

# AINSUSE SISSEÜTLEVA VORMI VALIKU SEOS MORFOSÜNTAKTILISTE JA SEMANTILISTE TUNNUSTEGA – MATERJALI NING MEETODI SOBIVUS KORPUSANALÜÜSIKS

ANN SIIMAN

**Annotatsioon.** Käsitluse eesmärgi on välja selgitada, millistest morfosüntaktilistest ja semantilistest tunnustest on eesmärgi pika ja lühikese sisseütleva vormi valik. Artikkel jätkab käsitlust, kus uuriti, millistest morfofonoloogilistest tunnustest on eesmärgi sisseütleva vormi valik. Siinses käsitluses muudetakse eelmise uurimuse materjalikogumise põhimõtteid – erinevalt eelmisest kasutatakse tasakaalustatud valimit ja tulemused saadakse korduva vorme arvestamata. Uuritakse, kas sisseütleva pika või lühikese vormi valik on morfosüntaktilistest tunnustest, nagu sõnaliik, lauseliige, reksioonilisus ja püsiühendilisus. Semantiliste tunnuste juures vaadeldakse nimelisust – päris- ja üldnimisid. Pärissõna- ja üldnimed rühmitatakse semantiliselt ning uuritakse, kas mingisse semantilisse rühma kuulumine võiks mõjutada nende sisseütleva vormi valikut. Viimase tunnusega vaadeldakse ka uuritava vormi juurde kuuluva verbi tähendust. Materjal pärineb Keeleveebis kättesaadavatest korpusdest. Meetodina kasutatakse ühetunnuselise analüüsi. Selgub, et vaadeldud tunnustest võib sisseütleva vormieelistuste uurimisel olulisteks pidada reksioonilisust, püsiühendilisust, nimelisust ning nii päris- kui ka üldnime semantilist rühma. Pigem pikka sisseütlevat kasutatakse reksioonistruktuurides ning pärisnimede puhul (isiku- ja kohanimed); pigem lühikest sisseütlevat püsiühendites ning üldnimelistes koha- ja seisundifraasides.

**Võtmesõnad:** morfoloogia, morfosüntaks, semantika, korpuslingvistika, keele varieerumine, illatiiv, aditiiv, eesti keel

## 1. Sissejuhatus

Kuigi eesmärgi pika ja lühikese sisseütleva varieerumist on käsitlenud varem mitmed uurijad (nt Sõnajalg 1956; Raag 1998; Hasselblatt 2000; Kio 2006; Kaalep 2009), pole siiani eesti keele grammatika kirjeldustes üksmeelt,

kuidas sisseütleva varieerumist selgitada – milliste kvantitatiivsete meetoditega eesti keele morfoloogilist varieerumist uurida ning mis tunnuseid käändsõnavormistiku varieerumise kirjeldamisel kindlalt arvesse võtta.

Oma eelmises artiklis uurisin ainsuse pika ja lühikese sisseütleva valiku olenemist morfofonoloogilistest tunnustest (Metslang 2015) ning siinne käsitlus on sellest välja kasvanud. Eelmise artikli meetod oli ühetunnuseline analüüs – hii-ruut-test ja standarditud Pearsoni jäägid. Analüüsiks kasutasin morfoloogiliselt ühestatud korpuse materjali, kuhu kuulus 188 pika ja 1522 lühikese sisseütleva vormi. Mõlemat vormi võtsin uurimismaterjali hulka proportsionaalselt ehk sellest lähtudes, kui palju neid korpuses tegelikult oli, mitte võrdset. Kokkuvõttes leidsin, et sõna morfofonoloogilistest tunnustest võivad sisseütleva valikut mõjutada astmehalduslikkus, astmehalduse liik, lõpuvahelduslikkus, lõpuvaheldusmall, algvormi lõpphäälik ja omastavatüve silpide arv.

Nende kahe artikli eesmärk on välja selgitada, mis põhimõttel on grammatilise alternatsiooni uurimiseks kõige otstarbekam andmeid koguda ning milliseid morfofonoloogilisi, morfosüntaktilisi ja semantilisi tunnuseid võtta arvesse edaspidises mitmetunnuselises analüüsis. Siinses käsitluses uuritakse suuremat hulka materjali, kontrollitakse varasemate uuringute tulemusi ja kasutatakse uudset analüüsimeetodit. Artiklis ei keskenduta sisseütleva käände ajaloole (vt nt Hasselblatt 2000; Kio 2006) ega tehta kokkuvõtteid varasematest uurimustest (vt nt Metslang 2015).

Selles käsitluses kasutatakse sama terminoloogiat mis morfofonoloogiliste tunnuste uurimisel: pikk sisseütlev on *illatiiv* ja lühike sisseütlev *aditiiv*. Samu termineid kasutati ka Keeleveebis, kust materjal pärineb.

Artikli eesmärke on välja selgitada, millistest morfosüntaktilistest ja semantilistest tunnustest oleneb ainsuse pika ja lühikese sisseütleva vormi valik. Veel soovitakse lahendada eelmise artikli materjali tasakaalustamatuse probleem (Japkowicz, Stephen 2002). Nagu mainitud, oli eelmise artikli valim proportsionaalne ehk tasakaalustamata. Kasutada võiks ka võrdseteks hulkadeks tasakaalustatud valimit või alternatsioonide suhte järgi tasakaalustatud valimit (Japkowicz, Stephen 2002: 440). Eelmise artikli valimi 188 illatiivi- ja 1522 aditiivivormi järgi võiks illatiivide ja aditiivide suhe olla vastavalt 1 : 8. Tahetakse ka välja selgitada, kas otstarbekam on koguda materjali korduvate vormide kaupa või korduvaid vorme arvestamata. Nende materjali kogumise põhimõtetega tegeletakse peatükis 2.

Tulemusi võrreldakse varasemate uurimustega, mille põhjal saab püstitada esialgsed hüpoteesid, et reksioonistruktuurid võiksid olla eelistatud illatiivis (EKK 2007: 247); püsiühendifraasid, üldnimelised kohad ning liikumis- ja olemisverbide laiendid aditiivis (EKG I: 56–57; Kio 2006: 112–113, 126; EKK 2007: 247). Meetodina kasutatakse ühetunnuselise analüüsi: hii-ruut-testi ja standarditud Pearsoni jääke. Nullhüpotees on, et kui võtta valimisse 50% illatiivivorme ja 50% aditiivivorme ning eraldada valimist mistahes sõltumatu tunnuse (näiteks sõnaliigi) alusel alamhulk, siis on selles alamhulgas illatiivide ja aditiivide jaotus samuti 50% ja 50% (vt Arppe 2008: 75). Juhul, kui alamhulga jaotus ei ole võrreldes tervikuga statistiliselt oluline, võib teha järelduse, et tunnus mõjutab sisseütleva vormi valikut. Sellist meetodit kasutatakse ja hüpoteesi kontrollitakse peatükis 5.9, kus tulemuste kontrolliks määratakse ka efekti suurus.

Mitme järjestikuse artikliga soovitakse üht eesti keele morfoloogia probleemi käsitledes arendada eesti keele morfoloogia uurimiseks sobivaid meetodeid. Võrreldakse materjali kogumise viise ning korpusanalüüsi kvantitatiivseid ja kvalitatiivseid võimalusi, püüdes anda panuse eesti keele morfoloogia uurimisse ja metoodika täpsustamisse.

## 2. Materjali kogumise põhimõtted

Materjali kogumiseks otsiti peaaegu kõigist Keeleveebi korpustest kõiki ainsuse sisseütleva vorme. Korpustest jäid välja vaid tasakaalus korpus, kuna selle sisu kajastub teistes korpustes, ja murdekorpus, mille materjali on liiga vähe, et võrrelda sisseütleva kasutust kirja- ja murdekeeles. Esmalt tehti Keeleveebi kaudu päringud korpusest etTenTen, mis on internetist alla laaditud eestikeelsete veebilehtede korpus. etTenTen sisaldab 270 miljonit tekstisõna. Lisaks tehti Keeleveebi kaudu päringud kõigist korpustest, mis kajastuvad ka eesti keele koondkorpuses. Eesti keele koondkorpus sisaldab u 240 miljonit tekstisõna. Seega tehti otsingud kokku u 510 miljoni tekstisõna seast.

Täpsustatud otsinguga leiti ainsuse illatiivi vormid (otsinguga *sg ill*) ja aditiivivormid (otsinguga *adt*) – praegu saab Keeleveebist otsida vaid kõiki ainsuse või mitmuse sisseütleva vorme korruga, mitte eraldi aditiivivorme (*sg ill* või *pl ill*). Täpsemate korpusotsingute piiratus teeb soovitud morfoloogilise materjali kogumise üsna keeruliseks. Teine võimalus on rakendada soovitud failide peal automaatset morfoloogilist märgendust.

Samas on automaatsel rakendusel vormihomonüümia tõttu sisseütlevat käänat raske ära tunda. Paljude selliste vormide kirjpilt kattub omastava- või osastavavormide omaga (nt nim *pass* : om *passi* : os *passi* : sisseü *passi*). Seega tuleb valimi kõik vormid nende kontekstis üle vaadata ja vajaduse korral käsitsi märgendada – see on aeganõudev töö, mis seab piirangud valimi mahule.

Illatiivvormidest jäeti alles vaid need, millega võiks kasutada paralleelselt lühikest sisseütlevat. Esmalt arvati varasemate tähelepanekute järgi teoreetiliselt varieeruvateks sisseütleva vormideks näiteks *Eesti* ja *Eestisse*, *Hiina* ja *Hiinasse*, *riide* ja *riidesse*, *alla* ja *alasse* ning kõik liitsõnad, kus *ala* on põhisõna (*looduskaitseala*, *vastutusala* jne). Selgus aga, et korpuse andmed kinnitavad kirjakeele normingu andmeid – korpuses ei leidu aditiivivorme *Eesti*, *Hiina*, *riide* ega *alla*. Nende negatiivsete näidete põhjal otsustati lähtuda ainult kirjakeele normingu järgi lubatud paralleelsusest, toetudes 2013. aasta ÕSile. Näiteks muuttüüpide puhul, kus „sobival juhul võib keelearvitaja lühivorme ka juurde moodustada“ (ÕS 2013: 20–22), peetakse lubatud paralleelsusega sõnadeks ainult neid, kus aditiivivorm on sõnaartiklis toodud.

Materjali kogumisel kerkis küsimus, kas kasutada binaarse grammatilise alternatsiooni uurimisel tasakaalustamata või tasakaalustatud valimit. Eelmises artiklis kasutatud tasakaalustamata valim põhjustas ootuspäraseid probleeme: kuna juba varasemate uuringute järgi kasutatakse eesti keeles tunduvalt rohkem aditiivi (vt nt Hasselblatt 2000), tuli ka korpusepäringu vastuseks aditiive rohkem kui illatiive, vastavalt 1522 ja 188. Viimaste vormide arv on aga statistiliste testide jaoks tihtipeale liiga väike – kui 188 vormi omakorda tasemetele jaotada, võib isegi juhtuda, et ühele tasemele jääb üks liige – näiteks üks laadivaheldusliku sõna illatiivivorm (Metslang 2015: 135). Väikese valimi korral on juba väike tasakaalustamatus probleem. Piisavalt suur andmestik kõrvaldaks tasakaalustamata valimi probleemi, kuid seejuures peab ajakulu olema mõistlik. (Japkowicz, Stephen 2002: 435–436) Nimelt on teoreetilise paralleelvõimalusega illatiivivorme korpusest üsna keeruline leida – suure osa illatiivivormide puhul pole varieerumine võimalik. Neid otsides tuleb aga märgendada ka aditiivivorme, näiteks et teada saada korpusvarieerumiste hulk. Nii kulub käsitsi märgendamisele väga palju aega. Suure tasakaalustamata valimiga korpuse loomine oleks väga ajamahukas, mistõttu tuleb proovida teist meetodit – valimi tasakaalustamist. Tasakaalustatud valimit on eelistanud ka

Jane Klavan (2012) oma doktoritöös, kus ta kasutab binaarse alternatsiooni uurimiseks Gaëtanelle Gilquini ja Stefan Griesi definitsioonile vastavat prototüüpset korpust. Üks selle korpuse omadusi on tasakaalustatus, n-ö teoreetiline ideaal. (Gilquin, Gries 2009: 6) Seega peaksid siinse materjali teoreetilise ideaali järgi eesti keeles olema illatiiv ja aditiiv võrdselt kasutusel ehk mõlema esinemistõenäosus võiks olla 50%.

Teine küsimus materjali kogumisel on, kas koguda materjali korduvad vormid materjali hulka jättes või iga vormi ainult üks kord arvestades, st kas näiteks vormi *kätte* võtta materjali hulka nii palju kordi, kui see esineb, või ainult üks kord. Korduvate vormide puhul tuleb arvestada, et sõna omadused mõjutavad vormivaliku eelistusi, iseäranis sagedaste sõnade puhul, mida võib olla valimis väga palju, näiteks *kätte*. Et ühe sõna omaduste mõju valimile minimeerida, valitakse igast paralleelvormist mõlemat vaid üks kord: esimesed 420 ettejuhtuvat illatiivi- ja 420 aditiivivormi. Valimis ei esine ükski sõnavorm kaks korda, nt kui vorm *majasse* on juba esinenud, siis teist korda *majasse* valimisse ei satu. Vorm *majja* läheb aga valimisse, juhul kui see korpuses esineb. Nii võib ühest sõnast olla valimis kõige rohkem kaks vormi: illatiiv ja aditiiv. Sagedastel sõnadel võiks nii valimisse jõuda mõlemad vormid, et kajastada nende võrdset tõenäosust. Kui üks vormidest on sage ja teine mitte, siis on suurem tõenäosus, et valimisse satub sagedasem vorm ja harvem võib-olla mitte. On selge, et valimisse on suurem tõenäosus sattuda sagedasematel sõnadel nende sagedamates vormides, aga et neid on vaid üks, siis ei suuda ühegi sõna individuaalsed omadused tervikut kuigivõrd kallutada.

### 3. Lõplik materjal ja meetod

Tasakaalustatud valimi põhimõtte järgi vaadeldi Keeleveebi korpustest 1095 teoreetiliselt varieeruvat illatiivi- ja 1095 aditiivivormi ehk kokku 2190 sisseütleva vormi. Selle materjali hulka on arvestatud ka kõik kordused. 2190 vormi hulgas oli 45 varieeruvat sõna, mida leidis nii aditiivis kui ka illatiivis. Sellised sõnad olid näiteks *ajalugu*, *bussipeatus*, *eelmine* (vt täielikku loetelu lisast). Kokku 45 sõna 90 varieeruva vormi puhul kasutati samasugust ühetunnuselise analüüsi nagu lõpliku materjali puhul, kuid ühegi vaadeldud tunnusega ei saanud statistiliselt olulisi tulemusi või ei saanud *p*-väärtust arvutada, kuna vaadeldavaid vorme on meetodi kasutamiseks liiga vähe. Nii ei saa esitada eraldi tulemusi, mis tunnustest

oleneb korpusvarieeruvate sõnade illatiivi või aditiivi eelistus. Seega on põhjendatud valik, mille järgi valimisse võetakse ka teoreetiliselt varieeruvaid, mitte ainult korpusvarieeruvaid sõnu. Ei saa väita, et kui mingit paralleelvõimalust korpusel ei leidu, pole seda olemas – ükskõik kui suur korpus annab infot ainult seal leiduva kohta (vt nt Atkins, Levin 1995: 87, 108).

Kordusi arvestades oli esialgses materjalis 1095 illatiivi- ja 1095 aditiivivormi. Sellest materjalist arvestati iga vormi ainult üks kord seni, kuni materjal hakkas ennast ammendama, st edasisel otsimisel kahanes lisanduvate uute enne esinemata vormide hulk sedavõrd, et töömahtu arvestades ei olnud enam mõistlik valimi suurendamist jätkata – näiteks 800 sisseütleva vormi läbivaatamisel leiti viis uut teoreetiliselt varieeruvat illatiivivormi. Selle tulemusel peaks valimist välja jääma äärmiselt väikese sagedusega sõnade illatiivi- ja aditiivivormid, mis moodustavad korpusel sisalduvatest illatiivi- ja aditiivivormidest väga väikese osa. Äärmiselt väikese esinemissageduse tõttu on nende mõju uuritavatele tendentsidele eeldatavasti kaduvväike. Lõpliku valimi suurusks saadi 420 illatiivi- ja 420 aditiivivormi. Nagu mainitud, käsitletakse siinse materjalivaliku põhimõtte järgi iga vormi edaspidises analüüsis ainult üks kord. Uuritakse mittekorduvaid vorme, st nt vorm *majasse* võis algse 1095 illatiivi hulgas esineda mitu korda ning lõpliku 420 hulka pääses see vorm ainult korra. Seejuures ei sattunud lõpliku materjali hulka kõik 45 varieeruvat sõna, vaid 41 neist. Kokku saadi 840 vormi.

Töövahendina kasutatakse statistikaprogrammi R ja meetodina ühetunnuselise analüüsi. See on hea esmaste tulemuste väljaselgitamiseks, et teada saada, milliseid tunnuseid arvestada põhjalikumal mitmetunnuselises analüüsis. Uuritav tunnus on sisseütleva vormi valik: illatiiv või aditiiv. Seletavad tunnused on sõnaliik, lauseliige, rektsoonilisus, püsiühendilisus, nimelisus, pärisnime semantiline rühm, üldnime semantiline rühm ja verbi tähendus. Iga tunnus jaguneb 2–5 tasemeks. Neli tunnust on binaarsed, neljal on rohkem tasemeid. Kõik tunnused on nominaalsed, st nende tasemetel (nt sõnaliikide) vaheline kaugus pole arvuliselt mõõdetav. Seletava ja uuritava tunnuse vahelise seose olulisuse uurimiseks kasutatakse hii-ruut-testi. Iga seletava tunnuse alapeatükis on seletava ja uuritava tunnuse kohta tabel, kus on uuritavad sagedused ja sulgudes protsendid (mitu protsenti aditiividest või illatiividest kuulub mingile tasemele). Tabeli all tuuakse hii-ruut-testi tulemused. Hii-ruut-statistik  $X^2$  näitab, kas sisseütleva vormi

valiku ja vaadeldava tunnuse vahel on seos olemas, määramata sealjuures seose tugevust. Hii-ruut-statistiku järel on sulgudes vabadusastmete arv ja valimi suurus. Lõpus on statistilist olulisust kirjeldav  $p$ -väärtus. Hii-ruut-testi tulemus on statistiliselt oluline, kui  $p < 0,05$ . Selle järgi on võimalus, et tulemus oli juhuslik, väiksem kui 5%. Kui tunnus osutus statistiliselt oluliseks, on alapeatükis ka standarditud Pearsoni jääkide tabel, mis esitab tunnustevahelise seose tugevuse ja suuna. Statistiliselt oluliste tulemustena arvestatakse siinses artiklis standarditud Pearsoni jääkide tulemusi alates absoluutväärtusest 2 (Agregsti 2013: 81). Analüüsi lõpus (ptk 5.9) kontrollitakse hii-ruut-testi ja standarditud Pearsoni jääkide tulemusi veel  $n$ -ö osa ja terviku meetodiga ning efekti suurust hindava Craméri seosekordajaga  $V$ . Oma väitekirjas on grammatilise alternatsiooni uurimisel ühe meetodina ühetunnuselise analüüsi kasutanud ka nt Antti Arppe (2008) ja Jane Klavan (2012).

#### 4. Seletavad tunnused

Materjali uuritakse kaheksa tunnuse põhjal. EKG (I: 56) järgi oleneb sisseütleva vormi valik sõna fonoloogilis-derivatiivsest struktuurist, aga ka näiteks rektsioonistruktuurist ja leksikaalsest tähendusest. Sisseütleva valik võiks oleneda ka konkreetse verbiga koos esinemisest või püsiühendi koosseisu kuulumisest (nt Viitso 1976: 152; Kio 2006: 126; EKK 2007: 247). Seega oleneb grammatika kirjelduste järgi sisseütleva vormi valik peale morfofonoloogiliste tunnuste ka morfosüntaktilistest ja semantilistest tunnustest. Tabelis 1 esitatakse kaheksa seletavat tunnust, mille all kirjeldatakse neid kõiki lähemalt.

**Tabel 1.** Seletavad tunnused

<b>Morfosüntaktilised tunnused</b>	<b>Semantilised tunnused</b>
Sõnaliik	Nimelisuus
Lauseliige	Pärisnime semantiline rühm
Rektsioonilisuus	Üldnime semantiline rühm
Püsiühendilisuus	Verbi tähendus

## Morfosüntaktilised tunnused

**Sõnaliik.** EKSSi (2009) abil määratakse iga analüüsitava vormi sõnaliik. Tasemed on käändsõna liigid EKG (I: 18) järgi: 'arvsõna', 'asesõna', 'nimisõna' ja 'omadussõna'.

**Lauseliige.** EKG järgi on eesti keele lauseliikmed öeldis, alus, sihitis, öeldistäide, määrus ja täiend. Sisseütlevas käändes sõnad võivad esineda määruse või täiendina. (EKG II: 9–11) Niisiis on selle tunnuse üks tase 'määrus' ja teine 'täiend'. Täiendite puhul vaadati ka analoogiat, et uurida, kas laiend võib mõjutada põhisõna vormieelistust või vastupidi.

**Rektsioonilisus.** Järgmise kahe tunnuse tasemed on 'ei' ja 'jah'. Rektsioonilisuse üle otsustatakse selle järgi, kas uuritav vorm on verbi rektsiooniline laiend. Aluseks võetakse Sirje Mäearu (2011) loend „Valik rektsioone“, kus on nt levinumad (millessegi) *puutuma, suhtuma, uskuma*. Kui loetelus on *suhe* ja *suhtuma*, arvestatakse rektsioonistruktuuriks ka (millessegi) *suhtuja, suhtumine* jne.

**Püsiühendilisus.** Rektsioonilisus määratakse ühe sõna järgi, mille juurde sisseütleva käände vorm kuulub, kuid püsiühendilisus terve fraasi järgi – eesti keele verbikesksete püsiühendite andmebaasi põhjal. Andmebaasi tutvustuses selgitatakse, et silmas peetakse „verbi ja tema seotud laiendi püsivaid ühendeid – *jalga laskma, läbi saama, kõnet pidama*“. Sisseütlevas käändes laiendiga fraasidest on püsiühendite andmebaasis nt *jõusse jääma, põhja kõrbema, riidesse panema* jne.

## Semantilised tunnused

**Nimelisus.** Toimetatud tekstides on nimelisust hõlbus määrata sõnaalgulise suurtähe järgi, nt kui *Keskus* või *Riigikogu* on suure tähega, kuuluvad need tasemele 'pärisnimi' (väikesetähelised *keskus* ja *riigikogu* mitte). Toimetamata tekstide puhul loetakse pärisnimedeks ka nt *kopli, rate*. Teine tase on 'üldnimi'.

**Pärisnime semantiline rühm.** Siin vaadeldakse eraldi eelmise tunnuse pärisnimesid ja üldnimesid. Mõlemad jaotatakse semantilisteks rühmadeks. Pärisnimed jaotatakse tasemetele 'INSTITUTSIOON' (suurtähelised *Klubi, Gümnaasium, Riigikogu* jne), 'ISIK' (kõik suurtähelised isikunimed, sh *Jesus Kristus*) ja 'KOHT' (kõik suurtähelised kohanimed, aga ka vähesed kodulehed ja raamatud), neljandale tasemele jääb 'üldnimi'.



**Üldnime semantiline rühm.** Üldnimede semantiliste rühmade tase-  
med on 'KEHAOSA' (nt *kätte*, aga ka mentaalne *meelde* ja üldisem *kehasse*),  
'KOHT', kuhu koondatakse nii realselt olemasolevatele kohtadele osutavad  
nimisõnad, nt *kodu*, *muuseum*, *põuetasku*, kui ka sihtkoha tähenduses  
abstraktsematele kohtadele osutavad sõnad, nt *keelde*, *valdkonda*, *üksu-  
sesse*. Tase 'SEISUND' hõlmab kehalisi (nt *riidesse panema*) ja vaimseid (nt  
*jõusse jääma*, *toime tulema*) seisundeid, mis saavad oma täieliku tähenduse  
alles fraasi koosseisus. Sellise jaotuse puhul on mõningane subjektiivsus  
vältimatu, kuna abstraktset kohta ja seisundit on raske eristada, nt kas *rak-  
kesse asuma* või *vangi mõistma* on seisundid või kohad. Tasemele 'muu'  
kuuluvad rektsioonistruktuurid (nt *millesegi puutuma*, *suhtuma*), täiendid,  
ajamäärangud (nt *perioodi*, *praegusesse*) jm. Viimane tase on 'pärisnimi'.

**Verbi tähendus.** Verbid jagatakse tasemetele 'liikumisverb', 'olemis-  
verb', 'tegevusverb' ja 'ilma verbita'. 'Liikumisverb' tähendab, et toimub  
mingi ruumis ümberpaiknemine (liikumine) või kirjeldatakse liikumise  
viisi, nt *jõudma*, *lahkuma*, *sõitma*. 'Olemisverb' on staatilisem, nt *jääma*,  
*kuuluma*, *olema*. Tasemele 'tegevusverb' kuuluvad kõik aktiivsed, aga  
ka mentaalsed tegevused, nt *lisama*, *suhtuma*, *töötama*, millel on tead-  
likult toimiv, oma tegevust kontrolliv osaline. 'Ilma verbita' tähendab,  
et analüüsitav vorm ei kuulu ühegi verbi juurde. Sama jaotust on kasu-  
tanud nt Klavan (2012), kelle töös on ka 'asendiverb', mille näidetest  
leidub valimis *asuma*. See verb väljendab siinses materjalis aga aktiivset  
tegevust – *rakkesse asuma*, *õppima asuma*, *võitlusse asuma* –, nii et see  
arvatakse tegevusverbiks.

## 5. Materjali analüüs

### 5.1. Sõnaliik

Esimene morfosüntaktiline tunnus on sõnaliik. Tabelist 2 nähtub, et mater-  
jalis on enim nimisõnu, tunduvalt vähem on arv-, ase- ja omadussõnu.

**Tabel 2.** Sõnaliik ja ainsuse sisseütlev

Sisseütlev/Sõnaliik	Arvsõna	Asesõna	Nimisõna	Omadussõna	Kokku
Aditiiv	4 (50%)	2 (25%)	389 (50%)	25 (57%)	420
Illatiiv	4 (50%)	6 (75%)	391 (50%)	19 (43%)	420
<b>Kokku</b>	<b>8 (100%)</b>	<b>8 (100%)</b>	<b>780 (100%)</b>	<b>44 (100%)</b>	<b>840</b>

Hii-ruut-testi  $p$ -väärtuse arvutamiseks on mõnel tasemel liiga vähe liikmeid. Seetõttu ei saa esitada ka tulemusi standarditud Pearsoni jääkide kohta.

Protsentide järgi selgub aga, et asesõna puhul kasutati rohkem illatiivi (75%) ja omadussõna puhul aditiivi (57%). Selle põhjal on seos asesõna ja illatiivi (*kõigesse, mõnesse, sellesesamasse*) ning omadussõna ja aditiivi vahel (*alatisse, alumisse, eelmisse*). Arvsõna puhul kasutati aditiivi ja illatiivi sama palju, nii ei ole arvsõnal tugevat seost aditiivi ega illatiivi valikuga. Enamik vaadeldud sõnadest on nimisõnad ning on huvitav, et esinemuse järgi on nimisõnalisi illatiivivorme materjalis kahe võrra rohkem kui aditiivivorme. Tuleb aga tähele panna, et protsentide järgi on nimisõnadel aditiivi- ja illatiivivorme võrdselt.

## 5.2. Lauseliige

Sisseütleva käände vormid võivad lauses olla määrused või täiendid, viimaste all peetakse silmas eestäiendeid, nt *ajaloolisesse (fakti), alumisse (asendisse), arhailisesse (maaõhkkonda)*. Ootuspäraselt on materjali hulgas määrusi tunduvalt rohkem kui täiendeid.

**Tabel 3.** Lauseliige ja ainsuse sisseütlev

Sisseütlev/Lauseliige	Määrus	Täiend	Kokku
Aditiiv	396 (50%)	24 (53%)	420
Illatiiv	399 (50%)	21 (47%)	420
<b>Kokku</b>	<b>795 (100%)</b>	<b>45 (100%)</b>	<b>840</b>

Tulemus ei ole siiski statistiliselt oluline:  $\chi^2(1, N = 840) = 0,21, p = 0,65$  ja standarditud Pearsoni jäägid jäävad alla absoluutväärtuse 2. Protsentide järgi võiks täiendi puhul olla pisut eelistatum aditiiv (53%), nt *alumisse (asendisse), atlantilisse (kliimaperioodi), eelmisse (korterisse)*. Määruse puhul eelistus ei ilmne.

Täiendite puhul vaadati ka analoogiat – kas laiend mõjutab põhisõna või vastupidi. Uuriti, kas illatiivis või aditiivis laiendi eeskujul eelistatakse käänata ka põhisõna vastavalt illatiivis või aditiivis. Või vastupidi – kas põhisõna vormi eeskujul eelistatakse laiendi käändena illatiivi või aditiivi. 45 täiendist jäeti kõrvale need, mis kuulusid ilma paralleelvormita põhisõna juurde, nt *kolmetoalisesse (korterisse), mõnda (saatesse), uude*

(*lainesse*) – *korter*, *saade* ja *laine* ei varieeru. Alles jäi 29 täiendit: neist 13 puhul oli laiendi ja põhisõna vorm erinev, nt *endisesse* (*NSV Liitu*), *järgmisesse* (*paika*), *pisikesesse* (*vaatesilma*). 16 puhul olid laiend ja põhisõna samas vormis, nt *nimelisse* (*linnaossa*), *pisukesse* (*maailma*), *täielikku* (*valgustusse*). Selgub, et siinse materjali järgi on täiend ja põhisõna pigem samas vormis ning seega võiksid teineteist mõjutada. Siiski olid kõigil analoogiajuhtudel nii laiend kui ka põhisõna aditiivis ning on liiga vähe näiteid uurimaks, kas analoogial on mõju sisseütleva vormi valikule ja kas laiend mõjutab põhisõna või vastupidi. Edaspidi tuleks eraldi uurida sisseütleva vormianaloogiat, st kas laiendit ja põhisõna kasutatakse ühtmoodi illatiivis või aditiivis. Tuleks selgitada, kas laiend ja põhisõna mõjutavad teineteist – kui jah, siis kumb kumba mõjutab.

Sõnaliigi ja lauseliikme kohta selgub, et statistiliste meetodite põhjal ei ole need tunnused olulised. Põhjus võib olla selles, et nendel tunnustel on väga suured tasemed 'nimisõna' ja 'määrus'. Nendele tasemetele koonduvad enamik vaadeldavaid vorme ning suurimal tasemel on mõlemal juhul protsentide järgi pooled vormid illatiivis ja pooled aditiivis. Järgmistes alapeatükkides uuritakse tunnuseid, mida on sisseütleva vormi valiku kirjeldamisel ka varem välja toodud: rektsioonilisust (EKK 2007: 247) ja püsiühendilisust (Kio 2006: 126).

### 5.3. Rektsioonilisus

Rektsioonilisus on olemas, kui uuritav vorm on verbi rektsiooniline laiend, nt (millessegi) *investeerima*, *puutuma*, *suhtuma*. Rektsioonilisi laiendeid otsiti 45 sõnaga (Mäearu 2011), millest enamik olid verbid, kuid ka nendega moodustuslikult seotud nimisõnad (nt *uskuma* ja *usk*).

**Tabel 4.** Rektsioonilisus ja ainsuse sisseütlev

Sisseütlev/Rektsioonilisus	Ei	Jah	Kokku
Aditiiv	388 (55%)	32 (25%)	420
Illatiiv	323 (45%)	97 (75%)	420
<b>Kokku</b>	<b>711 (100%)</b>	<b>129 (100%)</b>	<b>840</b>

Tulemus on statistiliselt oluline:  $\chi^2(1, N = 840) = 38,7, p < 0,001$ . Tabeli 4 protsentide põhjal selgub, et mitterektsioonilised laiendid on pigem aditiivis (55%) ja rektsioonilised laiendid illatiivis (75%). Ka standarditud

Pearsoni jäägid (tabel 5) näitavad, et reksioonilised laiendid on sagedamini illatiivis, nt (*armunud*) 32aastasesse, (*siseneda*) *administreerimiskeskusse*, (*suhtumist kodumaisesse*) *alkoholitööstusesse*.

**Tabel 5.** Standarditud Pearsoni jäägid tabeli 4 kohta

Reksioonilisus	Aditiiv	Illatiiv
Ei	6,22	-6,22
Jah	-6,22	6,22

Nii standarditud Pearsoni jääkide kui ka esinemissageduse järgi saab väita, et reksiooniliste laiendite puhul on tõenäolisem illatiivi kasutamine. Varem on käsitletud eraldi nt ainult suhtumist väljendavate verbide reksioonilisi laiendeid või üksikute verbide reksioone. EKG (I: 56) järgi võib suhtumist väljendavates reksioonistruktuurides eelistatum olla aditiiv. Tiit-Rein Viitso (1976: 152) nimetab aga verbe *segunema*, *suhtuma*, mis peaksid tingima illatiivi kasutuse, ning *kalduma*, *kukkuma*, *minema*, *tulema* jt, mida kasutatakse eelistatult aditiiviga. EKK (2007: 247) järgi peaksid aga verbide *suhtuma*, *kiinduma*, *armuma* jt reksioonilised laiendid olema eelistatult illatiivis. EKK väite lükkab ümber Heiki-Jaan Kaalep (2009: 416), leides korpusest nende verbide reksioonilisi laiendeid lühikeses sisseütlevas. Ka Kati Kio (2006: 87) kinnitab, et nt ilmeka näite *suhtuma* puhul ei saa kindlalt väita ühe või teise vormi eelistamist. Kuna siinses materjalis ei vaadata eraldi hinnanguverbe verbe ega üksikuid verbe, on tulemusi raske võrrelda. Siinse analüüsi tulemuste põhjal on reksioonilise laiendi käändena eelistatud illatiivi, kuid on selge, et kõik reksioonilised laiendid ei käitu ühtmoodi ja tegu on vaid tendentsiga.

#### 5.4. Püsiühendilisus

Püsiühendilisus on olemas, kui uuritav vorm kuulub mõne püsiühendi koosseisu. Püsiühendisse kuulumine otsustatakse eesti keele verbikesksete püsiühendite andmebaasi järgi. Nagu arvata, leidub püsiühendeid vähem kui mittepüsiühendeid.

**Tabel 6.** Püsiühendilisuus ja ainsuse sisseütlev

Sisseütlev/Püsiühendilisuus	Ei	Jah	Kokku
Aditiiv	381 (48%)	39 (95%)	420
Illatiiv	418 (52%)	2 (5%)	420
<b>Kokku</b>	<b>799 (100%)</b>	<b>41 (100%)</b>	<b>840</b>

Hii-ruut-testi põhjal on tulemus statistiliselt oluline:  $\chi^2(1, N = 840) = 35,1, p < 0,001$ . Protsentide põhjal on näha seost mittepüsiühendite ja illatiivi (52%) ning püsiühendite ja aditiivi (95%) vahel. Tabelis 7 toodud standarditud Pearsoni jääkide järgi on samuti seos mittepüsiühendite ja illatiivi ning püsiühendite ja aditiivi vahel.

**Tabel 7.** Standarditud Pearsoni jäägid tabeli 6 kohta

Püsiühendilisuus	Aditiiv	Illatiiv
Ei	-5,92	5,92
Jah	5,92	-5,92

Tabelitest saab järeldada, et püsiühendite puhul on tõenäolisem aditiivi kasutamine, nt *hätta (jääma)*, *joonde (ajama)*, *kaela (panema)*. Selline tulemus pole kuigi üllatav, sest materjalis on ainult kaks juhtu, mil püsiühendi koosseisus kasutati illatiivivormi.

Pearsoni jääkidest selgunud seost kinnitab ka EKG, mis nimetab aditiivi mõne sõna korral konkreetse kohatähenduses ainuvõimalikuks, ning sellised kohaväljendid on kujunenud n-ö püsiühendiks. Sellisel juhul on sisseütleva vormid juba adverbistunud – vrd (*pani pähe*, (*võttis kätte*, EKG esitab püsiühendina ka (*kukkus jökke*. (EKG I: 56–57) Samas näiteks EKSSi määratluse järgi loetakse *pähe* ja *kätte* adverbideks, kuid *jökke* mitte.

Ka Kio (2006: 126) väidab oma materjali põhjal, et aditiiv n-ö kivineb aja jooksul ja kaotab oma põhitähendust. Siinses materjalis on püsiühendites vähem adverbistunud vorme, mis on säilitanud oma algse tähenduse, nt *hätta (jääma)*, *koju (kutsuma)*, *pankrotti (minema)*. Nendel juhtudel tundub, et *hätta*, *koju*, *pankrotti* pole täielikult adverbistunud, need on sisseütleva vormid algvormidest *hāda*, *kodu* ja *pankrot*. Siia sobiks ka EKG näide (*kukkus jökke*. Kuna tulemuste põhjal on püsiühendites eelistatum aditiiv, tekib küsimus, kas sellistes ühendites on võimalik vaba varieerumine – *hādasse jääma*, *kodusse kutsuma*, *pankrotisse minema*. Paari näite

põhjal on raske väita, kas neid vorme kasutatakse eelistatavalt aditiivis struktuuritunnuste pärast või püsiühendisse kuulumise tõttu. Siinsete tulemuste põhjal oleks näiteks rektsioonilised laiendid eelistatavalt illatiivis ja püsiühendi koosseisus laiendid aditiivis. Kui vormide *hätta*, *koju* ja *pankrotti* puhul võib püsiühendilisuse üle vaielda, siis selgem on olukord adverbistunud vormidega *kätte*, *pähe*, *silma* jne. Viimaste sõnadega ei leidu ühtegi illatiivivormi. Nende näidete puhul võib kehaosasõnade tähendus vormi mõjutada – näiteks arvukategooria valimisel on tähendus oluline tunnus (K. Õim, A. Õim 2015: 97–98). Tundub, et kehaosasõnad on sisseütleva käände vormis muude semantiliste rühmade sõnadest rohkem adverbistunud või n-ö vormilt kivinenud.

Just sagedased kehaosasõnad võiksid olla põhjus, miks rektsioonides eelistatakse illatiivi ja püsiühendites aditiivi (27 kehaosasõnast 8 on püsiühendis, kuid mitte ühtegi rektsiooni koosseisus, vt ptk 5.7). Püsiühendid on fraasid, nagu *kätte saama*, *pähe panema*, *silma hakkama* jne. Kehaosasõnu kasutatakse sellistes fraasides aditiivis ja seega ongi püsiühendites eelistatum aditiiv. Rektsioonilisust ei määra aga kogu fraas, vaid ainult verb, nt verbide *puutuma*, *suhtuma*, *uskuma* (millesegi) puhul eelistati illatiivivorme.

## 5.5. Nimeliskus

Kõik käsitletud vormid on jaotatud päris- ja üldnimedeks.

**Tabel 8.** Nimeliskus ja ainsuse sisseütlev

Sisseütlev/Nimeliskus	Pärisnimi	Üldnimi	Kokku
Aditiiv	43 (36%)	377 (52%)	420
Illatiiv	77 (64%)	343 (48%)	420
<b>Kokku</b>	<b>120 (100%)</b>	<b>720 (100%)</b>	<b>840</b>

Hii-ruut-testi järgi on tulemus statistiliselt oluline:  $\chi^2(1, N = 840) = 11,24$ ,  $p < 0,001$ . Protsentide põhjal on seos pärisnimede ja illatiivi (64%) ning üldnimede ja aditiivi (52%) vahel. Ka standarditud Pearsoni jäägid (tabel 9) näitavad, et üldnimede sisseütleva käände vormina kasutatakse tõenäoliselt aditiivi, nt *ajajärku*, *ajakirjandusse*, *ajavahemikku*, ja pärisnimedel illatiivi, nt *Amblasse*, *Ateenasse*, *Audentesesse*.

**Tabel 9.** Standarditud Pearsoni jäägid tabeli 8 kohta

Nimelisus	Aditiiv	Illatiiv
Pärisnimi	-3,35	3,35
Üldnimi	3,35	-3,35

Materjalist paistab, et keekekasutajad on üksmeelsed, kas kasutada mõnd pärisnime eelistatult aditiivis, nt *Berliini*, *Dakari*, *Haapsallu*, või illatiivis, nt *Aakresse*, *Amblasse*, aga ka *Eestisse*, *Hiinasse*. Seejuures on nimesid, mida võidakse kasutada nii aditiivis kui ka illatiivis, siinsest materjalist nt *Soome* ja *Soomesse*.

'Üldnimi' on väga paljude esinemisjuhtudega tase, nii võiks arvata, et aditiivi kasutamine on üldisem. Siiski on ainuüksi nimelisuse järgi raske määrata, kas eelistatakse illatiivi või aditiivi. Seetõttu vaadatakse seda taset üksikasjalikumalt, täpsemaks analüüsiks on päris- ja üldnimed semantiliselt rühmitatud.

### 5.6. Pärisnime semantiline rühm

Pärisnimesid on kokku 120 ning need jaotatakse omakorda kolmele tasemele, üldnimed jäävad neljandale.

**Tabel 10.** Pärisnime semantiline rühm ja ainsuse sisseütlev

Sisseütlev/ Pärisnime semantiline rühm	INSTITUTSIOON	ISIK	KOHT	Üldnimi	Kokku
Aditiiv	15 (54%)	0 (0%)	28 (33%)	377 (52%)	420
Illatiiv	13 (46%)	7 (100%)	57 (67%)	343 (48%)	420
<b>Kokku</b>	<b>28 (100%)</b>	<b>7 (100%)</b>	<b>85 (100%)</b>	<b>720 (100%)</b>	<b>840</b>

Hii-ruut-testi tulemusi ei saa arvutada, kuna paljudel tasemetel on vähe liikmeid. Niisiis saab oletusi teha vaid protsentide põhjal. Rühma 'INSTITUTSIOON' kuuluvate sõnade puhul on eelistatum aditiiv (54%), nt *EBSi*, *Euroliitu*, *Keskerakonda*. Siiski on aditiivivormis institutsioone vaid kahe võrra rohkem, nii et see pole piisav järelduste tegemiseks. Rühma 'ISIK' kuuluvaid sõnu ei esine aditiivis üldse ning vähesed on illatiivis, nt *Jeesusesse*, *Kristusesse*, *Mozartisse*. Rühma 'KOHT' sõnad on pigem

illatiivis (67%), nt *Audrusse, Egiptusesse, Elvasse*. Sellised tulemused on kooskõlas eelmise pärisnimede alapeatükiga, kus pärisnimed olid eelistatult illatiivis. Enamik pärisnimedest on kohanimed ja seega on ka need eelistatult illatiivis.

## 5.7. Üldnime semantiline rühm

Ka 720 üldnime jaotatakse omakorda tasemetele.

**Tabel 11.** Üldnime semantiline rühm ja ainsuse sisseütlev

Sisseütlev/ Üldnime semantiline rühm	KEHAOSA	KOHT	SEISUND	Muu	Pärisnimi	Kokku
Aditiiv	17 (63%)	258 (56%)	41 (67%)	61 (55%)	43 (36%)	420
Illatiiv	10 (37%)	202 (44%)	20 (33%)	111 (65%)	77 (64%)	420
<b>Kokku</b>	<b>27</b> <b>(100%)</b>	<b>460</b> <b>(100%)</b>	<b>61</b> <b>(100%)</b>	<b>172</b> <b>(100%)</b>	<b>120</b> <b>(100%)</b>	<b>840</b>

Hii-ruut-testi tulemused näitavad, et tulemus on statistiliselt oluline:  $\chi^2(4, N = 840) = 40,03, p < 0,001$ . Protsentide põhjal on rühma 'KEHAOSA' kuuluvad sõnad pigem aditiivis (63%), samuti rühmadesse 'KOHT' (56%) ja 'SEISUND' (67%) kuuluvad sõnad. Vormid rühmast 'muu' (65%) on eelistatult illatiivis. Standarditud Pearsoni jääkide järgi (tabel 12) on rühmade 'KOHT' ja 'SEISUND' sõnade puhul tõenäolisem aditiivi kasutamine. Rühma 'KEHAOSA' puhul statistiliselt olulist seost ei selgu ning rühm 'muu' on eelistatult illatiivis, aga liiga hägune, et järeldusi teha.

**Tabel 12.** Standarditud Pearsoni jäägid tabeli 11 kohta

Sisseütlev/Semantiline silt	Aditiiv	Illatiiv
KEHAOSA	1,37	-1,37
KOHT	3,88	-3,88
SEISUND	2,79	-2,79
Muu	-4,28	4,28
Pärisnimi	-3,35	3,35



Tase 'КОИТ' hõlmab nii konkreetseid kui ka abstraktseid kohti. EKK (2007: 247) põhjal on sihtkohad peaaegu ainuvõimalikult aditiivis, nt *läks parki* (mitte *pargi/sse*), *kukkus vette* (mitte *vee/sse*), *tõlkis eesti keel/de* (mitte *keele/sse*). EKG (I: 56) järgi on paljude sõnade korral konkreetse koha tähenduses võimalik nii pikk kui ka lühike sisseütlev – vrd (*läks muuseumi* ~ *muuseumisse*). Standarditud Pearsoni jäägid kinnitavad EKK väidet, mille järgi kohad on eelistatult aditiivis, nt *adventkogudusse*, *aeda*, *boksi*. Ka seisundifraasid eelistavad aditiivi, nt *põhja* (*kõrbema*), *vabadusse* (*jõudma*), *vastuollu* (*minema*). Seejuures on enamik püsiühendeid seisundifraasid: 41 püsiühendist on 23 seisundifraasid. Nii on ootuspärane, et nii püsiühendites kui ka seisundifraasides kasutatakse rohkem aditiivi. Nagu mainitud, on tase 'muu' järelduste tegemiseks liiga üldine. Taseme 'pärisnimi' seos illatiiviga kinnitab nimelisuse juures saadud tulemusi, et pärisnimede puhul kasutatakse rohkem illatiivi ja üldnimede puhul aditiivi.

Kokkuvõtteks ilmnevad selged semantilised rühmad, millel on olemas seos sisseütleva vormi valikuga. Kõige konkreetsem ja suurima liikmete arvuga on rühm 'КОИТ', mille puhul on tõenäolisem aditiivi kasutamine. Samamoodi on aditiivi kasutamine tõenäolisem rühma 'SEISUND' puhul, kuigi sinna kuuluvad väga erinevad fraasid.

## 5.8. Verbi tähendus

Kõige rohkem leidub vorme koos tegevusverbiga, vähem liikumis- või olemisverbiga. Kõige vähem on vorme, mis ei kuulu ühegi verbi juurde.

**Tabel 13.** Verbi tähendus ja ainsuse sisseütlev

Sisseütlev/ Verbi tähendus	Liikumis- verb	Olemis- verb	Tegevus- verb	Ilma verbita	Kokku
Aditiiv	110 (49%)	52 (53%)	242 (50%)	16 (44%)	420
Illatiiv	115 (51%)	46 (47%)	239 (50%)	20 (56%)	420
<b>Kokku</b>	<b>225 (100%)</b>	<b>98 (100%)</b>	<b>481 (100%)</b>	<b>36 (100%)</b>	<b>840</b>

See tulemus ei ole statistiliselt oluline:  $\chi^2(3, N = 840) = 0,94, p = 0,82$ . Protsentide järgi esinevad koos liikumisverbidega eelistatult illatiivivormid (51%) ja olemisverbidega aditiivivormid (53%). Ilma verbita vormidega aga jällegi illatiivivormid (56%), kuigi neid on ainult nelja võrra rohkem.

Protsentide järgi ei saa öelda, kas tegevusverbidega eelistatakse aditiivi- või illatiivivorme.

Varem on verbide semantiliste rühmade järgi sisseütleva eelistusi põhjendanud Kio (2006). Tema käsitluses on verbide semantilisi rühmi tunduvalt rohkem – suuremad rühmad on sihilised ja sihitud verbid. Sihilised on protsessiverbid, kausatiivid, tunnetusverbid, hinnanguverbid, suhtlusverbid ja valdamisverbid. Sihitud verbid on protsessiverbid, liikumisverbid ja hinnanguverbid. Siinne meetod aga nii palju tasemeid ei võimalda ja ehk ei ole nii üksikasjalik jaotus ka otstarbekas. Kio järgi on nt *jõudma*, *saabuma*, *sattuma* protsessiverbid, siinses jaotuses aga liikumisverbid. Kio (2006: 113) järgi kasutatakse liikumisverbidega pigem lühikest sisseütlevat, siinses materjalis aga pikka – eelistus toetub siiski vaid viiele vormile. Mõõndustega võiks võrrelda Kio protsessiverbe ja sinseid olemisverbe. Enamik Kio (2006: 112–113) jaotuse protsessiverbidest on sisseütleva valiku suhtes neutraalsed, kuus soosivad lühikest sisseütlevat. Ka siinses materjalis eelistavad olemisverbid pigem aditiivivormi.

## 5.9. Tulemuste kontrollimine

Hii-ruut-testi ja standarditud Pearsoni jääkidega saadud tulemuste kontrollimiseks rakendati ka n-ö osa ja terviku meetodit. Kuna valimis on uuritava nähtuse ehk sisseütleva vormide jaotus 50% illatiive ja 50% aditiive, siis nullhüpoteesi kohaselt peaks selle valimi mistahes alamhulgas olema illatiivi- ja aditiivivormide jaotus samuti võrdne ehk 50% ja 50% (vt Arppe 2008: 75). Uurimaks, kas see on nii, vaadeldi eraldi iga tunnuse taset ja võrreldi seda võrdjaotuselise tervikuga (420 illatiivi- ja 420 aditiivivormi). Näiteks esimese tunnuse, sõnaliigi esimese taseme 'arvsõna' puhul võrreldi uuritava taseme jaotust (4 illatiivi- ja 4 aditiivivormi) jaotusega kogu valimis (420 illatiivi- ja 420 aditiivivormi). Ootuspäraselt saadi *p*-väärtuseks 1 ehk erinevus pole statistiliselt oluline. Kui mõne taseme puhul on erinevus statistiliselt oluline, tähendab see, et üks sisseütleva vorm on eelistatum.

Tulemused kinnitasid paljuski varasema meetodiga saaduid: sõnaliigi, lauseliikme ja verbi tähenduse tunnuse puhul polnud ühegi taseme ja terviku võrdlus statistiliselt oluline. Ka üldnime semantilise rühma puhul saadi sama tulemus: üldnimelistes koha- ja seisundifraasides eelistatakse aditiivi, rühma 'muu' ja pärisnimede puhul illatiivi. Rektsioonilisuse vaatlusel leiti, et rektsioonide jaotuse (97 illatiivi- ja 32 aditiivivormi) erinevus

terviku jaotusest on statistiliselt oluline, kuid mitterektsioonide jaotuse erinevus mitte. See tähendab, et rektsioonide puhul eelistatakse illatiivi, kuid ei saa väita, et mitterektsioonide puhul eelistatakse aditiivi, nagu saadi tulemuseks varasema analüüsiga. Samamoodi leiti püsiühendilise puhul, et püsiühendite jaotuse erinevus terviku jaotusest on statistiliselt oluline, kuid mittepüsiühendite jaotuse erinevus mitte. Nimelisuse puhul leiti, et pärisnimede jaotuse erinevus terviku jaotusest on statistiliselt oluline nagu varasemas analüüsiski, aga üldnimede jaotuse erinevus mitte. Samast tulemusest lähtub ka kvalitatiivne analüüs, kuna mitterektsiooni, mittepüsiühendi ja üldnime tasemele kuuluvad liiga erinevate omadustega sõnad, et nende kohta midagi üldistavat väita. Pärisnime semantiliste rühmade osa-terviku jaotuste järgi erineb isiku- ja kohanimede jaotus terviku jaotusest ning mõlema puhul on eelistatud illatiiv.

Selle kontrollmeetodiga saadi nelja tunnuse puhul samad tulemused. Need tunnused on sõnaliik, lauseliige, üldnime semantiline rühm ja verbi tähendus. Nelja tunnusega saadi aga täpsustatud tulemused: rektsioonilisuse puhul on statistiliselt oluline seos ainult rektsioonistruktuuride ja illatiivi vahel; püsiühendilise puhul ainult püsiühendite ja aditiivi vahel; nimelisuse puhul ainult pärisnimede ja illatiivi vahel; pärisnime semantiliste rühmade puhul on aga isiku- ja kohanimed eelistatud illatiivis.

Tabelis 14 tuuakse kontrollmeetodina ka efekti suurus (ingl *effect size*, ES). Tabeli esimeses tulbas on tunnus, teises ja kolmandas tase(med), mille seotus aditiivi või illatiiviga osutus statistiliselt oluliseks, st mille standarditud Pearsoni jääkide absoluutväärtus oli vähemalt 2. Neljandas tulbas on  $p$ -väärtused ja viiendas ES seose tugevuse hindamiseks.  $2 \times 2$  tabelite puhul või iga kahe rea või tulbaga tabeli puhul on ES võrdne Craméri seosekordajaga  $V$ , mis iga tunnuse puhul arvutati. Siinse materjali väikese, keskmise ja suure ESi väärtuseks sobivad vastavalt 0,1; 0,3 ja 0,5. Selle korpusuuringu valim ( $N = 840$ ) on piisavalt suur, et leida keskmised ja suured efektid. (Cohen 1988: 216–227, 1992: 156, 158, tabel 2; ülevaade Arppe 2008: 86–87; Klavan 2012: 133–136, 162–163) Siiski on selle andmestiku suurim ES ainult 0,22 – tunnusega üldnime semantiline rühm. Tabel on järjestatud viimase tulba järgi suurimast ESist väiksemani.  $p$ -väärtust pole toodud, kui mõnel tasemel oli liiga vähe liikmeid selle arvutamiseks. Osa-terviku meetodiga saadud tulemus on sulgudes. Kui ühetunnuseline analüüs ja osa-terviku meetod andsid erinevad tulemused, siis teine neist on sulgudes. Tabeli järgi on kaheksast tunnusest

viis statistiliselt olulised – esimesel viiel on  $p$ -väärtus alla 0,05 või väike ES. Viimase kolme  $p$ -väärtused ja ESid ei osutunud statistiliselt oluliseks.

**Tabel 14.** Morfosüntaktiliste ja semantiliste tunnuste ühetunnuselise analüüsi kokkuvõte

Tunnus	Adiitiivi eelistav(ad) tase(med)	Illatiivi eelistav(ad) tase(med)	$p$ -väärtus	Efekti suurus (Craméri V)
Üldnime semantiline rühm	'KOHT', 'SEISUND'	'muu', 'pärisnimi'	$p < 0,001$	0,22
Rektsioonililus	mitte-rektsioonililus/ (-)	rektsioonililus	$p < 0,001$	0,21
Püsiühendililus	püsiühendililus	mitte-püsiühendililus/ (-)	$p < 0,001$	0,20
Pärisnime semantiline rühm	–	('ISIK', 'KOHT')	–	0,15
Nimelilus	üldnimed / (-)	pärisnimed	$p < 0,001$	0,12
Sõnaliik	–	–	–	0,06
Verbi tähendus	–	–	$p = 0,82$	0,03
Lauseliige	–	–	$p = 0,65$	0,02

Kokkuvõttes annavad nii  $p$ -väärtus, osa-terviku meetod kui ka ES sarnased tulemused ning vajaduse korral täiendavad üksteist. Tabelis toodud esimesest viiest tunnusest neli olid  $p$ -väärtuse järgi statistiliselt olulised. Pärisnime semantilise rühma puhul oli isikunimesid vaid seitse, mida oli usaldusväärsete tulemuste saamiseks ehk liiga vähe. See-eest saab tunnuse paigutada skaalale ESi järgi. Tulemusi kontrolliti ka osa-terviku meetodiga, mis andis kasulikku lisainfot, et pärisnime semantilise rühma puhul võiks rühmad 'ISIK' ja 'KOHT' eelistada illatiivi ning rektsioonililuse puhul ei ole mitterektsioonidel, püsiühendililuse puhul mittepüsiühenditel ja nimeliluse puhul üldnimedel sisseütleva eelistust. Unustada ei tohi kvalitatiivset analüüsi: väga suure taseme, nagu mitterektsiooni, mitte-püsiühendi ja üldnime kohta ongi raske järeldusi teha. Nendele tasemetele kuuluvad erinevate omadustega sõnad ning sisukamaid järeldusi saab teha vaid väiksema ja konkreetsema taseme kohta. Samuti ei saa teha järeldust, et üldnimesid rühmast 'muu' eelistatakse kasutada illatiivis, sest keelekasutajad ei tee vormivalikut uurija loodud 'varia' rühmast lähtudes.

## 6. Järeldused ja kokkuvõte

Siinne artikkel oli järg käsitlusele, kus uuriti ainsuse pika ja lühikese sisseütleva valiku olenemist morfofonoloogilistest tunnustest. Praeguses artiklis uuriti morfosüntaktilisi ja semantilisi tunnuseid. Materjali kogumisel oli võrreldes eelmisega kaks erinevust: esiteks kasutati eelmises uurimuses tasakaalustamata, siinses tasakaalustatud valimit; teiseks arvestati eelmises uurimuses materjali hulka ka korduvad vormid, selles artiklis mitte.

Tasakaalustatud materjali eelis on, et see annab ülevaate ka muidu harvem esinevast võimalusest – siin illatiivivormidest. Kui eelmises artiklis oli illatiivivorme vaid 188, siis siinses 420. Eelmises artiklis sai uurida vaid tendentse – mingite tunnuste puhul on illatiiv standarditud Pearsoni jääkide põhjal eelistatum. Siiski kasutati materjalis kõigi tunnuste puhul rohkem aditiivivorme. Tuleb silmas pidada, et sellised probleemid kerkivad väikese valimi korral ning kui materjali on piisavalt, ei ole tasakaalustamatus probleem, vaid annab tasemete jaotusest terviklikuma pildi. Nii võiks valimi suuruse järgi otsustada, kumb viis sobib materjali kogumiseks paremini – kummalgi pole selget eelist teise ees ja viisid täiendavad teineteist.

Koos kõigi kordustega oli algselt materjalis 2190 vormi: 1095 illatiivi- ja 1095 aditiivivormi. Nendega tehti analüüsid – hii-ruut-test ja standarditud Pearsoni jäägid. Tulemuseks saadi, et sisseütleva valikul on olulised kõik tunnused peale pärisnime semantilise rühma. Kas selle järgi võib oletada, et sõnade omadused on valimit mõjutanud, st mida rohkem kordusi on materjalis, seda tõenäolisem, et rohkem tunnuseid on sisseütleva valikul statistiliselt olulised? Eelmise artikli tulemuseks saadi, et ühetunnuselise analüüsi põhjal on statistiliselt olulised kaheksast tunnusest kuus, siinses kaheksast tunnusest neli. Kontrollimaks, kas tulemusi on mõjutanud korduvate vormide jätmine materjali hulka, tuleks analüüsida eelmise artikli materjali mittekorduvate vormide kaupa.

Siinne artikkel oli metoodiline katse uurimaks, millised morfosüntaktilised ja semantilised tunnused mõjutavad ainsuse sisseütleva valikut. Vaadeldi 840 sisseütleva vormi, mille puhul võivad illatiiv ja aditiiv teoreetiliselt varieeruda. Hii-ruut-testi ja standarditud Pearsoni jääkidega vaadeldi kaheksast tunnust, mida kontrolliti osa-terviku meetodiga ja järjes-tati efekti suuruse järgi. Need tunnused olid sõnaliik, lauseliige, rektsiooni-lisus, püsiühendilisus, nimelisus, pärisnime semantiline rühm, üldnime semantiline rühm ja verbi tähendus. Hii-ruut-testi põhjal olid statistiliselt

olulised rektsioonilisus, püsiühendilisus, nimelisus ja üldnime semantiline rühm. Osa-terviku meetodi ja efekti suuruse järgi võiks oluliseks pidada ka pärisnime semantilist rühma. Standarditud Pearsoni jääkide põhjal eelistatakse aditiivi mitterektsioonide, püsiühendite, üldnimede ning üldnimeliste koha- ja seisundifraaside puhul. Illatiivi eelistatakse rektsioonide, mittepüsiühendite, pärisnimede ning üldnime semantiliste rühmade 'muu' ja 'pärisnimi' puhul. Osa-terviku meetodi järgi pole eelistusi mitterektsioonide, mittepüsiühendite ja üldnimede puhul. Pärisnime semantilistest rühmadest ilmneb isiku- ja kohanimede puhul illatiivi eelistus.

Seega said kinnitust varasemate uuringute põhjal püstitatud hüpoteesid, et rektsioonistruktuurid on eelistatult illatiivis; püsiühendid ning üldnimelised kohad aditiivis. Erinevalt hüpoteesis oletatust ei selgunud vormieelistusi verbi tähenduse järgi.

Artiklis püüti välja selgitada, milliseid morfosüntaktilisi ja semantilisi tunnuseid kaasata edaspidisesse mitmetunnuselisse analüüsi. Kvantitatiivse ja kvalitatiivse analüüsi põhjal võiks oluliseks pidada rektsioonilisust, püsiühendilisust, nimelisust ning nii päris- kui ka üldnime semantilist rühma. Pigem illatiivi kasutatakse rektsioonistruktuurides ning pärisnimede (täpsemalt isiku- ja kohanimede) puhul. Aditiivi eelistatakse püsiühendites ning üldnimelistes koha- ja seisundifraasides. Edaspidi tuleks ka kõik tunnused koos kaasata mitmetunnuselisse analüüsi, sest tunnuseid korraga mudeldades võib üks tunnus teise osalusel oluliseks osutada. Lisaks peab vaatama tunnuste omavahelisi seoseid, näiteks püsiühenditesse kuuluvad tihti kehaosasõnad. Niisiis ei saa väita, et mingi sisseütleva vorm võiks eelistatud olla ainult ühe tunnuse põhjal, pigem on eelistus mitme tunnuse koosmõjust. Ühetunnuseline analüüs näibki olevat hea meetod esmaste tulemuste väljaselgitamiseks, kuid tuleb arvestada, et tulemused ei pruugi olla nii ühesed, kui meetod järeldada lubab. Edasistes uuringutes tuleks morfofonoloogilisi, morfosüntaktilisi ja semantilisi tunnuseid kasutada ka eksperimentides, et täpsustada, millistest tunnustest keekekasutaja valik sellisel juhul on.

## Kirjandus

- Agresti, Alan 2013.** Categorical Data Analysis. 3rd Ed. New Jersey, Hoboken: John Wiley & Sons Inc.
- Arppe, Antti 2008.** Univariate, Bivariate, and Multivariate Methods in Corpus-based Lexicography – a Study of Synonymy. University of Helsinki. Helsinki: Helsinki University Print.
- Atkins, Beryl T. S., Beth Levin 1995.** Building on a corpus: A linguistic and lexicographical look at some near-synonyms. – International Journal of Lexicography 8 (2), 85–114.
- Cohen, Jacob 1988.** Statistical Power Analysis for the Behavioral Sciences. 2nd edition. Hillsdale: Lawrence Erlbaum Associates. <http://dx.doi.org/10.1037/0033-2909.112.1.155>.
- Cohen, Jacob 1992.** A power primer. – Psychological Bulletin 112 (1), 155–159.
- EKG I** = Mati Erelt, Reet Kasik, Helle Metslang, Henno Rajandi, Kristiina Ross, Henn Saari, Kaja Tael, Silvi Vare 1995. Eesti keele grammatika. I. Morfoloogia. Sõnamoodustus. Peatoim. Mati Erelt, toim. Tiiu Erelt, Henn Saari, Ülle Viks. Tallinn: Eesti Teaduste Akadeemia Eesti Keele Instituut.
- EKG II** = Mati Erelt, Reet Kasik, Helle Metslang, Henno Rajandi, Kristiina Ross, Henn Saari, Kaja Tael, Silvi Vare 1993. Eesti keele grammatika. II. Süntaks. Lisa: Kiri. Peatoim. Mati Erelt, toim. Tiiu Erelt, Henn Saari, Ülle Viks. Tallinn: Eesti Teaduste Akadeemia Keele ja Kirjanduse Instituut.
- EKK** = Mati Erelt, Tiiu Erelt, Kristiina Ross 2007. Eesti keele käsiraamat. Kolmas, täiend. trükk. Tallinn: Eesti Keele Sihtasutus.
- EKSS** = Eesti keele seletav sõnaraamat 2009. „Eesti kirjakeele seletussõnaraamatu“ 2., täiend. ja parand. tr. Toim. Margit Langemets, Mai Tiits, Tiia Valdre, Leidi Veskis, Ülle Viks, Piret Voll. Tallinn: Eesti Keele Sihtasutus.
- Gilquin, Gaëtanelle, Stefan Th. Gries 2009.** Corpora and experimental methods: a state-of-the-art review. – Corpus Linguistics and Linguistic Theory 5 (1), 1–26.
- Hasselblatt, Cornelius 2000.** Eesti keele ainsuse sisseütlev on lühike. – Keel ja Kirjandus 11, 796–803.
- Japkowicz, Nathalie, Shaju Stephen 2002.** The class imbalance problem: A systematic study. – Intelligent Data Analysis 6, 429–449.
- Kaalep, Heiki-Jaan 2009.** Kuidas kirjeldada ainsuse lühikest sisseütlevat kasutamismetega kooskõlas? – Keel ja Kirjandus 6, 411–425.
- Kio, Kati 2006.** Sisseütleva käände kasutus eesti kirjakeeles. Magistritöö. Käsikiri Tartu Ülikooli eesti keele osakonnas.
- 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.

- Metslang, Ann 2015.** Ainsuse pika ja lühikese sisseütleva valiku olenemine morfofonoloogilistest tunnustest – korpusanalüüs. – Emakeele Seltsi aastaraamat 60 (2014). Peatoim. Mati Erelt. Tallinn, 127–147. <http://dx.doi:10.3176/esa60.06>.
- Mäearu, Sirje 2011.** Valik rektsioone. Tartu: Keelehooldakeskus. <http://keeleabi.eki.ee/artiklid/rektsioone.html> (23.09.2015).
- Raag, Virve 1998.** The effects of planned change on Estonian morphology. (= *Studia Uralica Upsaliensia* 29.) Uppsala.
- R Development CoreTeam 2013.** R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. <http://www.r-project.org/> (15.04.2016).
- Sõnajalg, Ingrid 1956.** Illatiiv eesti kirjakeeles. Diplomitöö. Käsikiri Tartu Ülikooli eesti ja üldkeeleteaduse instituudis.
- Viitso, Tiit-Rein 1976.** Eesti muutkondade süsteemist. – Keel ja Kirjandus 3, 148–162.
- Õim, Katre, Asta Õim 2015.** Kehaosasõnade arvukategooriast. – Keel ja Kirjandus 2, 88–99.
- ÕS 2013** = Eesti õigekeelsussõnaraamat ÕS 2013. Toim. Maire Raadik. Koost. Tiiu Erelt, Tiina Leemets, Sirje Mäearu, Maire Raadik. Eesti Keele Instituut. Tallinn: Eesti Keele Sihtasutus.

## Korpused

- Eesti keele koondkorpus. <http://www.cl.ut.ee/korpused/segakorpus/index.php> (27.09.2015).
- Eesti keele verbikesksete püsiühendite andmebaas. <http://www.cl.ut.ee/ressursid/pysiyhendid/> (17.09.2015).
- etTenTen. <http://www2.keeleveeb.ee/dict/corpus/ettenten/about.html> (27.09.2015).
- Keeleveeb. <http://www.keeleveeb.ee/> (10.09.2015).



**How the choice between the singular long and short  
illative case is related to morphosyntactic  
and semantic variables – which material and methods  
are suitable for a corpus analysis**

ANN SIIMAN

The article examines the choice between the singular long and short illative case. It attempts to find out which morphosyntactic and semantic variables are statistically significant for choosing the long or short illative case. The variables considered are part of speech, part of sentence, government, fixed word combination, proper or common noun, proper noun semantic group, common noun semantic group and meaning of the verb lemma.

The material investigated comes from the Estonian web *corpus etTenTen* and the Estonian Reference Corpus. In the final material, a total of 840 illative case forms were analysed. The material was balanced (420 long illative forms and 420 short illative forms) and all word forms were included only once. With the help of the computer software program R, a chi-square test and standardized Pearson residuals were performed. The results were controlled with a so-called part-whole method and using Cramér's V effect size method.

Based on the quantitative and qualitative analyses, five variables were found to be important: government, fixed word combination, proper or common noun, proper noun semantic group and common noun semantic group. It was found that the government structures, proper nouns and the levels 'person names' and 'place names' from the variable proper noun semantic group prefer the long illative case. The short illative case is more preferred with fixed word combinations and the levels 'place phrases' and 'state phrases' from the variable common noun semantic group. In this univariable corpus-based study, the variables government, fixed word combination, proper or common noun, proper noun semantic group and common noun semantic group were statistically significant, but further studies should test if these variables remain statistically significant when applying multivariable analysis or in experiment-based studies.

**Keywords:** morphology, morphosyntax, semantics, corpus linguistics, language variation, illative, short illative, Estonian

**Lisa.** Esialgse materjali varieeruva illatiivi ja aditiiviga sõnad  
(2190 vormi seast)

ajalugu	küla	praegune	teadvus
bussipeatus	list	raamatukogu	teine
eelmine	loetelu	Rakvere	tööstus
elu	Läti	reaalsus	valdus
esimene	maja	Riia	Valga
kaks	muuseum	r/Riigikogu	vastuolu
keskus	mõni	rivi	võitlus
kodakondsus	m/Mägi	seadus	üks
kodu	mälu	Soome	ümbrus
kolmetoaline	osa	Tallinn	
koondis	pere	tasku	
koostis	perse	teadus	

Ann Siiman  
eesti ja üldkeeleteaduse instituut  
Tartu Ülikool  
Jakobi 2  
51014 Tartu  
ann.siiman@ut.ee