurimuse tarvis on ilukirjandus- ja õppijakeele valimist leitud kõik verbist vasakule jäävad konstruktsioonid, mis koosnevad kolmest

komponendist, kusjuures viimane on alati verb (nt ILU: *metsa vari ulatus, veel ei liikunud*; Ö: *ise pidid tegema, kord mul oli* jne). Analüüsitud on konstruktsioone, mis esinesid tekstikasutuses viis ja enam korda. Ilukirjandustekstidest on leitud kokku 904 verbist vasakule jäävat kolmest komponendist koosnevat konstruktsiooni, mis jagunevad 18 klatri ja 6 morfoloogilise klassi vahel. Mahult suurema õppijakeele esseedevalimi alusel tuli esile 13941 kolmest komponendist koosnevat konstruktsiooni, mis jagunesid 38 klatri ja 8 morfoloogilise klassi vahel. Siinne ilukirjandus- ja õppijakeele analüüs piirdub vastavalt 6 morfoloogilise klassi 18 klatri ning 8 morfoloogilise klassi 38 klatri võrdlemisega (vt tabel 2). Konstruktsioonitasandi varieeruvust, selle morfosüntaktilisi ja leksikaalseid piiranguid ilukirjandus- ja õppijakeeles artiklis ei kirjeldata, nende kohta tuuakse vaid mõned illustreerivad näited.

2. Mõistete seletusi

Klatri all mõistetakse uurimuses sarnaste objektide kogumit, mis kujutab endast sageduse alusel hulkadest leitud olulist informatsiooni. Siinses analüüsis on see leitud morfo- ja süntaksimärgendite lineaarse koosinemise ja varieerumise põhjal tekstis. Iga klaster erineb teiste klatri objektidest. Klasterdamine on objektide rühmitamine nimetatud ühistunnuste ilmnemise regulaarsuse alusel, mis toob esile kolme liiki klastreid: morfoloogilisi, süntaktilisi ja morfosüntaktilisi.

Morfoloogiline klaster kirjeldab keelekasutusele omast sõnaliigi-morfoloogiat. Näiteks: verb (V) + adverb (D) + adverb (D) ehk VDD (*on veel vara*); adverb + adverb + substantiiv ehk DDS (*väga palju emotsioone*); pronoomen + verb + adverb ehk PVD (*see on praegu*) jne¹. Sageduse alusel tulevad esile keelekasutuse sõnaliigilised tunnusjooned.

Süntaktiline klaster kirjeldab lauseliikmete järjendeid, nt @ NEG @ +FMV @ ADVL tähistab järjendit verbi eitus (üldeitus) + finiitverb

¹ Juhised eesti keele morfo- ja süntaksimärgendite kohta on kättesaadavad Tartu Ülikooli arvutilingvistika töörühma kodulehel, vt math.ut.ee/~kaili/Korpus/pindmine/labels.pdf (4.1.2013).

öeldise funktsioonis + adverbiaal (*ei tule enam*). Süntaktilised klastrid toovad esile keelekasutusele omased sõnajärjemallid.

Morfosüntaktiline klaster kirjeldab morfo- ja süntaksimärgendite samalaadset kooskasutamist, nt subordineeriv ehk alistav sidesõna + substantiiv ainsuse nominatiivis aluse funktsioonis + finiiitverb indikaatiivi preesensi ainsuse 3. pöördes öeldise funktsioonis: *_J_ sub // ** CLB @ J _S_ com sg nom // @ SUBJ _V_ main indic pres ps3 sg ps af #FinV #Intr // @ + FMV (et riietus on, et loeng kestab)* jm. Morfosüntaktiliste klastrite alusel tulevad esile keelekasutuse morfosüntaktilised piirid.

Morfoloogiliselt samalaadselt vormistatud alguse-lõpuga klastrid koonduvad morfoloogilistesse klassidesse. Klass on ülemmõiste, mis ühendab klastreid, järjestab need esinemissageduse alusel, toob välja keelekasutuse morfoloogilise hierarhia ja varieerumise piirid. Näiteks substantiivialguline klass S–V koosneb klastritest substantiiv + kaassõna + verb ehk SKV (*maja juures on*), substantiiv + adverb + verb ehk SDV (*süda natuke kripeldas*), substabiiv + substantiiv + verb ehk SSV (*metsa vari ulatus*) jne, kus konstruktsiooni keskmine komponent on morfoloogiliselt varieeruv.

Morfoloogiliste klastrite alamliigitus tugineb samalaadsete morfosüntaktiliste konstruktsioonide regulaarsusele. Teisisõnu: klaster on ülemmõiste, mis ühendab samalaadselt vormistatud konstruktsioonid ja järjestab need sageduse alusel. See võimaldab kirjeldada keelekasutuse morfosüntaktilist hierarhiat ning piiranguid konstruktsioonide morfosüntaktilises ja leksikaalses varieeruvuses.

Kõige sagedamini esinenud klass, klaster või konstruktsioon on invariant. Morfosüntaktilise ja leksikaalse varieeruvuse alusel ilmnevad keelekasutusele iseloomulikud leksikaalgrammatilised mustrid: sõnavormide stereotüüpsed, dominantsed, tüüpilised, vähem iseloomulikud ja ebatüüpilised kooskasutused. Tüüpiline klass, klaster või konstruktsioon esineb teistest selgelt sagedamini, selle osakaal valimis võib ulatuda 50%-ni. Kui klassi, klatri või konstruktsiooni osakaal valimis ületab 50% piiri, siis on tegu keelekasutuse dominandiga. Kui mõnda tüüpilistest või dominantsetest morfosüntaktilistest struktuuridest on

ühes kindlas leksikaalgrammatilises variandis kasutatud valdavalt, siis saab rääkida stereotüübist.

Sõna- ning vormikasutust klassides, klastrites ja konstruktsioonides iseloomustab kitsam/avaram varieeruvus või ainukordsus. Sageduskaala alusel on võimalik saada ettekujutus keelekasutaja valikutest ja eelistustest, hinnata neid olulisuse/ebaolulisuse, ratsionaalsuse/ebaratsionaalsuse, struktuuri keerukuse/lihtsuse, sõnavara rikkuse/kasainuse alusel. Loetletud tunnused on suhtelised ega kirjelda absoluutseid väärtusi. Miski saab olla keerukas või lihtne võrdluses millegi teise samaaadsega, mida ühel või teisel põhjusel sobib võrdlusalusena kasutada. Siinses uurimuses on selleks valim tekste eesti ilukirjandusest, sest kirjanikud on sõnastusega vaeva näinud, toimetajad seda oma parema äranägemise järgi kohendanud ning korrektorid lihvinud. Õppija kirjutatud esseede keelekasutuse hindamisel on ilukirjanduskeel omamoodi standard, mille poole võiks püüelda.

3. Uurimuse lähted

Uurimuse teoreetilised seisukohad on seotud kasutuspõhise keelekäsitlusega (nt Langacker 2000: 1–63; Barlow & Kemmer 2000: 7–28; Bybee 2007; 2010; Bybee & Eddington 2006; Bybee & Hopper 2001; Verhagen 2009: 119–152 jt), mis rajaneb konstruktsioonilisuse (Croft 2003; Goldberg 2006) ja funktsionaalse grammatika (Praha lingvistiline koolkond, vene keele grammatikateooria) sidususel. Sellest on välja kasvanud hulki uusi teooriaid, sh optimaalsuse (ülevaade Haspelmath 1999; Wunderlich 2004) ja markeerituse teooria (ülevaade Moravcsik & Wirth 1986; Eckman 1977; Wurzel 1999; Remes 2009). Ka eesti keele uurimisel on rakendatud kasutuspõhise keelekäsitluse ja radikaalse konstruktsioonigrammatika põhimõtteid (vt Sahkai 2008; Penjam 2008; Veismann 2009; Klavan 2012; Uibo 2013 jt).

Martin Haspelmath on märkinud, et keelekasutusest lähtuvad uurimused vajavad niisugust teooriat, mis seletaks, kuidas on omavahel ühendatud keelestruktuurid, nende varieerumise piirid ja keele

kasutamine (Haspelmath 2007). Analüüsinud Haspelmathi varasemaid selleteemalisi mõttearendusi (vt Haspelmath 1999), on William Croft pidanud oluliseks just keekekasutuse varieerumise piiride määramist, kuna see toob esile mitte ainult keeleüksuste kombineerimise võimalused, vaid ka põhimõtted, milles tema arvates peitubki lingvistika olemus. Seetõttu on Croft pidanud Haspelmathi sünteesi formaalse ning funktsionaalse lähenemise vahel väga perspektiivseks (Croft 1999: 206). Küsimus on selles, kuidas ja millega keekekasutaja eelistusi põhjendada, kuidas seletada kirjalike tekstide ja kõne produtseerimist, milles keeleüksusi ning sõnu omavahel valikuliselt kombineeritakse. Traditsiooniliselt on räägitud keeleökonomiast tulenevatest motiveeritud valikutest, mis on Haspelmathi väitel ratsionaalsed ja tuginevad keele sisemistele kategoriaalsetele seostele. Neid saab kirjeldada keeleökonomiast motiveeritud universaalsete morfosüntaktiliste asümmeetriatena. Näiteks esinevad Ø-tunnusega vormid sagedamini kui lisatud tunnustega vormid (nt ainsuse vorm *book* + Ø-lõpp on inglise keeles sagedam kui mitmuse vorm *book* + *s*-lõpp); lühemaid keeleüksusi kasutatakse sagedamini kui pikemaid (nt vene keele sünteetiline refleksiivverb *мыть* + *ся*-postfiks on lühem kui samatähenduslik analüütiline konstruktsioon *мыть* + enesekohane asesõna *себя* – P. Esloni lisatud näide); lihtsad konstruktsioonid esinevad sagedamini kui keerulised (nt saksa keele lihtlause *Ich will spielen* on lihtsam kui põimlause *Ich will, dass du spielst*); standardne on tavaline ja sage, mittestandardset keekekasutust tuleb ette harvem (nt standardne *the book* on keekekasutuses tavaline, samas kui mittestandardne **the my book* võib ette tulla, kuid harva) jne. (Haspelmath 2007: 188–189.) Kuigi kasutussagedus pole keelesüsteemi asümmeetrisuse esiletoomisel ainus faktor, seletab see siiski, et sagedamini kasutatud keelendeid tajub kõneleja tavalise, lihtsa ja seega loomulikuna, harvem kasutatud keelendeid aga suhteliselt keeruka ja vähem loomulikuna, harva esinevaid keelendeid peetakse kas hälbeks või individuaalseks nähtuseks (Haspelmath 2006). Suurem osa universaalsetest morfosüntaktilistest asümmeetriatest on ajalooliselt kujunenud ning neid saab seletada diakrooniliste muutustega (nt

fonoloogiline reduktsioon, grammatisatsioon, analoogiapõhine muutus): “Economical patterns are created by speakers in language use, and when innovative patterns spread through the community, they are manifested in the results of language change” (Haspelmath 2007: 213). Kuigi Haspelmath on kasutanud mõistet *economical patterns* ‘ökonoomised mustrid’, on ta silmas pidanud keelekasutusmustreid, mille aluseks on emakeelekõneleja suulises või kirjalikus produktsioonis ilmnunud *rationally selecting* ‘ratsionaalsed valikud’.

Wolfgang Ullrich Wurzel leiab, et siinkohal tuleks lahus hoida grammatika koos sellele omaste süsteemsete seostega ning keelekasutuse optimaalsus, ökonoomsus ja ratsionaalsus. Nende kahe poole sidusus avaldub universaalses vastanduses loomulik/markeeritud: “This markedness principle thus favors forms that structurally and typologically fit into the appropriate morphological system. It is crucial that the markedness principle has a universal character, but that it favors forms with different properties in different languages” (Wurzel 1999: 246). Frederick J. Newmeyer on sama seisukohta väljendanud veelgi radikaalsemalt: grammatika on grammatika ja kasutus on kasutus: “We have grammar and we have usage. Grammar supports usage, but there is a world of difference between what a grammar is and what we do – and need to do – when we speak” (Newmeyer 2003: 695).

Eelnev on jätkuks periooditi elustuvatele mõttevahetustele keele/kõne dihhotoomiast. Üldiselt ollakse seisukohal, et lingvistika tegeleb keele kui hierarhiliselt struktureeritud süsteemiga. Kõne või tekst on lingvistika huviobjekt vaid seetõttu, et tugineb keeleelementide lineaarsele kooskasutusele, mille regulaarsus ja varieeruvus on keeleökonoomiast motiveeritud ning väljendub esinemise sageduses. Joseph H. Greenberg on väitnud, et kui tahame teha täielikku keeleanalüüsi, siis peab see sisaldama ka kasutussageduse analüüsi, sest sagedus on keele kaasnähtus, nt käändesageduse seos lekseemi semantikaga. Selles aspektis on sageduse mõistega opereerimisel ka teoreetiline väärtus (Greenberg 1990: 219–220). Analoogselt on Joan Bybee ja David Eddington vaadeldud hispaania keele alusel ‘saama’ tähendusega verbide kooskasutust

adjektiiviga ja elus subjektiga ning jõudnud järeltulele, et selle konstruktsiooni alusel ei produtseerita mitte ainult muutust tähistavaid idiomaaatilisi terviküksusi, vaid ka samatähenduslikke mitteidiomaatilisi üksusi (Bybee & Eddington 2006). Analoogne nähtus on esile tulnud eesti õppijakeele Ø-markeeritusega objektifraasis, kus tuumverbid *tegema*, *andma*, *saama* ja *võtma* moodustavad elutu objektiga nii idiomaaatilisi terviküksusi (nt *nalja tegema*, *nõu andma*, *asja saama*, *sõna võtma*) kui ka vabu sõnaühendeid (nt *leiba tegema*, *tööd saama*, *ravimeid võtma*), mida produtseeritakse lekseeme vabalt varieerides ning mida võib sünkrooniliselt lähenedes pidada idiomaaatiliste terviküksuste struktuurianaalooiks. Aja jooksul võivad struktuurianaaloo alusel tekkinud vabad sõnaühendid järk-järgult idiomatiseeruda. (Vt Esilon 2012: 23–26.) Verbi kanda jääb sel juhul vaid grammatiline informatsioon, tähendust märgib substantiiv. Niisugused kinnistunud morfosüntaktilise struktuuriga terviküksused, mille alusel regulaarselt produtseeritakse uusi, moodustavad keelevaramu mobiilse osa. Selles võib näha mõningast analoogiat geneetikanähtusega, kus suurem osa genoomist kui geneetilise informatsiooni varamust töötab n-ö tühikäigul (*junk DNA*) ega ole organismi elutegevuseks otseselt vajalik. Samas on kindlaks tehtud, et *junk DNA* pole täiesti kasutu genoomi osa. Selle koosseisu kuuluvad mobiilsed geneetilised elemendid, mis on nii organismi arengu, adapteerumisvõime kui ka genoomi eksistentsi tagatis (Markov jt 2010: 4 jj). Midagi sarnast võib näha grammatika kui hierarhilise süsteemi ja keele elementide tekstikasutuses. Esiteks ei lähe tekstikasutuses suuremat osa grammatikast üldse vaja. Teiseks on grammatika kui hierarhilise süsteemi iseloomulikuks jooneks liiasus, mida sarnaselt *junk DNA*-ga ei saa samuti pidada kasutuks. Kolmandaks, teatud osa keelesüsteemist on väga mobiilne. Sageli kasutatavate keelestruktuuridega kaasnevad semantilised ja morfosüntaktilised piirangud, mis tagab keelesüsteemi funktsioneerimiseks vajaliku regulaarsuse ja on vajalik keele säilimiseks.

Eelnevat mõttearendust kokku võttes: paradigmat ja keelesüsteemi hierarhiline ülesehitus on grammatikakirjelduse osa, lineaarse korruvuse alusel tekstikasutuses esile tulnud konstruktsioonid on aga

keelekasutus. Grammatika seoseid keelekasutusega võib kirjeldada erinevatest aspektidest ja erinevatelt alustelt, sh sageduse põhjal.

Funktsionaalsetes, kasutuspõhistes, konstruktsioonilistes ja kognitiivsetes uurimustes on tekstisagedust vaadeldud korrelatsioonis keelestruktuuride keerukusega. Zipfi seaduse kohaselt on konstruktsioonide lingvistiline keerukus pöördvõrdeline nende sagedusega. Näiteks invariantsuse seos esinemissagedusega toob välja valikuskaala, mille ühes otsas on kõige enam kasutatud morfosüntaktiline struktuur ja teises otsas sama struktuuri variandid, mida invariandiga võrreldes on tunduvalt harvem kasutatud. Ühelt poolt toob see skaala esile valikud ning teisalt keelekasutaja morfosüntaktilised ja leksikaalsed eelistused, mille alusel saab teha järeltõlke valikute ratsionaalsuse ja keelekasutuse optimaalsuse kohta.

4. Analüüsi tulemused

Järgnev kasutusest tulenev keelekirjeldus tugineb invariantsete/variantsete klastrite ja klasside eristamisele, millel on ilukirjandus- ja õppijakeele valimites erinev esinemissagedus ja osakaal. Invariantsete ja seega oluliste morfoloogiliste koosluste varieeruvus kummaski keelekasutusvariandis on võrreldav nende läheduse-erinemise alusel.

4.1. Ilukirjandus- ja õppijakeele morfoloogiliste klasside võrdlus

Morfoloogiliste klasside kasutusmustrite võrdlus arvestab klasside hulka, sagedust ning osakaalu ilukirjandus- ja õppijakeeles (vt tabelid 1 ja 2).

TABEL 1. Morfoloogiliste klasside osakaal ilukirjandus- ja õppijakeele valimites

Klassid	J-V	P-V	S-V	V-V	D-V	A-V	K-V	G-V
ILU	25%	6%	17%	2%	45%	5%	0%	0%
Õ	26%	20%	17%	17%	16%	2%	1%	1%

Morfoloogiliste klasside tasandil on eesti keele õppija ja emakeelekõneleja keelekasutusmuustrite ühisjooned seotud substantiivi- ja konjunktsioonialguliste klassidega S–V ja J–V, mille osakaal on võrreldavates valimites sama (ILU 17%, Õ 17%) või väga lähedane (ILU 25%, Õ 26%).

Adjektiiv-, pronoomeni-, verbi- ja adverbialguliste klasside (A–V, P–V, V–V ja D–V) osakaal võrreldavates valimites erineb. Õppija eelistab pronoomeni- ja verbialgulisi klasse, mida esineb emakeelekõneleja tekstikasutuses harva. Selge eelistus kuulub adverbialgulisele klassile, mida õppija keelekasutuses on ligi kolm korda vähem (ILU 45%, Õ 16%).

Adpositsiooni- ja käändumatu adjektiivialgused klassid K–V ja G–V on esindatud vaid õppijakeeles, kuid sealgi väga harva.

Morfoloogiliste klasside kasutusmuustrite võrdlusest ilmneb, et emakeelekõneleja rakendab põhiliselt kolme klassi: adverbialguline klass on tüüpiline ja seega kõige olulisem, substantiivi- ja konjunktsioonialgulisi klasse on kasutatud sageli. Pronoomeni-, adjektiiv- ja verbialgulised klassid on küll esindatud, kuid nende kasutamist pole emakeelekõneleja pidanud oluliseks. Niisiis on ilukirjanduskeeles klasside varieeruvus nähtav, kuid olulised on vaid kolm klassi: D–V, J–V ja S–V. Tegu on selgelt piiritletud ratsionaalse valikuga – suhtlusvajadusteks piisab kolmest morfoloogilisest klassist. Seevastu õppija katsetab oma esseedes talle teadaolevaid võimalusi ja seetõttu on morfoloogiliste klasside varieeruvus avaram ning klasside osakaal valimis ühtlasem. Õppija väldib ilmselgelt adverbialgulist morfoloogilist klassi. Selle asemel rakendab ta kõige sagedamini konjunktsioonialgulist, pidades samavõrd oluliseks kõiki viit klassi. Seega ei oska õppija teha emakeelekõnelejaga sarnaseid optimaalseid valikuid. Konkreetsemad õpiraskustega seotud nähtused peaksid ilmnema iga klassi klastrivarieeruvuses.

4.2. Ilukirjandus- ja õppijakeele klastrivarieeruvus

Võrreldes klastrite osakaalu ja nende hulka igas klassis (vt tabel 2), saab kindlaks teha klasside sõnaliigilise varieeruvuse piirid, tuua esile stereotüüpsed, domineerivad, tüüpilised ja ebatüüpilised klastrid, mis iseloomustavad emakeelekõneleja või õppija tekstiloomet. Kahe valimi klastrikasutuse võrdlemisel liigume läheduse-erinemise skaalal ühisosast järjest suuremate erinevuste suunas. Morfoloogiliste klasside võrdluse põhjal eeldame, et emakeelekõneleja ja õppija tekstiloomed ühisjooned võiksid olla seotud substantiivi- ja konjunktsioonialguliste klassidega S–V ja J–V. Klasside A–V, P–V, V–V ja D–V kasutusmustrid erinevad, klasse K–V ja G–V pole ilukirjanduskeele valimis esile tulnud (vt 4.1) ning need jäävad seetõttu kirjeldusest välja. Iga klassi klastrivarieeruvus on seotud klastri keskmise komponendi sõnaliigilise varieerumisega (vt tabel 2).

TABEL 2. *Klastrite varieerumine morfoloogilistes klassides*

ILU klassid	ILU klastrid	Sagedus	Õ klassid	Õ klastrid	Sagedus
D–V 412	DDV	206	J–V 3590	JPV	1558
	DVV	116		JSV	977
	DJV	69		JVV	582
	DSV	11		JDV	473
	DAV	10			
J–V 230	JDV	114	P–V 2803	PVV	1881
	JPV	67		PSV	558
	JSV	49		PDV	177
				PKV	67
				PPV	64
S–V 150			S–V 2430	PAV	56
	SDV	45		SVV	1329
	SVV	43		SSV	412
	SKV	29		SKV	273
	SSV	20		SJV	204
	SJV	13		SPV	116
			SDV	96	

P-V 53	PDV	28	V-V 2402	VVV	1025
	PVV	15		VDV	576
	PSV	10		VJV	305
				VPV	264
				VSV	211
				VAV	21
A-V 44	ASV	44	D-V 2169	DVV	832
				DPV	590
				DDV	468
				DSV	224
				DJV	42
				DAV	13
V-V 15	VDV	15	A-V 347	ASV	308
				AVV	17
				AJV	16
				ADV	6
–	–	–	K-V 124	KSV	50
				KVV	46
				KJV	18
				KDV	5
				KPV	5
–	–	–	G-V 76	GSV	76

4.2.1. Klass S–V

Substantiivialgulise klassi S–V osakaal langeb ilukirjandus- ja õppijakeeles kokku (vt tabel 1), samal ajal kui klastrivarieeruvus, esinemissagedus (vt tabel 2) ja osakaal (vt tabel 3) on võrreldavates valimites osaliselt vastupidine. Järelikult erinevad ka klastrite kasutusmustrid. Nii näiteks on adverbis sisaldav klaster SDV ilukirjanduskeeles tavaline, kuid õppijakeelekasutusele mitte omane (ILU 30%, Õ 4%). Verbis sisaldav klaster SVV on ilukirjanduskeeles osakaalult teine, õppijakeeles aga dominantne (ILU 29%, Õ 55%). Substantiivi sisaldav klaster SSV on ilukirjanduskeele valimis osakaalult neljas, õppijakeeles teine (ILU 13%, Õ 17%). Konjunktsiooni sisaldava klasteri SJV osakaal on võrreldavates

valimites lähedane ning samalaadselt vähe oluline (ILU 9%, Õ 8%). Pronoomenit sisaldav klaster SPV tuli esile vaid õppijakeeles (Õ 5%).

TABEL 3. *Klastrite osakaal ilukirjandus- ja õppijakeeles: klass S–V*

Klastrid	SVV	SSV	SKV	SJV	SPV	SDV
ILU	29%	13%	19%	9%	0%	30%
Õ	55%	17%	11%	8%	5%	4%

Niisiis tuleb ilukirjanduskeeles klatri keskmise komponendi varieeruvuse alusel substantiivialgulisel klassis S–V esile viis ja õppijakeeles kuus klatri. Emakeelekõneleja eelistab kasutada keskmise komponendina adverbi ja verbi; kaassõnu ning substantiive sisaldavate klastrite osakaal on väiksem, kuid enam-vähem sama; kõige harvemini on kasutatud konjunktsiooni. Õppijakeeles on esile tulnud üks dominantne klaster, mille keskmine komponent on verb. Vähem on kasutatud substantiivi ja kaassõna, harvem konjunktsiooni sisaldavaid klastreid (analoogia emakeelekõnelejaga). Õppija on täiesti vältinud adverbi sisaldavat klatri, samal ajal kui emakeelekõnelejal on see ootuspäraselt kõige sagedam. Teisalt on õppijakeeles esile tulnud pronoomenit sisaldav klaster, mis emakeelekõneleja kasutusmustris puudub. Kokkuvõttes moodustub substantiivialgulisel klassi S–V klatrikasutuses ilukirjandus- ja õppijakeele vahel läheduse-erinemise skaala. Kõige lähedasem on selle skaala alusel konjunktsiooni sisaldava klatri SJV kasutamine, mida õppijakeeles on kergelt alakasutatud. Vähem sarnane on substantiivi sisaldava klatri SSV kasutus, mida õppijakeeles on ülekasutatud. Vaid ligilähedaseks saab pidada võrreldavates valimites kaassõna sisaldava klatri SKV kasutamist, mida õppijakeeles on selgelt alakasutatud. Kõige suurem erinevus ilmneb adverbi ja verbi sisaldavate klastrite puhul: esimest on õppijakeeles oluliselt alakasutatud ning teist oluliselt ülekasutatud (tegu on õppijakeele dominantse morfoloogilise klatriga, mille kasutus ilukirjanduskeeles piirdub vaid eitava kõne näidetega). Seega ei ava morfoloogiliste klasside üldised kasutusmustrid klastrite tasandil ilmnevate valikute ja eelistuste sarnasust ega erinevusi.

4.2.2. Klass J–V

Konjunktsiooniga algava klassi J–V klastrivarieerumine ja esinemisagedus (vt tabel 2) ning klastrate osakaal klassis (vt tabel 4) on võrreldavates valimites otse vastupidine (osaline analoogia substantiivialgulise klassiga). Ilukirjanduskeeles esineb domineeriv adverbi sisaldav klaster JDV, samal ajal kui õppijakeeles on selle klastri osakaal kõige väiksem (ILU 50%, Õ 13%) – analoogia klastri SDV kasutusega substantiivialgulises klassis. Õppija keelekasutusele on seevastu tüüpiline pronoomenit sisaldav klaster JPV, mille osakaal ilukirjanduskeeles on aga poole väiksem (Õ 44%, ILU 29%). Substantiivi sisaldava klastri JSV osakaal näitab võrreldavate valimite vahel mõningast lähedust, kuid on siiski erinev (ILU 21%, Õ 27%). Verbi sisaldav klaster JVV pole ilukirjanduskeeles valimis üldse esile tulnud, õppijakeeles on aga selle klastri osakaal üsna suur (Õ 16%).

TABEL 4. *Klastrate osakaal ilukirjandus- ja õppijakeeles: klass J–V*

Klastrid	JPV	JSV	JVV	JDV
ILU	29%	21%	0%	50%
Õ	44%	27%	16%	13%

Niisiis tuleb klastrate osakaalu läheduse-erinemise alusel konjunktsioonialgulises klassis esile klastrikasutuse erinemise skaala: klastri JSV puhul on kõige enam lähedust, kuid õppijakeeles on seda klastrit siiski ülekasutatud; klastrate JPV ja JVV kasutamine erineb suuresti, sest õppijakeeles on mõlemat klastrit märgatavalt ülekasutatud; klastri JDV kasutus erineb kardinaalselt, sest õppija on seda suurel määral alakasutanud. Niisiis ei toeta ka klassi J–V klastrikasutuse võrdlev analüüs klas-side tasandil esile tulnud samalaadsust, vaid toob välja otse vastupidiseid tendentse.

4.2.3. Klass A–V

Ilukirjandus- ja õppijakeeles on adjektiivialgulise klassi A–V osakaal lähedane (vt tabel 2). Selle klassi klastrikasutuses tuleb esile substantiivi sisaldav dominantne klaster ASV, mida võrreldavates valimites on selgelt eelistatud (ILU 100%, Õ 89%), kuid osakaalu poolest siiski silmanähtavalt erinev (vt tabel 5). Teine eristav nähtus on klastrite varieeruvus, mida leidis vaid õppijakeeles, kuigi variantide osakaal valimis ei ole suur (AVV 5%, AJV 4% ja ADV 2%). Niisiis ei moodustu klastrite osakaalu läheduse-erinemise alusel skaalat, kuid adjektiivialgulise klassi A–V klastrikasutus on võrreldavates valimites nii klastrite osakaalu kui ka varieeruvuse poolest suuresti vastandlik.

TABEL 5. *Klastrite osakaal ilukirjandus- ja õppijakeeles: klass A–V*

Klastrid	ASV	AVV	AJV	ADV
ILU	100%	0%	0%	0%
Õ	89%	5%	4%	2%

4.2.4. Klass P–V

Pronoomenialgulise klassi P–V kasutussagedus on võrreldavates valimites lähedane (vt tabel 2), klastrite osakaal klassis näitab nii samalaadsust, lähedust kui ka suuri erinevusi (vt tabel 6). Nii on substantiivi sisaldava klatri PSV osakaal samalaadne (ILU 19%, Õ 20%); adpositsiooni, pronoomenit, adjektiivi sisaldavad klastrid PKV, PPV ja PAV pole küll ilukirjanduskeeles esile tulnud, kuid ka õppijakeeles on nende osakaal ühtlaselt väike (vastavalt 3%, 2% ja 2%). Oluline erinevus tuleb sisse domineerivate klastrite kasutamises: õppijakeeles on see seotud verbi sisaldava klastriga PVV (Õ 67%, ILU 28%), mida ilukirjanduskeeles esines üle kahe korra vähem, ning ilukirjanduskeeles adverbi sisaldava klastriga PDV (ILU 53%, Õ 6%), mida õppijakeeles oli harva kasutatud. Niisiis on õppijakeeles klastrit PVV selgelt ülekasutatud ja klastrit PDV

suuresti alakasutatud. Seega on võrreldavates valimites klassi P–V domineerivate klastrite kasutuseelistused vastupidised.

Klastrite osakaalu läheduse-erinemise alusel moodustub läheduse-erinemise skaala (analoogia substantiivialgulisel klassiga S–V). Võrreldavate valimite dominantsed klastrid erinevad omavahel kardinaalselt, kuid vastanduvad ühtmoodi ülejäänud klastritele. Ilmselt sellele tuginemki ilukirjandus- ja õppijakeele klastrikasutuse analoogsus klassiga S–V, vahe on vaid selles, et õppijakeeles on avaram klastrivarieeruvus. Esile tuleb õppijale iseloomulik tendents eelistada pronoomenit sisaldavaid struktuure, samal ajal kui ilukirjanduskeeles on eelistatud adverbi sisaldavaid struktuure.

TABEL 6. *Klastrite osakaal ilukirjandus- ja õppijakeeles: klass P–V*

Klastrid	PVV	PSV	PDV	PKV	PPV	PAV
ILU	28%	19%	53%	0%	0%	0%
Õ	67%	20%	6%	3%	2%	2%

4.2.5. Klass V–V

Verbialguline klass V–V on ilukirjanduskeeles esindatud adverbi sisaldava dominantse klastriga VDV (100%), mida õppijakeeles on alakasutatud (24%) – siin on analoogia eespool kirjeldatud adjektiivialgulisel klassiga A–V (vrd tabelid 5 ja 7).

TABEL 7. *Klastrite osakaal ilukirjandus- ja õppijakeeles: klass V–V*

Klastrid	VVV	VDV	VJV	VPV	VSV	VAV
ILU	0%	100%	0%	0%	0%	0%
Õ	42%	24%	13%	11%	9%	1%

Lisaks esineb õppijakeeles selgelt nähtav klastrivarieeruvus, mis on adjektiivialgulisel klassiga võrreldes avaram ja osakaalu poolest suurem (VJV 13%, VPV 11%, VSV 9% ja VAV 1%). Õppijakeele tüüpiline

klaster sisaldab verbi VVV (42%), ilukirjanduskeelee dominantne klaster adverbil. Niisiis on klassis V–V võrreldavate valimite klastrikasutus täiesti vastandlik ja üldjoontes sarnane eespool kirjeldatud klassiga A–V.

4.2.6. Klass D–V

Adverbialguline klass D–V on tüüpiline ilukirjanduskeeleele; ülejäänud klassidega võrreldes on selle osakaal valimis kõige suurem. Seevastu õppija keelekasutuses on klassi D–V osakaal oluliselt väiksem (vt tabel 1). Niisiis peaks just selle klassi klastrivarieeruvuse, sageduse ja osakaalu alusel (vt tabel 8) ilmne kahe valimi keelekasutusmustrite maksimaalne erinevus.

TABEL 8. *Klastrite osakaal ilukirjandus- ja õppijakeeles: klass D–V*

Klastrid	DVV	DPV	DDV	DSV	DJV	DAV
ILU	28%	0%	50%	3%	17%	2%
Õ	38%	27%	22%	10%	2%	1%

Esiteks on ilukirjanduskeelee klastrivarieeruvus kitsam kui õppijakeeles: viiest klastrist on kasutatud põhiliselt kolme – adverbil, verbi ja konjunktsiooni sisaldavaid klastreid, õppijakeeles kuuest nelja – verbi, pronoomenit, adverbil ja substantiivi sisaldavaid klastreid. Teiseks, ilukirjanduskeelee domineeriv klaster DDV sisaldab adverbil, õppijakeeles on seda klastrit suuresti alakasutatud (ILU 50%, Õ 22%), mis on tõendus sellest, et emakeelekõneleja keelekasutuses on adverbil kesksel kohal. Õppijakeele tüüpilist verbi sisaldavat klastrit DVV on ilukirjanduskeelega võrreldes selgelt ülekasutatud (Õ 38%, ILU 28%), nagu ka substantiivi sisaldavat klastrit DSV (Õ 10% , ILU 3%). Adjektiivil sisaldav klaster DAV esines mõlemas valimis võrdselt harva (ILU 2%, Õ 1%). Pronoomenit sisaldav klaster DPV ei tulnud ilukirjanduskeeles üldse esile, küll aga on seda rohkesti kasutatud õppijakeeles (Õ 27%, ILU 0%), mis on tõendus õppijakeele subjektifraasi pronoomenikesksusest. Niisiis vastandub ilukirjanduskeeles domineeriv adverbil sisaldav klaster DDV sageli esinevatele konjunktsiooni ja verbi **sisaldavatele klastritele**

DJV ja DVV. Õppija keelekasutuses domineeriv klaster puudub, tüüpiline on verbi sisaldava klatri DVV kasutamine, samuti üsna avar klastrivarieeruvus (DVV–DPV–DDV–DSV). Klastrite osakaalu läheduse-erinemise alusel on klassi D–V klatrikasutus võrreldavates valimites märgatavalt vastandlik. Ilmselt on tegu ühe raskemini omandatava nähtusega eesti keeles.

5. Järeldusi

On loomulik, et keeleõppija ei jäljenda emakeelekõneleja valikuid täiel määral, sest kummalgi on oma kaasasündinud või omandatud keelekompetentsus, teksti produtseerimise strateegiad ja sellest tulenevad keelekasutuseelistused. Samas on Nick C. Ellis rõhutanud, et keele omandatust iseloomustab just see, kuivõrd ratsionaalselt õppijad keelt kasutavad ning kui optimaalne on see protsess (Ellis 2006).

Eesti ilukirjandus- ja õppijakeele võrdleva klasteranalüüsi tulemused kinnitavad seda seisukohta. Morfoloogiliste klasside tasandi võrdlus tõi selgelt esile emakeelekõneleja valikute ratsionaalsuse. Eelistatakse põhiliselt kolme morfoloogilist klassi: adverbialguline on tüüpiline (nt *väga tublisti kosunud, seal varemgi olnud, veel ümber riietunud*) ja seega kõige olulisem klass, mis sageduse poolest vastandatud samuti sagedastele substantiivi- (nt *süda natuke kripeldas, põlv ebameeldivalt tuikas, vee peale valanud, mehe üle võtnud*) ja konjunktsioonialgulisele klassidele (nt *ja kindlasti oli, ning siis kadus, aga nüüd sündis*). Seevastu õppijakeele morfoloogiliste klasside varieeruvus on avaram ja klasside osakaal valimis ühtlasem. Ilmselgelt väldib õppija adverbialgulist morfoloogilist klassi, eelistades konjunktsioonialgulist (nt *et see on, sest see toob, ja ma saan*) ning kasutades samaväärselt kõiki viit morfoloogilist klassi. Ent kui tahetakse keelt omandada, siis tuleb õppida keelesüsteemi funktsionaalset potentsiaali sama osavalt kasutama kui emakeelekõneleja. Selle mehhanismi mittetäielik valdamine põhjustabki õppija keelekasutuse nihestatuse emakeelse inimesega võrreldes. Mõlemad räägivad sama keelt, kuid teevad seda erinevalt: emakeelekõneleja tugineb ilmselt

keelekompetsusest tulenevale loomulikkusele, õppija aga ülekasutusele, lihtsustamis- ja vältimisstrateegiatele. Näiteks väldib õppija emakeelekõneleja jaoks tavalist morfoloogilist klassi D–V ning ülekasutab klasse V–V ja P–V.

Morfoloogiliste klasside täpsema kirjelduse annab klastrivarieeruvus ja klastrate osakaal valimis. Analüüsi alusel selgus, et klasside A–V ja V–V klastrikasutus on täiesti vastandlik; klassi D–V klastrikasutus on märgatavalt vastandlik; klassi J–V klastrikasutus on erinev; klasside S–V ja P–V klastrikasutus näitab läheduse/erinemise skaalat. Kui aga arvestada klasside klastrisagedust ja morfoloogilise varieeruvuse piire klassi sees, siis selgub, et kõige olulisemad erinevused emakeelekõneleja ja õppija keelekasutuses koonduvad adverbialgulise klassi ümber, erinevuste poolest järgneb konjunktsioonialguline klass. Substantiivi- ja pronoomenialguliste klasside klastrivarieeruvuses on rohkem lähedust kui erinevusi. See järjestus võimaldab ühtlasi seletada mõningaid õpiskusi, seda teadmist saab rakendada eesti keele õppes. Kuna klasside A–V ja V–V osakaal on võrreldavates valimites suhteliselt väike, siis ei ole nende vastandlikul klastrikasutusel keelekasutustendentside seisukohalt olulist tähendust.

Et selgitada, miks on eesti keele õppijal raske kasutada adverbialgulist morfoloogilist klassi nii, nagu seda teeb emakeelekõneleja, toome näite klassi D–V invariantse klatri DDV invariantse konstruktsiooni `_D_ // @ ADVL _D_ // @ ADVL _V_ main partic past ps #Intr // @ -FMV` (nt *seal varemgi olnud, väga tublisti kosunud*) morfosüntaktilise ja leksikaalse varieeruvuse piiridest ning võrdleme seda õppijakeele sama klassi invariantse klatri DVV invariantse konstruktsiooniga `_D_ // @ADVL _V_ aux neg // @NEG _V_ main indic pres ps neg #FinV #NGP-P // @+FMV` (nt *sageli ei vali, kohe ei saa, rohkem ei tee, nagunii ei saa*).

Ilukirjanduskeele klaster DDV koosneb kahest leksikaalselt varieeruvast adverbist ning morfosüntaktiliselt ja leksikaalselt varieeruvast verbist. Verbivormide valik näitab üsna suurt keerukust: kõige sagedamini on kasutatud liitaegadesse kuuluvaid *nud-* ja *tud-*partitsiipe (valdavalt perfekti koosseisus); *da-*infinitiiv on tavaline kas infiniitse öeldise

või objekti funktsioonis; finiiitverbi vormivalik on piiratud indikatiivi ainsuse 3. pöördega imperfektis ja preesensis; *ma*-infinitiivi on kasutatud infiniitse öeldisena. Verbivormide valik viitab klatri DDV morfosüntaktilisele keerukusele, ajavormidest eelistatuim on perfekt.

Jaatavas kõnes toob verbi leksikaalne varieerumine esile rikkaliku sünteetiliste ja analüütiliste verbide kasutamise. Enamik analüütilisi verbe on ainukordsed ühendverbid: *ette tulema* (*teatama*), *edasi minema* (*tormama*), *järele kuulama* (*hõikama*), *kinni mähkima* (*panema*), *kokku liimima*, *kõrvale jätma*, *läbi astuma* (*pugema*, *uurima*), *lahti saama*, *maha jätma* (*jääma*, *võtma*, *vaatama*), *mööda käima*, *sisse mahtuma* (*minema*, *seadma*), *peale sattuma* (*vaatama*), *tagasi minema* (*tulema*, *võtma*), *täis kirjutama*, *vastu tulema*, *välja ilmuma* (*tulema*, *joonistama*, *tegema*, *üttelema*), *ülal pidama*, *üle lugema* (*kuulama*, *ümber kirjutama*, *vaatama*, *minema*, *saama*), *üles võpatama* (*kaisutama*, *leidma*), *ära ajama* (*sööma*, *lõpetama*, *minema*, *käima*, *sulama*, *mahtuma*, *sõitma*, *pidama*, *põletama*, *peitma*, *tegema*, *lendama*) jne. Väljendverbe esines vähe.

Silma hakkab ka adverbide suur leksikaalne varieeruvus, kuigi mõningaid adverbe on eelistatud korrata nii konstruktsiooni alguses (*enam*, *nüüd*, *siis*, *mitte*) kui ka sees (*põhjalikult*, *täpselt* ja *õigeaegselt*). Ülejäänud adverbid on ainukordsed: *ainiti*, *uuesti*, *pahasti*, *tublisti*, *kenasti*; *imestunult*, *tähelepanelikult*, *väliselt*, *hädavaevalt*, *ootamatult*, *täpselt*, *lihtsalt*, *sügavalt*, *lõplikult*, *teaduslikult*, *tõenäoliselt*, *naeruväärselt*, *päriselt*, *juhuslikult*, *tõsiselt*, *südamlikult*, *rahuldavalt*, *täht-tähelt*, *äärepealt*, *kõrgelt*.

Õppijakeeles oli morfoloogilise klassi D–V sagedam klaster DVV, selle invariantne konstruktsioon on *_D_ // @ADV_L _V_ aux neg // @NEG _V_ main indic pres ps neg #FinV #NGP-P // @+FMV* (nt *sageli ei vali*, *kohe ei saa*, *rohkem ei tee*, *nagunii ei saa*). Konstruktsiooni alguses on adverb määruse funktsioonis, järgneb eituspartikkel *ei* ja finiiitverbi eitava kõne indikatiivi preesens, harvem eitava kõne indikatiivi imperfekt.

Konstruktsiooni kasutusele on iseloomulik leksikaalne varieeruvus, ent konstruktsiooni alguses eelistab õppija siiski adverbe (k.a

modaaladverbid) *üldse, üldiselt, ka, kunagi, kindlasti, veel, isegi, lihtsalt, seal, siin, praegu, enam, täpselt*. Muuhulgas tõi klasteranalüüs esile ka õppijakeele tüüpvea **juba* <enam>, kus sõnade *juba* ja *enam* valikul pole mõistetud vaatepunktide erinevust. Ainukordseid adverbe on vähe.

Verbivalik on kesine: eitava kõne indikatiivi preesensis on sagedamad *ei ole, ei saa, ei ütle, ei tea, ei unusta, ei mõtle(gi), ei meeldi* ning eitava kõne imperfektis tavapäraselt *ei olnud*, harva *ei saanud (esinenud, ilmunud, jäänudki, kujundanud, meeldinud)*.

Niisiis on õppijakeele invariantse klatri DVV invariantse konstruktsiooni morfosüntaktiline varieerumine lihtne (finiitverb eitava kõne indikatiivi preesensis, harvem eitava kõne indikatiivi imperfektis), leksikaalne varieerumine piiratud. Esile tuleb õppija keelekasutusele tüüpiline eitava kõne leksikaalgrammatiline muster, mille keskmes on kas eksistentsiaalverb *olema* või modaalverb *saama*: *üldse (üldiselt, ka, kunagi, kindlasti, veel) ei ole / ei saa ~ ei olnud*.

Seevastu ilukirjanduskeele invariantse klatri DDV invariantse konstruktsiooni kasutust iseloomustab 1) verbivormide keerukas morfosüntaktiline varieerumine; 2) liitaegade ja analüütiliste verbide kasutus; 3) teatud adverbide (*enam, nüüd, siis, mitte*) kinnistumine konstruktsiooni alguses (seotud adverbiliste funktsioonide taandumisega); 4) idiomaatilist laadi analüütiliste verbide kasutamine (enamasti ainukordsed ühendverbid); 5) rikkalik leksikaalne varieeruvus.

Kirjeldatud võrdlusnäide morfoloogilise klassi D–V kasutusest ilukirjandus- ja õppijakeeles toob esile neid keelenähtusi, mis tekitavad õpiraskusi ning mille tõttu see morfoloogiline klass ei kuulu õppija kasutuseelistuste hulka. Eelistatud on hoopis konjunktsioonialgulist klassi J–V, selle pronoomenit sisaldavat invariantset klastrit JPV ja selle invariantset konstruktsiooni *_J_ sub // **CLB @J _P_ dem sg nom // @SUBJ _V_ main indic pres ps3 sg ps af #FinV #Intr // @+FMV* (nt *et see on, et see viib, sest see annab*). Invariantse konstruktsiooni alguses on tavaliselt subordineeriv sidesõna *et* (harvem *sest*), järgneb demonstratiivpronoomen *see* ainsuse nominatiivis ja finiitverb indikatiivi preesensi ainsuse 3. pöördes. Lisaks invariantsele konstruktsioonile on

klastris JPV veel 43 konstruktsiooni, mille varieerumine on koondunud erinevate personaalpronoomenite ümber (sagedamad neist on vastandus *ma-nad* ja *ta*). Subjekti funktsioonis on neid kasutatud nii ainsuses kui ka mitmuses ning tavapäraselt lühivormis – pikk vorm on olnud erand. Samas mainitakse sageli ühe tüüpveana just seda, et õppijad ei taju eesti personaalpronoomenite lühi- ja pika vormi kasutuserinevusi. Subjektifraasi klasteranalüüsi tulemused seda väidet ei toeta. Kui jätkata võrdlust eesti ilukirjanduskeelega, siis siin pole pronoomenit sisaldavat klastrit JPV eelistatud, vaid on kasutatud hoopis adverbi sisaldavat klastrit JDV ja selle invariantset konstruktsiooni `_ J _ crd // ** CLB @ J _D_ // @ ADVL _ V _ main indic impf ps3 sg ps af #FinV #Intr // @ + FMV` (nt *ja kindlasti oli, ning siis kadus, Kuid kahtlemata oli, Aga ilmselt jäi*). Subordineerivat sidesõna esineb konstruktsioonis võrdselt nii lause alguses kui ka sees, järgneb adverb ja finiiitverb öeldise funktsioonis: eelistatud on indikatiivi imperfekti ainsuse 3. pööret, sageduselt järgmine on preesensi ainsuse 3. pööre ja imperfekti mitmuse 3. pööre. Niisiis on eesti keele funktsionaalset potentsiaali erinevates keelekasutusvariantides rakendatud erinevate põhimõtete alusel.

Kokkuvõttes võib öelda, et morfoloogiliste klasside, klastrite ja nende eristamise aluseks olnud konstruktsioonide morfosüntaktiline ja leksikaalne varieeruvus võimaldab kirjeldada ning seletada, kuidas keelesüsteemi funktsionaalset potentsiaali rakendatakse ning missuguseid rohkem või vähem loomulikke piiranguid erinevates keelekasutusvariantides võib esineda. Selles peitubki kasutuspõhise keelekirjelduse väärtus ühelt poolt grammatika ja teisalt keeleõppe jaoks.

Tänuõnad

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)”; ETFi grant 8222 “Ülekantud tähenduses fraasid eesti keele korpustes (2010–2013)” ja ETFi grant 8240 “Lähtekeele mõju ja teise keele omandamine: korpuspõhine uurimus (2010–2013)”.

Lühendid

ILU ilukirjanduskeel
 Õ õppijakeel

Kirjandus

- Barlow, Michael, Suzanne Kemmer (Eds.) 2000. Usage-Based Models of Language. Stanford: CSLI.
- Biber, Douglas, Susan Conrad, Randi Reppen 1998. Corpus Linguistics: Investigating Language Structure and Use. Cambridge: Cambridge University Press.
- Bybee, Joan 2007. Frequency of Use and the Organization of Language. Oxford: Oxford University Press.
- Bybee, Joan 2010. Language, Usage and Cognition. Cambridge: Cambridge University Press.
- Bybee, Joan, David Eddington 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, Paul Hopper (Eds.) 2001. Frequency and the Emergence of Linguistic Structure. Amsterdam: John Benjamins Publ. Co.
- 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.
- Eckman, Fred R. 1977. Markedness and the Contrastive Analysis Hypothesis. – Language Learning 27 (2), 315–330. <http://dx.doi.org/10.1111/j.14671770.1977.tb00124.x>
- 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 2012. Objekti ja tegevuse leksikaalne markeeritus eesti õppijakeeles. – Lähivõrdlusi. Lähivertailuja 22, 15–42. <http://dx.doi.org/10.5128/LV22.01>
- Eslon, Pille, Katre Õim, Annekatrin Kaivapalu, Reili Argus, Erika Matsak 2010. Kuidas uurida esimese ja teise keele omandamist? – Lähivõrdlusi. Lähivertailuja 20, 11–48. <http://dx.doi.org/10.5128/LV20.01>
- 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, CA: 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 2006. Against markedness (and what to replace it with). – *Journal of Linguistics* 42 (1), 25–70. <http://dx.doi.org/10.1017/S0022226705003683>
- Haspelmath, Martin 2008. Creating economical morphosyntactic patterns in language change. – Jeff Good (Ed.). *Linguistic Universals and Language Change*. Oxford: Oxford University Press, 185–214.
- Klavan, Jane 2012. Evidence in linguistics: Corpus-linguistic and experimental methods for studying grammatical synonymy. *Dissertationes linguisticae Universitatis Tartuensis* 15. Tartu: University of Tartu Press.
- Langacker, R. W. 2000. A dynamic usage-based model. – Michael Barlow, Suzanne Kemmer (Eds.). *Usage-Based Models of Language*. Stanford: CSLI Publications, 1–63.
- Markov jt 2010 = Марков Александр Владимирович, Валерий Анатольевич Анисимов, Андрей Витальевич Коротаев 2010. Взаимосвязь размера генома и сложности организма в эволюционном ряду от прокариот к млекопитающим. – *Палеонтологический журнал* № 4, 3–14.
- Moravcsik, Edith A., Jessica R. Wirth 1986. Markedness – an overview. – Fred R. Eckman et al. (Eds.). *Markedness*. New York: Plenum Press, 1–11.
- Newmeyer, Frederick J. 2003. Grammar is grammar and usage is usage. – *Language* 79 (4), 682–707. <http://dx.doi.org/10.1353/lan.2003.0260>
- Ots, Sander 2012. Statistikapõhise tarkvara loomine morfoloogiliste kollokatsioonide eraldamiseks eesti keele tekstidest. *Bakalaureustöö*. Tallinn: TLÜ informaatikainstituut.
- Penjam, Pille 2008. Eesti kirjakeele *da-* ja *ma-*infinitiiviga konstruktsioonid. *Dissertationes philologiae Estonicae Universitatis Tartuensis* 23. Tartu: Tartu Ülikooli Kirjastus.
- Remes, Hannu 2009. Muotot kontrastissa. Suomen ja viron vertailevaa taivutusmorfologiaa. *Acta Universitatis Ouluensis B* 90. Oulu: Oulun yliopisto.
- Sahkai, Heete 2008. Konstruktsioonipõhine keelemudel ja sõnaraamatumudel. – *Eesti Rakenduslingvistika Ühingu aastaraamat* 4, 177–186. <http://dx.doi.org/10.5128/ERYa4.11>
- Uiboaed, Kristel 2013. Kollostruktsioonilised meetodid ja konstruktsioonilise varieerumise tuvastamine. – *ESUKA / JEFUL* 4 (1), 185–204.

- Veismann, Ann 2009. Eesti keele kaas- ja määrsõnade semantika võimalusi. 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>
- Wunderlich, Dieter 2004. Optimality theory in morphology and syntax. – <http://user.phil-fak.uni-duesseldorf.de/~wdl/Optimality%20theory.pdf> (3.2.2013). [To appear in: *Encyclopedia of Language & Linguistic*, 2nd ed. Oxford: Elsevier.]
- Wurzel, Wolfgang Ullrich 1999. Principles of evaluation, change and related issues. – *Zeitschrift für Sprachwissenschaft* 18 (2), 242–250. <http://dx.doi.org/10.1515/zfsw.1999.18.2.242>

Pille Eslon

Tallinna Ülikooli eesti keele ja kultuuri instituut

Narva mnt 25

10120 Tallinn, Estonia

pille.eslon@tlu.ee

The comparative study of language use: Morphological classes and clusters

PILLE ESLON

Tallinn University

This study makes use of cluster analysis to describe the functional mobility of a language system. It also juxtaposes the use of language in a context occurring to the left of the verb in Estonian belletristic and learner language texts. All constructions consisting of three units were identified, grouped according to morphosyntactic similarity and frequency, and described as a hierarchy of morphological clusters and classes. While classes are indicative of more general tendencies characteristic of language usage, clusters are used to describe the constraints on morphological variability of classes. Specific constructions highlight the constraints on morphosyntactic and lexical variability of clusters. It is presumed that the comparison of Estonian belletristic and learner language will reveal how rationally the functional potential of a language system is used by native speakers and language learners. Results of the analysis show that native speakers make optimal choices, which entail morphosyntactic simplicity but also complexity and vocabulary richness. Learners' language usage deviates in this sense and, even though it may be correct, sounds slightly awkward for a native speaker. On the one hand the structures used in learner language usage show a broader range of variability than those used in belletristic language, but on the other hand, learner language usage reveals oversimplification based on stereotypical language structures, overuse and underuse of morphological classes and clusters and poor vocabulary. Therefore, broader variability is not a sign of complexity of language usage, but of the fact that learner in his/her language usage has not acquired the native speaker's innate ability to make optimal choices and use the language rationally. Different speakers of language thus apply the functional potential of language relying on different sets of principles, which prompt the activation of different systemic relations and, subsequently, the potential co-occurrence of linguistic forms.

Keywords: cluster analysis; constraints and variability on language use; Estonian belletristic language; Estonian learner language; usage-based language description