

TARTU ÜLIKOOL  
HARIDUSTEADUSTE JA KUNSTIDE VALDKOND  
EESTI JA ÜLDKEELETEADUSE INSTITUUT  
ÜLDKEELETEADUSE OSAKOND

Mariann Proos

MIDA ÜTLEB KORPUS TÄHENDUSE KOHTA? KÄITUMISPROFIILI  
ANALÜÜSI JA KLASTERANALÜÜSI MEETOD TAJUVERBI *NÄGEMA*  
TÄHENDUSE UURIMISEL

Magistritöö

Juhendaja Jane Klavan

Tartu 2016

## Sisukord

Sisukord .....	2
Sissejuhatus .....	4
1. Polüseemia kognitiivses semantikas .....	7
2. Korpuspõhine kognitiivne semantika.....	10
2.1 Kognitsioon ja korpus .....	10
2.2 Käitumisprofili analüüsi rakendamisest .....	11
3. Magistritöö materjal ja meetod .....	15
3.1 Materjal .....	15
3.2 Meetod.....	16
3.2.1 Semantilised ID-sildid .....	18
3.2.2 Süntaktilised ja morfoloogilised ID-sildid.....	26
3.2.3 Väljajäänud ID-sildid.....	29
3.2.4 Hierarhiline klasteranalüüs .....	31
4. Tulemused .....	34
4.1 Üldised tulemused .....	34
4.2 Hierarhilise klasteranalüüsi tulemused.....	38
4.2.1 Valideerimine siluettehnikaga (ingl <i>Silhouette plot</i> ) .....	41
4.2.2 <i>Bootstrap</i> -tehnika (ingl <i>bootstrapping</i> ) .....	43
4.3 Moodustunud klastrite iseloomustus .....	45

4.4 <i>Nägema</i> tähendusvõrgustikust.....	52
5. Sorteerimiskatse (Proos 2014) ja käitumisprofiili analüüsi tulemuste võrdlus....	57
6. Järeldused ja arutelu.....	60
Kokkuvõte.....	63
Kirjandus.....	66
A behavioral profile analysis of the Estonian verb <i>nägema</i> 'to see'. Summary .....	70
LISA 1. Verbi <i>nägema</i> käitumisprofiil (%) .....	72

## Sissejuhatus

Magistritöö tegeleb ühe kognitiivses semantikas enim uuritud nähtuse, polüseemiaga. Uuritavaks verbiks valisin eesti keele tajuverbi *nägema*, mille polüseemiaga tegelesin ka oma bakalaureusetöös (Proos 2014). Tajuverbid laiemalt on uurimisteemana väga huvitavad, kuna just need verbid on peamine keeleline vahend, mille kaudu väljendame oma kõige vahetumat kogemust maailmast. Samas aga on tajuverbid paljudes keeltes väga polüseemsed, nende kaudu väljendame ka hoopis teistsuguseid, abstraktseid kogemusi. Kognitiivse keeleteaduse järgi lähtume abstraktsete kogemuste mõtestamisel füüsilistest kogemustest, seega on huvitav uurida, millist kogemust mõtestame kõige vahetumaid füüsilisi kogemusi väljendavate verbidega.

Maailma keeleteaduses on tajuverbid laialt uuritud teema, kuid eesti keele tajuverbe on uuritud väga vähe ning kitsamalt kognitiivse semantika valdkonnas pole tajuverbide polüseemiat keegi peale siinkirjutaja põhjalikult uurinud. Toomas Born on oma diplomitöös „Eesti keele tajuverbid. Tähenalusallikad ja tähendusmuutused“ (1995) uurinud tajuverbe ning nende tähenduse laienemise mehhanisme diakroonilises mõõtmes. Regilaulude semantika uurimise raames on Kanni Labi (2006) oma doktoritöös pühendanud peatüki ka tajuverbidele. Tajuverbid *nägema*, *paistma* ja *tunduma* on leidnud kajastust Maria-Maren Sepperi magistritöös „Indirektaal eesti 19. sajandi lõpu ja 20. sajandi aja- ja ilukirjanduskeeles“ (Sepper 2006)

Selles magistritöös lähenetakse polüseemiale läbi korpuse – täpsemalt kasutatakse eesti keele tajuverbi *nägema* uurimiseks ühte korpuspõhise keeleteaduse meetodit, käitumisprofiili analüüsi (ingl *behavioral profile*). Praeguseks on vajadus ka keeleteaduses kasutada empiirilisi meetodeid juba korduvalt kinnitust leidnud ning suur osa uurimustest

toetuvadki empiirilistele andmetele. Korpus on empiirilise uurimistöö jaoks väga hea allikas, kuna materjali hulk korpuses on väga suur ning samas on materjal olemuselt väga loomulik. Käitumisprofiili analüüs annab võimaluse anda põhjalik ülevaade sõna või konstruktsiooni distributiivsetest sagedustest, kusjuures lähtutakse eeldusest, et distributiivne sarnasus peegeldab funktsionaalset (sh tähenduslikku) sarnasust.

Eesti keele tajuverbi *nägema* uurisin oma bakalaureusetöös (Proos 2014) sorteerimiskatsega. Katseliste meetodite kasutamine polüseemia uurimiseks on üks levinumaid tehnikaid, kuid korpuspõhised lähenemised on populaarsust kogunud just viimasel ajal. Üks põhjuseid selleks on tehniliste vahendite kiire areng – korpused muutuvad aina paremaks ning samuti lihtsustub üha suuremahulisema andmestiku põhjal statistilise analüüsi tegemine, kuna uurija käsutuses olevad tehnilised vahendid arenevad kiiresti nii võimekuse kui ka võimaluste poolest.

Uurimuse materjal on pärit eesti keele tasakaalus korpuse aja- ja ilukirjandustekstidest. Kokku on analüüsi alla võetud 700 lauset, iga lause puhul määrati ära erinevaid nii semantilisi kui ka morfosüntaktilisi tunnuseid. Saadud materjali analüüsiti hierarhilise aglomeratiivse klasteranalüüsi meetodiga, klasteranalüüsi tulemuste valideerimiseks kasutati kaht tehnikat – *bootstrap*-tehnika (ingl *bootstrapping*) ja siluettehnika (ingl *Silhouette plot*). Klasteranalüüsi kasutamine võimaldab uurida, millised tähendused on omavahel tihedamalt seotud ning millised asuvad üksteisest kaugemal. Üks magistr töö eesmärke on ka analüüsida, kas ja kui sobiv on käitumisprofiili analüüsi tehnika polüseemia uurimisel. Meetodit analüüsitakse eelkõige võrdluses katseliste meetoditega, täpsemalt lähtun võrdlusel oma bakalaureusetöös tehtud sorteerimiskatsest ja selle tulemustest.

Töö jaguneb kuueks suuremaks peatükiks. Esimeses peatükis antakse laiem ülevaade polüseemia uurimisest kognitiivses semantikas, kuidas seda nähtust mõistetakse ning mis on uurimise juures oluline. Teises peatükis keskendutakse kitsamalt korpuspõhisele kognitiivsele lähenemisele. Antakse ülevaade sellest, kuidas on võimalik uurida keelelist tunnetust läbi korpuse ja miks seda teha. Samuti kirjeldatakse täpsemalt käitumisprofiili analüüsi meetodit ning selle meetodiga tehtud uurimusi.

Kolmandas peatükis kirjeldatakse magistritöö materjali ning täpsemalt käitumisprofiili analüüsi meetodit. Peatükis seletatakse lahti meetodi olulisemad komponendid, nagu ID-siltide valik ja nende määramine, hierarhilise klasteranalüüsi olemus ning valideerimistehnikate vajalikkus ja rakendamine.

Neljandas peatükis antakse põhjalik ülevaade magistritöö tulemustest. Kirjeldatakse nii üldisi tulemusi kui ka konkreetselt hierarhilise klasteranalüüsi tulemusi. Lisaks antakse ülevaade sellest, mida annavad analüüsile juurde valideerimistehnikad. Peatüki lõpus arutletakse verbi *nägema* tähendusvõrgustiku üle. Viiendas peatükis siirdutakse sorteerimiskatse ja käitumisprofiili analüüsiga saadud tulemuste võrdluse ja järelduste juurde, mida meetodiga saadud tulemuste põhjal teha võib. Viimases peatükis (peatükk 6) üldistatakse magistritöös saadud tulemuste abil, kas ja mil määral on võimalik korpuse kaudu tähendust uurida, arutletakse töö panuse üle eesti keeleteadusesse ning pakutakse välja võimalusi edasiseks uurimiseks.

## 1. Polüseemia kognitiivses semantikas

Kognitiivses keeleteaduses on tähendusel keskne roll, seega polüseemia kui ühe vormi tähenduste paljusus on suuna esindajatele alati huvi pakkunud. Tähendus ei ole kognitiivse keeleteaduse jaoks eraldiseisev nähtus, vaid suhestub laiemale taustsüsteemiga, mis hõlmab keelekasutaja kogemusi, tausta, entsüklopeedilisi teadmisi, uskumusi jms. Kognitiivsed ja kasutus põhised, laiemalt post-strukturalistlikud ja post-formalistlikud teooriad käsitlevad keelt ja selle struktuuri üldiselt kui dünaamilist ja holistilist nähtust, mis on saanud oma kuju kasutuses (Glynn 2014: 7) (ning on seetõttu ka pidevalt muutuv). Glynn (2014) on rõhutanud, et polüseemia ei ole kognitiivse keeleteaduse jaoks selgepiiriline nähtus, vaid seda peaks üldistama kui mõisteid ja funktsioone, mida tähistab või täidab üks vorm.

Polüseemia ning laiemalt tähenduse sellise käsitluse jõuti, ühendades teiste hulgas Wittgensteini (1997:31–32) idee perekondlikust sarnasusest ja psühholoogias tuntud Roschi teooria kategoriseerimisest (Geeraerts 2010a: 183–199). Mõiste defineeriti kui kategooria, mis koosneb omavahel perekondliku sarnasuse kaudu seotud semantilisest ja mõistelisest infost, kusjuures mingi osa sellest infost esildub rohkem kui muu (st on prototüüpilisem). Ka polüseemse sõna erinevad tähendused on omavahel seotud perekondliku sarnasuse teel ning üks (osa) tähendus(i) on prototüüpsem kui teised. (Cuyckens, Zawada 1997: xiii)

Kategooria keskmest e prototüübist lähtub keelekasutaja ka otsustades, kas mingi element kuulub kategooriasse või mitte. Kõik kategooria liikmed peaksid jagama teatud hulka tunnuseid, et keelekasutaja suudaks neid analüüsida samasse kategooriasse kuuluvatena. Sarnasus aga ei ole kergesti defineeritav mõiste. Sarnasuse aste on hinnatav,

kuid mille järgi? Sarnasus ja selle hindamine on alati subjektiivne, sõltub hinnangu tegijast (Taylor 1995: 60–61) ning seeläbi kõigest, mis mängib rolli tähenduse kujunemisel.

Sarnasus, mis sõna polüseemseid tähendusi seob, on tekkinud selliste semantiliste protsesside teel nagu metafoor, metonüümia, üldistus jms. Kuigi, nagu mainitud, on kategoorias mingit sorti prototüüp, ei tähenda see, et ülejäänud tähendused on sellest otseselt tuletatud. Pigem on tegemist tähendusvõrgustikuga, kus ei saa välja tuua „pärimisliini“, vaid kõik tähendused on omavahel vastastikku seotud. (Cuyckens, Zawada 1997: xiv) Polüseemia on seega kompleksne keelenähtus, mille struktuuri kirjeldamisel tuleb arvesse võtta vähemalt kõiki ülaltoodud tegureid: prototüüpne tähendus, sarnasuse skaalad, semantilised protsessid; üldisemal tasandil keele kasutuskontekst, keelekasutaja entsüklopeedilised teadmised jne. Kõige olulisem on aga see, et neid tegureid ei saa üksteisest lahutada.

Kuna keel ja selle kasutus ei ole teineteisest lahutatavad, keele kujunemisel ja muutumisel mängivad rolli nii mitmed omavahel seotud tegurid, tekib paratamatult küsimus, milliste vahenditega on selles tegurite paljususes võimalik keelt uurida? Siin saabki oluliseks empiirika roll keele uurimisel. Uurijal ei ole võimalik kõrvutada keelt mingi *a priori* konstruktiga, et keele ülesehitust analüüsida, küll aga on tal võimalik koguda võimalikult palju loomuliku keele kasutuses esinevaid näiteid, et nende põhjal teha järeldusi keele struktuuri kohta. (Glynn 2014: 7–11) Ehkki keelt on võimalik koguda nii eksperimentaalselt (nt katsed) kui vaatluslikult (nt korpused), on kasutuspõhisuse tähtsust silmas pidades korpusandmetel teiste ees vaieldamatu eelis: loomulikkuse poolest vastavad korpusandmed hea keelematerjali nõuetele kõige paremini (Geeraerts 2010: 72).

Praeguseks on empiiriline keeleteadus ilmselt normiks saanud ning teadusliku meetodi eelised ei vaja uuesti välja toomist. Geeraerts (2010) kergitab aga hoopis huvitavama küsimuse: kas teadusliku meetodi kasutamine tähendab, et intuitsioonil ei ole enam keeleteadusel kohta? Tõesti, statistiliste meetodite olulisust keeleandmete analüüsimisel rõhutatakse pidevalt ning kahtlemata ongi see oluline. Kui keeleandmete põhjal on võimalik teha järeldusi keele struktuuri kohta, on vajalikud tööriistad näitamaks, millise tõenäosusega tehtud järeldused ja üldistused rakenduvad, samuti kinnitamaks, et need tulemused on üldse



võimelised rääkima üldisemal tasandil kui seda on valim (Glynn 2010: 9). Samas ei ole võimalik ka statistilistest meetoditest intuitsiooni rolli täielikult kõrvaldada (ning seda ei peagi tegema). Intuitsioon on vajalik hüpoteeside loomisel, operatsionaliseerimisel ja tulemuste tõlgendamisel (Geeraerts 2010b: 75–76). Intuitsiooni olulisus saab nähtavaks ka selles magistritöös kasutatud meetodi, käitumisprofili analüüsi puhul.

Töö eesmärk on anda ülevaade tajuverbi *nägema* polüseemiast ning teha seda lähtuvalt lauselisest kontekstist, kus verbi erinevad tähendused esinevad. Käitumisprofili analüüs on selleks hea vahend, kuna võimaldab korraga käsitleda suurt hulka materjali nii semantilisest kui ka morfosüntaktilisest küljest. Järgmises peatükis keskendutakse lähemalt korpuspõhisele kognitiivsele semantikale ning käitumisprofili analüüsile.

## 2. Korpuspõhine kognitiivne semantika

### 2.1 Kognitsioon ja korpus

Üks erinevusi korpusmeetodite ja katseliste meetodite vahel on n-ö inimlikkuse või reaalaja mõõde. Mõneti on katseliste meetodite puhul vahetum uurija ja uuritavate keeleandmete vahel vahetum, katse puhul on uurija käsutada vahetu ja originaalne andmete allikas – inimene ise. Kui käsitleda inimest ennast kui igasuguse keeleandmestiku originaalallikat, on korpuse puhul võimatu selle allikani tagasi minna. Samas räägib korpuse kasuks andmete kõrgem „naturaalsus“ ehk loomulikkus – korpus ja selles sisalduv andmestik ise ei ole mõjutatud andmestiku kogumise viisist. Millise meetodiga on võimalik kätte saada võimalikult autentsed andmed, mis võimaldaksid jõuda võimalikult tõelähedaste tulemusteni?

On väidetud, et keelelise kognitsiooni uurimiseks on korpusest paremaks allikaks katseliste meetodite teel saadud andmed. Samas eeldab kasutuspõhine keeleteooria, et väljendi / keele elemendi kinnistumise, meelde jätmise ja taastootmise vahel on tugev seos. Siit järeldeb, et kui mingi väljend esineb keeles sageli ja samas naabruses, võib uurija selle põhjal teha järeldusi ka selle kasutamise kognitiivse reaalsuse kohta. Korpus pakub suurt hulka materjali, mille põhjal uurida keele elemendi üldist sagedust ning selle koosinemist teiste elementidega, seega peaks korpust saama pidada vahendiks, mille kaudu pääseb ligi kognitiivsele reaalsusele. Zeschel (2010) väidabki, et pole veenvat põhjust, miks katsete teel kogutud andmeid peaks pidama korpusandmetest otsesemaks, tõestatumaks või privileegeritumaks. (Arppe jt 2010: 10)

Just eeldusel, et keele elemendi korpuses käitumise alusel saab midagi väita selle elemendi organiseerituse kohta inimese meeles, on välja töötatud käitumisprofiili analüüsi (ingl *behavioral profile analysis*) meetod. Käitumisprofiili analüüsi tutvustasid esmakordselt aastal 2006 Dagmar Divjak ja Stefan Th. Gries, kes rakendasid sellist meetodit uurimaks vene keele lähisünonüüme, mis väljendavad mõistet PROOVIMA (Divjak, Gries 2006). Alates sellest on käitumisprofiili analüüsi kasutatud mitmete keelte polüseemia ja lähisünonüümia uurimisel, tuntumad uurimused on näiteks inglise verbi *get* (Berez, Gries 2008) või *run* (Gries 2006) kohta, üks viimaseid tehtud uurimusi käsitles hispaania keele verbi *sentir* polüseemiat (Jansegers jt 2015). Eesti keelt, täpsemalt verbi *seisma* polüseemiat on käitumisprofiili abil uurinud Kristiina Kask (2014).

## 2.2 Käitumisprofiili analüüsi rakendamisest

Käitumisprofiili analüüsi meetodit on rakendatud erinevate keelenähtuste puhul ning erinevas mahus. Seda, kas ja mis mahus mingi meetod adekvaatne on, saab välja selgitada näiteks võrreldes tulemusi teiste meetoditega saadud tulemustega. Nii on käitumisprofiili analüüsi meetodi tulemusi võrreldud nii katsete teel saadud tulemustega kui ka näiteks WordNeti tähendusvõrgustikuga. Hoolimata tehtud uurimuste materjalihulga, korpuse suuruse ja märgendatud ID-siltide hulga erinevustest, on tulemused kinnitanud, et käitumisprofiili analüüs on nii polüseemia kui ka lähisünonüümia uurimiseks väga kasulik tööriist.

Nagu öeldud, aitab meetodit valideerida võrdlus teiste meetoditega saadud tulemustega. Üks näitajaid on saadud tulemuste ühisosa – ehkki on pea võimatu, et kaks meetodit annaksid täpselt sama tulemuse, leiab ühisosa kaudu lisakinnitust tulemus ise, aga samas annab see lisatugevust ka meetoditele endile. Andrea L. Berez ja Stefan Th. Griesi käitumisprofiilil põhinev analüüs inglise verbi *get* 'saama' kohta (Berez, Gries 2008) näitas, et leidus kindel ühisosa nende uurimuse tulemustel ja tulemustel, milleni jõudis Jarno

Raukko (1999) küsitluse teel. Berezi ja Griesi jaoks oli oluline näidata, et Raukko (1999) väide korpusest kui semantika (või vähemalt tähenduse kognitiivse reaalsuse) uurimiseks täielikult sobimatust vahendist on vale. Seega võrreldi käitumisprofiili analüüsi tulemusi Raukko uurimusega.

Berezi ja Griesi uurimuse materjal oli pärit korpusest ICE-GB (*International Corpus of English, British component*) ning 3668-st *get* vastest moodustati 600-lauseline juhuvalim. Tähendused/tähendusrühmad moodustati käsitsi ja kokku eristati uurimuses 47 erinevat *get* tähendust, mis jagunesid 9 tähendusrühma vahel. Käitumisprofiili koostamiseks kasutati 55 erinevat ID-sildi taset, nt abstraktne, konkreetne, aktiiv, passiiv, minevik, olevik jms. Saadud sagedustabeliga viidi läbi hierarhiline klasteranalüüs. (Berez, Gries 2008: 161–162)

Kuigi korpusuurimuse andmete hulk oli küllaltki väike ning korpuse andmestiku eripära tõttu ka mürarikas, näitas dendrogramm siiski küllalt selgelt moodustuvat struktuuri. Nii näiteks moodustus klaster tähendustest *get* kui omamise erinevatest variatsioonidest, eraldi klaster *get* kui saama tähendustest, klaster *get* kui liikumist väljendav verb ning eraldi klaster *get*'i grammatiseerunud vormidest. Raukko (1999) saadud tähendusrühmad olid sarnased, ehkki veidi teisiti nimetatud. Samuti osutusid sarnaseks rühmade „*get* kui saama“ ja „*get* kui liikumisverb“ esinemissagedused. (Berez, Gries 2008: 160–163)

Tähendusrühmi või laiemalt, tähendusvõrgustikku, saab moodustada mitmeti ning üht selliseid võimalusi esindab WordNet, kus võrgustik on üles ehitatud sõnade ja tähenduste nii struktuurilistele kui ka tähenduslikele seostele. Arvesse on võetud sünonüümiat, hierarhilisi suhteid, semantilisi välju jms. (WordNet 2016) Seega on WordNeti tähendusvõrgustik veel üks võimalikke käitumisprofiili analüüsi meetodiga saadud tulemuste võrdluskohti.

Stefan Th. Gries on käitumisprofiili analüüsi meetodiga uurinud inglise keele verbi *run* 'jooksma' polüseemiat ning leidnud, et saadud tähendusrühmad tõesti peegeldavad WordNetis toodud tähendusvõrgustikku. Griesi materjal pärines kahest korpusest – ICE-GB-st (*International Corpora of English, British component*) ja Browni ameerika inglise keele korpusest (*Brown Corpus of American English*). Kokku analüüsis Gries 815 *run* juhtu, tähendused määrati käsitsi WordNeti liigituse abil. Laiemalt jagas Gries tähendused kolme

rühma: transitiivsed, intransitiivsed ja idiomaatilised. Materjal märgendati lähtuvalt 252-st erinevast ID-sildi tasemest. (Gries 2006: 63–75)

Polüseemia puhul on alati küsimus ka sõna prototüüpses tähenduses – milline see tähendus on ning mille põhjal seda väita saab? Sellele pakub lahendust ka WordNet, Gries toob välja, et WordNeti järgi on verbi *run* tähendusvõrgustiku keskmes *motion* 'liikumine'. Käitumisprofili analüüsi põhjal saab Gries aga väita, et täpsemalt on keskne tähendus „kiire jalgsi liikumine“ (ingl *fast pedestrian motion*). Selle tähenduse sagedus on kõige suurem, lisaks on see kõige vabamalt käituv tähendus ning seda saab teiste tähendustega võrreldes pidada kõige vähem markeeritud tähenduseks. Näiteks lubab see tähendus fraasides kõige rohkem erinevaid peasõnu, lisaks on selle tähenduse puhul ID-siltide varieeruvus kõige suurem. (Gries 2006: 75–76) Griesi uurimuse põhjal saab väita, et käitumisprofili analüüs on küllalt hea vahend keskse või prototüüpse tähenduse väljaselgitamiseks. Veel enam – kui saadud dendrogramm peegeldab sõna tähendusvõrgustikku, on see kinnitus selle kohta, et erinevate tähenduste lauseline (korpus-) käitumine peegeldab ka tähenduste organiseeritust lähtuvalt suhetest teiste vastavasse leksikaalsesse ja semantilisse välja kuuluvate sõnadega.

Prototüüpse tähenduse leidmine ei ole sugugi lihtne ülesanne, ilmselt on see ka põhjus, miks selle küsimuse juurde alati tagasi tullakse. Kristiina Kask on oma 2014. aasta magistritöös näidanud, et käitumisprofili analüüs võib prototüüpse tähenduse probleemi lahendada seal, kus katseline meetod seda ei suuda. Kask kasutas magistritöö „*Seisma* verbi polüseemia. Korpuspõhine käitumisprofiil ja klasteranalüüs“ materjali eesti keele tasakaalus korpusest, lemma järgi andis korpus 7000 vastet, millest uurimuses kasutati 500 juhuslikult valitud lauset. Töös võeti analüüsiks 117 ID-sildi taset, nagu näiteks konkreetne, abstraktne, elus agent, eluta agent, olevik, minevik, tulevik jpm. Erinevaid verbi *seisma* tähendusi oli töös kokku 23. Käitumisprofili põhjal tehti klasteranalüüs ning analüüsiti selle tulemusel saadud dendrogrammi (Kask 2014: 30–31, 53–56)

Hierarhilise klasteranalüüsi tulemusena moodustasid *seisma* tähendused kaks suuremat klastrit, ühes esinesid konkreetset ja teises abstraktset tähendused, millest viimane oli heterogeensem. Kuigi töö eesmärk ei olnud verbi keskse tähenduse väljaselgitamine, peab Kask selleks tähendust „seisma kui inimese püsti seismine“ – see

tähendus esines korpuses kõige sagedamini ning samuti esines see tähendus koos kõige suurema arvu ID-sildi tasemetega. Kask (2014) toob võrdluse Pajusalu (2001) tehtud katsega, kus samuti prooviti muu hulgas välja selgitada *seisma* keskne tähendus. Katsest (Pajusalu 2001) selgus aga, et katseisikud ei osanud enamasti verbi keskset tähendust määrata. Nii katselistel kui ka korpusmeetoditel on oma eelised, kuid nagu Kask (2014: 56, 63–64) välja toob, ei pruugi katsete teel saada piisavalt infot hinnangu andmiseks. Nagu ka Griesi (2006) uurimuse puhul, näitas ka verbi *seisma* käitumisprofiili analüüs (Kask 2014), et meetod sobib hästi keskse tähenduse määramiseks samas, kui katseliste meetoditega ei pruugi see õnnestuda.

Eelnevast lähtudes võib öelda, et käitumisprofiili analüüsi meetodi kasulikkus on juba kordi kinnitust leidnud, samuti on korduvalt tõendusmaterjali esitatud väite kohta, et korpuse kaudu on võimalik uurida ka niivõrd tabamatut nähtust kui seda on tähendus ning selle realiseerumine inimese meeles. Samas pakuvad kõik meetodid keelenähtuse-spetsiifilist uut infot, seega on erinevate meetodite rakendamine kindlasti tervitatav. Ka selle magistritöö üks eesmärke on võrrelda katselist ja korpusmeetodit ning seda tehakse töö viimases peatükis. Ehkki selle töö raames saab võrdlust pakkuda rangelt võttes ainult tajuverbi *nägema* uurimise koha pealt, pakub võrdlev analüüs ainet ka laiemale meetodite võrdlusele.

## 3. Magistritöö materjal ja meetod

### 3.1 Materjal

Uurimismaterjal on pärit eesti kirjakeele tasakaalus korpusest (<http://www.cl.ut.ee/korpused/grammatikaliides/>), kusjuures valimisse on võetud ilu- ja ajakirjandustekstid. Korpuses on kokku 15 miljonit sõna, iga liiki teksti 5 miljonit sõna. Teadustekstid ei ole analüüsi kaasatud, kuna nendes on verbi polüseemide varieerumine võrdlemisi väike, seega jäi korpuse suuruseks 10 miljonit sõna. Korpuse kasutajaliidesest (kasutati liidest aadressil <http://www.keeleveeb.ee/>) lemma järgi otsides leiti korpusest 18 679 lauset, mis sisaldasid verbi *nägema* kõigis võimalikes eri vormides. Juhusliku valikuga võeti analüüsitavaks materjaliks 700 lauset<sup>1</sup>. Kui lauses esines verbi kaks või rohkem korda, lähtuti analüüsil alati esimesest esinemisjuhust. Algsest valimist kõrvale jäänud ühendverbidega laused asendati samuti juhusliku valiku teel nii, et kokku oleks uuritavaid lauseid 700.

Ühendverbid nagu *välja nägema*, *läbi nägema* jms ei ole käitumisprofiili analüüsi kaasatud, kuna neid ei käsitleta töös kui verbi *nägema* polüseeme. Ühendverbid on tihti ka ise polüseemsed, seega on selguse eesmärgil otstarbekas need põhiverbi polüseemide analüüsil välja jätta. Samuti on varasemad uurimused (Kask 2014) näidanud, et oma konstruktsioonilise spetsiifika tõttu moodustavad ühendverbid sellisel analüüsil eraldi

---

<sup>1</sup> Eeldati, et käitumisprofiili analüüsi rakendamiseks on 700-lauseline valim piisav, kuna ka tähendusi on suhteliselt vähe (võrreldes teiste sama meetodit kasutavate uurimustega), seega peaks valim vastama representatiivsuse nõuetele.

tähendusrühma, mis aga ei kajasta niivõrd tähenduste ühtekuulumist, kuivõrd vormilist eripära.

### 3.2 Meetod

Tajuverbi *nägema* polüseemiat on magistritöös uuritud käitumisprofiili analüüsi meetodiga, kuna see lubab analüüsida korraga suurt hulka materjali nii semantilisest kui ka morfosüntaktilisest küljest ning annab seega põhjaliku ülevaate tähenduste lauselisest käitumisest. Andmete analüüsimisel lähtutakse eeldusest, et sagedusandmetel põhinev info peegeldab tähenduslikku infot ning selle meetodiga on võimalik vaadelda, millised polüseemse sõna tähendused on omavahel rohkem või vähemal määral sarnased või erinevad. Lisaks on käitumisprofiili analüüs üks võimalusi sõna prototüüpse tähenduse väljaselgitamiseks.

Gries ja Divjak käsitlevad käitumisprofiili kui elementide kogu, mis esinevad koos uuritava keeleüksusega loomuliku keele kõnes ja/või kirjas fraasis või lauses. Lähenemine põhineb eeldusel, et sõna tähendused kuuluvad võrgustikku, kus tähenduste omavahelised seosed on tugevuselt erinevad. Mida tugevamad on sidemed, seda seotumad on tähendused. Eeldatakse, et sõnad, mis käituvad sarnaselt, st millel on sarnane käitumisprofiil, on tihedamalt seotud ning seega on tähendused lähedasemad. (Gries, Divjak 2009 :61)

Meetod koosneb neljast etapist:

- (representatiivse ja juhusliku) valimi hankimine korpusest;
- sõnavormide analüüsimine (ID-siltide (ingl *ID tags*) määramine);
- koosinemist näitava tabeli tegemine;
- tabeli hindamine ja analüüsimine erinevate statistiliste meetoditega. (Gries, Divjak 2009: 61)



Kui andmed on kogutud, järgneb nende märgendamine, mis toimub kolmel tasandil (ID sildi tüüp, ID-silt, ID-sildi tase). Esiteks jagunevad ID-siltide tüübid selle järgi, mis sorti silt on, st jagunevad keele komponentide järgi. Nii võib eristada morfoloogilisi, süntaktilisi, semantilisi ja leksikaalseid ID-sildi tüüpe. Edasi toimub jagunemine konkreetse nähtuse järgi – morfoloogilised võivad jaguneda näiteks aja, kõneviisi, tegumoe, arvu või transitiivsuse ID-siltideks, semantilised näiteks objekti semantilise tüübi, noomeni loendatavuse jms järgi. Konkreetselt määratakse iga korpuse vaste puhul ära ID-sildi tase, st kõneviisi ID-sildil on tasemed imperatiiv, indikatiiv jne; semantilise tüübi ID-sildil tasemed konkreetne, abstraktne, elus, elutu. (Gries & Divjak 2009:62) Magistritöös kasutatud ID-sildid ja nende tasemed on välja toodud tabelis 1. Kokku kasutati töös 17 ID-silti ja 89 ID-sildi taset.

Tabel 1. Magistritöös kasutatud ID-sildid ja nende tasemed

<b>ID-SILDI TÜÜP</b>	<b>ID-SILT</b>	<b>ID-SILDI TASE</b>
<b>ALLIKAS</b> <b>SEMANTILINE</b>	ALLIKAS	Ilukirjandus, ajakirjandus
	TÄHENDUS	N1, N2, N3, N4, N5, N6, N7, N8, N9, N10, N11, N12
	ALGUSPUNKTI ELUSUS	Elus, eluta, ei kehti
	ALGUSPUNKTI ABSTRAKTSUS	Abstraktne, konkreetne, varjatud
	SIHTPUNKTI ABSTRAKTSUS	Abstraktne, konkreetne, varjatud, ei kehti
	SIHTPUNKTI TÜÜP	Objekt, sündmus, situatsioon, ei kehti
	SIHTPUNKTI LEMMA TÜÜP	Füüsiline elutu, inimene, pärisnimi, loom, muu elusolend, nimetus, mitte füüsiline elutu konkreetne SP, tunne, muu abstraktne mõiste, asesõna, ei kehti
	SEMANTILINE AEG	Minevik, olevik, tulevik
	SEMANTILINE ISIK	Isikuline, üldisikuline, varjatud
	<b>SÜNTAKTILINE</b>	LAIEND
AHELVERB		Olema, võima, tahtma, pidama, saama, tohtima, tulema, suutma, muu verb+da-infinitiiv,

		muu verb+supiin, muu verb+vat-vorm, ei kehti
	KÕRVALLAUSE LIIK	et-KL, tingimuslik kui-KL, ajaline kui-KL, viisimääruslik KL (kuidas), objektile suunatud KL (kes/mis/kus), rinnastusseos, ei kehti
MORFOLOOGILINE	AEG	Olevik, lihtminevik, enneminevik, täisminevik, üldminevik, ei kehti
	VERBI VORM	1Sg, 2Sg, 3Sg, 1Pl, 2Pl, 3Pl, ma-infinitiiv, da-infinitiiv, des-vorm, nud-kesksõna, tud-kesksõna, umbisikuline olevik, umbisikuline lihtminevik, ei saa määrata
	KÕNEVIIS	Kindel, tingiv, käskiv, kaudne
	TEGUMOOD	Isikuline, umbisikuline, ei kehti
	KÕNELIIK	Jaatav, eitav, ei kehti

### 3.2.1 Semantilised ID-sildid

Iga korpuselause kohta määrati, millises tähenduses verb *nägema* esineb. Magistritöös on arvesse võetud *nägema* 12 erinevat tähendust. Tähenduste eristamisel on lähtunud autori bakalaureusetöös (Proos 2014) analüüsitud tähendustest, kuid magistritöös on tähendusi ühe võrra vähem. Bakalaureusetöös kasutatud tähenduste eristamisel lähtuti EKSS-i *nägema* kirjest, kuid mõni EKSS-is kirjeldatud tähendus jäi välja ning osa tähendusi, mis olid EKSS-is märgitud kui alatähendused, olid bakalaureusetöös käsitletud eraldi tähendustena. Selles töös käsitletud tähendused on järgmised<sup>2</sup> (näitelaused on autori poolt välja mõeldud):

<sup>2</sup> Tähenduste esitusjärjekord on juhuslik ning tähenduste „nimed“ on selliselt toodud, et teha nende eristamine lugejale võimalikult lihtsasti jälgitavaks.

- N1 füüsiline\_nägemine: füüsiline, nägemismeele kaudu tajumine. *Nägin laual kahte raamatut.*
- N2 nägema\_jagamatu: nägemiskogemus, mis on subjektiivne; millegi (näiliselt) tajumine nägemisaistingutega, kuid puudub füüsilisuse piirang. *Nägin unes valget hobust enda poole traavimas. Nägi viirastusi, kummitusi, nägemusi.*
- N3 nägema\_kohtuma: samuti füüsiline nägemiskogemus, kuid oluline on kontakti lisandumine, st nägemisele lisandub enamasti suhtlus. *Rõõm sind jälle näha! Pole teda aastaid näinud.*
- N4 nägema\_arusaama: millegi üldise mõistmine, tajumine. *Ma näen, et koolitusest ei ole kasu olnud.*
- N5 nägema\_ohte: midagi abstraktset tajuma; tähendused N4 ja N5 on omavahel sarnased, kuid N5 on keskendub abstraktsele sihtpunktile. *Ta nägi selles ohtu oma egole.*
- N6 nägema\_kogema: nägemise objekt on enamasti abstraktne, kuid väljendatakse kogemise/läbi elamise tähendust, mitte ainult tajumise. *Karin oli terve päeva toiduga vaeva näinud.*
- N7 nägema\_tunda\_saama: ka selle tähenduse puhul on oluline kogemise tähendus, kuid „kogeja“ on elutu. *Tahvel pole kuid lappi näinud.*
- N8 nägema\_nentima: tähelepanu juhtides millelegi osutama. *Näed, jõudsidki siia! Näe, mõte läks jälle kaduma.*
- N9 nägema\_soovima: väljendab ootusi mingi tulemuse / asjade seisu suhtes. *Näeksin hea meelega, et mu lemmikmeeskond jõuaks finaali.*
- N10 elame\_näeme: midagi tulevikus kogema, millestki tulevikus aru saama; oluline on vastandus teadmatuse (praegu) ja teadlikkuse (tuleviku) vahel. *Saab näha, kas sellest projektist ka asja saab.*
- N11 nägema\_vaatama: midagi süvenenult, tähelepanuga vaatama; jälgima. *Olin seda etendust juba mitu korda näinud.*
- N12 nägema\_otsustama: millegi visuaalselt tajutava põhjal hinnangut andma; otsuseid tegema. *Nägin ta näost, et tema sisemus kees vihast.*

Bakalaureusetöös (Proos 2014) oli eraldi tähendusena välja toodud *nägema* kui *märkama*, *tähele panema*, kuid tehtud sorteerimiskatse näitas, et keelekasutajad ei erista sellist tähendust konkreetsest füüsilisest tähendusest. Et neid kahte tähendust eristab suures osas semantiline aspekt (vt ka ptk 3.2.3), siis ei olnud võimalik korpusmeetodi põhjal ka vastupidist väita. Samas on tähendused N6 ja N7 käsitletud eraldi tähendustena ka magistritöös, kuigi ka neid ei tajunud keelekasutajad erinevana. Magistritöös on tähendused siiski lahus hoitud eelkõige seepärast, et neid kaht tähendust eristab kogeja elusus või elutus, mida on võimalik korpusandmetega võrdlemisi objektiivselt analüüsida. Samuti võib väga palju sarnasusi leida tähenduste N4 ja N5 ning tähenduste N4 ja N8 vahel, kuid ka neid tähendusi tajusid keelekasutajad eri tähendustena, seega on mõistlik neid eraldi hoida ka magistritöös. Korpuse lausetele tähenduse määramine põhines ainult töö autori arvamusel ja otsustel.

Semantiliste tunnuste märgendamisel lähtuti esiteks sellest, mis tunnused võiksid verbi *nägema* puhul olulised olla ning teiseks sellest, et analüüsi saaks kaasatud ka info, mis rangelt morfoloogilisel märgendamisel jääks välja. Nii märgendati eraldi ka semantiline isik ja semantiline aeg. Semantilisel märgendamisel ei lähtutud ilmtingimata ainult verbi *nägema* vormist, vaid võeti arvesse ka laiemat konteksti, kus verb esines. Laused jaotati semantilise isiku järgi kolmeks: isikulised, üldisikulised ja varjatud isikuga laused. Isikulisena märgendati laused, kus isik on väljendatud kas *nägema* verbi pöördelise vormiga, sellega koosineva verbi pöördelise vormiga või mõne muu vahendiga (isikuline asesõna, noomen jms). Näites<sup>3</sup> 1 ei väljenda verbifraas *tasub näha* isikulisust, seda väljendab hoopis tegijamäärus *kassapidajal*.

- (1) *Et kalleid ja vaevarikkaid dieete ennetada, tasub kassapidajal näha pisut vaeva ja hankida oma lemmikule vahelduseks valmistoidule värsket silku, hõbeheiki, turska või kasvõi poes müüdavaid külmutatud kalakuubikuid.* (EE\_1999)

---

<sup>3</sup> Kõik töös kasutatud näited on pärit kasutatud korpusvalimist, iga näite järel on sulgudes toodud näite allikas.

Üldisikulistena märgendati laused, kus vormiliselt on küll kasutuses isikuline verbivorm, kuid tähendus ei viita siiski mainitud isikule. Näitelausestes 2 ja 3 viidatakse tegelikult üldisele isikule: näite 2 puhul puudub kolmas isik, kelle kohta midagi öeldakse, näites 3 aga ei viidata mitte teisele isikule, vaid pigem hoopis iseendale.

(2) *Teiseks **näeb** pea igal tänavanurgal logelevaid kohalikke.* (EPL\_2002)

(3) *See situatsioon, kus sa istud inimese vastas ja sa **näed** ja sa tajud, et tal on kohutavalt valus, on õudne.* (Ilukirjandus)

Üldisikulistena märgendati ka laused, kus isik oli väljendatud modaali või mõne muu verbi pöördelie vormiga *nägema* infinitiivi ees. Ka näite 4 puhul väljendub modaali pöördelises vormis ainsuse kolmas isik, kuid see ei viita konkreetsele isikule, vaid käitub kui mingi olukorra või fakti kirjeldaja ning ei piiritle ära, kes midagi näha saab:

(4) *Esimesel oktoobril algaval FIBITil **saab** Viru Keskuse aatriumis **näha** kolmes eri stiilis moeetendusi, kus Eesti parimad modellid esitlevad kokku enam kui 300 rõivakomplekti.* (EPL\_2004)

Varjatud isiku sildi said laused, kus oli kasutatud umbisikulist tegumoodi, aga ka need laused, kus *nägema* verb esines infinitiivis koos abiverbiga *on* (vt näited 5 ja 6). Kuigi ka konstruktsioonis *on näha* väljendab pöördeline verbivorm kolmandat isikut, on konstruktsioon siiski teise tähendusega kui eelnevalt üldisikut väljendavaks märgendatud konstruktsioonid. *On näha* on sihtpunktikeskne, st fookustab seda, mis näha on, kusjuures nägija suhe nähtavaga pole oluline. Modaalkonstruktsioonid nagu *saab näha* või *võib näha* lähtuvad aga algpunktist. Et modaalsus väljendabki suhet rääkija ja väljendatava propositsiooni vahel, on tegija alati vaikumisi olemas.

(5) *Koopa tagaosas lirtsib, lõhnab kõntsa järele, aga vett silmaga **näha ei ole**, kõrb on ikkagi kõrb.* (Ilukirjandus)

(6) *Piltidel **on** kaelale langevad pikad läikivad juuksed hästi **näha**.* (Ilukirjandus)

Kuna eesti keeles puudub morfoloogiliselt väljendatud tulevik, ostus vajalikuks tuleviku märgendamine semantilistest tunnustest lähtuvalt. Nii märgendati semantiliselt kolm aega: minevik, olevik, tulevik. Näites 7 on verb morfoloogiliselt küll oleviku vormis, kuid semantiliselt on tegevus paigutatud tulevikku. Tulevikule viitab nii ajamäärus *varsti* kui ka partikkel *las*:

(7) *Noh, see õnnis naeratus kaob varsti, mõtles Saša — las mees **näeb** hinnakirju, mis ma talle tõin!* (Ilukirjandus)

Et kõiki *nägema* erinevaid tähendusi on võimalik siduda TEEKONNA kujundskeemiga (Proos 2014), siis pöörati märgendamisel tähelepanu ka skeemi osalistele, st teekonna alguspunktile (AP) ja sihtpunktile (SP). Nägemise alguspunkt on märgendatud elususe (elus-eluta) ja konkreetsuse (konkreetne-abstraktne) järgi. Lisaks esines korpuses lauseid, kus nägemise alguspunkti polnud väljendatud, see oli varjatud või polnud seda võimalik määrata. Et loogiliselt on nägemisel alati olemas alguspunkt (vastupidiselt sihtpunktile, mis ei ole kohustuslik), ei kasutatud märgendamisel varianti, kus alguspunkt puuduks. Juhtumid, kus alguspunkti ei olnud võimalik määrata või oli see varjatud, viidi kokku ühe ID-sildi taseme alla. Nii märgendati näiteks semantiliselt varjatud isikuga ja üldisikulised laused, aga ka laused, kus verb esines *da*-infinitiivis koos verbiga, milles ei sisaldunud viidet tegelikule alguspunktile, nagu näidetes 8–10.

(8) *Kindralite erinevused **on näha** ka Jakarta tänavatel.* (EPL\_1998)

(9) *Vastustest koorunud edetabelit **saab näha** esmaspäevase lehe vahel ilmuvas kõrgkoolilisas Kooliristmik.* (EPL\_2001)

(10) *Lootus pääseda Euroopa Liitu, milles **nähakse** võimalust tsiviliseeruda, paneb kõiki kiiremini liigutama.* (EE\_1999)

Alguspunkti elusus määrati ära kõikidel lausetel, eeldades jällegi, et nägemisel peab olema loogiline alguspunkt ka siis, kui seda väljendatud ei ole. Seega lause 11 määrati koosinevaks ID-sildi tasemega „alguspunkt elus“, kuna on keeruline ette kujutada, et selles kontekstis võiks nägemise allikas olla mitte-elus.

(11) *Piltidelt on näha, et fotoaparaate on mitu, ja enne röövitud kauplusi, olen sellest kuulnud külaski.* (Ilukirjandus)

Näidetes 12–15 on välja toodud laused, kus nägemise alguspunkt on elutu. Enamasti osutusid mitte-elusateks nägemise alguspunktideks siiski nägemisega seotud „agendid“ nagu silmad või pilk. Teine suurem rühm sama ID-sildi tasemega märgendatud allikaid on institutsioonid ja rühmitused. Institutsioonide ja rühmituste korral võib muidugi vaielda, et tegemist on siiski inimestest koosnevate rühmitustega, seega reaalselt on „nägijad“ siiski elus, kuid töös on tähelepanu pööratud just sellele, et harv pole juhus, kus tegelikult elutu institutsioon väljendatakse tegema midagi, mida saaks teha ainult elus agent.

(12) *"Vaata" – Addy paiskab talle liiva vastu silmi – , "kuni silmad näevad!"*  
(Ilukirjandus)

(13) *Elias seisab mu ees, ja isegi läbi ta kinnise suu näeb mu kõikeuskuv naiivne pilk, kui mitu tooni ta hambad valgemaks on läinud.* (Ilukirjandus)

(14) *Riskide maandamist näeb keskpank eelkõige pankade kapitaliseerituse kasvus.*  
(EPL\_1997)

Esines ka üks juhtum, kus elutu alguspunkt oli abstraktne, kuid polnud seotud nägemisega:

(15) *Rõõmud põgenevad sind nähes ning haarates sädeluse järgi, kustutad sa selle.*  
(Ilukirjandus)

Sihtpunkti puhul märgendati ära selle konkreetsus/abstraktsus, sihtpunkti olemus (objekt, sündmus, situatsioon) ning lemma tüüp. Nagu ka alguspunkti puhul, osutus vajalikuks ID-sildi tase, mis väljendaks varjatud sihtpunkti. Sihtpunkti puhul ei märgendatud eraldi sihtpunkti elulisust või elutust. Et sihtpunkte leiti laiemalt olema kolme tüüpi: objekt, sündmus ja situatsioon, siis elususe kategooriat oleks mõistlik rakendada nendest vaid esimest tüüpi sihtpunktide puhul. Lisaks märgendati sihtpunktina esinenud lemmad eraldi tüüpide kaupa, kust elus-eluta kategooria välja tuleb. Seega sisuliselt märgendati sihtpunkti elusus-elutus täpsemalt, kui alguspunkti oma.

Konkreetse sihtpunktina esinesid nt lemmad *asi, London, kunstisaal, kõbi, sabotaažnik*; abstraktsena nt *vaev, seos, vastuhakk, hoiak, põhjus, korrapära*. Et sihtpunkt võib oma olemuselt olla palju erinevam kui alguspunkt, siis märgendati ka ära, kas sihtpunkt on objekt, sündmus või situatsioon. Sihtpunkti märgendamisel kas objekti, sündmuse või situatsioonina lähtuti lausest tervikuna, lauses kasutatud konstruktsioonidest ning muudest verbidest, mis lauses esinesid. Just verbid pakkusid tihti erinevaid tõlgendusvõimalusi, nii võis suuresti samamoodi konstrueeritud sihtpunkt saada kord tõlgenduse kui situatsioon (näide 16), teine kord kui sündmus (näide 17):

(16) *Kord, kui mul sinna asja oli, vaatasin ukse vahelt sisse ja nägin nahkjoppides turde mehi keset kontorit seismas ja seinale projitseeritud numbreid uurimas.*(EE\_1998)

(17) *Ajad, mil ka talviti meie vanasõidukiklubide põnevat ja mitmekesisist veoautoparki veeremas näeb, jäävad ilmselt kaugemasse tulevikku.* (Maaleht\_2004)

Peamine vastandus, millest sihtpunkti olemuse märgendamisel lähtuti, oli staatilisuse-dünaamilisus. Näites 16 väljendab verb staatilisust, seega käsitleti seda situatsiooni, mitte sündmusega. Võrreldes näitega 16 on näide 17 palju dünaamilisem ning *veoautopargi veeremist* käsitleti kui sündmust, mitte situatsiooni. Tihti oli eristust raske teha, kuid siin toetus autor oma keelevaistule ning otsuste tegemisel jäädi võimalikult truuks staatilisuse-dünaamilisuse vastandusele. Lisaks sellisele sihtpunkti jagunemisele osutus vajalikuks ka ID-sildi tase, mis väljendaks olukorda, kui sihtpunkt on varjatud või ei ole seda võimalik ühegi kategooria alla jagada. Siia alla liigitusid eelkõige laused, kus küll sihtpunkt oli olemas (vrd näitega 20), kuid oli tähistatud näiteks küsisõna (näide 18) või mõne muu vahendiga, mille põhjal ei olnud võimalik sihtpunkti klassifitseerida (näide 19).

(18) *Mida nägi või kuulis eesti talurahvas katolikuaegsetes kirikutes, kloostrites?* (Maaleht\_2004)

(19) *Veebist on näha märkimisväärselt rohkem kui otseülekanDES televiisoris - ekraanile tuleb reziipuldist miksituna kokku neli erinevat pilti: peakanal, info, tehnika ja backstage.*(EPL\_2002)



Erinevalt alguspunktist ei ole sihtpunkt skeemi kohustuslik element, seega ID-sildi tase „ei kehti“ väljendab siin just selliseid tähendusi, kus sihtpunkti ei ole, nagu näites 20:

(20) "*Ostan need prillid, millega hästi näen.*" (EE\_1999)

Sihtpunktid, mis märgendati ära kui objektid, märgendati omakorda lemma tüüpide alusel täpsemalt. Ülejäänud laused said selle ID-sildi raames märgendi „ei kehti“. Erinevaid sihtpunkte analüüsides kujunes 10 erinevat tüüpi sihtpunkte: füüsiline elutu SP, inimene, pärisnimi, nimetus, lind, muu elusolend, mitte-füüsiline konkreetne SP, tunne, muu abstraktne SP, asesõna. Füüsilise elutu SP all mõisteti selliseid lemmasid nagu *asi, raamat, tramm, põld*. Sihtpunkti tüüp „inimene“ koondas enda alla inimese üldnimetused nagu *mees, naine, vend, tüdruk*. Et lemmadena esines küllalt palju ka pärisnimesid, märgendati ka need eraldi. Samuti märgendati eraldi tüüp „nimetus“, mille all mõisteti lemmasid, mis küll viitavad inimesele, kuid teevad seda mingi spetsiifilise tunnuse, ameti või muu kategoriseeriva elemendi kaudu, nt *muusikasõbrad-melomaanid, modell, teisik, loomaarst, kõbi*. Elusolendi tüüpidest märgendati veel ära „lind“ ja „muu elusolend“. Tüübiks võeti „lind“ eelkõige seetõttu, et ühtegi looma valimis sihtpunktina ei esinenud, kuid „muude elusolendite“ alla liigitati olendid, keda küll käsitletakse elusana, kuid kes/mis ei kuulu traditsioonilisse faunasse, nagu *haldjas, näkk, kratt*.

Kõige keerukamaks ja mitmekesisemaks tüübiks osutus sihtpunkti tüüp „mitte-füüsiline konkreetne sihtpunkt“. Selle all on mõeldud objekte, mis ei ole küll abstraktsed, kuid mida on raske pidada ka füüsiliseks. Siia ala liigutavad sellised lemmad nagu *sündmus, laul, külvikevad, painaja, vikerkaar, sõda*. Nende ja sarnaste lemmade abstraktsuse aste on küll veidi suurem kui eelnevalt kirjeldatud asjade ja olendite oma, kuid siiski ei saa neid pidada abstraktseteks mõisteteks. Seda näitab ka jagunemine tüüpidesse „tunne“ ja „muu abstraktne mõiste“, kuhu kuuluvaid lemmasid ei saa kokku panna „mitte-füüsiliste konkreetsete sihtpunktidega“. Tüüp „tunne“ koondab enda all ennekõike tähenduse N6 põhilist sihtpunkti *vaev*, aga ka lemmasid nagu *rõõm, joovastus, õhin*. Ülejäänud abstraktsed sihtpunktid koondati tüübi „muu abstraktne sihtpunkt“ alla. Siia kuuluvad lemmad nagu *seksuaalsus, võimalus, hingepõhi, tähendus, mõte, põhjus*.

Eraldi märgendati ka sihtpunktid, mis olid tähistatud asesõnaga. Asesõna märgiti eraldi tüübiks, kuna tihti ei olnud lausest aru saada, millisele objektile asesõnad viitavad. Asesõnadena esines nii umbmääraseid asesõnu nagu *miski, keegi*, näitavaid asesõnu *see, too*, määratlevaid asesõnu (*kõik*), isikulisi asesõnu *mina, nemad, sina* jpmis. Isikulised asesõnad peaksid viitama elusale objektile, kuid esiteks ei olnud tihti selge, millist tüüpi elusolendile asesõna viitab ning teiseks esines ka juhtumeid, kus ei saanud olla kindel, kas isikuline asesõna üldse viitab elusale objektile, nagu näites 21. Kuigi see ei ole eesti keeles norm, on võimalik isikulise asesõnaga viidata ka elutule objektile, eriti kui tegemist on jutu fookuse või inimesesarnase objektiga. Näites 21 ei ole raske ette kujutada, et jutt käib mingit laadi mannekeenist või kujust, mida ehitakse.

(21) *Aga enne ei näe teda tervikuna, kui ta on õiges kohas ja kui tal on need vidinad kõik küljes - mõõk ja piigid. (EPL\_2003)*

### 3.2.2 Süntaktilised ja morfoloogilised ID-sildid

Lisaks semantilistele tunnustele märgendati ära ka hulk süntaktilisi ja morfoloogilisi tunnuseid. Kui morfoloogilised ID-sildid väljendavad kitsamalt verbi *nägema* käitumist lauses, siis süntaktiliste siltide abil saab ülevaate ka laiemast kontekstist, kus verb esineb.

Valimis esines verb *nägema* näiteks ka ahelverbide koosseisus. Ahelverbe moodustasid nii modaalverbid kui ka muud erinevad verbid. Iga ahelverbi kohta märgendati, kas ahelverbi pöördeline osa on abiverb *olema*, modaalverb *võima, tahtma, pidama, saama, tohtima, tulema, suutma* või mõni muu verb (nt *tasuma, hakkama, õnnestuma*). Modaalverbide esinemine ahelverbides märgiti ära modaalverbi järgi, kuna enamik modaalverbe esinesid ahelverbides korduvalt, samas kui muude verbide (v.a *olema*) esinemine jäi enamasti ühekordseks. Lisaks märgendati ahelverbide puhul ära, milline oli verbi *nägema* infiniitne vorm.

Lause/konstruksiooni tasandil märgendati ära ka see, kas *nägema* esineb mingi laiendiga. Laiendina peeti silmas, kas nägemine, selle suund vms on täpsustatud muude vahenditega peale algus- ja sihtpunkti. Laiendid jagati nende tüübi järgi vahendiks, allikaks, kohaks, määruseks ja trajektoori täpsustuseks. Sellised esmapilgul süsteemitud tüübid kategoriseeriti kogutud materjalist lähtuvalt. Vahendina esinesid nägemisega seotud lemmad (*oma*) *silm*, (*oma*) *ihusilm*, (*uus/muundunud*) *pilk*, *prillid*<sup>4</sup>, nagu näidetes 22 ja 23.

(22) *Absoluutselt ammendav informatsioon olendite kohta, keda igäüks vähemalt ühe korra elus oma ihusilmaga on näinud.* (Postimees\_1998)

(23) *Ma nägin teda nüüd muundunud pilguga.* (Ilukirjandus)

Allikana peeti silmas juhtumeid, kus oli täpsustatud mitte nägemise e pilgu allikas, vaid see, millest või kust midagi nähti, nt *rõõm ja joovastus on näha põhiliselt nende silmist, nägin ta rusikasse tõmmatud kätest, piltide peal oli näinud*. Kategooria „koht“ alla kuuluvad laiendid, mis täpsustavad kohta, kus asub sihtpunkt, mitte alguspunkt. Seega ei ole siin laiendite alla arvestatud sellised kohatäpsustused nagu *Londonis nägin, oli seal näinud* jms. Märgendatud laiendites võib tajuda konstruksiooni polüseemsust, laiemalt jagunevad ka koha-laiendid konkreetseks ja abstraktseks. Nii võib näha pisaraid kellegi *silmis*, aga ka probleemi *toimuvas*. Siin saab tõmmata paralleeli *nägema* erinevate tähenduste abstraktsuse-konkreetsuse skaalaga: skaala ühes otsas paiknevad füüsilised „kohad“, keskmes abstraktse ja konkreetse vahepealsed kohad, nagu *hämärus* ja *uni* ning teises otsas abstraktsed „kohad“ nagu *reklaamimises, sisimas, seisukorras*.

Määruse alla liigitusid laiendid, mis täpsustasid mingil viisil nägemise kvaliteeti, viisi või hulka. Trajektoori täpsustus hõlmas juhte, kus oli täpsustatud pilgu liikumise teekond, selliseid täpsustusi esines ainult koos tähendusega N1. Trajektoori saab täpsustada mitmeti – nii alguskohta (*näen mina oma katusekambrist*), vahepealset teekonda (*läbi akna*), suunda (*maski tagant välja*) kui ka sihtkohta (*läbi akna tupp*). Siin on oluline vahet teha alguskoha

---

<sup>4</sup> Täpsustus: siin on mõeldud lauset *Samas mu bändikaaslane ostis juhuslikust kohast väga lahedad prillid, aga kui päike tuli välja, ei näinud ta nendega enam midagi.*, seega *prillid* on siin lemma, millele *nendega* tagasi viitab.

ja alguspunkti vahel – alguspunkti all mõistetakse töös pilgu alguspunkti, skemaatiliselt TEEKONNA kujundskeemi osa, punkti A, kust teekond alguse saab. Alguskoht seevastu on seotud teekonna taustaga, st see on koht/ümbrus, kuhu alguspunkt on paigutatud.

Lausetes, kus *nägema* sihtpunkt oli vormistatud kõrvallausena, määrati ära selle kõrvallause tüüp. Lähtuvalt andmetest märgendati ära 9 erinevat kõrvallause tüüpi. Materjalis leidis nii rinnastus- kui ka alistusseoses kõrvallauseid. Alistusseoses kõrvallased jagati omakorda kaheksaks erinevaks tüübiks: *et*-kõrvallause, tingimuslik *kui*-kõrvallause, ajaline *kui*-kõrvallause, adverbiaalne *kui*-kõrvallause, objektile suunatud kõrvallause (kes/mis/kus), viisimääruse kõrvallause (kuidas), adjektiivne kõrvallause (missugune) ning viisimääruslik *nagu*-kõrvallause. Näites 24 on toodud viisimääruslik kõrvallause ning näites 25 objektile suunatud kõrvallause.

(24) *Nüüd näeb, kuidas edasi.* (Maaleht\_2004)

(25) *Näed mis on heade inimeste palk.* (Ilukirjandus)

Ehkki enamik eraldi analüüsitud tüüpidest kuuluvad eriküsilausete hulka, on töös ka need jagatud eraldi selle alusel, kas sidendina kasutatud küsisõna eeldab noomenifraasilist, adverbifraasilist või adjektiivifraasilist vastust. Sidendiga *kui* vormistatud kõrvallased on määratud kolme erineva ID-sildi tasemega sõna polüseemsuse tõttu. Kõik rinnastusseose laused on koondatud ühe ID-sildi taseme alla.

Morfoloogilisel märgendamisel lähtuti lause peaverbi vormist ning märgendati ära verbi aeg, kõneviis, tegumood, kõneliik, pööre ja verbi infinitiivsed vormid. Kui verb esines ahelverbi koosseisus, märgendati lause seega verbi pöördelise osa järgi. Valiti selline märgendamissüsteem, kuna see annab rohkem infot ka verbi *nägema* käitumise kohta lauses, kui ainult verbi *nägema* vormist lähtumine. Ainult verbi *nägema* järgi märgendades oleks näites 26 toodud lause saanud vaid sildi *da*-infinitiiv – et infinitiivi puhul ei saa rääkida muudest verbi kategooriatest, oleks muu info kõrvale jäänud. Samas põhiverbi järgi märgendades sai näites toodud lause endale ID-sildi tasemed „lihtminevik“, „kindel“, „jaatav“, „3Sg“. Et ahelverbid märgendati niikuinii eraldi, siis tuli süntaktiliste siltide all

välja ka see, mitu korda millises infinitiivis verb esines, seega ei jäänud ka see info märgendamata.

(26) *Ta polnud aimanudki, et Marit võis teda tüdrukuga **näha**.* (Ilukirjandus)

Andmete hulgas leidus ka n-ö puhtaid infinitiive ning need märgendati juba morfoloogilisel märgendamisel nii. Näiteks lauses 27 esineb *nägema da*-infinitiivis, kuid mitte ahelverbi koosseisus, seega lisati lausele ID-silt *da*-infinitiiv.

(27) *Enne trepikotta astumist vaatasin kõrvalakendesse, et **näha** seal tüdrukut, kes istus aknal, käsi põsakil, ning toetudes korraks teisele käele, lehvitas mulle põgusalt ning naeratas mahedalt.* (Ilukirjandus)

Lisaks osutus vajalikuks eraldi silt nendele lausetele, kus verbi vorm oli küll pöördeline, kuid polnud võimalik määrata, milline see pööre on. Eesti keeles on palju vormilisi kattuvusi ning ka vorme, mis tingimata pöördelõppu ei nõua, seega ei saa kindel olla, milline vorm konkreetsetes lauses väljendub. Näiteks vormistati selle sildi alla näited 28 ja 29.

(28) *Selles kerguses oli midagi nii suvist, et hetkeks tundus, nagu **peaks** seal taevas **nägema** pääsukesti lendamas.* (Ilukirjandus)

(29) *Päikest ennast **ei näe** siit korterist enne kui ükskord kevadel, ma ei mäleta enam, vist alles aprillis.* (Ilukirjandus)

### 3.2.3 Väljajäänud ID-sildid

Algsest märgendamissüsteemist jäid välja mõned ID-sildid, mida otsustati lõplikus märgendamissüsteemis mitte arvestada. Näiteks plaaniti alguses märgendada lemma tüübi järgi nii sihtpunkt kui ka alguspunkt, kuid alguspunkt töös nii märgendatud ei ole. Kui valim oli mitu korda läbi käidud, pandi tähele, et alguspunkti tüübi osas suurt varieerumist ei ole – hinnanguliselt 80% juhtudest oli alguspunkt väljendatud kas isikulise asesõna, muu inimest

tähistava väljendi või lihtsalt verbi vormiga. Et alguspunkti puhul määrati ära selle konkreetsus/abstraktsus ja elusus/elutus, on nende kahe ID-sildi abil võimalik välja tuua ebatüüpilisemad juhud. Vähest variatiivsust näitab ka see, et elutuid alguspunkte oli valimis kokku 13. Alguspunkti tüüpide väljatoomist ei peetud ka oluliseks, kuna verbi *nägema* semantika tingib kindlat tüüpi (tahtliku) tegutseja ning seda kõikide tähenduste puhul, samas kui sihtpunktile sarnaseid semantilisi tingimusi ei ole, vastupidi – vahel tingib just sihtpunkti semantika selle, millise tähenduse verb saab.

Süntaktilisel tasandil plaaniti märgendada ka lause kommunikatiivne tüüp. Seda peeti eelkõige vajalikuks, kuna verbi tähenduste hulgas on selliseid, mille kontekst on piiratud ning seda loodeti näidata kommunikatiivse tüübi määramisega. Märgendamisel aga selgus, et kommunikatiivset tüüpi on korpuseandmete puhul keeruline märgendada. Korpuseandmete puhul on tegemist üksikute lausetega, kuid suhtluseesmärgi hindamine eeldab ka laiema konteksti arvestamist. Võib ka öelda, et mõnes mõttes on piiratud ükskõik millise kirjaliku teksti suhtluseesmärgi määramise võimalus, kuna suhtluse olulised osad on ka intonatsioon jms, mis kirjalikku teksti üle ei kandu.

Semantiliste ID-siltide hulka taheti alguses lisada ka silt aspekti kohta. Aspekti määramine osutus aga keeruliseks, kuna eesti keeles puudub üheselt väljendatud aspekti kategooria ning seega prooviti aspekti määrata erinevate semantiliste tunnuste toel. Kategooria laialivalgusus aga saigi märgendamisel takistuseks. Et oli keeruline leida kriteeriume, mis oleksid rakendatavad kõikide lausete aspekti määramise puhul, oleks märgendamine põhinenud n-ö ühe lause kaupa hindamisel, mis aga oleks niigi subjektiivse semantilise märgendamise muutnud autori arvates ülemäära subjektiivseks. Tuleb aga mainida, et kõik siinmainitud tööst välja jäänud tunnused võivad verbi *nägema* käitumise kohta pakkuda uudset informatsiooni ning nende kasutuse või kasulikkuse kohta saab hinnangu anda ainult pärast nende tegelikku rakendamist.

### 3.2.4 Hierarhiline klasteranalüüs

Käitumisprofili analüüsimiseks on võimalik kasutada nii ühemõõtmelisi kui ka mitmemõõtmelisi meetodeid (Gries, Divjak 2009: 64), kuid enamasti rakendatakse analüüsil siiski viimast, täpsemalt hierarhilist klasteranalüüsi. Klasteranalüüs „korrastab“ andmed etteantud parameetrite järgi – milliseid parameetreid kasutada, oleneb uuritavast materjalist ja uurimiseesmärkidest. Käitumisprofili meetodi kontekstis on nendeks parameetriteks ID-sildid ja nende tasemed. Eelkõige on klasteranalüüs tööriistaks väga varieeruva andmestiku korrastamiseks, käitumisprofili puhul on see „toores“ andmestik sõna käitumisprofiil, mis näiteks selles magistritöös on vormiliselt 11 x 89 suurune andmematriks – see on suurus, mis ilma statistilise meetodite abita oleks uurijale palja silmaga hoomamatu.

Klasteranalüüs töötab sarnasuse (või erinevuse) põhimõttel<sup>5</sup> – koosinemismatriksit analüüsitakse nii, et omavahel kõige sarnasemad objektid sorteeritakse kokku, st sarnasus kahe ühes klastris esineva objekti vahel on maksimaalne ning sarnasus kahe erinevasse klastrisse kuuluva liikme vahel minimaalne. Ei ole olemas üht kindlat klasteranalüüsi meetodit – võimalik on kasutada erinevaid mõõte ja klastrite moodustamise tehnikaid, konkreetne valik sõltub uuritava andmestiku iseloomust.

Nagu öeldud, on klasteranalüüsi sisend sarnasustel või erinevustel põhinev matriks ning esimene valik tulebki teha kaugusemõõtude vahel. Kaugusmõõte on erinevaid, levinumad on Eukleidiline, Manhattani ja Canberra mõõt (Divjak, Fieller 2014: 416). Andmete eripära tõttu on käitumisprofili analüüsil tihti kasutatud Canberra mõõtu, kuna see käsitleb teistest paremini andmekogumit, kus on palju nullväärtusi (Divjak 2010: 132). Canberra mõõtu on kasutanud näiteks Divjak ja Gries (2006), Kask (2014), Berez ja Gries (2008); Eukleidilist aga Gries (2006). Kuna puuduvad reeglid, millist mõõtu valida, on kasulik vaadata just nimelt eelnevalt tehtud uurimuste poole, õnnestunud uurimuses

---

<sup>5</sup> Sarnasused ja erinevused esinevad mõnes mõttes koos, st need mõlemad on teisest tuletatavad. Üks põhjuseid ühe teisele eelistamiseks on nõuded ja piirangud, mida seavad statistikaprogrammid (Divjak & Fieller 2014:214). Kõige levinumalt kasutuses olev statistikaprogramm R nõuab sisendina näiteks erinevusmatriksit.

kasutatud mõõt võiks sarnase meetodi puhul olla ennast tõestanud ning võiks pakkuda häid tulemusi ka uues uurimuses.

Lisaks kaugusemõõdule<sup>6</sup> on oluline ka see, kuidas klastrid moodustatakse. Hierarhilise aglomeratiivse meetodi puhul alustatakse üksikust elemendist, liidetakse järgmine kõige sarnasem element jne, kuni kõik elemendid on liidetud, moodustades ühe suure klatri. Võimalik on kasutada ka meetodit, mis alustab ühest suurest klastrist ning struktuur moodustub jagunemise põhimõttel, või meetodit, kus antakse ette moodustatavate klastrite arv. Määrata tuleb ka objektide liitmise meetod.

Hierarhilise klasteranalüüsi üks graafilisi väljundeid on dendrogramm e jagunemispuu, mis kujutab visuaalselt, millised tähendused moodustavad sarnasuste põhjal rühmi. Nagu öeldud, on klasteranalüüs eelkõige vajalik andmetesse struktuuri loomiseks. Iseenesest ei paku analüüsi tulemus lahendusi küsimusele, miks just selline tulemus saadi – uurija osaks jääb leida seletused, miks on tähendused dendrogrammil just nii rühmitunud. Tulemust mõjutavad valitud mõõt ja meetod, kuid tagasi peaks vaatama ka oma andmetele ja arvestama nende iseärasustega. Siin võib uuesti mainida käitumisprofili analüüsi esimest osa, andmete märgendamist. Mida rohkem ja erinevamaid tunnuseid märgendamisfaasis määratakse, seda ülevaatlikumat pilti pakub ka dendrogramm. Kui näiteks piirduda ainult morfoloogiliste tunnustega, jaguneksid erinevad tähendused ka dendrogrammil selle järgi, millises vormis verb lauses esines. Ehkki see pakub kahtlemata väärtuslikku infot, ei ole võimalik ainult sellise info põhjal otsustada tähenduste jagunemise üle. Teisest küljest ei tohiks märgendamisprotsessi kaasata midagi üleliigset, kuna see muudab üldpildi liiga kirjuks.

Kuna klasteranalüüsi puhul tuleb uurijal teha palju valikuid, on olemas erinevaid statistilisi tehnikaid, mis aitavad hinnata, kui hästi klasteranalüüs on õnnestunud, mitut klastrit oleks mõistlik dendrogrammil eristada jms. Siluetitehnika (ingl *Silhouette plot*) näitab, kui hästi iga liige talle määratud klastrisse sobib. Visuaalselt näitavad seda nn

---

<sup>6</sup> Head ja põhjalikku ülevaadet klasteranalüüsi kasutamisest keeleandmete töötlemisel pakuvad nt Divjak ja Fieller (2014).



„siluetid“, iga klastrit kajastab üks siluett, mis näitab selle sidusust või killustatust. Kõikide liikmete peale arvutatakse kokku klastri keskmine väärtus. Mida lähemal on silueti laius 1-le, seda paremini on klasterdamine õnnestunud (seda sarnasem on objekt oma klastri teiste liikmetega ja seda erinevam teistesse klastritesse kuuluvatest liikmetest). Objektid/klastrid, mille silueti laius jääb 0 lähedale, võiksid sama hästi kuuluda mõnda teise klastrisse, ning kui väärtus on skaala miinuspoolel, on elemendid halvasti klasterdunud ning asetsevad kuskil klastrite vahel (st neid on keeruline mingisse klastrisse määrata). Siluettehnika on kasulik ka otsustamiseks, mitut klastrit dendrogrammil eristada. (Divjak, Fieller 2014: 432) Siluettehnika puhul saab analüüsi läbi viia  $n$ -õ mitmes variandis, määrates erineva klastrite arvu, millega algoritm arvestab.

Veel üks võimalus klasteranalüüsi tulemuste analüüsimiseks on hinnata nende kindlust. Seda on võimalik teha *bootstrap*-tehnikaga (ingl *bootstrapping*), arvutades klastritele välja  $p$ -väärtused<sup>7</sup>, mis näitavad, kui hästi andmestik klastreid toetab. Meetod  $n$ -õ simuleerib andmestikku, st tekitab olemasoleva andmestiku põhjal uued, sarnased andmestikud, võrdleb neid originaaliga ning hindab, millise tõenäosusega moodustuksid samasugused klastrid. (Robinson 2014: 99)

*Nägema* käitumisprofiili analüüsi tulemusel saadud dendrogrammi, selle tõlgendamist, klastrite arvu ja valideerimismeetodite tulemusi kirjeldatakse täpsemalt peatükkides 4.2 ja 4.3.

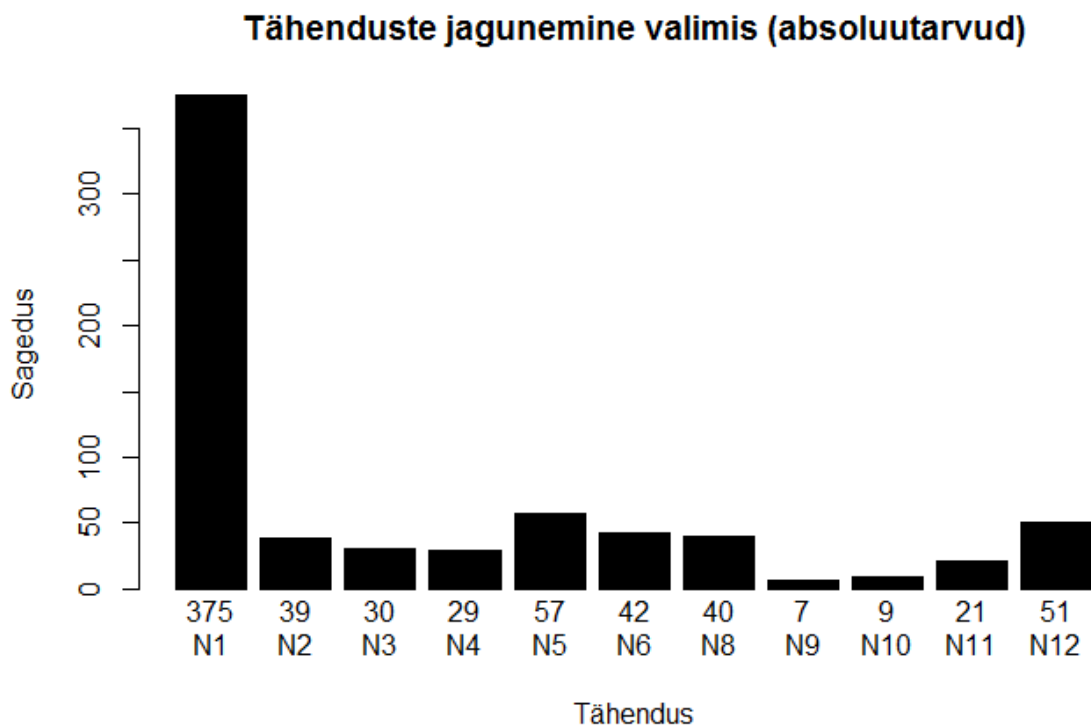
---

<sup>7</sup> Desagulier (Desagulier 2014:164) on välja toonud, et  $p$ -väärtusteks on neid nimetanud R-i paketi *pvclust* loojad, kuid pigem on tegemist kindluse hinnangu, kui klassikalise  $p$ -väärtusega.

## 4. Tulemused

### 4.1 Üldised tulemused

Ehkki magistritöö põhieesmärk on kirjeldada, mis infot annab tajuverbi *nägema* käitumisprofiili analüüs verbi polüseemia kohta, leidis juba meetodi II etapis (andmestiku märgendamise) huvitavaid seadus- ja isepärasid, mis väärivad eraldi välja toomist. Üldise struktuuri kirjeldamiseks lähtuvalt erinevatest tähendustest on joonisel 1 välja toodud kõikide tähenduste jagunemine korpuses absoluutarvudes. Jagunemine on oluline ka seepärast, et selle abil saab kirjeldada sõna polüseemia olemust/struktuuri (täpsemat kirjeldust vt ptk 4.4).



Joonis 1. Täheenduste jagunemine valimis absoluutarvudes

Jooniselt 1 on näha, et vaieldamatult kõige sagedasem *nägema* tähendus korpusvalimis oli N1 *nägema\_füüsiline*, moodustades 54% koguvalimist. Esinemise sageduse järgi järjestuvad ülejäänud tähendused järgmiselt (sulgudes on toodud, mitu protsenti koguvalimist tähendus moodustas): N5 *nägema\_ohte* (8,1%), N12 *nägema\_otsustama* (7,3%), N6 *nägema\_kogema* (6%), N8 *nägema\_nentima* (5,7%), N2 *nägema\_jagamatu* (5,6%), N3 *nägema\_kohtuma* (4,3%), N4 *nägema\_arusaama* (4,1%), N11 *nägema\_vaatama* (3%), N10 *elame\_näeme* (1,3%) ja N9 *nägema\_soovima* (1%). Üks oluline tulemus on, et tähendust N7 *nägema\_tunda\_saama* ei esinenud valimis kordagi, ehkki eeldati, et korpuses tuleb see tähendus välja kui eraldi tähendus. Selle tulemuse üle arutatakse pikemalt peatükis 5.

Verbi *nägema* süntaktiline käitumine avab nii mõndagi huvitavat nii verbi enda kui ka konstruktsioonide kohta, kus see esineb. Näiteks esines verb kokku 157 korda ahelverbi koosseisus. Peatükis 3.2.2 kirjeldatud ahelverbidega seotud ID-sildi tasemed valiti sellised abiverbide esinemissageduse tõttu. Modaalverbid ning abiverb *on* esinesid valimis läbivalt, samas kui tasemes „muu verb“ esinenud verbe kohtas materjalis vaid üksikutel kordadel. Modaalverbidest esines valimis ühe korra *tohtima* ja *tulema*, kuid näiteks *võima* 23 ja *tahtma* 28 korda. Peale modaalverbide ja *olema* esines *nägema* ahelverbis koos 23 muu verbiga, millest korduvalt verbidega *harjuma* (4 korda), *püüdma* (2 korda) ja *soovima* (2 korda). Näidetes 30–32 on toodud mõned näited ahelverbidest, mis on moodustatud muude verbidega peale modaalverbide või verbi *olema*.

(30) *Seniste ürituste ja tegevuse põhjal on Võru hakatud mingil määral nägema kogu Eestit folkloorikeskusena.* (EPL\_1995)

(31) *Politsei palub kõigil, kes juhtusid kiriku ümbruskonnas midagi kahtlast nägema, või kui kellelgi on infot juhtunu kohta, teatada sellest politseile telefonil 7 308 652.* (SLÕhtuleht\_2006)

(32) *Setod on ajast aega mõistnud muredega toime tulla, elamises ilusat näha ja ka kehvadel päevadel nalja teha.* (Postimees\_1996)

Konstruksiooni *on + da*-infinitiiv esines valimis 52 korda ehk 1/3 juhtudest (kokku oli valimis 157 lauset, kus *nägema* esines *ma*-supiinis või *da*-infinitiivis). Mõneti ei ole tulemus üllatav, kuna *olema* kuulub eesti keele tuumverbide hulka ning on eesti kirjakeele kõige sagedasem sõna (Tragel 2003:27). Konstruksioon *on + INF* on eesti keeles küll produktiivne, kuid piirangutega. EKG II järgi moodustavad ahelverbi verbiga *olema* ennekõike (kognitiiv)verbid *teadma, nägema, märkama, kuulma, tundma, kartma, arvama, lootma, ootama* jne. Enamasti väljendab selline konstruksioon perifrastilist passiivi või impersonaali. (Erelt jt 1993) Andmestikus esines *on + da*-infinitiiv kõige sagedamini koos tähendusega N12 *nägema\_otsustama* (27,5%) ning passiivilisus väljendubki mingi üldise seisundi kirjeldamises, pigem on rõhutatud staatilisus kui impersonaalsus, nagu näites 33, kus ei ole oluline agendi peitmine, vaid olukorra kirjeldus.

(33) *Muidugi on see juba näost näha, kes on kes.* (Ilukirjandus)

Verbiga *saama* moodustas *nägema da*-infinitiiv ahelverbi kõige sagedamini tähendustega N10 *elame\_näeme* (11%) ja N11 *nägema\_vaatama* (14%). Trageli ja Habichti järgi on *saama* tihedalt kasutatav ja polüsemne sõna, mis võib väljendada õnnestumise tähendust, kuid see *saama* tähendus ilmneb ainult kognitiivsete verbidega. Tegija on millekski võimeline, teisest küljest puuduvad ka välised asjaolud, mis õnnestumist takistaksid (st välised asjaolud on soosivad). (Habicht, Tragel 2014: 840–841) Korpusnäited illustreerivad seda ideed hästi, näites 34 on õnnestumist soosivaks tingimuseks näituse avatus, näites 35 aga asjaolu, et tuuril on plaanis nurgataguseid näidata.

(34) *Saab näha ka sajandi jooksul kolimise otstarbeks kasutatud vahendeid ja visata pilk pööningule ...* (Postimees\_1999)

(35) *Ma olen ise kaua aega Tallinnas elanud , aga kui see tuur kaasa sõita, saab näha ka nurgataguseid, kuhu muidu ei satu.*(SLÕhtuleht 2007)

Nägemise sihtpunktina hakkas korduvalt silma konstruksioon *noomen(part) + ma*-supiin(inessiiv) ehk *händkakke saaki varitsemas* või *Odidit õhus hõljumas*. Sihtpunktina esines sellist konstruksiooni 35 korral, kusjuures alati esines konstruksioon koos

tähendusega N1 nägema\_füüsiline. Supiini inessiiv väljendab mingi muu tegevusega samal ajal toimuvat tegevust (EKG), töö andmestikus alati nägemisega samal ajal toimuvat tegevust. Siin on huvitav just sellise konstruktsiooni valik – sama sisu oleks võimalik väljendada ka adjektiiviga kasutades, nt *on näha saaki varitsevaid händkakke*. See, et tihti on valitud just supiiniga versioon näitab, et oluline on nähtava sündmuse dünaamilisus. Niimoodi on lauses fookustatud tegevus, adjektiivilise konstruktsiooni puhul asetub fookusesse objekt. See on ka põhjus, miks see konstruktsioon sihtpunktina märgendati töös kui sündmus või situatsioon (vt alaptk 3.2.1), mitte kui objekt, kuigi vormiliselt on siin sihtpunktiks siiski noomeniga väljendatud objekt.

Huvipakkuv on ka vaadata laiendeid, mis erinevate tähendustega koos esinesid. Kokku esines laiendeid 85 lauses. Konstruktsioonisarnasust võib märgata laiendi „koht“ juures – nägemise sihtpunkt võib asuda nii konkreetsetes kohas nagu *silmis*, aga ka sellistes „kohtades“ nagu *hämärus* või *reklaamimine*. Kuigi intuiitiivselt on *silmis* ja *reklaamimises* väga erinevad kohad, võib neis siiski leida midagi sarnast. See sarnasus on seletatav ANUMA kujundskeemiga<sup>8</sup>. Oluline on siin sihtpunkti asetsemine millegi sees – füüsilise tähenduse puhul on näiteks *pisarad* silmas kui anumad. Millegi sees olemine tähendab ka seda, et anum ümbritseb objekti, seda näitlikustab hästi *hämäruuses*, kus on fookustatud just see, et *hämärus* on kõikjal objekti ümber. Abstraktse ja konkreetse piiri peal asub *uni*, mis oli ka oodatavalt selle alarühma kõige populaarsem laiend (13/32). *Une* puhul on keeruline öelda, kas selle „sisse“ paigutub ainult sihtpunkt või ka alguspunkt. Füüsiliselt asetub alguspunkt küll unest väljapoole, kuid samas ei vaadelda sihtpunkti tegelikult väljastpoolt.

---

<sup>8</sup> Kujundskeemide kohta saab täpsemalt lugeda Mark Johnsoni teosest „The Body in the Mind. The Bodily Basis of Meaning, Imagination, and Reason“ (Johnson 1987). Samuti on *nägema* polüseemia analüüsimisel kasutatud kujundskeeme autori bakalaureusetöös (Proos 2014).

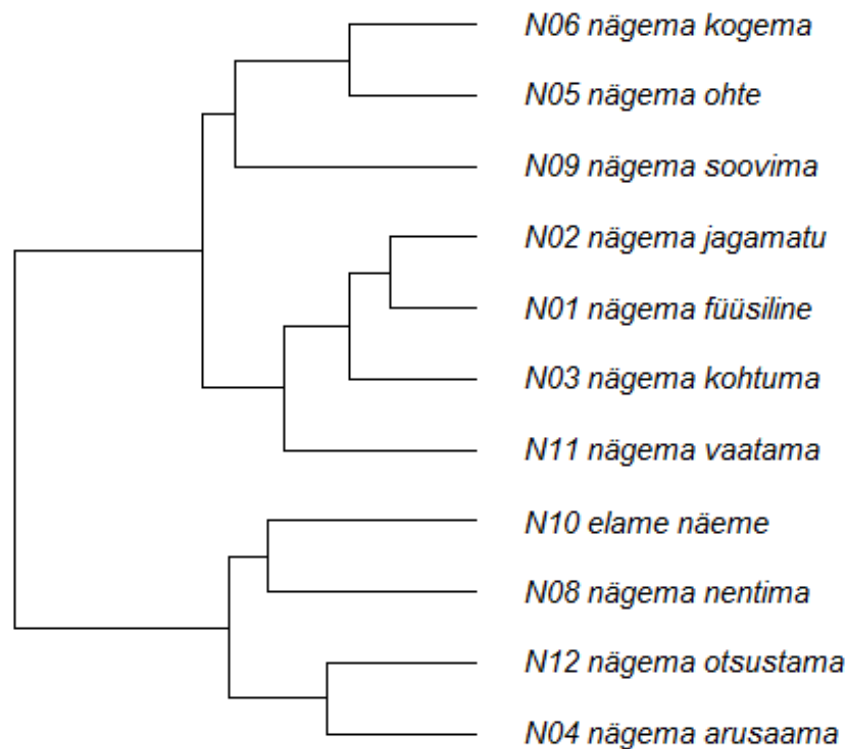
## 4.2 Hierarhilise klasteranalüüsi tulemused

Nagu mainitud peatükis 3.2.4, tuleb uurijal langetada klasteranalüüsi kasutamisel mitu otsust, millest olulisemad on erinevuse/sarnasuse mõõt ning klastrite moodustamise meetod. Otsuste langetamisel tuleb arvestada konkreetse andmestiku eripäraga, näiteks on käitumisprofiili analüüsi puhul tihti kasutatud Canberra mõõtu, kuna andmestikus esineb palju nullväärtusi. Ka selles magistritöös esines omajagu nullväärtusi, niisiis kasutati klasteranalüüsi läbiviimisel esialgu Canberra mõõtu. Kahjuks ei pakkunud Canberra mõõduga klasteranalüüs mõistlikult tõlgendatavaid tulemusi. Klasteranalüüsi puhul on oluline ka andmete intuitiivne tõlgendatavus, seega ei olnud Canberra mõõt õnnestunud valik. Canberra mõõduga tehtud klasteranalüüsi tulemusi analüüsiti ka *Silhouette ploti* valideerimistehnikaga, mis näitas samuti, et klasterdamine ei ole õnnestunud (keskmine siluetiväärtus oli 0). Palju parema tulemuse andis Eukleidilise kauguse kasutamine ning selles magistritöös ongi jäädud selle mõõdu juurde.

Liitmismeetodiks on magistritöös valitud Wardi meetod, kuna see annab tulemuseks võrdlemisi palju väikese liikmete arvuga klastreid, mis on hea just väikese hulga analüüsitavate objektide puhul (selle eelise on välja toonud Divjak (2010)). Tajuverbi *nägema* käitumisprofiilil põhineva aglomeratiivse hierarhilise klasteranalüüsi tulemused<sup>9</sup>, kasutades Eukleidilist mõõtu ning Wardi liitmise meetodit, on toodud dendrogrammina joonisel 2.

---

<sup>9</sup> Kogu statistiline analüüs (k.a valideerimine alapeatükkides 4.2.1 ja 4.2.2) tehti statistikaprogrammiga R (R Core Team 2015).



Joonis 2. Verbi *nägema* käitumisprofiilil põhinev dendrogramm

Dendrogrammi lugemisel on oluline kõrgustelg, telje suhtes kõige madalamal asetsevate klastrite liikmed on omavahel kõige sarnasemad ning kõige kõrgemal liituvad kõik tähendused kokku üheks klastriks, seega tähenduste kaugus kasvab alt üles. Joonisel 2 toodud dendrogramm on lihtsama loetavuse huvides ümber pööratud, nii et kõrgus kasvab paremalt vasakule.

Nii nagu ei näita dendrogramm ära seda, miks tähendused nii jagunevad, ei ütle meetod ka ette klastrite arvu. See, milliseid jagunemisi lugeda üheks klastriks, jääb samuti uurija enda otsustada. Nii võib joonisel 2 eristada kaht, seitset, kolme, nelja jne klastrit.

Õnnestunud dendrogrammi üks kriteeriume on selle intuiivselt tõlgendatavus, st moodustuvad klastrid peavad olema seletatavad piisava, kuid mitte liiga suure üldistusvõimega. Lisaks võib ka klastrite arvu tõlgendamisel abi olla valideerimistehnikatest (vt alaptk 3.2.1). Neid tingimusi arvesse võttes eristatakse magistritöös dendrogrammil analüüsiks järgmised neli klastrit:

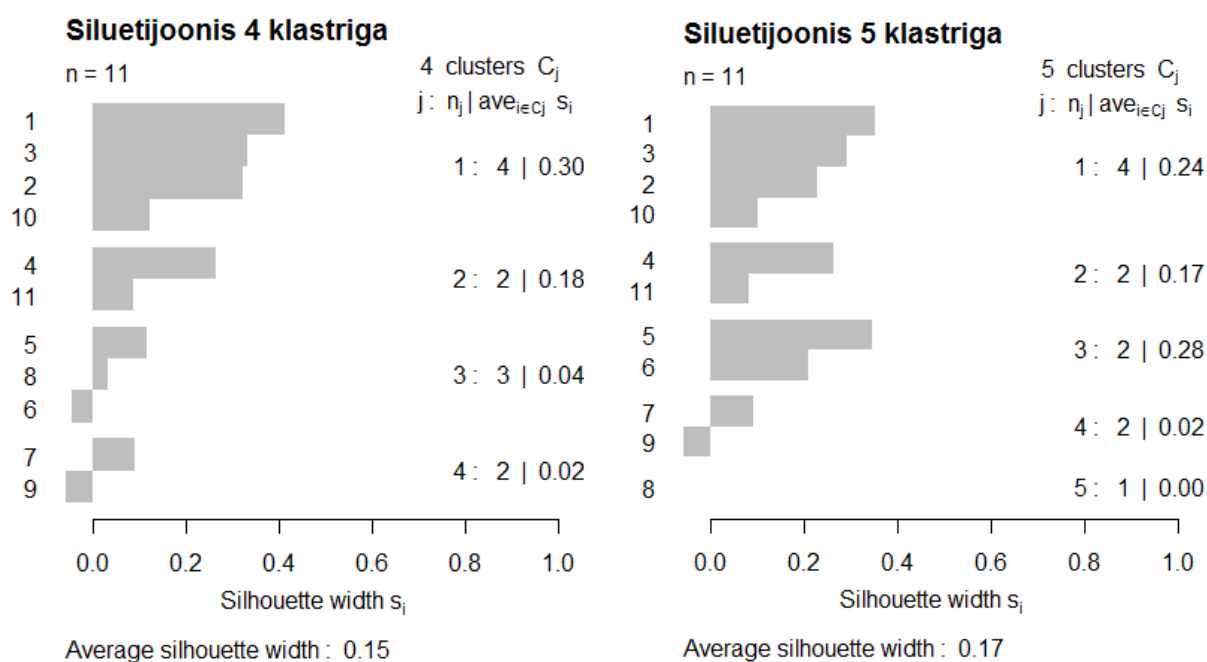
- a) N4 nägema\_arusaama: millegi üldise mõistmine, tajumine  
N12 nägema\_otsustama: millegi visuaalselt tajutava põhjal hinnangut andma; otsuseid tegema
  
- b) N10 elame\_näeme: midagi tulevikus kogema, millestki tulevikus aru saama; oluline on vastandus teadmatuse (praegu) ja teadlikkuse (tuleviku) vahel  
N8 nägema\_nentima: tähelepanu juhtides millelegi osutama
  
- c) N11 nägema\_vaatama: midagi süvenenult, tähelepanuga vaatama; jälgima  
N3 nägema\_kohtuma: füüsiline nägemiskogemus, kuid oluline on kontakti lisandumine, st nägemisele lisandub enamasti suhtlus  
N1 nägema\_füüsiline: füüsiline, nägemismeele kaudu tajumine  
N2 nägema\_jagamatu: nägemiskogemus, mis on subjektiivne; millegi (näiliselt) tajumine nägemisaistingutega, kuid puudub füüsilisuse piirang.
  
- d) N9 nägema\_soovima: väljendab ootusi mingi tulemuse / asjade seisu suhtes  
N6 nägema\_kogema: nägemise objekt on enamasti abstraktne, kuid väljendatakse kogemise/läbi elamise tähendust, mitte ainult tajumise.  
N5 nägema\_ohte: midagi abstraktset tajuma; tähendused N4 ja N5 on omavahel sarnased, kuid N5 on keskendub abstraktsele sihtpunktile.

Järgmises kahes alapeatükis antakse ülevaade kahe valideerimistehnika – siluetijoonise ja *bootstrap*-tehnikate – tulemustest. Valideerimistehnikad aitavad uurijal dendrogrammi tõlgendada ja/või leida kinnitust enda tõlgendusele.



#### 4.2.1 Valideerimine siluettehnikaga (ingl *Silhouette plot*)

Peatükis 3.2.1 mainiti, et siluettehnikaga on võimalik katsetada, millise klastrite arvuga dendrogrammi tõlgendamise lahendus on kõige optimaalsem. Joonisel 3 on toodud neljaklastriline ning joonisel 4 viieklastriline lahendus<sup>10</sup>.



Joonis 3. Neljaklastriline siluetijoonis *nägema* käitumisprofili põhjal

Joonis 4. Viieklastriline siluetijoonis *nägema* käitumisprofili põhjal

Mida kõrgem on siluettide üldine keskmine laiuse väärtus, seda paremini on klasterdamine õnnestunud ehk sellise arvu klastritega arvestamine peaks aitama kaasa dendrogrammi edukale tõlgendamisele. Kõrgem oli keskmine siluettiväärtus 5 klastriga analüüsil (0,17), mille puhul moodustus n-ö füüsilise tähenduse klaster tähendustest N1

<sup>10</sup> Kuna materjalis ei esinenud tähendust N7 ühtegi korda, on siluetijoonised alates 7 elemendist ühe võrra nihkes, st number 7 esindab tähendust N8 jne.

nägema\_füüsiline, N2 nägema\_jagamatu, N3 nägema\_kohtuma ja N11 nägema\_vaatama ja kaheliikmelised klastrid tähendustest N4 nägema\_arusaama ja N12 nägema\_otsustama, N5 nägema\_ohte ja N6 nägema\_vaeva ning N8 nägema\_nentima ja N10 elame\_näeme. Tähendus N9 nägema\_soovima jääb selle joonise järgi teistest klastritest eraldi.

Nelja klastriga lahenduse puhul on jäänud samaks esimesed kaks klastrit, samuti on nende väärtus peaaegu sama. Erinevus ilmneb joonistel tähenduste N5, N6 ja N9 jaotumises. Kui viie klastriga lahenduse puhul moodustas N9 n-ö eraldi klaster, siis neljaklastrilise lahenduse puhul on N9 liidetud ühte klastrisse tähendustega N5 ja N6. Silmatorkav on ka klastrite silueti väärtuste erinevus, kui klaster moodustavad ainult tähendused N5 ja N6, on klaster keskmine silueti väärtus 0,28, mis on ka kogu joonise kõige kõrgem. Samas juhul, kui klaster on liidetud tähendus N9, on klaster silueti väärtus ainult 0,04 – see on piisavalt lähedal nullile ning väärtus näitab seega, et liikmed oleksid võinud olla ka teisiti sorteeritud. Kusjuures on huvitav, et neljaklastrilise lahenduse puhul saab negatiivse väärtuse hoopis tähendus N6, mitte tähendus N9. On selge, et N9 klasterdamisel on siin määrav roll. Nelja klastriga lahenduse puhul ongi madala klasteri väärtuse põhjus selles, et N9 viib klasteri n-ö tasakaalust välja, leitakse küll mingi hulk sarnasusi, kuna ette on antud klasterite arv 4, kuid seeläbi kaotab ka viie klastriga lahenduse puhul tugevalt toetatud klaster oma väärtust.

Ka selle tehnika puhul tuleb tulemusi kriitiliselt analüüsida – lähtuda tuleb endiselt sellest, et klasterid peavad olema optimaalselt tõlgendatavad. Kuigi dendrogrammil viie klasteri tõlgendamine oli siluettehnika järgi optimaalsem, jäädigi töös siiski neljaklastrilise lahenduse juurde. Üldine keskmine väärtus on viieklasterilise vähe väiksem ning tähenduse N9 klasterdamine kokku tähendustega N5 ja N6 on samuti intuiitselt seletatav. Teisest küljest pakub näiteks viieklasteriline lahendus kinnitust, et N5 ja N6 kahese klasterina on võrdlemisi hästi õnnestunud ning see on ka andmete järgi lihtsalt seletatav. Üks siluettehnika kasulikke väärtusi ongi, et see näitab, milliste tulemuste suhtes olla „ettevaatlik“. Nagu ka klasteranalüüs ise, ei paku ka siluettehnika põhjendust, miks on klasterid nii jaotatud ning miks väärtused just sellised on. N9 puhul tasub kindlasti meele hoida, et seda tähendust esines terve valimi peale vaid 7 korda, kindlasti mängis ka väike sagedus väärtuse

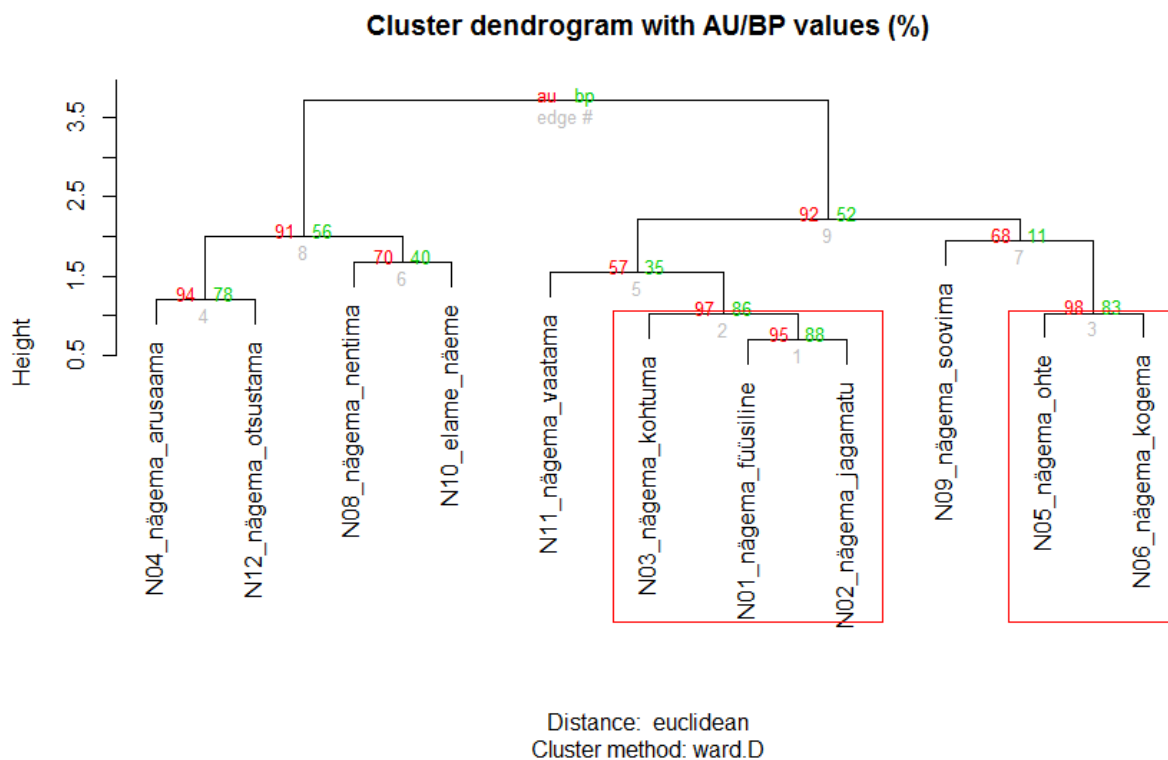
määramisel rolli. Seega kuigi enda intuitsioonile toetumine on tulemuste tõlgendamisel omal kohal, tuleks arvestada ka siluetijoonisega.

Kõige kindlamalt saab siluetijoonise järgi väita, et hästi on õnnestunud n-ö füüsiliste tähenduste klaster, mille väärtus on 0,30. Kusjuures siin on märgata, et tähenduse N11 silueti väärtus on klastris kõige väiksem, mis on ka silueti üldist väärtust ilmselt madaldanud, kuna teistel klastri liikmetel on kogu siluetijoonise õige kõrgemad väärtused. Klaster tähendustest N4 ja N12 on samuti võrreldes ülejäänutega küllalt hästi õnnestunud, selle väärtus on 0,17. Kõige madalam siluetiväärtus (0,02) on klastril, kuhu kuuluvad tähendused N8 ja N10, kusjuures viimase enda silueti väärtus on skaala miinuspoolel. Ilmselt tuleks ka siin arvesse võtta, et tähendust N10 esines valimis väga vähe.

#### 4.2.2 *Bootstrap*-tehnika (ingl *bootstrapping*)

*Bootstrap*-tehnika on klasteranalüüsi valideerimistehnika, mis näitab *p*-väärtuste abil, kui hästi on moodustunud klastrid andmestiku poolt toetatud. Joonisel 5 (järgmine lk) on toodud *nägema* käitumisprofiilil põhinev dendrogramm *p*-väärtustega. Analüüsil kasutati R-i paketti *pvclust*.

Iga dendrogrammi liitumiskoha juures on märgitud kolm arvu – vasakul AU (*Approximately Unbiased*) *p*-väärtus, paremal BP (*Bootstrap Probability*) *p*-väärtus ning liitmike all on kirjas klastri moodustamise järjekorranumber. Klastri headuse hindamiseks on parem AU-väärtus, mida lähemal on väärtus sajale, seda tugevamalt on klaster andmestiku poolt toetatud. (Divjak, Fieller 2014: 434) Joonise 4 järgi on kõige õnnestunud kahene klaster tähendustest N5 *nägema\_oh*te ja N6 *nägema\_kogema*, mille AU-väärtus on 98 (need klastrid on esile toodud kastidega klastrite ümber). Samuti on kõrge AU-väärtusega (97) klaster, mis moodustub tähendustest N1 *nägema\_füüsiline*, N2 *nägema\_jagamatu* ja N3 *nägema\_kohtuma*. Oluline on silmas pidada, et väärtus 97 iseloomustab N3 liitumist teise kahe tähendustega, mitte N2 ja N1 moodustatud klastri tugevust, kuigi ka viimaste omavahelise liitumise AU-väärtus on küllalt kõrge (95).



Joonis 3. *p*-väärtustega *nägema* käitumisprofiilil põhinev dendrogramm

Teistest märgatavalt väiksemad väärtused on klastritel, kuhu kuuluvad tähendused N9 *nägema\_soovima*, N10 *elame\_näeme* ja N11 *nägema\_vaatama*. Klasteri väärtus, kuhu liitub N9, on 68, N10 puhul on väärtus 70 ja N11 puhul 57, mis on ka terve dendrogrammi kõige madalam väärtus. Siluetijooniste ja *p*-väärtustega dendrogrammi vahel võib märgata palju sarnasusi. Mõlema valideerimistehnika järgi on kõige paremini klasterdunud tähendused N1, N2 ja N3 ning tähendused N5 ja N6. Kui siluetijoonistest võis järeldada, et N9 klasterdamine ei ole eriti kindel, siis sama kinnitavad ka *p*-väärtused. N9 liitumine tähendustega N5 ja N6 on 68% kindel, samuti on selle tähenduse liitumise koht dendrogrammi *y*-telje suhtes kõige kõrgemal, mis näitab, et N9 ei olnud teiste klasterliikmetega väga sarnane. Sama tendentsi on näha ka N11 liitumisel, sel juhul on tõenäosus ainult 57%. Samuti peegeldab siluetijooniste tulemusi tähenduste N8 ja N10 liitumise tõenäosus, mis on 70%.

Mõlemad valideerimistehnikad näitasid, et mõne tulemuse suhtes tuleks olla ettevaatlik, st mitte anda nendele liiga suurt tõlgenduslikku väärtust. Samas, nagu korduvalt mainitud, ei saa täielikult lähtuda ka statistiliste testide tulemustest, uurija intuitsioon on oluline nii dendrogrammi tõlgendamisel kui ka valideerimistehnikate tõlgendamisel ja hindamisel. Kahtlemata avavad aga valideerimistehnikad tulemusi rohkem, kui esmapilgul lihtsalt dendrogrammi vaadates näha on. Oluline on leida tasakaal statistika ja intuitsiooni vahel, kuna keeleandmete tõlgendamisel on mõlemal mängida oluline roll. Järgmises peatükis näidatakse nii sagedusandemete kui ka muu põhjal, kas ja milliseid seletusi moodustunud klastritele on võimalik leida.

#### 4.3 Moodustunud klastrite iseloomustus

Dendrogrammi järgi jagunevad tähendused kõigepealt kaheks:

- 1) N4 nägema\_arusaama, N12 nägema\_otsustama, N8 nägema\_nentima, N10 elame\_näeme;
- 2) N11 nägema\_vaatama, N3 nägema\_kohtuma, N2 nägema\_jagamatu, N1 nägema\_füüsiline, N9 nägema\_soovima, N5 nägema\_ohte, N6 nägema\_vaeva.

Jagunemist on võimalik põhjendada, lähtudes sihtpunkti tüübist – ühte harusse jäävad tähendused, mille puhul nägemise sihtpunkt on enamasti piiritletud, st tegemist on kas konkreetse või abstraktse objektiga; teise harusse jäävates tähendustes on *nägema* sihtpunkt üldisem ja „hajusam“, st tegemist on mingi situatsiooni, sündmuse või muu sarnasega. Tabelis 2 on välja toodud tähenduste ning sihtpunkti tüüpide koosinemised protsentuaalselt.

Tabel 2. Nägema tähenduste jagunemine sihtpunkti tüübi kaupa (%)

TÄHENDUS	SP objekt	SP situats.	SP sünd.	ei kehti <sup>11</sup>	KOKKU
N1 nägema_füüsiline	66,9 %	6,6%	11,7%	14,6%	100%
N2 nägema_jagamatu	76,9%	10,2%	5,12%	7,6%	100%
N3 nägema_kohtuma	<b>93,3 %</b>	0%	0%	6,6%	100%
N4 nägema_aru saama	3,4 %	<b>89,6%</b>	<b>6,8%</b>	0%	100%
N5 nägema_ohte	<b>96,4 %</b>	0%	0%	3,5%	100%
N6 nägema_kogema	78,5%	2,3%	4,7%	14,2%	100%
N8 nägema_nentima	15%	<b>42,5%</b>	<b>22,5%</b>	20%	100%
N9 nägema_soovima	85,7%	0%	14,2%	0%	100%
N10 elame_näeme	11,1%	<b>44,4%</b>	<b>11,1%</b>	33,3%	100%
N11 nägema_vaatama	<b>90,4%</b>	0%	4,7%	4,7%	100%
N12 nägema_otsustama	<b>50,9%</b>	<b>41,7%</b>	5,8%	1,9%	100%

Tabelis 2 on paksus kirjas märgitud tähendused, mille puhul oli suurim protsentuaalne vahe sihtpunkti ID-sildi tasemetel „SP objekt“, „SP situatsioon“ ja „SP sündmus“ vahel. Tabeli 2 ja dendrogrammi vahel on kerge paralleele tõmmata – tähendused N4 nägema\_arusaama, N8 nägema\_nentima ja N10 elame\_näeme asetsevad dendrogrammil ühes harus, tähendused N3 nägema\_kohtuma, N5 nägema\_ohte ja N11 nägema\_vaatama teises harus. Erandina tuleb siin välja tuua tähendus N12 nägema\_otsustama, mille sihtpunktid jagunevad situatsiooni ja objekti vahel enam-vähem võrdselt, koosinemine objekti-sihtpunktiga on veidi sagedasem, kuid dendrogrammil asetseb N12 siiski koos tähendustega N4, N8 ja N10.

Dendrogrammil (joonis 2) visuaalselt üleval olev suurem klaster jaguneb üheks nelja- ja üheks kolmeliikmeliseks klastriks. Neljaliikmelise klastrit moodustavad tähendused N11, N3, N1 ja N2 ning kolmeliikmelise klastrit tähendused N9, N6 ja N5. Klastrite sellist jagunemist saab põhjendada näiteks sihtpunkti konkreetsuse või abstraktsuse kaudu. Huvitav on siinjuures ära märkida, et tähendus N2 (jagamatu nägemiskogemus) ja tähendus N1 (füüsiline nägemine) hargnevad dendrogrammil kõige madalamalt, mis tähendab, et need

<sup>11</sup> Siin ja edaspidi tabelites toodud kategooria *ei kehti* väljendab juhte, kus konkreetne ID-silt ei olnud määratav, sellised juhud on vaja märgendada meetodi arvutusliku poole õigeks toimimiseks. Nt tabelis 3 tähendab *ei kehti*, mitmel protsendil lausetest sihtpunkt puudus või oli see varjatud.

tähendused on omavahel kõige sarnasemad (või täpsemalt, nende erinevuste kaugus on kõige väiksem).

Kui mõelda sellele, mis eristab jagamatut nägemiskogemust füüsilisest nägemiskogemusest, ei ole konkreetseid tunnuseid eriti lihtne välja tuua. Unes või vaimusilmas võib näha täpselt samasuguseid asju nagu füüsilises maailmas, tihti ongi kahe nägemiskogemuse väljendamine erinevus see, et täpsustatakse, et midagi nähti unes. Tabelis 3 on välja toodud laiendi ja tähenduste koosinemise sagedused protsentides.

Tabel 3. Tähenduste N1 ja N2 koosinemine ID-sildiga „laiend”.

TÄHENDUS	allikas	koht	määrus	trajektoori täpsustus	vahend	ei kehti	KOKKU
N1 nägema_füüsiline	3,20%	2,1%	0,80%	1,3%	2,7%	89,9%	100%
N2 nägema_jagamatu	0%	33,3%	2,6%	0%	0%	64,1%	100%

Nagu tabelist 3 näha, esines tähendusega N2 kõige sagedamini koos laiendiga „koht“, samas kui tähendusega N1 esines see laiend vaid 2,1% juhtudest. Väga suure osa nendest laienditest moodustabki laiend *unes*, ning kindlasti võib sellist laiendit pidada üheks tähendust N2 eristavaks elemendiks. Kusjuures on huvitav märkida ära, et kokku esines erinevaid laiendeid tähendusega N2 rohkem kui tähendusega N1.

Ka N3 nägema\_kohtuma sobitub tähendustega N1 ja N2 hästi samasse klastrisse, kuna sisuliselt kannab N3 endas siiski konkreetset füüsilise nägemise tähendust, kuid see tähendus on piiritletum. Piiritletus väljendub N3 puhul just selles, milline on lemma tüüp ehk nägemise sihtpunkt, N3 puhul on see alati inimene, kuid väljendatud võib see olla erinevalt (nt asesõna, üldnimetus, pärisnimi – vt täielikku tabelit töö lisast nr 1). Tähendused N1, N2, N3 ja N11 eristuvad teistest tähendustest ka sihtpunkti abstraktsuses-konkreetsuses. Tabel 4 näitab, kuidas jagunesid tähendused ID-sildi „sihtpunkti abstraktsus“ raames.

Tabel 4. Tähenduste koosinemine sihtpunkti tüübi (abstraktsus/konkreetsus) kaupa<sup>12</sup>.

TÄHENDUS	SP abs.	SP konk	SP varjatud	ei kehti	KOKKU
N1 nägema_füüsiline	<b>0,5%</b>	<b>84,8%</b>	8,8%	5,8%	100%
N2 nägema_jagamatu	<b>17,9%</b>	<b>74,3%</b>	7,6%	0%	100%
N3 nägema_kohtuma	<b>0%</b>	<b>93,3%</b>	3,3%	3,3%	100%
N4 nägema_arusaama	66%	34,4%	0%	0%	100%
N5 nägema_ohte	80,7%	15,7%	3,5%	0%	100%
N6 nägema_kogema	61,9%	23,8%	11,9%	2,3%	100%
N8 nägema_nentima	20%	60%	2,5%	17,5%	100%
N9 nägema_sooxima	28,5%	71,4%	0%	0%	100%
N10 elame_näeme	33,3%	33,3%	22,2%	11,1%	100%
N11 nägema_vaatama	<b>4,7%</b>	<b>90,4%</b>	4,7%	0%	100%
N12 nägema_otustama	52,9%	45%	1,9%	0%	100%

Tähendused N1, N3 ja N11 eristuvad teistest tähendustest abstraktsete ja konkreetsete sihtpunktide jagunemise poolest. On väga selgelt näha, et nende tähenduste puhul on abstraktsete sihtpunktide osakaal nullilähedane (täheenduse N3 puhul koguni 0%). Samas esineb just nende tähendustega kõige tihemini konkreetseid sihtpunkte. N2 puhul on küll konkreetsete ja abstraktsete sihtpunktide osakaalu vahe väiksem, kuid ka need tähendused jälgivad sama tendentsi.

Kuigi N11 nägema\_vaatama käitub eelkõige tähenduste N1 ja N3 sarnaselt, siis valideerimistehnikad näitasid, et tähendus ei ole klastrisse väga hästi sobitunud. Nagu peatükis 4.2.1 mainiti, võis oluliseks saada tähenduse võrdlemisi väike sagedus. Siin ongi oluline vaadata ka seda, millega ja kui palju tähendus koos esines – sarnasus teiste klastriliikmetega on väga suur. Samas, kuna tähendust esines vähe, siis ei ole ka selle konstruktsiooniline varieeruvus valimi põhjal eriti suur, st võrreldes teiste klastriliikmetega esines sagedustabelis rohkem nullväärtusi.

<sup>12</sup> Tasemetel „SP varjatud“ ja „ei kehti“ kirjeldus on toodud peatükis 3.2.1.



Teise „objektile suunatud“ tähenduste klasteri tähendused N6 nägema\_kogema, N5 nägema\_ohte ja N9 nägema\_soovima. Dendrogrammi üldises pildis on selle klasteri liikmed N5 ja N6 eelnevalt mainitud paari N1 ja N2 järel sarnasuselt teisel kohal, st ka nende erinevuste vaheline kaugus on väike (liitumine asub dendrogrammil madalal). Ka valideerimine näitas, et nende kahe tähenduse liitmine on hästi õnnestunud. Vaadates eespool toodud tabeleid, selgub, et tähenduste N5 ja N6 käitumisprofiil on sarnane, mõlema sihtpunkt on enamasti abstraktne objekt. Et aga dendrogramm väljendab, et sarnasus ei ole täielik, võib andmetest otsida neid kaht tähendust kõige paremini eristavat faktorit. Nagu mainitud peatükis 3.2.1 märgendati seal, kus võimalik eraldi ka sihtpunkti lemma tüüp. Tähenduste koosinemine erinevate lemma tüüpidega on toodud tabelis 5.

Tabel 5. Tähenduste N5 ja N6 koosinemine lemma tüüpidega

LEMMA TÜÜP	N5 nägema_ohte	N6 nägema_kogema
asesõna	14%	16,7%
füüsiline elutu	7%	7,1%
inimene	0%	0%
lind	0%	0%
mitte füüsiline elutu	24,6%	16,7%
muu abstraktne	50,9%	9,5%
muu elus	0%	0%
nimetus	3,5%	0%
pärisnimi	0%	0%
tunne	0%	45,2%
ei kehti	0%	4,8%
KOKKU	100%	100%

Tabelist 5 nähtub, et tähendused N5 ja N6 esinevad enamike sihtpunkti lemma tüüpidega koos enam-vähem võrdsel määral. Küll aga erineb koosinemine kahe lemma tüübiga märkimisväärselt – need tüübid on „muu abstraktne“ ja „tunne“. Tähendus N5 esineb 50% juhtudest koos muu abstraktse lemmaga, samas kui N6 puhul on see protsent vaid 8,5. Kuid lemma tüübiga „tunne“ esineb jällegi koos tähendus N6 45% kordadest, kuid N5 mitte ühtegi korda. Sihtpunkti tüübis peitubki üks põhjuseid, miks need tähendused erinevad on.

N9 on aga mõneti füüsilise ja abstraktse tähenduse piirimail – ehkki N9 puhul on tendents esineda koos konkreetse elusa sihtpunktiga, on võimalikud ka abstraktsed sihtpunktid, mis teevad selle tähenduse vähem piiritletuks kui näiteks tähendused N1 ja N3, aga samas ka N5 ja N6. Valideerimistehnikad näitasid, et see tähendus on teistest kõige rohkem „klastrite vahel“, st selle määramine ükskõik millisesse klastrisse on keeruline. Jällegi – tuleb meeles pidada, et seda tähendust esines valimis kõigist tähendustest kõige vähem, lisaks on tegemist tähendusega, mis võib esineda koos nii abstraktsete kui ka konkreetsete sihtpunktidega.

Kõige kõrgemalt jagunev teine klaster jaguneb omakorda kaheks kaheliikmeliseks klastriks, ühe klatri moodustavad tähendused N4 nägema\_arusaama ja N12 nägema\_otsustama ning teise N10 elame\_näeme ja N8 nägema\_nentima. N10 ja N8 on kõikidest tähendustest kõige vähem varieeruvad, st neid võib pidada konstruktsioonitähendusteks. Seda tendentsi on hea vaadata just morfosüntaktilise märgendamise pealt. Et konstruktsioonitähenduse tähendus ilmnebki ainult teatud konstruktsioonides, näitab tendents esineda ainult koos mõne kindla morfoloogilise kategooriaga, et kõnealune tähendus on piiratud. Tabelis 6 on toodud tähenduste N8 ja N10 esinemine koos erinevate verbivormidega, võrdluseks on toodud tähenduse N1 füüsiline\_nägemine ja N5 nägema\_ohte koosesinemised.

Tabel 6. Tähenduste N8, N10, N1 ja N5 koosesinemine ID-sildi „verbi vorm“ tasemetega

VERBI VORM	N1 nägema_füüsiline	N5 nägema_ohte	N8 nägema_nentima	N10 elame_näeme
pl1	3,5%	1,8%	0%	22,2%
pl2	1,6%	1,8%	20%	0%
pl3	5,1%	14%	0%	0%
s1	25,9%	26,3%	5%	11,1%
s2	4,3%	5,3%	37,5%	33,3%
s3	45,9%	31,6%	35%	33,3%
da-inf.	2,4%	5,3%	2,5%	0%
ma-inf.	0,5%	1,8%	0%	0%
des-vorm	3,5%	0%	0%	0%

<i>nud</i> -partitsiip	0,3%	0%	0%	0%
<i>tud</i> -partitsiip	1,6%	1,8%	0%	0%
ei saa määrata	5,1%	1,8%	0%	0%
ei kehti	0,5%	8,8%	0%	0%
KOKKU	100%	100%	100%	100%

Tabel 6 näitlikustab, kui vähe konstruktsioonilist varieeruvust esineb tähenduste N8 ja N10 vahel võrreldes tähendusega N1 ja N5. Kuigi tähendused N1 ja N5 ei asu samas klastris, esineb verb *nägema* nendes tähendustes väga erinevates vormides, tähendused N8 ja N10 aga vastupidi ei esine koos enamike ID-sildi „verbi vorm“ tasemetega. Sama võib märgata ka selliste ID-siltidega nagu „laiend“ (kumbki tähendus ei esinenud ühegi tasemega) või „ahelverb“ (N8 ei esine ühegi tasemega, N10 3 tasemega 11st (vt ka lisa 1)).

Käitumisprofiiliga saab ka iseloomustada, milline on see konstruktsioon, kus verb saab tähenduse N8 või N10. Kuigi keele variatiivsus on niivõrd suur, et alati võib leida erandeid (seda enam, et käitumisprofiili puhul on tegemist piiratud valimiga), saab välja tuua vähemalt tendentsid, mida konstruktsioonid jälgivad. Nii võib näiteks välja tuua, et tähenduse N8 *nägema\_nentima* puhul ei esine verb *nägema* ahelverbi koosseisus, vaid alati iseseisvalt, N10 *elame\_näeme* aga võib ahelverbi moodustada verbidega *on* ja *saab*.

Dendrogrammil visuaalselt kõige allpool asub klaster, kuhu kuuluvad N4 *nägema\_arusaama* ja N12 *nägema\_otsustama*. Nagu enne välja toodud, siis N4 sihtpunkt on enamasti kas sündmus või situatsioon, nagu ka tähendustel N10 ja N8, kuid tähenduse N12 puhul jagunevad sihtpunktid tüüpide vahel enam-vähem võrdselt. Samas tuleb siin tähele panna ka seda, et kõige kõrgemas jagunemises teise klastrisse jäävate tähenduste puhul on tugevalt ülekaalus sihtpunkt kui objekt, seega ei saa väita, et N12 asetumine oma positsioonile oleks vasturääkiv. Sarnast reeglipära on märgata ka kui vaadata teiste ID-siltide tasemete koosinemisi tähendusega N12. Mõne ID-sili puhul sarnaneb N12 üht, mõne puhul teist tüüpi ID-sildi tasemetega. Kui nüüd jällegi tagasi mõelda sellele, millist tähendust N12 üldse endas kannab, ongi see tähendus n-ö kahe maailma vahel. Näited 36–38 annavad aimu, kui erinevalt see tähendus võib esineda.

(36) *Et mina pole see, kes saaks enda kõrval õnnelikku Carli näha.* (Ilukirjandus)

(37) *Nähes temas nii palju mandumise märke, söandasin talle ka lõpuks silma vaadata.*  
(Ilukirjandus)

(38) *Viimasel ajal märgivad arstid ja ämmaemandad sünnitustoas seinale punaseid ja siniseid täppe - siis on näha, mitu tüdrukut ja poissi on ilmale tulnud.*  
(Postimees\_2000)

Näites 36 on sihtpunkt konkreetne inimene, kuid talle on omistatud teatud tundeseisund, mida ei ole võimalik näha muud moodi, kui tõlgendada seda läbi mingite teiste, füüsiliste tunnuste. Sama kehtib ka näite 37 puhul, kuid siin on lause vormistatud nii, et sihtpunkt iseenesest on abstraktne, ehkki tunnused ilmselt füüsilised. Samas aga näide 38 näitlikustab võimalust, et nii see, mille kohta järeldus tehakse kui ka see, mille abil seda tehakse, võivad olla konkreetset. Seega on siin formaalsete tunnuste alusel keeruline tähendust üksüheselt kuhugi jagada.

Klastranalüüsi tulemusel moodustunud klastritele on võimalik seletusi tuua nii toetudes konkreetsetele statistilistele andmetele kui ka laiemalt *nägema* erinevatele tähendusvarjunditele mõeldes. Käitumisprofiili analüüs pakub erinevate tähenduste kokku rühmitumisele alternatiivseid seletusi kui sorteerimiskatse (Proos 2014) tulemuste analüüs. Peatükis 6 tuuakse välja peamised erinevused sorteerimiskatse tulemuste ja käitumisprofiili analüüsi tulemuste vahel, ning pakutakse neile võimalikke seletusi.

#### 4.4 *Nägema* tähendusvõrgustikust

Peatükis 2.2 mainiti, et käitumisprofiili analüüs on hea tööriist prototüüpse tähenduse väljaselgitamiseks. Tähendusvõrgustiku prototüüpse tähenduse väljaselgitamine pakub endiselt keeleteaduses palju kõneainet, ühtset meelt, mis on selle jaoks parim meetod ning mida üldse pidada üheselt prototüüpseks tähenduseks, ei ole endiselt. Griesi (2006: 75–77) järgi peaks prototüüpne tähendus ehk tähendusvõrgustiku keskpunkt olema keeleteadustaja

meeles kõige tugevamini kinnistunud, produktiivne ning seega vormiliselt kõige rohkem varieeruvam, seega on seda sagedusandmete põhjal võimalik määrata. Teisalt aga on väidetud, et sagedusandmed ei näita tingimata seda, kuidas keelekasutaja tegelikult erinevaid tähendusi tajub, st sageduse põhjal ei saa kindlalt väita, milline tähendus on keelekasutaja jaoks keskne ning millised asuvad tähendusvõrgustiku äärealadel (Schmid 2010). Hea eestikeelse ülevaate prototüüpse tähenduse määramise seostest katseliste meetodite ja korpuse sageduse toel leiab näiteks Ann Veismanni (2009) doktoritööst.

Prototüüpse tähenduse toel peaksid olema kirjeldatavad teised tähendused, polüseemia üks tunnuseid on tähenduste omavaheline seotus ning see seotus peaks olema seletatav tähenduse ülekande mehhanismidega. Tähendusvõrgustiku illustreerimiseks ei ole kasutatud traditsioonilist graafilist meetodit, kuna joonisel on keeruline piisavalt hästi välja tuua kõikide tähenduste ja tähenduste muutumise protsesside omavahelised seosed.

Käitumisprofiili järgi võib väita, et *nägema* prototüüpne tähendus on *füüsilise aistinguga, silmadega tajuma* (N1). Seda tähendust leidis valimis ülekaalukalt rohkem kui teisi, moodustades 54% koguvallimist. Teine oluline aspekt on tähenduse konteksti varieeruvus, st prototüüpne tähendus peaks esinema võimalikult paljude erinevate ID-sildi tasemetega. 89-st ID-sildi tasemest ei esinenud N1 koos viie tasemega ning oli seega kõigi tähenduste raames kõige väiksema nullväärtuste arvuga tähendus. Kuna prototüüpne tähendus peaks olema produktiivne ning suhteliselt piirangutevaba, on ilmekas vaadata just neid ID-silte, millega see tähendus koos ei esinenud.

Morfosüntaktilistest ID-siltidest ja nende tasemetest ei esinenud N1 kordagi käskivas kõneviisis ega moodustanud ahelverbi modaalverbidega *tulema* ja *tohtima*. Käskiva kõneviisi kasutamine selles tähenduses ei ole mõneti võimalik – *nägema-vaatama* vastanduses on *nägema* n-ö mitte-sihilik kogemisverb ning mitte-sihilikku tegevust ei saa panna käskivasse vormi, samas kui *vaata!* on vastuvõetav vorm. Siin tuleb rõhutada, et imperatiiv pole võimalik just nimelt selle tähenduse puhul – nii pea kui *nägema* käskivas kõneviisis esineb,

on tegemist juba teise tähendusega, kusjuures enamasti tähendusega N8 nägema\_nentima<sup>13</sup>. Samas rõhutab ka EKSS selle tähenduse kirjes, et kuigi tegemist on imperatiivi vormidega, kasutatakse neid nõrgenenud tähenduses.

Nii imperatiiv, *tulema* kui ka *tohtima* väljendavad deontilist modaalsust. Konstruktsioon *tuleb* + da-infinitiiv väljendab kohustust, kusjuures tegemist on konstruktsioonimodaalsusega, kuna teistes vormides modaaltähendus ei väljendu (Pajusalu jt 2004: 31). Jällegi – millekski mitte-sihilikuks ei saa kedagi kohustada, seega konstruktsioon *tuleb näha*, kus nägema esineb tähenduses N1 nägema\_füüsiline, kõlab kummaliselt. Sama ei saa väita aga lubamise kohta ning kuigi valimis ei esinenud kordagi konstruktsiooni *tohib näha*, ei ole sellise ahelverbi kasutuskonteksti väga keeruline ette kujutada.

Lisaks ei esinenud tähendus N1 nägema\_füüsiline koos kahe semantilise ID-sildi tasemega, lemma tüübi ID-sildi tasemetega „tunne“ ja „muu abstraktne“. Nagu ka imperatiivi puhul, on siingi tegemist nähtusega, kus selle ID-sildi tasemega koos esinedes ei oleks enam tegemist tähendusega N1, vaid ühega teistest tähendustest. Kui postuleerida N1 kui prototüüpne *nägema* tähendus, siis koosinemise abstraktse sihtpunktidega on võimalikuks muutnud metafoorne ülekanne. Metafoorne ülekanne toimub füüsilisest valdkonnast abstraktsesse, kusjuures metafoori allikvaldkonna ja sihtvaldkonna vahel on hulk sarnasusi, mis võimaldavad keelekasutajal sellist ülekannet teha ning ülekanne on motiveeritud igapäevasest kogemustest (Lakoff 1987: 287).

Nägemistaju valdkond on kõige tihedamalt seotud teadmiste/intellekti valdkonnaga, põhiliselt seob neid kaht objektiivsus – nägemismeele kaudu saadav informatsioon on tajudest kõige objektiivsem, kuna see on „jagatav“, st üht ja sama asja võivad näha erinevad inimesed, samamoodi objektiivsena nähakse ka teadmisi. Sarnane vastandus tekib ka sellest, et nägemise kaudu on võimalik infot koguda distantsilt, samuti on teadmised „kaugel“ individuaalsest isiksusest, seda eriti vastanduses emotsionaalse või tunde-valdkonnaga.

---

<sup>13</sup> Käskivas kõneviisis esinevad andmestikus veel tähendused N10 elame\_näeme ja N6 nägema\_kogema.

(Sweetser 1990: 37–39) Ülekannet füüsilisest valdkonnast intellekti valdkonda esindavad *nägema* puhul väga selgelt tähendused N4 *nägema\_arusaama*, N5 *nägema\_ohte* ja N8 *nägema\_nentima*, mille kõigi puhul on võimalikuks saanud millegi abstraktne nägemine ja seda just millegi mõistmise või millestki aru saamise tähenduses.

Selle mõttearenduse taustal on aga huvitav käsitleda tähendust N6 *nägema\_kogema*. Ka füüsiline nägemine on teatud sorti kogemine, tähenduse N6 puhul on toimunud kitsenemine ning uue tähenduse keskmes on ainult kogemise aspekt. Kui aga nii füüsiline nägemine kui ka metafoorne, teadmistega seotud nägemine on mõlemad pigem objektiivsed, siis kogemise tähendus seda ei ole, või ei ole seda vähemalt samal viisil. Üldisel tasandil võib muidugi kogemusi tõesti mõista jagatuna – sarnaste kogemuste olemasolu ongi kognitiivses raamistikus üks eeldusi, kuidas me metafooridest (ning ka laiemalt keelest) aru saame.

Semantilise täpsustuse protsess on ka tähenduste N3 *nägema\_kohtuma* ja N11 *nägema\_vaatama* taga – esimese puhul on keskmesse asetatud ainult ühte tüüpi sihtpunkti nägemine ning lisandunud on ka suhtluslik aspekt; teise puhul on aga täpsustatud nägemise aspekt ja viis, st oluline on keskendumine sellele, mida nähakse, ning olemas on vältimatu kogemise lõppemise hetk (mis ei sõltu nägijast). Tähendust N10 *elame\_näeme* võib võtta kui täpsustusena nii tähendusest N6 kui ka N4/N5, fookustatud on see, et arusaamine või kogemine saab toimuma tulevikus. Tähendus N9 *nägema\_sooxima* on ühest küljest seotud küll tuleviku fookustamisega, teisest küljest aga ka tähendustega N4 ja N5 (N9 ei pea tingimata väljendama tulevikulisust). Samuti on omavahel seotud tähendused N9 ja N12 *nägema\_otsustama* – kummagi puhul ei ole ranget piirangut sihtpunkti konkreetsuse või abstraktsuse suhtes, mõnes mõttes võib nende tähenduste puhul näha nii semantilist täpsustumist kui ka üldistumist.

Tähenduse N2 alla koonduvad nn jagamatud nägemiskogemused, mis iseenesest on üldistus füüsilise nägemise tähendusest – läbi üldistuse on võimalikuks muutunud ka millegi nägemine, mis ei vasta ühele füüsilise nägemistaju kriteeriumile, jagatavusele. Ka N9 on seotud jagamatu nägemisega, kellegi isikliku ettekujutusega millestki. Tähendus N12 *nägema\_otsustama* on selles mõttes objektiivsem, et nähakse siiski füüsilisi tunnuseid, teisest

küljest põhinevad ka need otsused, mida tunnuste põhjal tehakse, inimese maailmakogemusele, mis jällegi on (teatud piirini) jagatud teadmine.

Ilmselt võib üsna julgelt väita, et tähendus N1 nägema\_füüsiline on verbi *nägema* prototüüpne tähendus. Eelkõige on selle taga tajuverbide eripära – tuleb uuesti rõhutada, et tajuverbidega väljendame kõige vahetumat füüsilist kogemust maailmast. See on midagi, mis on maailma mõtestamisel väga oluline, seega ei ole üllatav, et kuigi mõtestame nende verbidega ka teisi kogemusi, jääb selle vahetu kogemuse väljendamine siiski kõige olulisemaks. Käitumisprofiil pakub võimaluse välja tuua need kohad, kus tähendused üksteisest eristuvad ning läbi selle sõlmpunktid tähendusvõrgustikus. Oluline on aga meeles pidada, et võrgustiku sidemed ei ole ainult ühesuunalised ning kõik tähendused on ka omavahel seotud. Kuigi mõnel juhul on võimalik selgelt välja tuua, millise semantilise protsessi läbi tähendus laienenud on, siis ei ole ka need protsessid kunagi ühesed, vaid seotud teiste protsessidega.



## 5. Sorteermiskatse (Proos 2014) ja käitumisprofiili analüüsi tulemuste võrdlus

*Nägema* käitumisprofiili analüüsi teel saadud dendrogrammi võrdlemine sorteermiskatse põhjal tehtud klasteranalüüsi tulemustega (Proos 2014) pakub võimalust teha üldistusi ka selle kohta, mis on põhilised erinevused polüseemia korpus- ja katselise meetodiga uurimisel.

Eelkõige tekib meetodite vaheline vastandus sellest, et palju, mis on keelekasutaja jaoks tähenduste eristamisel oluline, ei ole võimalik tõlkida formaalsetesse tunnustesse. Tähenduste vahel leitakse intuiitviseid sarnasusi ja erinevusi, mis osutuvadki kõige tähtsamateks. Nii näiteks osutusid korpusanalüüsi puhul kõige sarnasemaks tähendused N1 *nägema\_füüsiline* ja N2 *nägema\_jagamatu*. Ka sorteermiskatse tulemused näitasid, et need tähendused kuuluvad samasse klastrisse, kuid N2 ja N1 vaheline kaugus oli tunduvalt suurem. Formaalselt on tõesti väga keeruline välja tuua, miks need tähendused erinevad on, tihti ongi ainsaks erinevuseks täpsustus, et midagi nähti unes. Keelekasutaja tajus võib rolli mängida ka see, et üks olulisi nägemise kui taju tunnuseid on selle objektiivsus ehk jagatavus (Sweetser 1990: 39). N2 puhul on see tunnus muutunud ning nägemiskogemus ei ole enam jagatav, mistõttu intuiitvisele ei ole need kaks tähendust samad. Kui aga proovida sellist tunnust väljendada formaalselt, põhineks see nii-öelda formaalne tunnuste määramine siiski uurija intuitsioonil ning viib liigse introspektsioonile toetumiseni.

Ka tähendus N3 *nägema\_kohtuma* asetuse käitumisprofiili põhjal samasse rühma N1 ja N2-ga. Sorteermiskatse põhjal aga võis väita, et N3 moodustab eraldiseisva rühma, kuna selle asetuse dendrogrammil ja katseisikute antud kommentaarid näitasid, et tähendust tajuti teistest kaugena ning ka osa katseisikutest moodustas eraldi üheliikmelise tähendusrühma.

Sellist valikut põhjendasid katseisikud tihti tähenduse suhtlusliku aspekti ja kontakti olemasoluga. (Proos 2014) Selliseid tunnuseid ei ole jällegi võimalik formaalseteks tunnusteks ümber tõlkida. Aimu andis sellisest tähenduserinevusest küll ID-silt „lemma tüüp“, kus tähendus N3 esines koos ainult inimest tähistavate ID-sildi tasemetega, kuid muud tunnused ühtisid suuresti tähendusega N1.

Tähendused N4 nägema\_arusaama ja N8 nägema\_nentima väljendavad suuresti sama tähendust, kuid N8 on konstruktsioonitähendus ning seega on selle kasutus piiratud, omane tähendus ilmneb just nimelt konstruktsioonilistes eripärades, mitte semantilistes. Tundub, et sorteerimiskatses lähtusid katseisikud semantilisest poolest, kuna need tähendused sorteeriti kokku 19 korral 20st (Proos 2014). Käitumisprofiili analüüsi tulemusel aga paigutus N8 kokku teise konstruktsioonitähendusega N10 elame\_näeme. Tuleb aga taaskord tähelepanu juhtida asjaolule, et valideerimistehnika järgi ei olnud klaster tähendustest N8 ja N10 eriti hästi moodustunud, ilmselt jäigi nende tähenduste puhul puudu võimalusi nende semantika kirjeldamiseks.

Peatükis 4.3 mainiti, et N12 nägema\_otsustama asub tähenduslikult piirialal, esinedes kord koos abstraktsete, kord koos konkreetsete mõistetega. Sarnane tendents tuli välja ka sorteerimiskatses, kus see tähendus liideti oma klastrile üpris kõrgelt, mis viitab sellele, et tähendus ei olnud teiste oma rühma liikmetega väga sarnane (Proos 2014). Sorteerimiskatse puhul oli jagunemise põhjus see, et kord sorteeriti tähendus kokku üht sorti, kord teist sorti tähendustega. Mõneti peegeldab ka käitumisprofiili analüüs sellist käitumist – kord esines N12 koos üht sorti ID-sildi tasemetega, kord koos teist sorti tasemetega. Katses kasutatud lause oli *Vaatad otsa ja näed, see mees on intellektuaal* – lauses ei ole vihjet sellele, kas järeldus tehti füüsiliste või abstraktsete tunnuste vahel. See, et lause sorteerimine oli nii varieeruv, viitab tähenduse potentsiaalile esineda koos mõlemat sorti sihtpunktidena ning nagu öeldud, seda näitas ka käitumisprofiil.

Sorteerimiskatse tulemusel moodustunud klastreid iseloomustati bakalaureusetöös kujundskeemidega ning seeläbi kinnitati töö hüpotees, mille järgi keelekasutajad lähtuvad sõna polüseemsete tähenduste mõistmisel kehalisest kogemusest ja selle mõistelisest struktuurist (Proos 2014). Samasugust analüüsi ei olnud võimalik magistrیتöös kasutatud

käitumisprofiili analüüsi tulemuste põhjal läbi viia. Siit ilmneb ka, et mõlema meetodi puhul tekkis *nägema* polüseemidest natukene erinev tähendusvõrgustik. Kuigi sorteerimiskatse tulemused ja käitumisprofiili analüüsi tulemused ei kattunud väga palju, ei saa öelda, et üks meetod oleks teisest parem. Käitumisprofiili analüüsil ilmnas paljut, mis katse puhul ei saagi välja tulla. Kahtlemata on tegemist teineteist täiendavate meetoditega, mille mõlema rakendamine aitab uurijal ligemale pääseda ühele keele kõige huvitavamale küljele, tähendusele.

## 6. Järeldused ja arutelu

Magistritöö üks eesmärke oli tuua välja korpusmeetodi eeliseid ning kitsaskohti tähenduse uurimisel, nii üldiselt kui ka võrdluses sorteerimiskatsega. Kahe meetodi abil saadud tulemusi võrreldi peatükis 5, peatükis 6 üldistatakse tulemusi meetodite võrdlemise eesmärgil.

Üks olulisemaid kitsaskohti käitumisprofiili analüüsi meetodi juures on ka üldiselt korpusmeetodi võimalik kitsaskoht. Peatükis 4.1 toodi välja tulemus, et tähendust N7 nägema\_tunda\_saama ei esinenud valimis kordagi. On võimalik, et 700-lauseline valim ei olnud uurimuseks siiski piisav, samas ei esinenud sellist probleemi teiste tähenduste puhul, seega kui lähtuda lähenemisest, mille järgi väljendi sagedus on otseses seoses selle kinnistatusega keelekasutaja meeles, on see tähendus lihtsalt väga harv. Teisest küljest eeldati, et kuigi sorteerimiskatse näitas, et seda tähendust ei tajutud eraldiseisvana, siis korpusega on võimalik näidata, et tegemist on siiski iseseisva tähendusega. Seda eeldust ei ole aga korpusandmete põhjal võimalik ei tõestada ega ümber lükata – nimelt ei saa valimi põhjal teha üldistust kogu populatsiooni kohta.

Meetodi miinusena võib välja tuua ka selle ringsuse. Ühest küljest on meetodi eesmärk näidata, kas ja millised tähendused erinevate tunnuste alusel kokku rühmituvad, teisalt aga on ka verbi tähendus juba üks tunnustest, mida uurija andmete märgendamise ajal igale lausele määrab. Seega tekib olukord, kus uurija määrab ära tähenduse, samal ajal ise nendest tunnustest lähtudes. Järgmine samm pärast käitumisprofiili analüüsi võiks olla näiteks mudeli loomine – mudelisse oleks kaasatud ID-sildid peale tähenduse ning nende järgi ennustab mudel, millises tähenduses verb lauses esineb. Mudeli tulemusi oleks seega võimalik kõrvutada käsitsi määratud tähendustega ning seeläbi välja selgitada, millised

tunnused on tähenduse määramisel kõige olulisemad. Magistritöö mahupiirangute tõttu selles töös siiski mudeleid kasutatud ei ole, kuid kindlasti pakub see ainet edasiseks uurimistööks.

Eelnevalt on ka korduvalt välja toodud, et palju inimese keeletunnetusest ei ole võimalik „tõlkida“ formaalsetesse tunnustesse. Nii näiteks oli tähenduse N3 nägema\_kohtuma puhul sorteerimiskatses (Proos 2014) keelekasutajate jaoks sorteerimisotsustel oluline suhtluse aspekt, mida on keeruline tunnustega väljendada. Erisus tuli küll välja koosinemisel ID-sildiga „lemma tüüp“, kuid see kajastab vaid väikest osa sellest, mis igapäevases kasutuses „naiivse“ keelekasutaja jaoks oluline on. Samuti ilmnes see puudus tähenduse N2 nägema\_jagamatu klastrisse määramise puhul, mis käitumisprofiili järgi oli tähendusega N1 nägema\_füüsiline väga sarnane, kuid sorteerimiskatses tajusid katseisikud seda siiski rohkem eraldiseisvana. Samas näiteks tähendus N11 nägema\_vaatama osutus sorteerimiskatses tähendusega N1 palju sarnasemaks kui käitumisprofiil seda näitab, ilmselt kuna valimis ei olnud see tähendus võrreldes teistega väga sage.

Samuti on piiratud statistika võime analüüsida tähendusi, mis oma tunnuste järgi võiksid kuuluda erinevatesse klastritesse. Tähendus N9 nägema\_soovima sorteeriti katses (Proos 2014) kokku tähendusega N5 nägema\_ohte, ka käitumisprofiili järgi kuuluvad need tähendused ühte klastrisse, kuid nagu näitasid valideerimistehnikad, on N9 arvestamine sellesse klastrisse tinglik. Samuti võib juhtuda, et käitumisprofiili analüüsi tulemusena määratakse kokku konstruktsioonitähendused ning seda lähtuvalt pigem konstruktsioonilistest kui semantilistest tunnustest. Kui eesmärk on siiski uurida polüseemia struktuuri tähendusest lähtuvalt, ei paku analüüs „õiget“ lahendust.

Samas ei pruugi ka katseisikud lauseid sorteerida ainult verbi enda tähenduse järgi, katse peab olema läbi mõeldud ääretult täpselt, et minimaliseerida võimalus, et keelekasutajad lähtuksid sorteerimisel teistest strateegiates, kui uurija ootab. Näiteks võib juba küsilause kasutamine sorteerimise tulemusi kallutada, kuna keelekasutaja võib näha tähenduse erinevust just lähtudes suhtluslikest eesmärkidest, kuna igapäevases keelekasutuses on see suhtluse oluline osa.

Siit tuleb välja aga ka korpusmeetodite suur pluss – materjali rohkus. Enamasti on materjali piisavalt, et väiksemad kõrvalekalded ei osutu statistiliselt oluliseks. Katses on aga

võimalik kasutada ainult väga piiratud hulka materjali, mis, nagu öeldud, peab olema väga täpselt läbi mõeldud. Hästi läbimõeldud katsematerjali olulisus on tähtis ka selle poolest, et valitud materjal esindaks erinevaid tähendusi võimalikult hästi – see võib aga viia liiga kunstliku olukorra tekkimiseni. Loomulikus keeles kasutatakse sõnu väga erinevates kontekstides ja tähendustes, siit tuleb välja ka järgmine ning võib-olla kõige olulisem korpusandmete eelis – korpus on teiste allikatega võrreldes kõige loomulikum keel.

Magistritöös näidati, et korpus on polüseemia uurimisel väärtuslik allikas, mille kaudu ilmneb palju sellist, mis katses kajastust ei leia. Kuigi korpusel on omad piirangud, on need ka katsel, seega ei saa pidada üht meetodit teisest paremaks. Kuigi töös kasutatud tulemuste tõlgendamise meetodi, hierarhilise klasteranalüüsi tõlgendamisel on oluline roll uurija intuitsioonil, näidati, et erinevate valideerimistehnikatega on võimalik subjektiivsust vähendada. Samas on ka kindel, et statistika ja intuitsioon ei ole teineteist välistavad lähenemised – vastupidi, need on teineteist täiendavad viisid andmestikule lähenemiseks.

Korpusmaterjale on eesti keeleteaduses kasutatud palju, kuid ühe sõna põhjalikku korpuspõhist analüüsi on rakendatud vähe, konkreetselt käitumisprofiili analüüsi meetodit on autorile teadaolevalt rakendanud ainult Kristiina Kask oma magistritöös (Kask 2014). Nii pakub käesolev magistritöö ka väärtuslikku lisa eesti keeleteadusmaastikule.

Autori bakalaureusetöö ja käesolev magistritöö pakuvad mõlemad põhjalikku sissevaadet eesti keele tajuverbi *nägema* polüseemiasse. Samas aga ilmnes korpusest väga palju huvitavat, mida edasi uurida. Tajuverbid laiemalt on endiselt eesti keeleteaduses põhjalikult uurimata teema, mis pakub rikkalikku ainet edasiseks uurimiseks. Võrdluses *nägema* verbiga pakub kindlasti huvitavaid uurimisteemasid kas näiteks teiste nägemisvaldkonna verbide uurimine, või verbi *nägema* asetamine võrdlusesse teiste tajuverbidega. Lisaks ilmnes korpusest ka palju huvitavat konstruktsioonide kohta, mis tajuverbidega (ja laiemalt kognitiivsete verbidega) koos esinedes saavad teistsuguse tähenduse, seega oleks väärtuslik ka selliste tähendusmuutuste edasine uurimine. Samuti pakuks edasist uurimismaterjali juba mainitud variant rakendada tähenduste sarnasuste ja erinevuste hindamiseks mudelit.

## Kokkuvõte

Magistritöös kasutati käitumisprofiili analüüsi meetodit, et iseloomustada eesti keele tajuverbi *nägema* polüseemiat. Verb *nägema* valiti, kuna tajuverbid pakuvad huvitavat sissevaadet sellesse, kuidas mõtestame keeles abstraktseid kogemusi läbi konkreetsete füüsiliste kogemuste. Tajuverbe on eesti keeles vähe uuritud, põhjalikumat ülevaadet *nägema* polüseemiast pakub autori bakalaureusetöö (Proos 2014), kus kasutati meetodina sorteerimiskatset.

Analüüsitav materjal oli pärit eesti keele tasakaalus korpusest, kust juhusliku valiku teel võeti analüüsiks 700 lauset, kus verb *nägema* esines eri vormides. Andmete analüüsimisel kasutati käitumisprofiili analüüsi meetodit. Iga lause puhul määrati hulk semantilisi ja morfosüntaktilisi tunnuseid, mida meetodi raames nimetatakse ID-siltideks. Kokku kasutati töös 89 erinevat ID-sildi taset. Nende ID-siltide erinevate tähendustega koosinemiste põhjal on võimalik teha järeldusi polüseemse sõna tähenduste sarnasuste ja erinevuste kohta. Töös on eristatud 12 verbi *nägema* tähendust ning töö eesmärk on selgitada, millised nendest tähendustest on omavahel rohkem sarnased, millised vähem.

Korpusandmete põhjal saadud koosinemistabelit, mis vormiliselt ongi verbi käitumisprofiil, analüüsiti hierarhilise klasteranalüüsi meetodiga. Klasteranalüüs on kirjeldav statistiline meetod, mis rühmitab uuritavad elemendid nii, et moodustuksid võimalikult sarnased rühmad e klastrid. Meetodit kasutatakse eelkõige struktuuri loomiseks suures hulgas varieeruvates andmetes. Klasteranalüüsi tarbeks arvutati Eukleidilist mõõtu kasutades koosinemismaatriksi põhjal välja kaugusmaatriksi. Hierarhilisel klasteranalüüsil kasutati liitmist ning Wardi meetodit. Selleks, et leida optimaalne klastrate arv ning rohkem avada klasteranalüüsi tulemuste sisu, kasutati kaht valideerimistehnikat. Siluettehnikaga näidati, kui hästi klastrate liikmed enda klastrisse sobituvad. Kuigi viieklastriline lahendus oli

siluetitehnika järgi optimaalsem, arvestati edasisel analüüsil siiski nelja klastriga. Klasteritele *p*-väärtuste määramiseks kasutatud *bootstrap*-tehnika näitas, kui hästi on iga klastriks liitumine andmete poolt toetatud, aidates seega hinnata, milliste tulemuste tõlgendamisel peaks olema ettevaatlik.

Laiemalt jagunesid tähendused käitumisprofiili põhjal kaheks – tähendused, mille puhul sihtpunkt on piiritletud, st mingit sorti objekt, ja tähendused, mille sihtpunkt on hajusam, st situatsioon või sündmus. Eraldi klasteri moodustasid „objektile suunatud“ konkreetsete tähendused N1 nägema\_füüsiline, N2 nägema\_jagamatu, N3 nägema\_kohtuma ja N11 nägema\_vaatama. Abstraktsete objektile suunatud tähenduste klasteri moodustasid tähendused N9 nägema\_soovima, N5 nägema\_ohte ning N6 nägema\_kogema. Valideerimistehnikad näitasid, et tähenduse N9 liitumine just selle klastriga ei ole väga kindel. Töös arutleti, et ilmselt on selle põhjuseks olnud tähenduse N9 vähene sagedus valimis. Teise suurde harusse jäi samuti kaks klasterit. Ühe neist moodustasid tähendused N4 nägema\_arusaama ja N12 nägema\_otsustama ning teise tähendused N8 nägema\_nentima ja N10 elame\_näeme. Viimased moodustasid ilmselt klasteri, kuna mõlemaid tähendusi võib pidada konstruktsioonitähendusteks – nii tähendus N8 kui ka tähendus N10 realiseeruvad vaid piiratud kontekstis.

Tähenduste rühmitumist käitumisprofiili analüüsi põhjal võrreldi sorteerimiskatse tulemusel saadud tähendusrühmadega. Nii oli võimalik välja tuua ka mõlema meetodi eeliseid ja puudusi. Näiteks ei ole võimalik käitumisprofiili analüüsi puhul märgendada selliseid semantilisi tunnuseid, mida ei ole võimalik formaalsetesse tunnustesse tõlkida. Näitena toodi tähenduse N3 nägema\_kohtuma rühmitumine – käitumisprofiili analüüsi puhul osutus see väga sarnaseks tähendusega N1 nägema\_füüsiline, kuid sorteerimiskatses pidasid katseisikud oluliseks tunnuseks „inimlikku aspekti“, mida ei ole võimalik objektiivse tunnusena kirja panna.

Üks olulisemaid tulemusi oli, et tähendust N7 nägema\_tunda\_saama ei esinenud valimis kordagi. Sorteerimiskatses selgus, et keelekasutajad peavad seda tähendust ning tähendust N6 nägema\_kogema niivõrd sarnaseks, et need sorteeriti kokku igal korral. Eeldati, et korpuseandmeid analüüsides tuleb nende tähenduste erinevus välja. See tulemus seoti ühe



korpusmeetodite üldise miinusega – valimi põhjal ei saa hinnangut anda kogu populatsioonile. Ühe suurema miinusena toodi välja ka käitumisprofiili analüüsi meetodi teatav ringsus – uurija eesmärk on uurida, kuidas tähendused korpusel tunnuste alusel rühmituvad, teisest küljest määrab uurija iga lause puhul ära verbi tähenduse lauses ning teeb seda just nende tunnuste alusel, mida märgendatakse.

Korpusmeetodite kasuks räägib aga vaieldamatult keelematerjali hulk, mis uurija käsutuses on. Katseliste meetodite puhul on maht väga piiratud, kuna katseisikud ei suuda töödelda ega toota väga suurt hulka infot adekvaatselt. Samuti on katse õnnestumiseks vaja see väga detailselt planeerida. Korpusel kasuks räägib ka keelematerjali loomulikkus, ei ole teist materjali allikat, mis oleks niivõrd suuremahuline ja samas loomulik.

Tulemuste analüüsil ja kahe meetodi tulemuste võrdlusel jõuti järeldusele, et sorteerimiskatse ja käitumisprofiili analüüs on teineteist täiendavad meetodid ning üht ei ole mõistlik pidada teisest paremaks. Samuti kinnitati, et statistika juures on oluline roll ka uurija intuitsioonil ning statistiliste meetodite kasutamine ei tähenda vaid numbrite esitamist.

Eesti keeleteadusmaastikul ei ole käitumisprofiili analüüsi eriti rakendatud, seega on lisaks verbi *nägema* polüseemia kirjeldamisele töö oluline väärtus meetodi tutvustamine. Analüüsi tulemusel ilmnes palju huvitavat, mis pakub ainet ka edasiseks uurimiseks. Lisaks on ka tajuverbid laiemalt eesti keeles põhjalikumalt uurimata valdkond – siinkirjutaja loodab ka tulevikus jätkata selle lünga täitmist.

## Kirjandus

- Arppe jt = Antti Arppe, Gaëtanelle Gilquin, Dylan Glynn, Martin Hilpert, Arne Zeschel 2010.** Cognitive Corpus Linguistics: five points of debate on current theory and methodology. *Corpora* 5 (1). 1–27.
- Berez, Andrea L., Stefan Th. Gries 2008.** In defense of corpus-based methods: A behavioral profile analysis of polysemous *get* in English. *Proceedings of the 24th Northwest Linguistics Conference*, Steven Moran, Darren S. Tanner & Michael Scanlon (eds.). University of Washington Working Papers in Linguistics, vol. 27, 157–166. Seattle: Departement of Linguistics.
- Born, Toomas 1995.** Eesti keele tajuverbid. Tähendusallikad ja tähendusmuutused. Diplomitöö. Tartu Ülikool.
- Cuyckens, Hubert, Britta Zawada 1997.** Introduction. *Polysemy in Cognitive Linguistics*, Hubert Cuyckens & Britta Zawada (eds.). Current Issues in Linguistic Theory IV, vol. 177, ix–xxvii. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Desagulier, Guillaume 2014.** Visualizing distances in a set of near-synonyms. *Corpus Methods for Semantics. Quantitative studies in polysemy and synonymy*. Human Cognitive Processing (HCP). Cognitive Foundations of Language Structure and Use, vol. 43, 145–179. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Divjak, Dagmar 2010.** *Structuring the lexicon: a clustered model for near-synonymy*. Berlin/New York: Walter de Gruyter GmbH & Co. KG.
- Divjak, Dagmar, Nick Fieller 2014.** Cluster analysis. Finding structure in linguistic data. *Corpus Methods for Semantics. Quantitative studies in polysemy and synonymy*,

Dylan Glynn & Justyna A. Robinson (eds.), 405–441. Amsterdam/Philadelphia: John Benjamins.

**Divjak, Dagmar, Stefan Th. Gries 2006.** Ways of trying in Russian: clustering behavioural profiles. *Corpus Linguistics and Linguistic Theory* 2-1. 23–60.

**EKG = Eesti keele grammatika I. Morfoloogia. Sõnamoodustus.** Toim Erelt, Mati, Reet Kasik, Helle Metslang, Henno Rajandi, Kristiina Ross, Henn Saari, Kaja Tael, Silvi Vare 1995. Eesti Teaduste Akadeemia Keele ja Kirjanduse Instituut.

**EKG II = Eesti keele grammatika II. Süntaks. Lisa: Kiri.** Toim Erelt, Mati, Reet Kasik, Helle Metslang, Henno Rajandi, Kristiina Ross, Henn Saari, Kaja Tael, Silvi Vare 1993. Eesti Teaduste Akadeemia Keele ja Kirjanduse Instituut.

**Geeraerts, Dirk 2010a.** *Theories of Lexical Semantics*. Oxford: Oxford University Press.

**Geeraerts, Dirk 2010b.** The doctor and the semantician. *Quantitative Methods in Cognitive Semantics: Corpus-Driven Approaches*. Cognitive Linguistics Research 46, Dylan Glynn & Kerstin Fischer (eds.), 63–79. Berlin/New York: Walter de Gruyter GmbH & Co. KG.

**Glynn, Dylan 2010.** Corpus-driven Semantics. *Quantitative Methods in Cognitive Semantics: Corpus-Driven Approaches*. Cognitive Linguistics Research, vol. 46. Dylan Glynn & Kerstin Fischer (eds.), 1–43. Berlin/New York: Walter de Gruyter GmbH & Co. KG.

**Glynn, Dylan 2014.** Polysemy and synonymy. Cognitive theory and corpus method. *Corpus Methods for Semantics. Quantitative studies in polysemy and synonymy*. Human Cognitive Processing (HCP). Cognitive Foundations of Language Structure and Use, vol. 43, 7–39. Amsterdam/Philadelphia: John Benjamins Publishing Company.

**Gries, Stefan Th. 2006.** Corpus-based methods and cognitive semantics: the many meanings of to run. *Corpora in cognitive linguistics: corpus-based approaches to syntax and lexis*, Stefan Th. Gries & Anatol Stefanowitsch (eds.), 57–99. Berlin & New York: Mouton de Gruyter.

**Gries, Stefan Th., Dagmar Divjak 2009.** Behavioral profiles: a corpus-based approach to cognitive semantic analysis. *New directions in cognitive linguistics*, In Vyvyan Evans & Stéphanie Pourcel (eds.), 57–77. Amsterdam/Philadelphia: John Benjamins.

- Habicht, Külli, Iлона Tragel 2014.** Verbiga saama väljendatud leksikaalsed kategooriad ja konstruktsioonid. *Keel ja Kirjandus* 11. 826–843.
- Jansegers jt = Jansegers, Marlies, Clara Vanderschueren, Renata Enghels 2015.** The polysemy of the Spanish verb *sentir*: a behavioral profile analysis. *Cognitive Linguistics* 26 (3), (Ed.) John Newman, 381–421.
- Johnson, Mark 1987.** *The Body in the Mind. The Bodily Basis of Meaning, Imagination, and Reason.* Chicago: The University of Chicago Press.
- Kask, Kristiina 2014.** *Seisma* verbi polüseemia: korpuspõhine käitumisprofiil ja klasteranalüüs. Magistritöö. Tartu: Tartu Ülikool.
- Labi, Kanni 2006.** Eesti regilaulude verbisemantika. Tartu Ülikool: Dissertationes Philologiae Estonicae Universitatis Tartuensis 18.
- Lakoff, George 1987.** *Women, Fire, and Dangerous Things: what categories reveal about the mind.* Chicago and London: The University of Chicago Press.
- EKSS = Eesti keele seletav sõnaraamat.** Toim Langemets, Margit, Mai Tiits, Tiia Valdre, Leide Veskis, Ülle Viks, Piret Voll 2009. Tallinn: Eesti Keele Sihtasutus.
- Pajusalu, Renate 2001.** Kas moos ja buss seisavad sarnaselt ehk väike katse verbiga seisma. *Keele Kannul. Pühendusteos Mati Ereli 60. sünnipäevaks.* Toim Reet Kasik, 250–274.
- Pajusalu jt = Pajusalu, Renate, Iлона Tragel, Ann Veismann & Maigi Vija 2004.** *Tuumsõnade semantikat ja pragmaatikat.* Tartu Ülikooli üldkeeleteaduse õppetooli Toimetised 5. Tartu: Tartu Ülikooli Kirjastus.
- Proos, Mariann 2014.** Eesti keele tajuverbi nägema polüseemia sorteerimiskatse põhjal. Bakalaureusetöö. Tartu Ülikool.
- Raukko, Jarno 1999.** An “Intersubjective” Method for cognitive-semantic Research on Polsemy: The case of GET. *Cultural, Psychological and Typological Issues in Cognitive Linguistics. Selected papers of the bi-annual ICLA meeting in Albuquerque, July 1995.* Masako K Hiraga, Chris Sinha & Sherman Wilcox (eds.), Current Issues in Linguistic Theory IV), 87–107. John Benjamins Publishing Company.

- R Core Team 2015.** *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Robinson, Justyna A. 2014.** Quantifying polysemy in Cognitive Sociolinguistics. *Corpus Methods for Semantics. Quantitative studies in polysemy and synonymy.* Human Cognitive Processing (HCP). Cognitive Foundations of Language Structure and Use, vol. 43, 87–117. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Schmid, Hans-Jörg 2010.** Does frequency in text instantiate entrenchment in the cognitive system? *Quantitative Methods in Cognitive Semantics: Corpus-Driven Approaches.* Cognitive Linguistics Research 46, 101–137. Berlin/New York: Mouton de Gruyter.
- Sepper, Maria-Maren 2006.** Indirektaal eesti 19. sajandi lõpu ja 20. sajandi aja- ja ilukirjanduskeeles. Magistritöö. Tallinn.
- Sweetser, Eve E. 1990.** *From Etymology to pragmatics.* (Cambridge Studies in Linguistics 54). Cambridge: Cambridge University Press.
- Taylor, John R. 1995.** *Linguistic Categorization. Prototypes in Linguistic Theory.* New York: Oxford University Press.
- Tragel, Iona 2003.** Eesti keele tuumverbid. Tartu Ülikool: Dissertationes Linguisticae Universitatis Tartuensis 3.
- Veismann, Ann 2009.** Eesti keele kaas- ja määrsõnade semantika võimalusi. Tartu Ülikool: Dissertationes Linguisticae Universitatis Tartuensis 11.
- Wittgenstein, Ludwig 1997.** *Philosophical Investigations.* Oxford: Blackwell Publishers Ltd.
- WordNet 2016.** WordNet. A lexical database for English. *About WordNet.* <https://wordnet.princeton.edu/>

## A behavioral profile analysis of the Estonian verb *nägema* 'to see'. Summary

The aim of this MA thesis is to determine the distribution of the different senses of the Estonian perception verb *nägema* 'to see' in the corpus and what this type of frequency information can tell us about the structure of the verb's polysemy. The background of the thesis lies in cognitive semantics, more specifically the logic of the analysis follows that of cognitive corpus-based semantics. The main reasoning behind using corpora to study how different senses may be coded in the language user's mind is that there is a strong connection between a lexical item's entrenchment and how language users remember and use it. It follows that when an item re-occurs often and in the same context, conclusions can be reached about the cognitive reality of its usage. A broader aim of the thesis was to compare the results of the corpus analysis with that of a sorting task that the author conducted in her BA thesis, and with that also shed light on to what extent polysemy can be analysed based on a corpus.

The method of the thesis is behavioural profile (BP) analysis. BP-analysis is a method that allows the researcher to analyse both semantic and morphosyntactic characteristics of the 12 different senses. The material of the MA thesis is taken from the written corpus of Estonian, included are texts from fiction and journalism. 700 sentences from a total of 18 679 occurrences of the verb *nägema* were randomly chosen for the analysis. A set of 17 ID-tags and 89 ID-tag levels were then assigned to each of the sentences. With the statistics program R, a frequency table was then constructed. Strictly speaking, the absolute and relative frequency tables are the behavioural profile of the word.

Hierarchical cluster analysis was then used to analyse the behavioural profile, using the Euclidean distance measure and Ward's amalgamation algorithm. The resulting

dendrogram was then analysed using two validation techniques – bootstrapping and the Silhouette plot validation. The results of the validation were of help in deciding on the optimal number of the clusters as well as determining how good the clustering solution is, i.e. which clusters are better supported by the data.

The results of the hierarchical cluster analysis showed that the senses formed two larger clusters – one where the object of seeing was a bounded entity i.e. an object; and the other where the object of seeing was a kind of situation or event. Within the bounded-object cluster, a cluster formed with all of the “physical” senses of *nägema*. A cluster was also formed, where the bounded object of seeing is an abstract entity, e.g. *problem, point, sorrow* etc. Within the second larger cluster also two distinct clusters were formed. One of the clusters combined the meanings which focus on the seeing of (in most cases) an abstract event or situation (*I see how you would think that, but...*) and the other cluster was formed from two construction meanings, that only occur in restricted syntactical contexts.

In comparison with the sorting task results, the results of the behavioural profile analysis showed a different kind of clustering. It was concluded that these differences are largely due to the fact that some characteristics that guide the language user in sorting the senses are simply not “translatable” into formal characteristics. It also has to be kept in mind that the frequency of the various senses in the material can influence the results of the analysis. The sense “physical seeing”, for example, had a frequency of 375 of a total 700, while the sense “to wish/hope for something” only 9.

The overall conclusion of the MA thesis is that experimental methods and corpus methods are mutually supportive, i.e. they both offer valuable information that the other one cannot. It was also emphasised that the application of statistical methods does not mean the researcher should not apply his or her own intuition – in the opposite, intuition still plays a crucial role when making decisions about which ID-tags to annotate or when making sense of the results.

As the field of Estonian perception verbs still remains rather unexplored, there is still much left to work with. Further research could include the comparison of *nägema* and other perception verbs as well as to other verbs of vision.

LISA 1. Verbi *nägema* käitumisprofiil (%)<sup>14</sup>

ID-SILT	ID-SILDI TASE	N1	N2	N3	N4	N5	N6	N08	N09	N10	N11	N12
ALLIKAS	AJA	26,4	12,8	6,7	41,4	59,6	42,9	17,5	85,7	33,3	61,9	33,3
	ILU	73,6	87,2	93,3	58,6	40,4	57,1	82,5	14,3	66,7	38,1	66,7
VERBI VORM	<i>da-inf.</i>	2,4	2,6	0	6,9	5,3	9,5	2,5	0	0	4,8	3,9
	<i>des</i>	3,5	2,6	0	3,4	0	0	0	0	0	0	3,9
	<i>ei saa</i>	5,1	0	13,3	0	1,8	4,8	0	0	0	4,8	2
	<i>ma-inf.</i>	0,5	0	0	0	1,8	0	0	0	0	0	0
	<i>nud</i>	0,3	0	0	0	0	2,4	0	0	0	4,8	0
	<i>pl1</i>	3,5	0	0	3,4	1,8	4,8	0	0	22,2	4,8	3,9
	<i>pl2</i>	1,6	0	0	6,9	1,8	7,1	20	0	0	0	0
	<i>pl3</i>	5,1	10,3	0	0	14	11,9	0	14,3	0	0	5,9
	<i>s1</i>	25,9	38,5	56,7	20,7	26,3	14,3	5	42,9	11,1	14,3	29,4
	<i>s2</i>	4,3	2,6	6,7	13,8	5,3	4,8	72,5	0	33,3	9,5	3,9
	<i>s3</i>	45,9	41	16,7	41,4	31,6	33,3	0	28,6	33,3	42,9	45,1
	<i>tud-vorm</i>	1,6	0	3,3	3,4	1,8	7,1	0	0	0	4,8	0
	<i>umbisik. lihtm.</i>	0,5	0	3,3	0	3,5	0	0	0	0	0	0
	<i>umbisik. olevik</i>	0	2,6	0	0	5,3	0	0	14,3	0	9,5	2
KÕNELIIK	<i>eitav</i>	24,8	17,9	26,7	10,3	38,6	23,8	0	14,3	11,1	9,5	9,8
	<i>jaatav</i>	69,1	76,9	73,3	79,3	56,1	69	97,5	85,7	88,9	76,2	82,4
	<i>ei kehti</i>	6,1	5,1	0	10,3	5,3	7,1	2,5	0	0	14,3	7,8

<sup>14</sup> Tabelis moodustub 100% iga ID-sildi taseme protsentide kokku liitmine ühe tähenduse ja ID-sildi raames. Näiteks ID-sildi „allikas“ puhul moodustub tähenduse N1 ID-sildi tasemetest kokku 100% e kõik valimis esinenud tähendust N1 sisaldunud laused.



TEGUMOOD	isikuline	91,5	92,3	93,3	86,2	80,7	85,7	97,5	85,7	100	81	90,2
	ei kehti	6,4	5,1	0	10,3	5,3	7,1	2,5	0	0	4,8	7,8
	umbisikuline	2,1	2,6	6,7	3,4	14	7,1	0	14,3	0	14,3	2
KÕNEVIIS	kaudne	1,1	2,6	0	0	0	0	0	0	0	0	0
	kindel	88,3	84,6	93,3	82,8	93	81	52,5	57,1	88,9	81	86,3
	käskiv	0	0	0	0	0	4,8	45	0	11,1	0	0
	ei kehti	6,9	5,1	0	10,3	5,3	11,9	2,5	0	0	14,3	7,8
	tingiv	3,7	7,7	6,7	6,9	1,8	2,4	0	42,9	0	4,8	5,9
AEG	ennemin evik	4	7,7	6,7	0	0	4,8	0	0	0	0	3,9
	lihtmine vik	42,9	38,5	30	27,6	33,3	14,3	2,5	0	0	23,8	43,1
	ei kehti	6,4	5,1	3,3	10,3	5,3	11,9	2,5	0	0	14,3	7,8
	olevik	34,7	30,8	36,7	48,3	52,6	40,5	95	100	100	47,6	35,3
	täismine vik	10,1	10,3	23,3	10,3	8,8	28,6	0	0	0	14,3	7,8
	üldmine vik	1,9	7,7	0	3,4	0	0	0	0	0	0	2
ALGUSPUNKTI ELUSUS	elus	78,9	89,7	90	72,4	66,7	83,3	80	85,7	66,7	61,9	66,7
	elutu	1,6	2,6	0	0	7	4,8	0	0	0	0	0
	ei kehti	19,5	7,7	10	27,6	26,3	11,9	20	14,3	33,3	38,1	33,3
ALGUSPUNKTI ABSTRAKT	abstraktsed	0,8	2,6	0	3,4	1,8	0	0	0	0	0	0
	konkreetsed	79,7	89,7	90	69	71,9	88,1	80	85,7	66,7	61,9	66,7
	varjatud	19,5	7,7	10	27,6	26,3	11,9	20	14,3	33,3	38,1	33,3
SIHTPUNKTI ABSTRAKT	abs	0,5	17,9	0	65,5	80,7	61,9	20	28,6	33,3	4,8	52,9
	konkreetsed	84,8	74,4	93,3	34,5	15,8	23,8	60	71,4	33,3	90,5	45,1
	varjatud	8,8	7,7	3,3	0	3,5	11,9	2,5	0	22,2	4,8	2
	ei kehti	5,9	0	3,3	0	0	2,4	17,5	0	11,1	0	0
SIHTPUNKTI	ei kehti	14,7	7,7	6,7	0	3,5	14,3	20	0	33,3	4,8	2
	objekt	66,9	76,9	93,3	3,4	96,5	78,6	15	85,7	11,1	90,5	51

	situats.	6,7	10,3	0	89,7	0	2,4	42,5	0	44,4	0	41,2
	sündmu s	11,7	5,1	0	6,9	0	4,8	22,5	14,3	11,1	4,8	5,9
LEMMA TÜÜP	asesõna	24	15,4	56,7	6,9	14	16,7	10	14,3	11,1	9,5	19,6
	fys elutu	29,3	23,1	0	0	7	7,1	2,5	0	11,1	57,1	3,9
	inimene	6,1	5,1	3,3	0	0	0	2,5	0	0	0	2
	lind	1,1	0	0	0	0	0	0	0	0	0	0
	mitte füs. elutu	5,9	28,2	0	3,4	24,6	16,7	2,5	14,3	0	28,6	17,6
	muu abs	0	7,7	0	0	50,9	9,5	0	0	0	0	5,9
	muu elus	0,5	0	0	0	0	0	0	0	0	0	0
	nimetus	6,9	0	20	0	3,5	0	0	42,9	0	0	7,8
	pärismim i	9,1	5,1	13,3	3,4	0	0	0	14,3	0	0	3,9
	tunne	0	0	0	3,4	0	45,2	0	0	0	0	3,9
	ei kehti	17,1	15,4	6,7	82,8	0	4,8	82,5	14,3	77,8	4,8	35,3
	SEMANTILJ NE AEG	sem minevik	60,5	59	63,3	41,4	31,6	52,4	2,5	0	0	42,9
sem olevik		34,7	35,9	23,3	58,6	63,2	31	97,5	28,6	22,2	33,3	41,2
sem tulevik		4,8	5,1	13,3	0	5,3	16,7	0	71,4	77,8	23,8	2
SEMANTI LINE ISIK	varjatud	12,3	7,7	16,7	17,2	22,8	14,3	2,5	14,3	22,2	28,6	33,3
	isikuline	77,9	89,7	83,3	69	73,7	83,3	97,5	85,7	55,6	57,1	62,7
	üldisik	9,9	2,6	0	13,8	3,5	2,4	0	0	22,2	14,3	3,9
AHELVERB	muu+inf	1,1	5,1	0	0	7	2,4	0	28,6	0	4,8	2
	muu+su piin	0,5	2,6	0	0	7	0	0	0	0	4,8	3,9
	muu+va t	0,3	2,6	0	0	0	0	0	0	0	0	0
	ei kehti	78,1	79,5	80	79,3	78,9	78,6	100	57,1	66,7	52,4	58,8
	on+inf	7,7	0	6,7	6,9	1,8	0	0	0	22,2	4,8	27,5
	pidama+ inf	1,3	0	0	3,4	0	7,1	0	0	0	0	0

	saama+inf	1,9	2,6	0	3,4	0	2,4	0	0	11,1	14,3	2
	suutma+inf	0,8	0	0	0	3,5	0	0	0	0	0	0
	tahtma+inf	3,7	2,6	13,3	0	0	4,8	0	14,3	0	19	3,9
	tohtima+inf	0	0	0	3,4	0	0	0	0	0	0	0
	tulema+inf	0	0	0	0	0	2,4	0	0	0	0	0
	võima+inf	4,5	5,1	0	3,4	1,8	2,4	0	0	0	0	0
LAIEND	allikas	3,2	0	3,3	6,9	0	0	0	0	0	0	13,7
	koht	2,1	33,3	0	0	14	2,4	0	0	0	0	3,9
	määrus	0,8	2,6	0	0	0	9,5	0	0	0	0	0
	ei kehti	89,9	64,1	96,7	93,1	86	88,1	100	100	100	100	76,5
	trajektoritapsustus	1,3	0	0	0	0	0	0	0	0	0	0
	vahend	2,7	0	0	0	0	0	0	0	0	0	0
KÕRVALLAUSE	kui aeg	0,3	0	0	0	0	0	0	0	0	0	0
	kui kond.	0,3	0	0	0	0	0	0	0	0	0	0
	rinnastus	0,3	0	0	10,3	0	0	32,5	0	0	0	0
	et	3,7	10,3	0	51,7	0	0	2,5	14,3	22,2	0	33,3
	ei kehti	90,4	84,6	100	6,9	100	95,2	55	85,7	33,3	100	58,8
	obj. suunatud	1,6	0	0	17,2	0	2,4	2,5	0	11,1	0	7,8
	viisimääruslik	3,5	5,1	0	13,8	0	2,4	7,5	0	33,3	0	0

**Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks**

Mina \_\_\_\_\_ Mariann Proos \_\_\_\_\_

(*autori nimi*)

(sünnikuupäev: \_\_\_\_\_ 26. september 1991 \_\_\_\_\_)

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) enda loodud teose „Mida ütleb korpus tähenduse kohta? Käitumisprofiili analüüsi ja klasteranalüüsi meetod tajuverbi *nägema* tähenduse uurimisel“ \_\_\_\_\_,  
(*lõputöö pealkiri*)

mille juhendaja on \_\_\_\_\_ Jane Klavan \_\_\_\_\_,

(*juhendaja nimi*)

- 1.1.reprodutseerimiseks säilitamise ja üldsusele kättesaadavaks tegemise eesmärgil, sealhulgas digitaalarhiivi DSpace-is lisamise eesmärgil kuni autoriõiguse kehtivuse tähtaja lõppemiseni;
  - 1.2.üldsusele kättesaadavaks tegemiseks Tartu Ülikooli veebikeskkonna kaudu, sealhulgas digitaalarhiivi DSpace´i kaudu kuni autoriõiguse kehtivuse tähtaja lõppemiseni.
2. olen teadlik, et punktis 1 nimetatud õigused jäävad alles ka autorile.
  3. kinnitan, et lihtlitsentsi andmisega ei rikuta teiste isikute intellektuaalomandi ega isikuandmete kaitse seadusest tulenevaid õigusi.

**Tartus**/Tallinnas/Narvas/Pärnus/Viljandis, \_\_\_\_19.05.2016\_\_\_\_ (*kuupäev*).