

TARTU ÜLIKOOL
EESTI JA ÜLDKEELETEADUSE INSTITUUT
EESTI KEELE OSAKOND

Kristiina Kask

SEISMA VERBI POLÜSEEMIA:
KORPUSPÕHINE KÄITUMISPROFIIL JA KLASTERANALÜÜS

Magistritöö

Juhendajad Jane Klavan ja Ann Veismann

TARTU 2014

SISUKORD

Sissejuhatus.....	4
1. Polüseemia kognitiivse lingvistika raamides	6
1.1. Kognitiivsed protsessid tähenduse kujundajana	6
1.1.1. Radiaalse tähendusvõrgustiku kujunemine.....	7
1.1.2. Kujundskeemide teke.....	9
1.2. Asendiverb seisma ja selle tähendused	10
1.3. <i>Seisma</i> verbi polüseemia uurimisest	14
2. Meetodi kirjeldus	16
2.1. Kasutuspõhine lingvistika	16
2.1.1. Korpuslingvistika.....	18
2.1.2. Moodsad korpused.....	19
2.2. Sõna käitumisprofiil (<i>Behavioral Profile</i>)	21
2.3. Klasteranalüüs	24
2.3.1. Objektidevahelise kauguse (sarnasuse) mõõtmine	25
2.3.2. Hierarhiline aglomeratiivne klasteranalüüs	26
3. <i>Seisma</i> verbi käitumisprofiili ID-sildid ja tasemed.....	30
3.1. Sõna käitumisprofiili koostamine	30
3.2. Üldtähendus ja allikas	31
3.3. <i>Seisma</i> verbi parafrasid ja tähendusvõrgustik	32
3.4. Verbiühendid.....	42
3.5. Agent ja sellega seotud ID-sildid	44

3.6. Taust ja sellega seotud ID-sildid	48
3.7. Morfoloogilised kategooriad	52
4. Klasteranalüüs <i>seisma</i> verbi tähendustega	53
4.1. Andmetöötlus	53
4.2. Klasteranalüüsi tulemused ja tõlgendus	55
4.3. Võrdlus eelnevalt tehtud uurimustega	63
4.4. Järeldused	65
Kokkuvõte	68
Kirjandus	70
LISA 1. Parafraasid, nende esinemise arv korpuses ja näitelause	76
LISA 2. Koosinemistabel parafraasi ja agenditüübi, elususe, vertikaalsusega	77
LISA 3. Koosinemistabel parafraasi aja-, koha- ja viisimäärusega ja määrusetüüp	79
The Polysemy of Estonian Posture Verb <i>seisma</i> 'Stand': Behavioral Profile and Cluster Analysis. Summary	80

Sissejuhatus

Käesolevas magistritöös on vaatluse all eesti keele asendiverbi *seisma* polüseemia. Kuigi polüseemia on keeles tavaline nähtus ega tekita suhtlemisel probleeme, on teine lugu selle uurimisega. Ühel meel elatakse, et polüseemse sõna tähendused moodustavad prototüübi ümber tähendusvõrgustiku, kuid see, kuidas täpselt tähendused paiknevad ja kui kaugel üksteisest asetsevad, on olnud keerulisem küsimus. Varem on sellekohased tööd lähtunud introspeksioonist või eri katsetest (nt sorteerimiskatse, vastuvõetavuse hindamise katse), ka on kasutatud korpusanalüüsi. Selles töös on samuti kasutusel korpuspõhine lähenemine – sõna käitumisprofiil.

Oma bakalaureusetöös uurisin asendiverbide *lebama*, *istuma*, *rippuma* ning *seisma* kasutust elutute entiteetide asendi kirjeldamisel, tollal kasutasin meetodina moodustuskatset (Kask 2012). Tööst selgus, et eesti keeles eelistatakse koopulat asendiverbile, kuid objektide asetsemisel on kõige vastuvõetavam just *seisma* verb, seda eriti juhul, kui vertikaalne telg on silmanähtavalt horisontaalsest pikem. Kuigi *seisma* verbi on eesti keeles küll põgusalt uuritud (Pajusalu 2001a, 2001b, Veismann 2012), siis põhjalikku korpusanalüüsi pole *seisma* verbi kohta tehtud. Selles magistritöös keskendungi vaid sellele ühele verbile ja uurin selle tähendusi ja kasutusvõimalusi eesti keeles.

Töö suurem eesmärk on aga katsetada üsna uudset Stefan Th. Griesi ja Dagmar Divjaki välja töötatud korpusanalüüsi *Behavioral Profile*'i ehk sõna käitumisprofiili meetodit ja vaadata milliseid tulemusi see annab eesti keele ühe polüseemse sõna uurimisel. Käitumisprofiil on radikaalselt korpuspõhine lähenemine, milles toetutakse sagedusandmetele ja funktsionaalsetele tunnustele (kasutuse sarnasus) palju rohkem, kui varasemates polüseemiat käsitlevates töödes.

Töö teoreetiline taust on kognitiivne ja kasutuspõhine lingvistika, st huvi keskmes on tegelik keelekasutus. Lähtudes sellest, et elektroonilistes korpusetes on kogutud loomuliku

keelekasutuse näiteid, on selles töös kasutatud korpusanalüüsi ja laused analüüsiks on võetud tasakaalus korpusest.

Töö aluseks on 500-lauseline juhuslik valim *seisma* verbiga lausetega. Seejuures on märgendatud nii palju erinevat semantilist, süntaktilist ja morfoloogilist infot *seisma* verbiga lausete kohta, kui võimalik ja seejärel jagatud tähendused hierarhilise aglomeratiivse klasteranalüüsi abil klastritesse, et vaadata kui kaugel või lähedal eri tähendused üksteisele asetsevad. Sellist analüüsimeetodit pole teadaolevalt Eestis varem kasutatud, kuigi eri variaablite märgendamine on korpusanalüüsi korral tavaline.

Magistritöö koosneb neljast peatükist. Esimeses kahes peatükis esitatakse töö teoreetiline taust ja meetodi kirjeldus. Teoreetilisse ossa kuulub selgitus tähendusvõrgustiku kujunemise aluseks olevatest kognitiivsetest protsessidest ning ülevaade *seisma* verbist ning varasematest uurimustest selle kohta.

Meetodi peatükis on kirjeldatud korpuslingvistikat ja tutvustatud käitumisprofiili meetodit ning antud teoreetiline ülevaade klasteranalüüsist.

Töö analüütiline osa jaguneb samuti kaheks. Kolmandas peatükis on kirjeldatud korpusanalüüsis märgendatud variaableid (ID-silte) ning *seisma* verbi korpuselauses esinevaid tähendusi. Neljas peatükk keskendub klasteranalüüsile ning selle tulemustele, samuti on neljandas peatükis võrreldud selle töö tulemusi varasemate uurimustega ning esitatud järeldused. Lisaks sisaldab töö kokkuvõtet, kirjanduse loetelu, ingliskeelset resümeed ja kolme lisa.

1. Polüseemia kognitiivse lingvistika raamides

1.1. Kognitiivsed protsessid tähenduse kujundajana

Kognitiivses lingvistikas lähtutakse sellest, et keelt saab uurida kui üht inimese kognitiivset võimet ja vormi asemel on peamine sisu ehk tähenduste uurimine ning see, millisel moel tähendused omavahel seotud on (Geeraerts 2006: 3).

Keelelises mõttes ei ole tähendus välismaailma peegeldus, vaid tõlgendus maailma tajumisest enda ümber. Tõlgendus oleneb aga vaatepunktist ehk inimesest endast. Enim võib seda tähele panna ruumi tajumisel. Olenevalt sellest, kus inimene ise asetseb, kirjeldab ta enda ümber toimuvat ja esemeid, mis tema vaatevälja jäävad. (Geeraerts 2006: 4)

Tähenduste kaudu tõlgendatakse ümbritsevat maailma, kuid maailm on pidevas muutumises, seega laienevad ka tähendusväljad ja pidevalt peab semantilisi kategooriaid uuendama. See tähendab, et midagi uut kogedes peab inimene kohandama kogemuse olemasolevatega. Kuna kogemusi väljendatakse keele kaudu, siis on oluline just keelekogemus ja see, kuidas seda omandatakse. Seetõttu nimetatakse kognitiivset lingvistikat ka kasutusühenduseks. (Geeraerts 2006: 4–5)

Kuna kognitiivses lingvistikas on kesksel kohal tähendus, siis on uurimisobjektiks polüseemia ehk see, kui ühel sõnal on mitu omavahel seotud tähendust. Sõna tähendused on omakorda ühendatud kindlate seaduspärasuste alusel. Kategoriseerimine on võrgustiku tekkimise aluseks ja on seega inimese üks üldistusvõime põhilisi avaldusi. Võib öelda, et nii moodustuvadki mentaalses ruumis tähendusvõrgustikud.

Kognitiivses keeleteaduses peetakse teedrajavaks polüseemiat uurivaks tööks Claudia Brugmani (1981) uurimust inglise sõna *over* 'üle' kohta (eesti keele kohta vt Veismann 2004). Brugman ja Lakoff väidavad, et polüseemse sõna tähendusi tuleb kirjeldada just

tähendusvõrgustiku abil, mitte neist lihtsalt nimekirja koostada (Brugman, Lakoff (1988) 2006: 138). Sellest alates on tähendusvõrgustike koostamisega ka tegeletud.

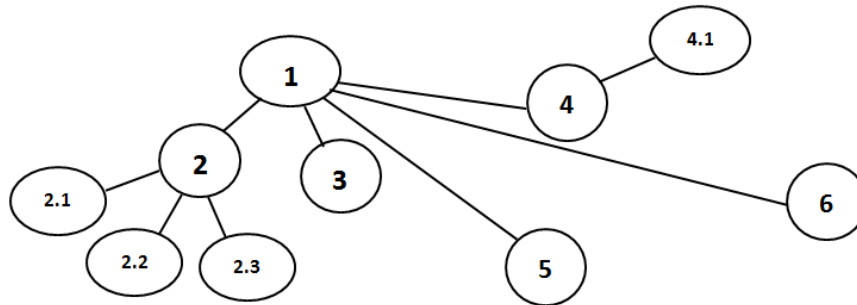
Järgnevates alapeatükkides on juttu sellest, kuidas on kategoriseerimine seotud tähendusvõrgustiku tekke ja prototüüpse tähendusega ning mis motiveerib prototüüpse tähenduse tekkimist.

1.1.1. Radiaalse tähendusvõrgustiku kujunemine

Psühholoog Rosch (1978) leidis, et inimesed kategoriseerivad objekte nende sarnasuse alusel ning esineb mingi keskne prototüüp, mis esindab selle kategooria omadusi kõige paremini. (Ravin, Leacock 2002: 13) Näiteks võib Põhja-Euroopas esindada õun kõige paremini puuvilja kategooriat, samal ajal võib lõuna pool olla prototüüp hoopis virsik.

Samamoodi esindab keeles iga esinevat kategooriat prototüüp (või mitu), millel on kategooriakesksed omadused ja radiaalsed, ähmaste piiridega äärealad, kus asuvad keeleüksuse need tähendused, mis on keskmest kõige kaugemal (vt joonis 1) (Ravin, Leacock 2002: 13–14). Nii tekib polüseemse sõna ümber radiaalne võrgustik, milles tähendused on seotud kognitiivsete printsiipide kaudu nagu metafoor, metonüümia ja kujundskeemid (Brugman, Lakoff (1988) 2006: 109–110).

Joonisel 1 on toodud näide radiaalsest võrgustikust, kus on sõna keskne ehk prototüüpne tähendus (1) ja sellest lähtuvad tähendused (2, 3 jne), millel võivad omakorda esineda alltähendused (2.1; 2.2). Tähendus 6 on joonisel kesksest tähendusest kõige kaugemal, seega äärealal ja ka kõige eraldiseisvam tähendus.



Joonis 1. Radiaalne tähendusvõrgustik.

Sellisel kategoriseerimisel esineb kaks eraldiseisvat tasandit: esmalt iga tähenduse sisestruktuur ja semantiline sisu, teisalt see, kuidas need tähendused võrgustikus asetsevad ja mentaalses leksikonis paiknevad. Seejuures esineb prototüüpne tähendus mõlemal tasandil. Võrgustiku tasandil on keskne tähendus sõna prototüüpne tähendus, sõnatasandil aga iga tähenduse semantiline sisu, kus prototüüpsus näitab, kui hästi sobib üks kindel tähendus mingisse teatud lausesse. (Brugman, Lakoff (1988) 2006: 112)

Selles magistritöös on oluline pigem teine tasand ehk kuidas asetsevad *seisma* verbi tähendused võrgustikus. Tähendused on parafraseeritud ning iga parafraasi illustreerimiseks on valitud korpusest laused, mis sobivad tähendusega kõige paremini ehk on selle tähenduse kesksed esindajad. Seega võib teoreetiliselt iga parafraasi alla kuuluvatest lausetest tekitada omaette võrgustiku.

See, kuidas tähendused võrgustikku asetada on aga keerulisem küsimus, millele lingvistid on vastust otsinud. Sõna keskse tähenduse välja selgitamiseks on kasutatud eri meetodeid: hindamistestid, sorteerimistestid, lapsel esimesena omastatav tähendus jne. Suuresti on sealjuures toetunud ka introspektsioonile. Viimasel ajal on aga üha enam kasutusel kasutuspõhised (*usage-based*) ja statistilised meetodid (vt ptk 2). (Gries, Divjak 2009: 60).

Peatükis 2.2. on esitatud üks uudne korpuspõhine lähenemine (*Behavioral Profile*), mille abil tähendusi võrgustikku asetada. Seejuures lähtutakse sõna kasutusest lauses ning kasutuse järgi asetatakse eri tähendused rühmadesse, tulemuseks on võrgustiku skemaatiline kujutus, dendrogramm ehk liigituspuu.

1.1.2. Kujundskeemide teke

Kognitiivse keeleteaduse seisukoht on, et inimese keelevõime on seotud üldiste kognitiivsete struktuuride ja kogemustega ning näitab seda, kuidas inimene maailmas toimuvat ja selles eksisteerivat keelelisse vormi paneb, seejuures on olulised kognitiivsed võimed, nagu võrdlemine, tähelepanu ja entsüklopeedilised teadmised.

Mõisteline struktuur on kehastatud ja tähenduse tekkimise aluseks on skemaatiline mõiste või mõistete kogum – kujundskeem (*image schema*, ka skeemkujutus). Näiteks tajume horisontaalsust-vertikaalsust oma keha põhjal (seistes on keha vertikaalne, lamades horisontaalne) ja kanname selle tunde üle teistesse valdkondadesse. Seega see, kuidas inimene oma keha seismise ajal tajub ja tunnetab, loob aluse mentaalse skeemi tekkeks. Kujundskeemid on üldised ega sisalda palju detaile. (Johnson 1987: xiv–xix)

Kognitiivse lingvistika raames tehtud uurimuste kohaselt eksisteerib rohkem kui 20 kujundskeemi, mis esinevad pidevalt inimese mõtlemises ja kujutluses. Põhilisemad on näiteks KONTEINER, TASAKAAL, TEEKOND, KESE/PERIFEERIA, SEOTUS. Nende abil mõistab inimene ka abstraktsemaid valdkondi. (Gibbs, Colston: (1995) 2006: 239)

Johnson (1987) toob näite TASAKAALU kujundskeemist. Tasakaalu tunnetame me oma igapäevases elus, kuid tajume seda teravamalt, kui näiteks laps õpib seisma ja kõndima, alguses värisedes ja kukkudes, kuid lõpuks tasakaalu leides, nii ka kahejalgse jalgrattaga sõitmist õppides. Samuti tajume tugevamalt tasakaalu olulisust, kui tunneme nõrkust ja silme eest läheb mustaks – tasakaal hakkab kaduma. (Johnson 1987: 74)

Seega kujuneb TASAKAALU kujundskeem inimese enda tasakaalu ja tasakaalutuse tunnetamisest ning me kanname selle ka abstraktsematesse valdkondadesse – tasakaalu korpus, eelarvepoliitika, jõud, energiad jne. Nendes tähendustes on tasakaalu kontseptsioon üle võetud meie enda kehalisest kogemusest.

Küsimus on aga selles, mis ühendab füüsilisi ja mittefüüsilisi tähendusi? On väidetud, et polüseemse sõna tähendusvälja saab iseloomustada metafoori, metonüümia ja kujundskeemide abil. Tähendusväli on seega struktureeritud kognitiivsete printsiipide

kaudu ning süsteemne. Need printsiibid kerkivad aga esile kehalisest kogemusest, see loob ka aluse prototüübile, mis on võrgustikus keskne. (Gibbs, Colston 2006: 243)

Seega on võimalus, et just kehaline kogemus on osaliselt alus sellele, kuidas me polüseemse sõna tähendusi laiendame ja tajume. Prototüüpse seismise kogemus loob aluse kujundskemide tekkimisele, mis omakorda motiveerivad semantilisi laiendusi teistesse valdkondadesse. (Lemmens 2012: 2).

1.2. Asendiverb seisma ja selle tähendused

Selles alapeatükis on juttu asendiverbist *seisma*, mis on koos verbidega *istuma* ja *lebama* üks põhiasendiverbidest (*cardinal posture verb*). Neist kolmest on *seisma* verb eesti keeles kõige polüseemsem ja sellel on „Eesti keele seletavas sõnaraamatus“ (EKSS) esitatud seitse põhitähendust koos hulga alltähendustega, samal ajal on *istuma* verbil viis põhitähendust ja *lebama* verbil üks.

Põhiasendiverbid viitavad inimese keha asendile, kuid nende tähendusväli laieneb ka teistele elusolenditele ja objektidele. Seismise puhul eeldame, et objekt on püstises asendis (vertikaalne) või seisab jalgadel (laud, tool). (Newman 2002: 7–8)

Asendiverbe ja nende polüseemiat on põhjalikumalt uuritud juba 1980. aastatest (van Oosten 1984; Gibbs jt 1994; Newman 2001; Newman, Rice 2004, Lemmens 2001, 2005 jpt.). Põhiliselt on huvi pakkunud asendiverbide polüseemia ja grammatiseerumise aste eri keeltes.

Keeled jagunevad asendiverbide kasutuse piirangute poolest – on keeli, kus asendiverbe kasutatakse ainult elusate objekti kirjeldamisel, näiteks prantsuse keel, samas inglise, nagu ka eesti keel lubab asendiverbe kasutada ka elutute entiteetide puhul. (Newman 2002: 7)

Täpsemalt kirjeldab Newman raamatus „The Linguistics of Sitting, Standing and Lying“ (2002) kolme põhiasendiverbi *seisma*, *istuma*, *lebama* prototüüpsemat tähendust.

Prototüüpne seismine on

- (1) kanooniline positsioon¹,
- (2) maksimaalselt vertikaalne,
- (3) jalgadele toetuv,
- (4) nõuab füüsilist pingutust,
- (5) (stardi)positsioon liikumise alustamiseks,
- (6) seostub võimu ja kontrolliga. (Newman 2002: 2, Lemmens 2012)

Newmani järgi (2002: 2–3) on sealjuures tegemist eri valdkondadesse kuuluvate omadustega. Näiteks ruumi-aja valdkond (2), aktiivne tsoon (3), jõudünaamika valdkond (4) ja sotsiaal-kultuurne valdkond (6). Alljärgnevalt kirjeldan neid valdkondi lähemalt.

Ruumi-aja valdkond viitab üldisele asetusele, mis säilib mingi aja vältel. Seismise (ja ka istumise ja lebamise) juures on oluline ajaline ulatus ja samal ajal mingis asendis püsimine. Seismise puhul on see asend vertikaalne ja jalgadele toetuv ning see tekitab omakorda ruumilise pildi inimese tajus ja mängib samal ajal rolli teiste (metafoorsemate) tähenduste motiveerimisel.

Jõudünaamika valdkonnas on oluline asendi säilitamiseks vajaminev jõud. Kuigi kõiki kolme põhiasendit säilitatakse ilma liigutamiset, peab mingil määral kasutama füüsilist jõudu, et asendis püsida. Seejuures vajab seismine kõige rohkem kontrolli. Seismise korral on „töös“ nii rindkere kui ka jalad, istumisel ainult rindkere ning lebamisel ei ole vaja jõudu rakendada. Seetõttu on seismine ka asend, mida on kõige raskem pikka aega säilitada.

Aktiivne tsoon on Langackerilt (1987) laenatud mõiste, mis viitab parajasti liikumise või ka seisundi säilitamisel aktiivsele osalisele, näiteks kudumisel on need käed. Seismisel on aktiivne osa eelkõige jalgadel, samas on oluline ka ülakeha vertikaalses asendis hoidmine.

¹ Mõiste on võetud Van Oostenilt (1984: 144), kes peab seismist, püstist asendit inimese normaalolekuks, lebamine ja istumine on tema järgi mittekanonilised.

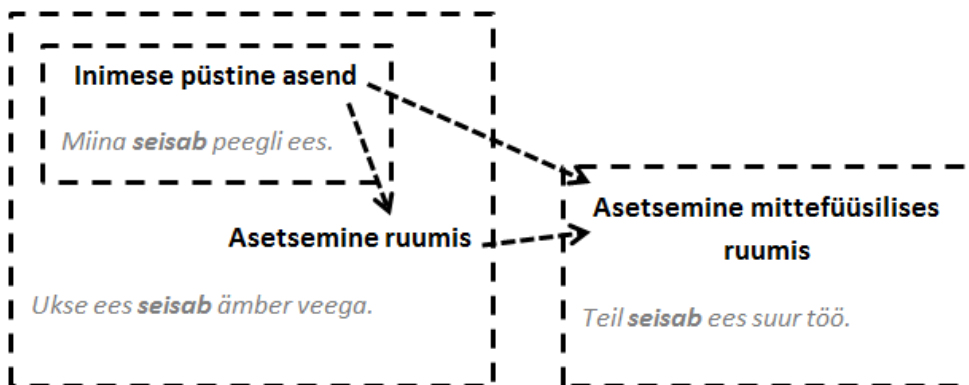
Sotsiaal-kultuurses valdkonnas võib seismise seostada võimu ja kontrolliga. Seda eelkõige võrreldes istumise ja lamamisega. Kui viimased on küllaltki mugavad asendid, istumine näiteks söömiseks, pikaajaliseks töötamiseks, seejuures ilma oluliselt väsimata, siis seismisel kasutatakse rohkem jõudu, vaade on kaugemale. Samuti on see positsioon liikumise alustamiseks ja ka lõpetamiseks. Lamamine seostub aga puhkamise, väsimuse ja isegi haigusega. (Newman 2002: 2–3)

Van Oosten peab aga seismist inimese kõige loomulikumaks olekuks. Selle üle võib vaielda, kuid Lemmens (2001) toob välja, et seismine on just see, mis eristab meie liiki primaatidest, samuti on seistes inimesed füüsiliselt tugevamad, kui teistes asendites. Seega võib inimene kõige paremini funktsioneerida, kui ta on seisvas asendis (Lemmens 2001: 1–2)

Kuigi *seisma* verbi prototüüpne kasutus viitab inimese asendile, siis paljudes maailma keeltes on asendiverbid üldist entiteedi asukohta väljendavad verbid, seda nii elus kui ka elutute objektide kirjeldamisel. Peale asendi väljendatakse nendega ka lokatiivsust (konkreetne entiteet asub füüsilises ruumis, prototüüpne seismine on alljuhtum) või metafoorsust, kus füüsiline ruum pole oluline (vt joonis 2). (Lemmens 2001:3–4)

Metafoorsed kasutused võib defineerida ka kui ülekandumised ühest valdkonnast (nt ruum) teise (nt aeg).

Joonisel 2 on toodud skemaatiliselt konkreetse tähenduse nihkumine metafoorsesse, sealjuures on oluline füüsiline ruum ja selle piirid. Kastides on toodud ka näited lausetest, mis illustreerivad *seisma* kasutust ruumis ja mittefüüsilises ruumis. Kui lauses *Miina seisab peegli ees* on tegemist konkreetse inimese (esimeses tähenduses) seismisega füüsilises ruumis ning lauses *Ukse ees seisab ämber veega* on üldisemalt tegemist lokatiivsusega, siis metafoorses lauses *Teil seisab ees suur töö* on nihkutud mittefüüsilisse ruumi ja aja valdkonda.



Joonis 2. Kolmeastmeline kontiinum inimese asendist füüsilisse ja abstraktsesse ruumi Lemmensi (2001: 6) järgi.

Selline lähenemine on seega sarnane Brugmani ja Lakoffi radiaalse võrgustiku teooriale – on üks keskne tähendus, millest hargnevad eri kaugustele teised tähendused, sh ülekantud tähendused.

Lemmens illustreerib seda kui teedevõrku, kus kõik teed on kuidagi pealinnaga ühendatud, kuid osadel pole otseühendust (Lemmens 2001: 6). Kui sellist teedevõrgu metafoori kasutada, siis on oluline ka see, et saab kasutada eri teid, et pealinna jõuda ning seetõttu võib olla raske otsustada, millist teed valida ehk milline on kõige lühem ja otsem distants.

Arvatavasti motiveerivad tähendusvõrgustiku laienemist ja metafoorsete tähenduste tekkimist kõik eelpool mainitud valdkonnad läbisegi, see teeb tähenduste vahel liiklemise raskemaks, kui ühe tähendust juures võib rolli mängida aja-ruumi valdkond (*seisin terve päeva järjekorras*), siis teisel näiteks hoopis sotsiaal-kultuurne valdkond või mõlemad (*seisab teadmiste poolest teistest kõrgemal*).

See, miks teatud mõtete väljendamise juures kasutatakse asendiverbi näiteks koopula asemel, tuleneb ka kinnistumisest (*entrenchment*) ehk sellest, mil määral mingi konstruktsioon on omaks võetud ja rutiiniks muutunud. Näiteks võib tuua *kirjas seisma*, mis on laialdaselt kasutusel igasuguse teksti kohta, mis kusagil eksisteerib. Selline kasutus on läbipaistmatu, kuid seda võib motiveerida see, et tähed seisavad kas mõttelisel või reaalsel joonel, nt vihikus. Viimane on Lemmensi (2001: 6) interpretatsioon hollandi *staan*

verbi kasutusest, kus ta nimetab *kirjas seismist* eriliseks asendiverbi kasutuseks. Kuid kuna see on kasutusel ka eesti keeles, siis arvatavasti on selline kasutus motiveeritud samadelt alustelt või tulnud kasutusse saksa keele kaudu.

Kui minna tagasi füüsilise ruumi valdkonda (ehkki on vaieldav, kas seisev tekst või tähed kuuluvad abstraktsesse või konkreetseesse ruumi), siis saab *seisma* verbi kasutada ka igasuguste esemete ja objektidega, mis toetuvad põhjale (*klaas seisab laual, jõulupuul seisab nurgas*).

Seisma verbi kasutust teistes valdkondades motiveerib kehaline kogemus ehk see, kuidas me ruumi tajume ja seda teistesse valdkondadesse üle kanname, seda saab selgitada eelmises alapeatükis mainitud kujundskeemide abil. Millised kujundskeemid on seismise juures olulised, sellest on juttu järgmises peatükis.

1.3. *Seisma* verbi polüseemia uurimisest

Gibbs jt (1994) on uurinud inglise verbi *stand* tähendusi ja nende tajumist nelja psühholingvistilise eksperimendiga, milles vaadeldi, kuidas kehaline kogemus mõjutab verbi tähenduste mõistmist. Eksperimentide käigus selgitati, et *seisma* verbi kasutus on mõjutatud põhiliselt viiest kujundskeemist.

Inimese seismise juures on kõige olulisemad TASAKAAL ja VERTIKAALSUS, peale nende kahe seostub seismisega veel KESE-PERIFEERIA (mingi objekt, sündmus on keskne, seda ümbritsev aga perifeerne), VASTUPANU (mingile välisele jõule, nt gravitatsioonile) ja SEOTUS (ühenduse tajumine objektide vahel). Kõige vähem seismisega seotuks pidasid katseisikud TEEKONNA kujundskeemi. Katse eesmärk oli näidata, et erinevad *stand* verbi kasutused pole arbitraarsed, vaid on motiveeritud kehalistest kogemustest maailmaga suhtlemisel ja inimesed tajuvad seoseid eri lingvistiliste aspektide vahel isegi siis, kui need on abstraktsed ja metafoorilised. (Gibbs jt 1994: 237–250)

Eestis on uurinud *seisma* verbi ja selle kasutust Renate Pajusalu (2001a, 2001b), Marek Kuk (2011), Ann Veismann (2012), oma bakalaureusetöodes on seda puudutanud ka Kleimann (2007) ja Kask (2012). Seejuures on enamik neist töödest toetunud „Eesti keele

seletavas sõnaraamatus“ esitatud *seisma* verbi põhitähendustele ja Pajusalu (2001b) sorteerimiskatsele nende lausetega (v.a. Kask ja Kleimann).

Pajusalu (2001a) toob välja, et inimese seismise juures on olulised TASAKAAL, VERTIKAALSUS ja (TOETUS)PÕHI. Objektide seismine on seotud näiteks nende täisnurkse asendiga aluse suhtes või kasutusvalmidusega (*tööriistad seisavad kuuris*). (Pajusalu 2001a: 171–183) Seega võivad objektide seismisel olla olulised samad määravad, mis inimestegi puhul (täisnurksus=vertikaalsus, kasutusvalmidus=(stardi)positsioon liikumise alustamiseks).

Sama tõdeb ka Lemmens (2001), kes toob näite (TOETUS)PÕHJA skeemi olulisusest ja väidab, et kujutlus objektist, mis toetub põhjale on tüüpiliselt seostatud seismisega. Seega on põhjale toetuva elutu objekti seismine nii kinnistunud, et laiendame seda ka teistele entiteetidele (mis pole vertikaalsed). (Lemmens 2001: 4)

Ka on seismine seotud võimega liikuma hakata või tähistab hetkelist liikumatust (*buss seisab peatuses*). Siin võivad koostööd teha mitu erinevat skeemi, esiteks toetub buss põhjale, rattaid võib pidada ka ülekantud tähenduses jalgadeks. Teiseks mängib rolli potentsiaalne liikumisvõime.

Korpuse materjali on sorteerimiskatse tulemuste võrdluseks kasutanud Veismann (2012). Analüüsitud on 205 lauset ajakirjanduskorpusest. Tegemist oli ajalehe Eesti Ekspress lausetega ja seal esines verb kõige rohkem (97 korda) abstraktses tähenduses ja 66 kasutust oli 'jalgadel, püstiasendis, elus' tähenduses ja 42 'paigal, elutu tähenduses'. Sellest võib järeldada, et ajakirjanduskeeles ületatakse aja-ruumi piire enim ja kaldutakse metafoorse poole peale.

Põhjalikku korpusuuringut pole *seisma* verbi kohta veel tehtud. Järgmises peatükis tutvustangi korpuspõhist meetodit, millega verbi kasutust uurin.

2. Meetodi kirjeldus

Selles peatükis tuleb juttu korpuspõhisest meetodist, mida selles töös on kasutatud. Kõigepealt on tutvustatud kasutuspõhist lähenemisest keeleteaduses ja selle ühte meetodit korpuslingvistikat. Lühidalt on peatutud ka kriitikal, mis korpusanalüüsi vastu on esitatud. Seejärel tutvustatakse tänapäeva moodsaid korpusi ja eri lähenemisi, mida korpusuuringutes on kasutatud.

Peatükis 2.2. on täpsemalt kirjeldatud töös kasutatud käitumisprofili meetodit – kuidas seda koostada ja mida sellega uurida. Viimasena on juttu klasteranalüüsist ning sellest, mida käitumisprofiilist saadud andmetega edasi teha ja millised valikud sealjuures võimalikud on.

2.1. Kasutuspõhine lingvistika

Kasutuspõhise lähenemise (*usage-based approach*) põhiseisukoht on, et grammatika tuleneb inimese kognitsioonist ja kogemustest keelega ning keelekasutusel on mõju keele struktuurile. Esmakordselt kasutas seda nimetust 1987. aastal Ronald W. Langacker, kes tõi välja, et „põhitähtsus omistatakse tegelikule keelesüsteemile ja kõneleja teadmisele selle kasutusest“. Seega vastupidiselt generatiivsele teooriale, milles keel on n-õ sisse programmeeritud ja õppimisvõime sellega minimaalselt seotud, on kasutuspõhises lähenemises oluline keele kasutus ja selle omandamine just läbi õppimise. (Langacker 2000: 91)

Praegu on sellel terminil palju laiem kasutus. Lähenemine on arenenud ja kasvanud omaette haruks nimega kasutuspõhine lingvistika, mis hõlmab erinevaid meetodeid ja uurimisvaldkondi, nagu grammatiseerumine, keele omandamine, sotsiolingvistika, fonoloogia jm. Otsapidi on see seotud konstruktsioonigrammatika (vt Goldberg 1995), esilekerkiva (*emergent*) grammatika (vt Hopper 1988) ja keele omandamise uurimisega (vt

Tomasello 2003). Joan Bybee sõnul annab kasutuspõhine lähenemine ja see kui me ei eralda struktuuri kasutusest meile võimaluse selgitada mingit lingvistilist fenomeni selgemalt ja täpsemini. (Bybee 2012: 1–2)

Kuna keelekasutus saab kasutuspõhises lähenemises keskse koha ja grammatika on selle keelekasutuse produkt, siis analüüsides konkreetseid kasutusjuhte ja nende esilekerkimist, saab pilgu heita üldisse keelesüsteemi. Kasutusjuhu all on siinkohal silmas peetud konkreetset üksikut keelekasutuse korda, mis siis vastavalt moduleerib inimese keelesüsteemi iga kord veidi ümber. Neid üksikuid kasutusjuhte analüüsides võib aga leida üldisemaid mustreid. (Tummers jt 2005: 228)

Sellega nihkub tähelepanu keelesüsteemi tundmisest selle kasutusse suhtlusprotsessis. Samuti kaob opositsioon leksikoni ja grammatika vahel ja selle asemel võib esitada hoopis skemaatilise võrgustiku, milles on esindatud nii abstraktsed grammatilised mustrid kui ka nende leksikaalsed esitlused. (Tummers jt 2005: 228)

Kasutuspõhises lingvistikas uuritaksegi kasutusmustreid, esinemissagedust, varieeruvust ja muutusi keeles. Selleks kasutatakse eri vahendeid nagu korpus ja diakroonilised andmed, samuti psühholingvistilisi eksperimente, keeltevahelist võrdlust ja uuritakse lapse keele arengut. (Bybee, Beckner 2009: 827) Seega on meetodite valik lai ja ühtegi meetodit ei jäeta kõrvale, et keele kohta midagi uut teada saada.

Kasutuspõhise lingvistika üks võimalik meetod on korpusuuring, mis võimaldab uurida loomulikku keelekasutust, leida kasutusmustreid ja neid analüüsida ning just seda on kasutatud ka selles magistritöös.

Tummers, Heylen ja Geeraerts (2005) arvavad koguni, et ei saa olla mingit kasutuspõhist lähenemist, kui ei uurita tegelikku keelekasutust ja nende sõnul on korpuslingvistika selleks üks parimaid meetodeid.

2.1.1. Korpuslingvistika

Korpuslingvistika ja korpuslingvistiliste meetodite rakendamise võib jagada kahte perioodi: varajane korpuslingvistika (ka Chomsky-eelne) ja moodne korpuslingvistika. Moodsa korpuse all mõeldakse elektroonilise keeleainese kogu, milles on spontaansed ja loomulikud keeleandmed. (McEnery, Wilson 1996: 2–3)

Chomsky ja generatiivse grammatika ajastul oli korpuslingvistika tagaplaanil, sest Chomsky ei pidanud korpust arvestatavaks vahendiks keele uurimisel. Tema arvates pidi keelt modelleerima keelepädevuse mitte -kasutuse järgi. Korpustega tegeldi küll vähemal määral ka generatiivse grammatika hiilgeajal, kuid alles 80ndatel hakkas huvi korpusuuringute vastu suurenema. Sel ajal võeti kasutusele arvutid ning nende abil sai koostada elektroonilisi andmebaase ja samas neid ka kergemini hallata ja analüüsida. (McEnery, Wilson 1996: 6, 24) Seega on kognitiivne lingvistika (mis sai samuti alguse 1980ndatel) ja korpuslingvistika arenenud külge külje kõrval teineteist toetades.

Korpuslingvistika ei ole lingvistika haru traditsioonilises mõttes, nagu süntaks, semantika või sotsiolingvistika. Tegemist on meetodikaga, mida viimaste uurimisel kasutada. (McEnery, Wilson 1996: 2) Ja nii nagu igal meetodil on pooldajaid ja vastaseid, siis pole erandiks ka korpuslingvistika.

Nagu juba eelnevalt mainitud, siis Chomsky ei pidanud korpusanalüüsi tõsiseltvõetavaks vahendiks keele uurimisel. Tema järgi oli keelekasutus hale vari keelepädevusest ja seega ei olnud see õige vahend, et keelt modelleerida. Samuti ei peegeldanud korpuse laused formalistide ideede järgi kogu keele kompleksust, mida meie keelepädevus võimaldab. (McEnery, Wilson 1996: 6) Formalistid pooldasid meetodina introspeksiooni ning eeldasid, et kogu keele uurimiseks vajaminev on sisemine teadmine, kuidas keelt kasutada.

Tänapäeval on meetodite rohkus ja kombineerimine pigem tavaks, seega on korpuslingvistika aktsepteeritud kui loomulik viis, kuidas keeleandmeid koguda ja analüüsida. Ometi on ka praegu keeleteadlasi, kes pigem kahtlevad korpuslingvistika võimalustes ja võimekuses.

Soome keeleteadlane Jarno Raukko on korduvalt korpuslingvistikat kritiseerinud (1999, 2003) ja on korpusanalüüsi osas skeptiline, seda eriti polüseemia uurimisel, sest väidetavalt toetub lingvist seal põhiosas – samuti nagu introspektsiooni puhul – omaenda intuitsioonile, kui analüüsib ja klassifitseerib sõnu ja nende tähendusi.

Tema sõnul oleks õigem kasutada eksperimentaalseid meetodeid, nt tema enda välja mõeldud tootmistesti (vt Raukko 1999). Sealjuures toetutakse tulemusi interpreteerides mitte-lingvistidest keelejuhtidele ja seega on saadud andmed tema sõnul „intersubjektiivsed“. Korpusi peab ta teatud määral eeltöödeldud tekstideks, kust lingvist leiab endale sobivad sõnakasutusjuhud ja seejärel tõestab oma hüpoteesi. (Gries, Divjak 2010: 335–337)

Selline arvamus kehtiks ehk tõesti korpusega illustreeritud analüüsi korral (vt ptk 2.1.2.), kuid korpusepõhise lähenemise juures ei ole see päris tõene. Korpuslingvistid Gries ja Divjak arvavad, et tegelikult on just vastupidi – vaadatakse kõiki korpuses esinevaid kasutusjuhte ja sageli võib nende hulgast leida selliseid, mille peale pelgalt introspektsiooniga ei tuleks. Oma intuitsiooni järgi märgendamise kaitseks võib öelda, et iga lingvist toetub mingil määral intuitsioonile, muidu ei saaks ju uurimisküsimusi püstitatagi. Korpuste puhul on vähemalt see hea, et tegeletakse loomulike tekstidega ja saab ülevaate sõna või väljendi kasutusest ja sagedustest. (Gries, Divjak 2010: 335–337)

2.1.2. Moodsad korpused

Praegused mahukad elektroonilised korpused võimaldavad uurida hõlpsalt tegelikku keelekasutust ja lekseemide või konstruktsioonide esinemissagedust keeles.

Kõige loomulikumat andmed on korpuses siis, kui salvestatakse emakeelse kõneleja juttu nii, et ta seda ise ei tea. See tõstatab aga eetilise küsimuse ehk kas niiviisi saadud materjali võib teadustöös kasutada ja seega on enamasti korpuses tegemist kirjalike tekstidega. Siiski võib korpuse tekste pidada paremaks algmaterjaliks, kui introspektsiooni, küsitluste või eksperimentide teel saadud andmeid, sest nende saamiseks pole kõnelejaid

kuidagi suunatud või pandud neid tegema midagi, mida nad keelega tavaliselt ei tee (nt sorteerimiskatse või vastuvõetavuse hindamise katse). (Tummers jt 2005: 229–230)

Samas on väga spetsiifilisi uurimisküsimusi ainult korpuseanalüüsi abil raske testida, kuna korpusest saadud andmetes ei pruugi alati leiduda seda, mida parajasti uuritakse. Seega täiendavad korpusanalüüs ja eksperimendid teineteist ja korpuslingvistid võiksid oma uurimustes neid rohkem kasutada (vt Gilquin, Gries 2009).

Et korpus oleks representatiivne ehk esinduslik peab selles olema esindatud igat liiki tekste. Kui korpus koosneb ainult ajakirjandustekstidest või ainult ilukirjandusest, siis ei saa olla kindel, et tulemused, mis nende andmete põhjal saadakse on üldistatavad kogu keelekasutusele. Ideaalis oleks korpuses esindatud nii suuline kui ka kirjakeel ja allkeeled, ja seda võrdses mahus. Tegelikkuses on sellist korpust raske koostada, sest keel areneb pidevalt ning muutub ajas. (Eslon 2010: 1)

Kasutuspõhised analüüsimeetodid korpuslingvistikas saab omakorda jagada kolmeks: korpusega illustreeritud (*corpus-illustrated*), korpuspõhised (*corpus-based*) ja korpusest tulenevad (*corpus-driven*) meetodid. (McEnery, Hardie 2012: 6, Tummers jt 2005: 227)

Korpusega illustreeritud meetodi korral kasutatakse korpusematerjali, et illustreerida introspektsiooni käigus tekkinud hüpoteese. Seega, kui teooria eeldab mingit nähtust ja see korpuses esineb, siis on see justkui kinnitus teooria õigsusest. Kui teooria eeldab, et mingit nähtust keeles ei esine, kuid see eksisteerib korpuses, siis on teooria kummutatud.

Korpuses esinevaid mustreid interpreteeritakse kui grammatilisi hinnanguid ja võetakse kui tõestusmaterjali teoreetilisele raamistikule. Samas ei arvestata korpuse sagedusandmetega ja alternatiivsete võimalustega, mis korpuses võivad esineda. Korrektne analüüs eeldaks, et ka teisi sarnaseid mustreid vaadeldakse süsteemselt ja sagedust hinnatakse statistika põhjal. Seega arvestades moodsate korpuste mahtu ja tehnilisi võimalusi, võib öelda, et „kasutuspõhine lingvistika vajab kvantitatiivset ja statistilist analüüsi“. (Tummers jt 2005: 234–235)

Korpuspõhine meetod ongi eelneva edasiarendus, kus empiirilised andmed mitte ainult ei toeta eelnevalt koostatud teoreetilist raamistikku, vaid korpusmaterjal moodustab põhiosa uurimusest ning seda analüüsides ja interpreteerides saadaksegi tulemused. Sealjuures on oluline ka statistiline analüüs, mis on korpusanalüüsi põhialus ja võimaldab leida keelekasutuses tendentse, mis pelgalt introspektsiooni abil oleks võimatu. (Tummers jt 2005: 235–236)

Korpusest tuleneva lähenemise puhul leitakse, et korpus peaks olema põhiallikas, mille põhjal hüpoteese püstitatakse. Ehk siis leitakse korpusest ühe või teise keelekasutusvariandi iseloomulikud nähtused, mille uurimist jätkatakse korpuspõhise analüüsiga. (McEnery, Hardie 2012: 6) Seega on see korpuspõhise meetodiga otseselt seotud, kuid kui eelneval on enne korpuse juurde pöördumist mingid eeldused või uurimisküsimused olemas, siis korpusest tuleneva lähenemise juures pööratakse esmalt korpuse juurde, et luua mingi teooria.

Kui korpusega illustreeritud lähenemine ei kasuta korpuse võimalusi täiel määral ja korpusest tulenev lähenemine vastupidi toetub liialt korpusele, siis kesktee võiks olla korpuspõhine lähenemine, mida on kasutatud ka selles magistritöös.

2.2. Sõna käitumisprofiil (*Behavioral Profile*)

Korpuslingvistid Stefan Th. Gries ja Dagmar Divjak on sõna tähenduste ja nende seotuse uurimiseks ühendanud Atkinsi 1987. aastal uurimistöös kasutatud lingvistilised parameetrid *ID-tag*'id (süntaktilised või leksikaalsed markerid, mis viitavad kindlale sõna tähendusele sõnaraamatus) ja Hanki 1996. aastal välja pakutud *Behavioral Profile*'i (BP) (vt Gries 2006; Divjak, Gries 2006; Gries, Divjak 2009, 2010; Berez, Gries 2008). Viimane kujutab täiendatud versiooni *ID-tag*'ide märgendamisest, kuhu on lisatud sõna semantiline roll lauses. Selle abil saab sõna käitumisprofiilis esinevate mustrite järgi uurida selle tähendusi ja vaadata, milline (ja kui tugev) seos mustrite vahel valitseb. (Gries 2006: 62)

Gries ja Divjak on lisanud Hanksi ja Atkinsi loodule juurde omapoolsed täiendused, eeskätt *ID-tag*'ide arvu ja sisu osas, samuti on täiendatud tulemuste (kvantitatiivset) analüüsi, nt klasteranalüüsi näol. Suurte korpuste olemasolu lihtsustab ka materjali kättesaadavust ning täiendatud arvutiprogrammid ja vabavarapaketid pakuvad võimalusi põhjalikuks statistiliseks analüüsiks.

Esmakordselt kasutas *Behavioral Profile*'i meetodit 2006. aastal Stefan Gries, kui uuris inglise keele polüseemse verbi *run* tähendusi. Griesi sõnul on selle meetodiga ühendatud lünk kognitiivse lingvistika ja korpuslingvistika vahel. Kuigi kognitiivlingvistidele oli selge, et tähendused moodustavad võrgustiku, siis toetuti enamasti oma töödes korpusest saadud sagedusandmetele. Griesi huvitas aga see, kuidas korpuse andmeid kasutades uurida, millisel moel need tähendused omavahel seotud on. (Gries 2006: 57)

Glynn (2012) märgib, et Griesi 2006. aastal ilmunud artikkel „Corpus-based methods and cognitive semantics: The many senses of *to run*“, on üks kõige olulisemaid polüseemia-alaseid uurimusi kognitiivses lingvistikas. Eriline pole mitte ainult kasutatud meetodika ja selle kasutusvõimalused, vaid et polüseemset sõna ja selle prototüüpeid tähendusi on kirjeldatud korpuse andmetele tuginedes selgelt ja ülevaatlikult. (Glynn 2012: 1)

BP annab uusi võimalusi, kuidas valimile läheneda, andmeid analüüsida ja tulemusi saada. Peale polüseemia uurimise saab BP-d kasutada ka sünonüümsete sõnade profiilide koostamiseks ja võrdlemiseks, samuti annab see häid tulemusi keeltevahelistes võrdlustes – nt märgendades eri keelte lausete ühiseid jooni. Alati võib saadud andmeid ka kontrollida või võrrelda teistsuguste meetoditega nagu näiteks katse. Veel saab uusi saadud tulemusi võrrelda juba varem läbi viidud katsete ja uurimustega.

Sõna käitumisprofili alus on andmepõhine operatsionaliseerimine² sõna või sõnavormiga ja selle „käitumise“ uurimine korpuses. Lähtutakse eeldusest, et sõnad või tähendused on osa tähendusvõrgustikust ja lähestikku asetsevad elemendid (täiendused)

² Andmete viimine sellisele kujule, et nendega on võimalik teha statistilist uuringut.

käituvad samamoodi nt fonoloogiliselt, süntaktiliselt või semantiliselt. (Gries, Divjak 2009: 63–64)

Käitumisprofili koostamisel märgendatakse võimalikult palju informatsiooni iga lause kohta, milles lekseem esineb ning seeläbi saadakse esmalt mingid sagedusandmed. Jaotussageduse järgi saab aga uurida ka funktsionaalset sarnasust – kuidas, mis rollides ja tähendustes funktsioneerib lekseem lauses. (Gries, Divjak 2010: 338)

Meetodil on neli osa:

1. Leida representatiivsest korpusest (kõik) sõna kasutusjuhud (terve lause)
2. (Pool)manuaalne märgendamine (ID-sildid):
 - morfoloogia (aeg, aspekt, kõneviis, tegumood jne)
 - süntaks (lausetüüp, pea- või kõrvallause)
 - semantika (sõna tähendused, semantiline roll lauses)
3. Koosinemistabeli koostamine
4. Analüüs (sagedus, korrelatsioon, klasteranalüüs) (Gries, Divjak 2010: 338)

Profili koostamiseks tuleb esiteks leida korpusest kõik sõna kasutusjuhud ja koostada neist representatiivne valim. Kindlasti on vaja konteksti ehk valimis peavad olema terved laused, mõningatel juhtudel on vaja ka laiemat konteksti. Seejärel tuleb laused märgendada, kasutades selleks *ID-tag*'e ehk ID-silte. Need võivad olla morfoloogilised omadused, nt aeg, kõneviis, pööre; süntaktilised omadused, nt kas lauses esineb kohamäärus, viisimäärus, ajamäärus; semantilised omadused – mis tähendustes sõna on kasutatud. Mida rohkem on ID-silte, seda parem. Märgendatud omadustest koosnebki profiil, mis on omakorda allikas edasisele (statistilisele) uuringule. (Gries, Divjak 2009: 64–66)

Kui laused on märgendatud, tuleb koostada koosinemistabel. Tabelis on kirjas kui mitu korda iga ID-sildi tase esines, see tähendab nt kui mitu korda kasutati uuritavat sõna küsilases, kui mitu korda esines mingi agenditüüp, kas verb esines peamiselt pea- või kõrvallases, palju esines aja- või kohamäärust jne. Nende esinemissageduste põhjal saab välja arvutada protsentuaalsed erinevused iga ID-sildi raames. (Gries, Divjak 2010: 339–340)

Polüsemse sõna puhul on olulised just erinevad tähendused ja milline on sõna keskne tähendus. Koosinemistabeli andmeid uurides saab vaadata, milline tähendus esineb näiteks erinevate ID-sildi tasemetega (nt minevikus, olevikus, kindlas ja tingivas kõneviisis, erinevate tegijatega (agentidega) jne). See tähendab, et see on kõige vähem piiratud tähendus ja kasutusel eri lausetüüpides, seega võib esitada hüpoteesi, et see võib olla keskne tähendus. (Gries, Divjak 2010: 339–340)

Kõik see annab juba mingi esialgse pildi verbi tavapärasest „käitumisest“ ja tähendustest lausetes. Edasi saab andmetega teha statistilist analüüsi (nt rakendada klasteranalüüsi).

2.3. Klasteranalüüs

Klasteranalüüs on andmeanalüüsi meetod, mis jagab objektid (nt sõnad, tähendused) rühmadesse vastavalt nende sarnasusele – sarnasus kahe samasse rühma või klastrisse kuuluva objekti vahel on maksimaalne ja eri rühmadesse kuuluvate objektide sarnasus on minimaalne. Klasteranalüüs on hea viis, et leida suures andmehulgas mingit struktuuri, seega saab seda kasutada, et leida andmetes esinevat süsteemi ilma selgituseta, miks see just selline on. Sageli on see kasutusel siis, kui ei ole mingit eelnevat hüpoteesi, vaid oma tööga ollakse uurivas faasis. (Divjak, Fieller *ilmumas*) Klasteranalüüs sobib hästi ka siia töösse, sest magistritöö eesmärk on uurida üldiselt *seisma* verbi (ja selle tähenduste) käitumist korpuses, mitte kontrollida mõne konkreetse hüpoteesi paikapidavust.

Klasteranalüüsi viiakse üldjuhul läbi valimi, mitte üldkogu andmete põhjal, kuna eeldatakse, et valimile toetudes moodustavad andmed piisavalt adekvaatsed klastrid, mis samas ka üldkogu esindavad. Seega peab uurija olema veendunud, et moodustatud valim on representatiivne ehk esinduslik. (Krillo 2010: 75)

Kõik klasteranalüüsid on algoritmilised, see tähendab, et klastrid on koostatud teatud algoritmi järgi. Neist õige valikul tuleb arvestada mitme olulise osaga. Esiteks kuidas arvutada vahemaad klastrate vahel. Teiseks tuleb otsustada, kas moodustada klastrid nii, et iga objekt on alguses omaette klaster ja viimaks moodustavad kõik kokku ühe klastri

(aglomeratiivne ehk liitmismeetod) või moodustavad kõik objektid algul ühe klatri, seejärel tuvastatakse kõige erinevamad objektid ning moodustatakse väiksemad klastrid (jaotusmeetod). (Divjak, Fieller *ilmumas*)

Järgnevates alapeatükkides on kirjeldatud, kuidas mõõta vahemaad klastrite vahel (ehk kuidas mõõta objektide sarnasust) ning hierarhilist aglomeratiivset klasteranalüüsi, mida oma töös kasutan. Mittehierarhilisi meetodeid selles töös ei kirjeldata.

2.3.1. Objektidevahelise kauguse (sarnasuse) mõõtmine

Selleks, et vaadata, kas kaks objekti on omavahel sarnased, tuleb hinnata nende sarnasust (*similarity*) või erinevust (*dissimilarity*). Sarnasus on seejuures arvuline mõõt sellest, kui sarnased kaks objekti on. Erinevus näitab vastupidi distantssi, kui erineva olemusega objektid on. Sealjuures viitavad suuremad väärtused ulatuslikumale erinevusele. (Divjak, Fieller *ilmumas*)

Sarnasus on enamasti kombineeritud mitmest variaablist või omadusest. Näiteks võib lingvistilises uurimuses ühendada mitu eri omadust: kasutussagedus, silpide arv, tähendus jne. Seejärel tuleb otsustada, kuidas variaablid numbrilisele skaalale viia. Sealjuures võib teha seda väga jämedakoeliselt ja kasutada vastavalt sarnasuse astmele lihtsalt numbreid 0, 1, 2, 3 või teisi numbrilisi väärtusi, näiteks protsente. See oleneb suuresti andmete iseloomust. Kuna lingvistiline materjal on enamasti väga mürarikas, siis tuleb olenevalt konkreetsest uuringust valida parim võimalus. (Divjak, Fieller *ilmumas*)

Kui andmed on viidud arvulisele kujule, tuleb järgmiseks otsustada, kuidas mõõta distantse nende vahel. Klasteranalüüsi raames on selleks erinevaid võimalusi, neist enimkasutatav on **Eukleidese distantss** (*Euclidean distance*), mis on geomeetriline vahemaa kahe punkti vahel. Lihtsamalt öeldes on see vahemaa, mis on linnulennul punktist A punkti B. Eukleidiline kaugus on üks kõige enam kasutatud mõõte lingvistilises analüüsis (Gries, Divjak 2009: 71).

Eukleidese kaugust (ja ka selle ruutu) arvutatakse reeglina algandmete pealt ning mitte standardiseeritud andmetega. Eukleidese distantss saadakse, kui kahe punkti vaheline

kaugus võetakse ruutu ja liidetakse omavahel ning võetakse saadud summast ruutjuur. Distanti ruutu kasutatakse näiteks Wardi meetodi korral (vt ptk 2.3.2.). (Krillo 2010: 77)

Kui Eukleidese distantsi arvutamiseks kasutatakse erinevuste ruutude summat, siis **City-block'i** ehk **Manhattani** ehk **absoluutse distantsi** korral kasutatakse erinevuse mõõdu arvutamisel näitajate absoluutsete erinevuste summat. Manhattani vahemaad mõõdetakse täisnurkselt, st et vahemaa pole mitte linnulennult, vaid peab minema justkui mööda täisnurga all asetsevaid tänavaid, et jõuda punktist A punkti B. Seega on Manhattani vahemaa alati pikem, kui Eukleidese oma. (Krillo 2010: 77–78)

Mõlema distantsi puhul on aga puuduseks see, et on ülioluline, mis ühikuid parajasti kasutatakse ja kuidas need teisendatud on. Kui näiteks kasutatakse tuhandete asemel miljoneid või gramme kilogrammide asemel, siis ei muutu mitte ainult vahemaa, vaid ka objektide järjestus klastrites. (Divjak, Fieller *ilmumas*).

Eelpool mainitud kaks vahemaa arvutamise võimalust on ühed sagedasemad, kuid peale nende on veel teisi, näiteks **Canberra** distant, mis on edasiarendus Manhattani distantist. Canberra meetodi korral lahutatakse kõigepealt kahe objekti väärtused ja saadakse erinevus, mis kahe objekti vahel esineb ning seejärel jagatakse tulemus samade objektide absoluutväärtuste summaga ($A-B/A+B$). Selline meetod sobib hästi, kui andmetes on palju 0-väärtusi. (Lance, Williams 1967)

Kuigi vahemaa kahe objekti vahel mõõdab arvutiprogramm, on siiski vaja andmetest lähtuvalt teha valik, millise meetodiga on parim viis vahemaad variaablite vahel mõõta. Selleks tuleb vaadata oma andmete iseloomu ja väärtusi, pole üht ja ainsat meetodit, mis iga kord toimiks ja parimaid tulemusi annaks.

2.3.2. Hierarhiline aglomeratiivne klasteranalüüs

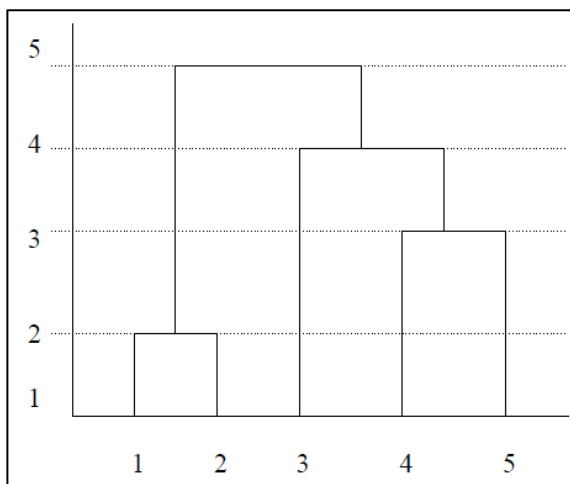
Kui objektidevaheline vahemaa on mõõdetud võib nende klastritesse jagamisel kasutada kas hierarhilist või mittehierarhilist meetodit. Hierarhilise meetodi korral ei ole klastrite arv eelnevalt kindlaks määratud ja uued klastrid moodustatakse alati olemasolevate põhjal. Aglomeratiivne ehk liitmismeetod on hierarhiline.

Aglomeratiivse klasterdamise juures on algselt iga indiviid omaette klaster. Esimese sammuna ühendatakse omavahel kõige sarnasemad klastrid ja seejärel liidetakse sammhaaval juurde sarnasuselt järgmised klastrid, kuni kõik objektid on omavahel ühendatud. Klastritesse jagamise protsessi saab kujutada graafiliselt ja üks võimalus selleks on dendrogramm ehk liigituspuu. (Divjak, Fieller *ilmumas*)

Dendrogrammil (vt joonis 3) on visuaalselt esitatud distants, kui kaugel objektid teineteisest asuvad. Dendrogrammi loetakse alt üles – horisontaalteljel on toodud objektid ja vertikaalteljel distantsi väärtus ehk seotuse kaugus. (Divjak, Fieller *ilmumas*)

Dendrogrammi tõlgendamisel tuleb arvestada, et need kaks tähendust, mis on kokku pandud, on omavahel kõige sarnasemad, kuid seda, kui sarnased need teistele (ja ka omavahel) on, näitab vertikaalse joone pikkus.

Jooniselt 3 on näha, et objektid 1 ja 2 on omavahel sarnasemad, kui kõik ülejäänud objektid üksteisele. Järgmise sammuna lisandub 3. objekt, mis moodustab omaette haru, seejärel lisatakse objektid 4 ja 5, mis on sarnasemad teineteisele, kui objektile 3.



Joonis 3. Dendrogramm ehk liigituspuu.

Kui omavahel kõige sarnasemad klastrid on moodustatud, tuleb määrata distants järgmiste klastritega. Selleks on vaja ühendamise reeglit (algoritmi), mille alusel on kaks klastrit piisavalt sarnased, et need omavahel kokku siduda.

Hierarhilistest meetoditest kasutatakse mitmeid meetodeid, mis eristuvad klastrite distantsi definitsiooni põhjal. Enimkasutatavad on

- **Minimaalse seose meetod** (*single linkage, nearest neighbour*) – klastritevaheline kaugus on arvatud nende lähima naabri järgi ehk distants kahe klastri vahel oleneb sellest, mis on vahemaa kahe lähima objekti vahel neis klastrites. Selle meetodi korral võivad ühte klastrisse kuuluda küllaltki erinevad objektid ja dendrogramm on välja venitatud nagu uss.
- **Täieliku seose meetod** (*complete linkage, furthest neighbour*) – klastritevaheline distants on kõige kaugemate liikmete vaheline distants. Täieliku seostatuse meetod võimaldab vältida pikkade klastrite teket ja tekkivad klastrid on kompaktsamad, sest iga uus liige peab asuma lähedal kõigile liikmetele klastris.
- **Keskmise seose meetod** (*average linkage*) – klastritevaheliseks distantsiks on klastri keskmine distants erinevates klastrites, tavaliselt kasutataksegi lihtsalt aritmeetilist keskmist. Kui minimaalse seose meetodiga moodustuvad väljaveninud, pikad objektide read ja täieliku seose meetodiga väga kompaktsed ja kokkusurutud klastrid, siis keskmise seosega meetodi puhul on ka tulemus nende kahe vahel.
- **Wardi meetod** – tegemist on aglomeratiivses analüüsis laialt kasutatud algoritmiga. Wardi meetodiga saab suhteliselt kompaktsed ja eri suurusega rühmad. Objekt liidetakse selle grupiga, mille puhul objektidevaheliste kauguste ruutude summa on minimaalne. Et vältida mõnede tunnuste domineerivat mõju on soovitatav muutujad enne sarnasuste arvutamist standardiseerida. (Divjak, Fieller *ilmumas*, Krillo 2010: 79)

Kui välja on valitud, millist distantsi mõõtu (Eukleidese, Manhattani, Canberra) kasutada ning kuidas klastrid omavahel siduda (nt täieliku, minimaalse või keskmise seose meetodiga) ja dendrogramm on valmis, siis järgmine samm on selle tõlgendamine. Interpretatsioonietapis tuleb analüüsida iga klastrit, et anda neile nende olemust võimalikult täpselt iseloomustav nimetus. Samuti tuleb analüüsida, kas saadud tulemused on kooskõlas teooriaga ja varasemate uurimuste tulemusega. Erinevuste (ja ka sarnasuste)

korral tuleb uurida nende võimalikke põhjusi. Samuti tuleb hinnata, kas tulemused on omased üldkogumile. (Krillo 2010: 81–82)

Põhjus, miks erinevatest meetoditest on ülevaade tehtud on see, et analüüsi tegema hakates tuleb uurijal palju valikuid teha – millised variaablid valida, kuidas need numbrilisele kujule viia ja mis nendega siis täpsemalt edasi teha, kuidas distantse mõõta, millise programmiga seda teha jne. Nagu eelpool mainitud, siis üht ja õiget ning kõige paremat valikut pole, seega võib uurija katsetada eri meetodeid, arvestada tuleb sealjuures oma andmete eripära.

Selles töös on kasutatud **Canberra distantssi**, kuna see sobib hästi juhul, kui andmehulgas on palju 0-lähedasi väärtusi³, Klastrid on omavahel ühendatud **Wardi meetodiga**, kuna sellega saab moodustada kompaktsed ja eri suurusega rühmad ning see on laialt kasutusel algoritm, mida paljudes lingvistilistes uurimustes kasutatakse. Samuti seetõttu, et andmeanalüüsis on kasutatud standardiseeritud andmeid.

Magistritöö analüütiline pool keskendub eesti keele polüseemse verbi *seisma* korpusanalüüsile ja klasterdamise tulemustele.

³ Kuna selles töös on märgendatud palju eri tasandeid (vt ptk 3.1.), siis on mõne *seisma* verbi tähenduse juures nende tasemete esinemissagedus väga väike või 0.

3. *Seisma* verbi käitumisprofili ID-sildid ja tasemed

3.1. Sõna käitumisprofili koostamine

Seisma verbi käitumisprofiil on koostatud eesti keele tasakaalus korpuse (<http://www.cl.ut.ee/korpused/grammatikaliides/>) lausete põhjal. Korpuse maht on 15 miljonit sõna ja seal leidub teadus-, aja- ja ilukirjandustekste, iga žanri viis miljonit sõna.

Kuna korpusest saab keeleveebi portaali (<http://www.keeleeveeb.ee/>) kaudu otsida sõnakasutust lemma järgi, siis on võimalik leida kõik verbi pöördvormid korraga. Korpusepäring andis vastuseks 7000 lauset. Töö aluseks on 500-lauseline juhuslik valim.

Käitumisprofiil on koostatud andmetöötlusprogrammis Microsoft Excel. ID-siltide märgendamine oli täiel määral manuaalne ja ID-siltide valimisel on lähtutud sellest, et neid ei tuleks ülemäära palju ning iga silt annaks analüüsietappi midagi juurde. ID-silte on kokku 17, lõppanalüüsil (klasteranalüüsil) kasutati neist 16 (välja jäi silt *allikas*).

Tabelis 1 on esitatud, mis variaablid märgendati (ID-silt) ja mis tasemed märgendamisel esinesid (ID-sildi tase). Kokku esines 117 erinevat taset, kõige rohkem tasemeid esines *käände* ja *kaassõna* ID-sildiga, kuna eri taseme alla on pandud kõik kaassõnad, mis esinesid (vt ptk 3.6.). Kuna mõnede ID-siltidega esines palju tasemeid (nt *agenditüüp*, *taustatüüp*), siis neist on kirjutatud eraldi alapeatükkides (ptk 3.5. ja ptk 3.6.)

ID-silte tekkis juurde vastavalt sellele, mis märgendamise käigus lauseid uuesti üle lugedes silma jäi ja oluline tundus. Algselt märgendasin üldtähenduse (konkreetne vs. abstraktne), seejärel täpsema tähenduse, mille hiljem parafraseerisin.

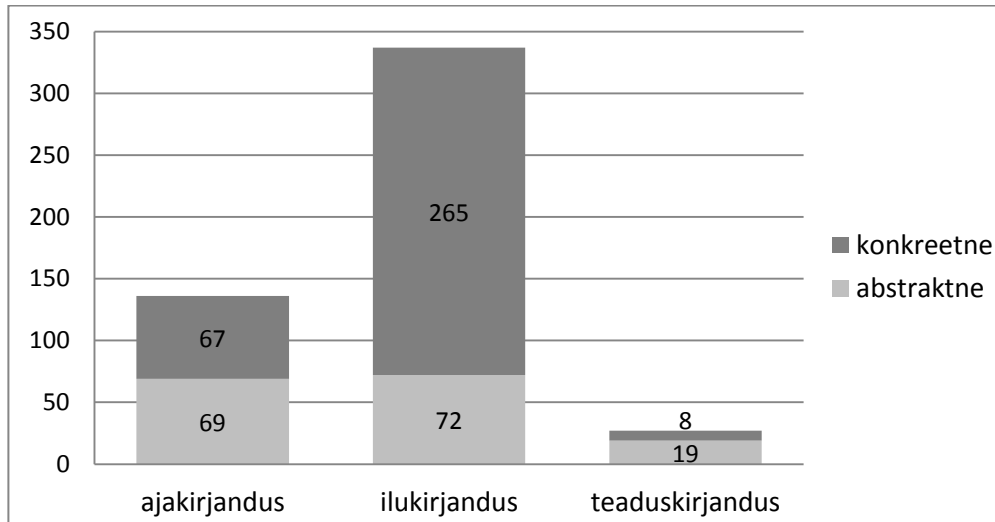
Tabel 1. *Seisma* profiilis esinenud ID-sildid ja nende tasemed.

ID-silt	ID-sildi tase
Tähendus	
Allikas	Ilukirjandus, ajakirjandus, teaduskirjandus
Üldtähendus	Konkreetne, abstraktne
Parafras	(vt ptk 3.3.)
Agent	
Agenditüüp	(vt ptk 3.5)
Agendi elusus	Elus, eluta, ei saa määrata
Agendi vertikaalsus	Jah, ei, ei saa määrata
Taust	
Taustatüüp	(Vt ptk 3.6)
Kohamäärus	Jah, ei
Määrusetüüp	Kaassõnafraas, nimisõnafraas, asemäärsõna
Kääne, kaassõna	Adessiiv, inessiiv, allatiiv, illatiiv ja erinevad kaassõnad
Ajamäärus	Jah, ei
Viisimäärus	Jah, ei
Morfoloogia	
Aeg	Olevik, minevik, tulevik ⁴
Pööre	Sg1, sg2, sg3, pl1, pl2, pl3, ma-inf, da-inf, -nud/tud, -tes, -mas, -vat
Kõneviis	Kindel, tingiv, käskiv, kaudne
Tegumood	Isikuline, umbisikuline
Kõneliik	Jaatav, eitav

3.2. Üldtähendus ja allikas

Joonisel 4 on esitatud, kuidas jagunesid laused ajakirjandus-, ilukirjandus- ja teadustekstide vahel. 500 lausest 136 olid ajalehtedest (27%), 337 ilukirjandusest (67%) ja 27 teaduskirjandusest (5%). Need rühmad on omakorda tähenduse järgi kaheks jaotatud – kas *seisma* verbi oli kasutatud konkreetsetes (kokku 360 korda) või abstraktsetes (140) tähenduses.

⁴ Tuleviku alla on märgendatud ühendverb *ees seisma* (mingi sündmus, tegevus toimub tulevikus)



Joonis 4. *Seisma* verbi kasutuse jagunemine teadus-, aja- ja ilukirjandustekstide vahel.

Kui ajakirjanduses on abstraktset ja konkreetset tähendust enam-vähem võrdselt (69 ja 67 korda), siis ilukirjanduses on abstraktse tähenduse osakaal vaid 21%, teaduskirjanduses on aga abstraktset tähendust 70%.

Konkreetsed ja abstraktsed tähendused jagasin omakorda 18 parafrasiks. Parafraaseerimata jäid viis verbiühendit: ühendverbid *ees seisma*, *eest seisma*, *koos seisma*, *vastu seisma* ja väljendverb *hea seisma*.

3.3. *Seisma* verbi parafrasisid ja tähendusvõrgustik

Peale süntaktiliste siltide tuli märgendada ka see, mis tähenduses verb lauses esineb. Polüseemse sõna tähenduse määramine on küllalt subjektiivne ja kohati keeruline, sest tähendused on omavahel seotud. Parafraaseerides on abiks olnud „Eesti keele seletav sõnaraamat“, milles *seisma* verbile on esitatud seitse eri tähendust koos hulga näidetega. Peale selle oli tähenduse määramisel abiks ka Gibbsi jt (1994) uurimus seismisega seonduvatel kujundskeemidel (TASAKAAL, VERTIKAALSUS, KESE-PERIFEERIA jne), mis viitavad eelkõige inimese seismisele ning teised peatükis 1.4. mainitud uurimused.

Algselt jagasin laused tähenduse järgi kuueks:

1. asuma/olema,
2. püstiasendis,
3. ühe koha peal,
4. kasutuseta olekus,
5. püsima,
6. abstraktne tähendus.

Seejärel analüüsisin iga tähendust täpsemalt ning koostasin parafraasid. Kuna peale nende kuue suurema tähendusrühma esines veel hulgaliselt lisatähendusi (olenevalt nt kas lauses esineb aja-, koha- või viisimäärus, kas asend on oluline jne), siis mõni parafraseering tuli pikk ja üsna täpne, näiteks EESMÄRGIGA OLEMA ÜHE KOHA PEAL või ÜHE KOHA PEAL PÜSTI ASUMA⁵.

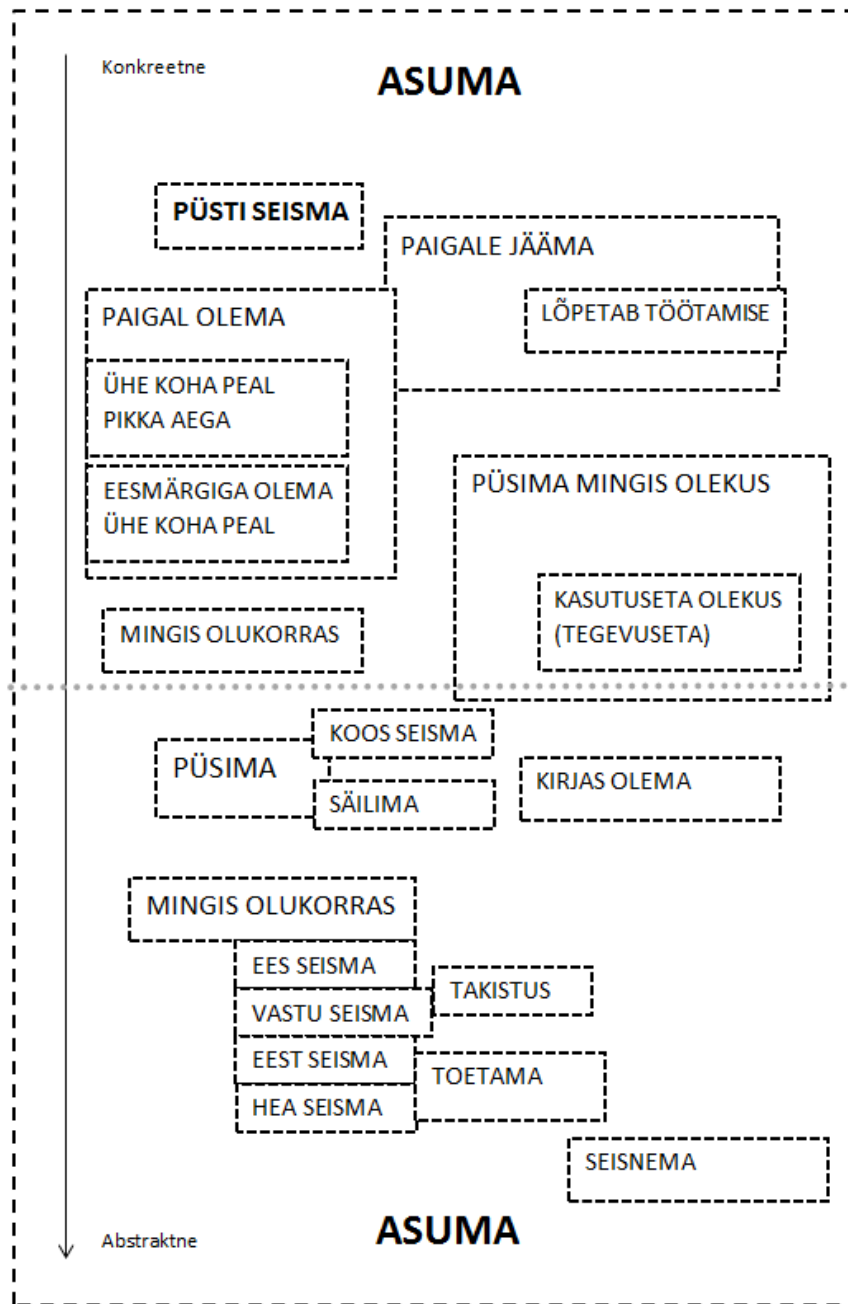
Nende parafraaside juures mängisid rolli mitu parameetrit. Esimese puhul on oluline eesmärgipärane kusagil asumine või olemine (nt ootamine) ja ühe koha peal asumine (pikemat aega), teise puhul on samuti tähtis ühel kohal asumine, kuid peale selle ka püstine asend, milles ühe koha peal oldi.

Järgnevalt on esitatud ülevaate parafraasidest, mille vahel laused on jagatud. Iga parafraasi kohta on toodud ka vähemalt üks näide korpuse lausetest ja mitu korda seda tähendust valimis oli kasutatud.⁶

Joonisel 5 on esitatud skeem, mis hõlmab korpusest leitud *seisma* tähendusi. Tähendused on ülalt alla liikudes järjest abstraktsemad. Kuna piirid eri tähenduste vahel võivad olla hägused, siis on tähenduste eraldamiseks kasutatud katkendjoont. Ka on mõni tähendus teisega tugevamalt seotud (kastid on ühenduses) ning ületab piire või on alltähendus.

⁵ Siin ja edaspidi on kõik parafraasid toodud suurtähtedes, et neid muust tekstist paremini esile tõsta.

⁶ Lisas 1 on ülevaatlik tabel kõigist parafraasidest ja verbiühenditest, mis valimis esinesid, nende esinemise sagedus ja näitelause.



Joonis 5. Seisma verbi tähendusväli ja parafrasid ülalt alla liikudes konkreetsest abstraktsemaks.

Keskelt jagab skeemi pooleks punktiirjoon, millest allapoole jäävad metafoorsed tähendused. Teatavasti väljendub metafoorne kasutus, kui ületatakse valdkonna piire, siinjuhul on kõige tugevam eristus liikumine füüsilisest ruumist mittefüüsilisse, abstraktsesse. Samal ajal on ka mõni piiripealne juhtum, nt PÜSIMA MINGIS OLEKUS.

Olen paigutanud parafraasi ASUMA kõige üldisemaks ja teisi hõlmavaks tähenduseks. See on n-õ taust, millelt teised tähendused esile kerkivad, samas ei tähenda see, et tegemist on tingimata *seisma* prototüüpse või keskse tähendusega.

Nagu eelpool mainitud, siis asendiverbid on paljudes maailma keeltes üldist entiteedi asukohta väljendavad verbid. Oluline on lokatiivsuse väljendamine ja asend jääb tagaplaanile või ei mängi üldse rolli. Tegemist on ka üldisema olemasolu ja eksistentsi väljendamisega, sel juhul võiks asendiverbi hõlpsasti asendada koopulaga, ilma et tähendus muutuks, vt näiteid 1 ja 2.

(1) *Kui su isal oli vana Zenit, siis küllap seisis tal kapi otsas ka fotosuurendi Leningrad?* [AJA]⁷

(2) *Kõrgemate kohtadel seisid mõned kuhjad.* [ILU]

Samal ajal kannab *seismine* lisaks asumisele lisatähendust, miks muidu neis lausetes seda koopulale eelistatakse. Kuna elutute objektide seismise puhul pole jalgadel asetsemine üldiselt oluline v.a. nt mööbliesemetel (voodi, laud), siis võib oletada, et rolli mängib pigem vertikaalsus.

Asumist pidasin korpuselausetes keskseks 36 korral ja neist 26-l oli tegemist vertikaalse agendiga. Enamasti oli tegemist objektidega (13 korda), millel vertikaalne telg pikem (raamat, uks), kuid esines ka neid, millel enamasti mitte (nt kiiver, diivan). Ka pidasin mõnikord inimese seismise juures esmaseks just kusagil asumist mitte asendit (näide 3).

(3) *Kõrbes seistes on ehk kõige põr pivam vaadata, kuidas kuiv jõesäng kaob liivas.*
[ILU]

Peale vertikaalsuse võib oluline olla ka ruumi-aja valdkond ehk mingis asendis teatud aja vältel püsimine (vt ptk 1.2.). Seega lisandub tähendusele füüsilise ruumi ja aja mõõde.

⁷ et siin ja edaspidi viitab AJA, ILU, TEAD vastavalt tasakaalus korpuse ajakirjandus-, ilukirjandus- ja teadustekstide allkorpustele, kust näitelause pärineb.

Kui vaadata METAFLOORSET ASUMIST, siis vertikaalsus on kadunud, ka ei oma tähtsust füüsiline ruum, kuid ajaline mõõde on säilinud. Näidetes 4 ja 5 on toodud korpuse laused, kus lokatiivsus on alles, kuid liigutud on mittefüüsilisse ruumi.

(4) *Soohundi vähemalt nelikümmend välilahingut kahekümnne aasta jooksul **seisid** siin koos temaga ega lasknud tal pead kaotada.* [ILU]

(5) *Maailm **seisab** põllumajandusliku revolutsiooni lävel.* [AJA]

Seismine väljendab ka siin tugevamat positsiooni, kui lihtsalt asetsemine või olemine. Näites 5 on oluline ka stardipositsioon, maailm on isikustatud ja inimese omadused sellele üle kantud, seega on maailm võimeline liikuma ja seisma.

Ülejäänud tähendused asetuvad nende kahe *asumise* alla (või sisse) ja moodustavad omaette rühmad, kus oluline on parajasti lokatiivsus, liikumatus, asend ning samal ajal kas ajaline või ruumiline mõõde. Seejuures tähendusvõrgustik järjest laieneb – kui keskse tähenduse juures on olulised kõik komponendid, siis järjest abstraktsemaks minnes kaotavad need tähtsust ja tähendus hakkab rohkem iseseisvuma, nii et ei pruugigi aru saada, kust mingi tähendus ja kasutus keelde on tulnud.

Keskne tähendus peaks seejuures olema ÜHE KOHA PEAL PÜSTI ASUMINE ehk inimese seismine, kuna asendiverb (*posture verb*) viitab esmajoones inimese asendile. Seejuures on määravad kõik eelpool mainitud omadused: lokatiivsus, liikumatus, vertikaalsus ja ruumi-aja valdkond. Korpusedlausetes esines püsti seismist ka kõige rohkem, 191 korda 500st. ID-siltide tasemete arvu järgi on see samuti keskne tähendus, mis esineb 83 ID-sildi tasemega 117st. Näites 6 on toodud korpuse lause, kus verbi on kasutatud just esimeses tähenduses.

(6) *Ta **seisis**, käed ristis, hing tühi, vaatas maad enda jalgade ees.* [ILU]

(7) *Minu mäletamist mööda tollel laual alkoholipudeleid **ei seisnud**.* [ILU]

Näites 7 on agendiks aga hoopis pudel, mitte inimene. Olen ka selle paigutanud seismise esimese tähenduse alla, sest täidetud on kõik samad kriteeriumid, mis inimese seismisel. Lemmensi (2001) järgi tajume vertikaalsust oma keha põhjal ja kui vaatleme

objekti alt üles ning näeme sarnasust, kanname samad omadused üle ka elututele objektidele. Seega võib pudel samamoodi seista kui inimene.

Võib oletada, et teised *seisma* verbi tähendused on motiveeritud just sellest prototüübsest seismisest ja seega võiks igal neist olla ka mõni iseloomulik omadus, mis seob seda keskse tähendusega.

PAIGAL OLEMA parafras esines valimis 25 korda ja hõlmas eelkõige agente, mille loomulik olek on liikumine, seismine on pigem hetkeline ja mööduv (nt sõidukid, õhk, vt näide 8). Parafras on seega tihedalt seotud liikumise vastandamise ja lokatiivsusega. Sellel on küllalt iseseisev tähendus, mis pole otseselt seostatud ei asendi ega vertikaalsusega, vaid just kitsalt n-õ peatumisega. Kõige enam esineski agendina sõiduk, seda 17 korda.

(8) *Maja ees seisis mingi tattide auto.* [ILU]

Sõidukite seismiselt võib aga tähendust laiendada ka üldisemalt entiteetide paigalolekule, mis normaalolekus on liikuvad (näited 9, 10 – *liiklus* ja *õhk*):

(9) *Endla ja Sõpruse puiestee ristmikul liiklus seisis, polnud aimugi, milles asi.*
[ILU]

(10) *Õhk seisab.* [ILU]

PAIGAL OLEMISE alla olen paigutanud parafraseid EESMÄRGIGA OLEMA ÜHE KOHA PEAL (18 korda korpuses) ja ÜHE KOHA PEAL PIKKA AEGA (19 korda). Neid kahte ühendab paigalolek, kuid rolli mängivad eri valdkonnad. Esimese tähenduse juures (näited 11, 12) pole niivõrd oluline liikumisele vastandumine, vaid püstises asendis olek, jõu rakendamine seismiseks, paigal püsimine.

(11) *Said jupikese vorsti, kui järjekorras seisid.* [AJA]

(12) *Pool päeva kulus kuskil polikliinikus "elavas järjekorras" oma õiguse eest õigel ajal kabinetti pääseda, seistes.* [AJA]

Enamasti oli tegemist järjekorras, sabas seismise ehk millegi ootamisega, seega on agendiks inimesed, enam on seost ka püstise asendiga, vertikaalsusega. Samuti esines kusagil rituaalsetel üritustel olemist, nt *matustel kirstu ees seismine*; tähenduse alla kuulub ka *valves seisma*. Võib olla, et kuna seismine on seotud füüsilise pingutusega, siis rõhutatakse selle tähenduse juures asendit eriliselt (näide 12). Seega on oluline jõudünaamika valdkond – jalgadel asendis püsimine.

Näidete 13 ja 14 juures ei ole aga tegemist hetkelise, mööduva paigalolekuga, vaid vastupidi rõhutatakse pikaajalisust. Selle parafraasi juures on oluline ajaline mõõde, mõnikord ka see, et agent asub tuttavas, harjumuspäras kohas või on tegemist agentidega (ehitised, kujud), mis on pikka aega liikumatult, st pole ise võimelised liikuma.

(13) *Riiulil seisis veel praegugi vana pruuniks roostetanud plekklehter, mille läbi vanaema tsaariajal vorstisoolikatesse verist tanguputru ajas. [ILU]*

(14) *Sulade ajal külmunud mullale seisma jäänud vesi kahjustas talirüpsi taimi. [TEAD]*

Paigalolekuga on seotud ka parafraas ÜHE KOHA PEALE JÄÄMA, mis esines korpuses 40 korda (näited 15,16), püsti seismise järel kõige rohkem. Siinkohal on tegemist muutuse väljendamisega. Verb *jääma* annab seejuures edasi liikumise seiskumist, protsessi lõppemist või tegevuse katkestamist, mille tulemus on püsiv seisund või paigalseis. (vt Soodla 2003: 58). Kui nt konstruktsioonid *jäi (edasi) lebama* ja *jäi (edasi) istuma* väljendavad duratiivsust, siis seisma verbi juures on võimalik vastandus liikumisega ja sel juhul oluline momentaansus, kuigi see ei tähenda keha täielikku staatilisust (näide 15).

(15) *Siis jäi ta seisma ning avas peo. [ILU]*

(16) *Jäin korraks seisma. [ILU]*

Samuti kui seismine on kinnistunud vertikaalse asendiga, võib see olla seotud ka just liikumatuse ja liikumise lõpetamisega, seda näitab ka järgmine parafraas, mille olen paigutanud SEISMA JÄÄMA alla.

LÕPETAB TÖÖTAMISE (esines 6 korda) on kitsam ja abstraktsem tähendus eelnevast. Siin on oluline just protsessi lõppemine ja tulemus on püsiv seisund.

(17) *Vedru käib maha ja masinavärk jääbki seisma!* [ILU]

(18) *Mul pidi syda seisma jääma – ta oli muutunud vanemaks!* [ILU]

Parafraasi juures kaob füüsilise ruumi mõiste ja olulisus, seega ületatakse valdkonna piire ja liigutakse abstraktsema poole peale. Mingi ese, objekt lõpetab töötamise, ei käi enam. Tegemist on agentidega, mille loomulik olek on pidevalt töötada, paigalolek viitab ebanormaalsele seisundile. Kuid erinevalt nt sõidukite liikumisest ja peatumisest viidatakse neis lausetes rohkem lõplikkusele mitte hetkelisele peatumisele.

Parafraas MINGIS OLUKORRAS (4 korda korpuses) on üks näide piiripealsest juhtumist, mille alla kuulub ka mingis vahekorras olek (näide 20). Näites 19 on agent reaalses olukorras ning seisab silmitsi konkreetse loomaga.

(19) [---] *seisis ta oma suureks ehmatuseks silm silma vastu suure oinaga.* [ILU]

(20) *Tema nägemuses on ametnikkond rahva hulgast pärinedes ja riigi teenistuses seistes hädavajalik ühenduslüli nende kahe vahel.* [TEAD]

Püsiühend *silm silma vastu* või *silmitsi seisma* on aga enamasti kasutusel abstraktse sõltuvusmäärusega (vt näide 21, 22). MINGIS OLUKORRAS (metafooriliselt) esines ka korpuse lausetes rohkem, 11 korda. *Silmitsi seisma* millegagi või mingi probleemi, dilemma *ees seisma* – st asuma olukorras, millele tuleks lahendus leida.

(21) *Sellest tulenevalt seisab praegune Eesti silmitsi sisemiste ohtudega.* [AJA]

(22) *Nüüd aga seisis Jaak oma naisega olukorra ees, mida keegi ei osanud lahendada, nii et hunt oleks söönud ja lambad ka alles.* [ILU]

Samamoodi on piirid hägused parafrasi PÜSIMA MINGIS OLEKUS (17 korda) juures. Asend ei ole enam nii konkreetset püstine, vaid on muutunud üldisemaks olekuks, seisundiks. Kui näites 23 on tegemist konkreetseesse valdkonda kuuluva lausega ja rõhutatakse asendit, siis näites 24 väljendatakse verbiga juba abstraktsemat püsimist.

(23) *Nende riided seisid higisoolast püsti [---].*[ILU]

(24) *Maja oli talve läbi kütmata seisnud.* [ILU]

KASUTUSETA OLEKU (18 korda) parafrasiga rõhutatakse seismist kui tegevusetat olemist, siin on seismine (pika)ajalisusele rõhutav verb. Sageli oli tegemist ruumide, ehitisega, millel lastakse seista tühjalt, lagunevana (näide 25). Inimeste puhul hõlmab kasutuseta olek ka tegevusetat seisundi (näide 26).

(25) *Ruumid on seisnud pikka aega pooltühjana.* [AJA]

(26) *Mes sa seisad siin jalus nigu üks pukk, mini parem vaata, mes teesed tüdrukud teevad.* [ILU]

Metafoorne PÜSIMA esines korpusel 15 korda. Siinjuures on liigutatud füüsilisest ruumist mittefüüsilisse ja joonisel 5 on järgneval arutluse all olevad parafrasiid allpool halli punktiirjoont.

(27) *Vaikus seisis meie vahel nagu tohutu, etteheitva pilguga eelajalooline loom.*
[ILU]

(28) *Sellised asjad ei seisa meeles lihtsalt, kui nii palju tööd on kogu aeg.* [ILU]

Tegemist on mentaalses ruumis (meeles, mõttes) püsimisega või on agent ehk tegija abstraktne (vaikus, uhkus). Enam pole oluline niivõrd olek kui (jätkuvalt) eksisteerimine. Esile tõuseb aja valdkond. Eriti määrav on see AJALISELT SÄILIMA⁸ parafrasi juures (näide 29), siin on metafoorne vastandus liikumisega (püsib–ei püsi, säilib–ei säili, läheb halvaks). Täendus kitseneb ja iseseisvub seega veelgi (vt Veismann 2012).

(29) *Ei need kaua seisa, aga või neid kauaks vaja ongi.* [ILU]

⁸ Viiesajast lausest esines vaid ühe korra *seisma* verb säilimise tähenduses. Kuna korpuselaused on ilma laiema kontekstita, siis tuli lause eraldi välja otsida ja eelnevatest või järgnevatest lausetest vaadata, millele või kellele pronoomen viitab. Selle lause puhul viitas demonstratiivpronoomen *need* jorjenipeadele, seega lilledele.

KIRJAS SEISMA (23 korda korpuses) on üsna kinnistunud konstruktsioon, milles agendiks on sisu – tekstid, artiklid, tsitaadid, reklaamid (näide 30). Nagu eelpool mainitud (ptk 1.3.), siis *seisma* verbi kastus võib selle tähenduse puhul olla motiveeritud sellest, et tähed seisavad (mõttelistel) joontel püsti. Rolli võib mängida ka ajaline mõõde, kui midagi on kirja pandud, siis püsib see kauem.

(30) *Seal **seisab** esimesel leheküljel artikkel, kus selges ja puhtas eesti keeles öeldakse, et ilmastikuolud heinatööd ei takista.* [ILU]

SEISNEMA (2 korda korpuses) puhul on samuti oluline sisu, kuigi see on minu arvates kõige eraldiseisvam ja läbipaistmatum kasutus, seega on huvitav vaadata, kuhu see käitumisprofiili järgi dendrogrammil paigutub ja mis tähendusega kõige enam sarnaneb.

(31) *Draama ja teatri erinevus **seisab** selles, et näidendis on tegemist kirjaliku, teatrilaval aga suulise kõnega.* [TEAD]

(32) *Probleem **seisab** ju selles, et kui oled kaotanud, siis tahad tagasi võita ega suuda mängulauast lahkuda.* [ILU]

TOETAMA (9 korda korpuses) parafraasi juures kasutatakse *seisma* verbi, et kellegi/millegi poolt või taga olla ja toetada (näide 33). Ühendverbi *taga seisma* otsustasin parafraseerida, kuna see sobis kokku kasutustega *kellegi poolt olema, kellegi/millegi kõrval seisma* (nt heas ja halvas) (näide 34).

(33) *Õpilaste kõrval **seisab** ta alates stuudiost kuni ajani, mil hakkab neile esinemisvõimalusi otsima.* [AJA]

(34) *See on natsionalism kitsamas mõttes, selle taga **seisavad** inimesed – rahvuslased.* [TEAD]

Ka ei saanud ühendverbi *taga seisma* eraldi jätta, sest tähenduse järgi jagunes see üldiselt kolmeks: keegi seisab kellegi (selja) taga, st toetab (näide 35). Millegi taga seisab keegi, st on selle algataja (näide 36). Mingi asi seisab millegi/kellegi taga, st on pidama,

kinni jäänud (näide 37). Näite 37 juures on tegemist aga juba TAKISTUSEGA. Takistuse tähendust esines korpuses kolm korda ja see on loetud eraldi parafrasiks (ka näide 38).

(35) [---] väike rühmitus, kelle *selja taga seisab* paremal juhul vaid väike osa töölistest. [AJA]

(36) Kurja kodanikualgatuse taga *seisab* kohalik elanik Tatjana Šugai [AJA]

(37) Pakkumine *seisab* andmebaaside arengu taga. [AJA]

(38) Maja hävitada, et ta *ei seisaks* enam elul ja arusaamisel põigiti ees, kaotaks oma ebaproportsionaalse tähtsuse. [ILU]

Vahekokkuvõtteks võib öelda, et korpuses esinesid kõik tähendused, mis EKSSis on mainitud, seega võib loota, et valim oli esinduslik. Kuigi abstraktseid tähendusi (parafrase) on arvult rohkem, leidub nende näiteid korpuses siiski vähem, nt AJALISELT SÄILIMA või TAKISTUS. Kuna aga korpus võiks peegeldada tegelikku keelekasutust, siis sellest tulebki välja, et kuigi abstraktsemad kasutused on tarvitusel, siis sagemini kasutatakse asendiverbi ikkagi füüsilises ruumis. Samas peab tõdema, et kasutusprotsent erineb ka tekstiliigiti.

3.4. Verbiühendid

Viis verbiühendit sõnaga *seisma* on jäetud parafraseerimata, need on ühendverbid *ees seisma*, *vastu seisma*, *koos seisma* ja *eest seisma* ning väljendverb *hea seisma*.

Kuigi verbiühendeid esines lausetes rohkem, nagu nt *silmitsi seisma* ja *taga seisma* siis ülejäänud otsustasin parafraseerida, sest need sobisid kokku mõne teise kasutusega.

Need viis verbiühendit on töösse jäetud, et näha, kuidas nad kasutuse järgi klasterdades parafraside vahele asetuvad ja millise parafrasiga (või teineteisega) tugevamalt seotud on. Samuti jäid just need ühendid parafraseerimata, kuna neid esines korpuses kõige rohkem, seega peaksid need olema keeles tavalisemad.

Kõige rohkem esines korpuses väljendverb *ees seisma*, seda 15 korda.

(39) *Ta sünnipäev, samuti ka lapse kolmekuu tähtpäev **seisid** alles ees, nii et see pidi tõesti lihtsalt kohvijoomine olema.* [ILU]

Selle ühendverbi juures on oluline (lähi)tulevikku suunatus, agendiks on enamikel juhtudel mingi sündmus. Lakoffi ja Johnsoni (1980) järgi on tulevik seotud metafooriga LÄHITULEVIKU SÜNDMUSED ON ÜLAL (JA EES). Eesti keeles ei saa tulevikku aga seostada ülalsuunaga, vaid pigem horisontaalse EES suunaga. Harilikult ühtib meie silmade vaatesuund tavapärase liikumissuunaga, mis on ettepoole, seega tulevik on ees ja lähisündmused on näha. (Lakoff, Johnson 2011: 48)

Kui lähtuda metafoorist ELU ON TEEKOND, siis seisavad sündmused teel ja meie liigume neile lähemale. Tegemist on fiktiivse asetsemisega, mingi pidepunktiga, millest mööduda. Aeg liigub ja meie liigume, aga sündmus on sama koha peal.

Vastu seisma esines korpuses seitse korda. Oluline on vastuseis, mida osutatakse ja jõud, mida kasutatakse.

(40) [---] *Anpilovi töölissalklased üritasid Kremlile jõuga **vastu seista**.* [AJA]

Kuigi ühendverb *koos seisma* oli korpuse 500 lauses vaid kaks korda, on see siiski eraldi välja toodud. Täenduselt on see lähedal parafrasile *püsima* ja seda saab kasutada kahes tähenduses: koosnema (näide 41) või ühes tükis, tervikuna *püsima* (näide 42).

(41) *Kuid, noormees, elu **ei seisa koos** ainult romantikast.* [ILU]

(42) *Etenduse väärtus **seisab koos** laval toimuva põhjendatusest, veenvusest ja kunstiküpsusest.* [AJA]

Hea seisma (5 korda; näide 43) ja *eest seisma* (7 korda; näide 44) väljendavad toetust, nagu abstraktses mõttes taga ja kõrval seisma.

(43) *Seaduste järgi peab arst **hea seisma** oma kohustuste korrektse täitmise ees.*
[TEAD]

(44) *Ainult neid, kes ise suudavad enda **eest seista**, suudab ka kentuura aidata.*
[ILU]

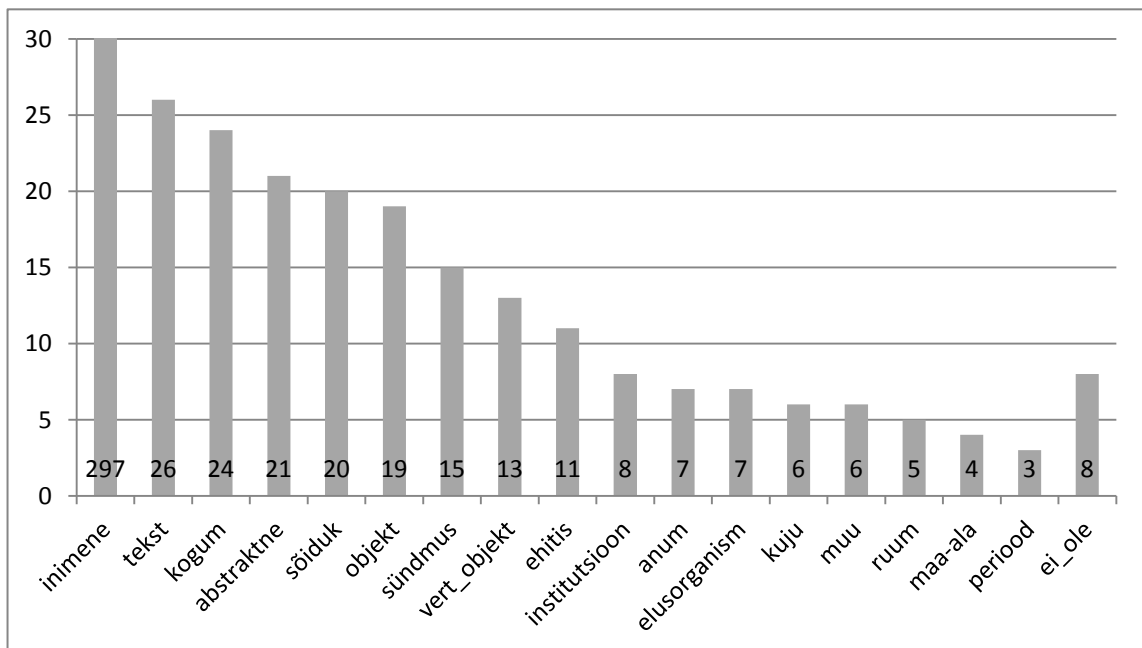
3.5. Agent ja sellega seotud ID-sildid

Ühe ID-sildina on märgendatud ka agendi ehk tegija tüüp. Seejuures pole tegemist semantilise rolli ehk teadlikult toimiva, oma tegevust kontrolliva osalisega, vaid mõeldud seisjat laiemas mõttes, olgu lauses tegemist elus või eluta osalisega.

ID-sildi märgendamise eesmärk oli vaadata, missugused entiteedid üldse seisavad ja kas neid on võimalik kuidagi rühmitada ning mis omadused peavad olema täidetud, et seista.

Kuna seismise juures on esmases tähenduses (inimese seismine) oluline vertikaalne asend, siis on märgendatud ka agendi vertikaalsus, samuti see, kas tegemist on elus või elutu agendiga (vt lisa 2).

Kokku esines **17 agentitüüpi**, ülekaalukalt kõige enam oli agendiks inimene (297 korral), seda nii pärisnime, pronoomeni, pöördelõpu kui ka erinevate elualade esindajate kujul. Kaheksas lauses agenti ei esinenud. Joonisel 6 on graafiliselt esitatud, kuidas agentitüübid lausete vahel jagunesid.



Joonis 6. Agentitüübid ja nende esinemise sagedus (N=500)

Inimese järel esines agendina kõige sagedamini *tekst*, seda 26 korda. Tekstiks märgendasin kõik laused, kus oli tegemist mingi kirjaliku väljendusega, mis kusagil asus – tsitaadid, teated, sildid ja kirjutised, mis kusagil seisid, aga ka nt sõnad (näide 45) või nimed (näide 46).

(45) [---], *nagu oleks esimene sõna seisnud eraldi järgnevatest* [TEAD]

(46) *Minu passis seisab nimi Neeme Truu.* [ILU]

24 korda esines lausetes agendina *kogum*. Kogumi alla on loetud agendid, kes algselt millegi muu alla ei sobinud, kuid üle jäädes moodustasid küllalt homogeense rühma. Sellesse rühma kuuluvad mitmusliku tähendusega sõnad nagu *kogudus, ametnikkond, perekond, jõuk, orkester, liiklus*. Samuti kogum millestki abstraktsemast – *maailm* (3 korda), *õhk, vesi* (2 korda), pronoomen *kõik*. Kuigi mõned agendid oleks võinud paigutada ka inimese alla (nt *perekond* või *jõuk*), siis otsustasin seda mitte teha ja *inimese* all ongi mõeldud konkreetselt ühte inimest.

Abstraktse (21 korda) agenditüübi moodustasid substantiivid nagu *aeg, elu, erinevus, otsus, probleem, uhkus, vaikus, väärtus, väljakutse*.

Sageduselt järgmisena esines agendina *sõiduk* 20 kasutuskorraga. Sõidukitest esinesid autod, rongid, laevad ja muud masinad.

Muud konkreetsed objektid ja vertikaalsed objektid otsustasin eraldi vaatluse alla võtta, et vaadata, kas elutute objektide juures on oluline vertikaalsus. Objekti all on mõeldud eelkõige väiksemaid konkreetseid esemeid, milledest ei saanud moodustada suuremaid rühmi, kuigi objekti alla võiks lugeda ka näiteks sõidukid, anumad või ehitised. Kuna viimaseid esines arvestatav hulk, siis otsustasin need eraldi välja tuua.

Mingi konkreetse objekti rühma kuulus 19 agenti, nt *maal, paber, paadimootor, kivi, riided* – need agendid mille kohta võib öelda asjad, kuid mida ei ühenda peale selle ükski eristuv omadus. Seevastu vertikaalse objekti puhul, mida esines 13 korda oli oluline, et vertikaalne telg oleks pikem kui horisontaalne, seega võiks olla selgem ka see, miks kasutatakse just *seisma* verbi. Sellesse rühma kuulusid nt *raamat, jõulupuud, jahirelv, uks, kuhi*.

15 korda oli agendiks mingi sündmus. Eelkõige kasutati seda ühendverbiga *ees seisma*. Esseisvad sündmused olid nt *arreteerimine, jutuajamine, salvestus, tähtpäev, võitlus* või *lahing*.

Ehitiste puhul (11 korda) oli enamasti tegemist kasutuseta olekuga ehk need seisid niisama, tühjalt, vahel ka asumine. Ehitise alla kuulusid nt *majad, kirik, kasarmu, bensiinjaam, võõrastemaja*.

Institutsiooni (8 korda) rühma tegin samuti eraldi, sinna alla läksid *riik, kirik, kolledž, liikmesriigid* jms.

Anum (7 korda) on küllalt konkreetne agent ja seetõttu võib olla kummaline, et see on eraldi välja toodud. Põhjus, miks otsustasin anumatele eraldi rühma tekitada ja mitte neid nt vertikaalse objekti või lihtsalt objekti alla panna on selles, et teiste keelte kohta tehtud uurimustes (nt hollandi keele kohta vt Lemmens 2001, 2012), kus asendiverbid on laialt kasutuses ka elutute entiteetidega, on kirjeldatud, et seisvad objektid on nt nõud, potid, kausid, mis toetuvad põhjale ja on pealt lahtised.

Seega on huvitav, et sama kehtib ka eesti keele puhul ja 500 lause hulgas oli *seisma* verbi kasutatud anumatega piisavalt palju, et see agent teistest eristuks ja selle saaks eraldi rühmas välja tuua. Sama peaks kehtima ka näiteks kingade, saabaste ja kastidega, sest neil on sarnased omadused, kuid selliseid agente valimis ei esinenud. Anumad, mis esinesid olid *ämber, plekklehter, puutünn, veeklaas, savikann*.

Elusorganismiks märgendasin linnud, loomad ja taimed. Kuna neid esines kokku vaid 7 korda, siis panin need ühte rühma, nt *kaktus, aaloed, kajakas, põder*. Elusorganisme esines valimi suurust arvestades väga vähe, eriti loomi, kelle seismist võib pidada sarnaseks inimese omale, eelkõige anatoomilises mõttes – seisavad nemadki ju jalgel.

Agendi *kuju* all on mõeldud isikute monumente ja ausambaid, mida esines kokku kuus korda. Kuna monumendid, ausambad on sageli inimese kujuga, siis see, et neile omistatakse seismist on üsna tavapärane, peale selle on kujude seismise juures aga oluline ka paigalolek, püsimine.

Esiialgu proovisin *muu* (6 korda) kategooriasse kuuluvad agendid teiste rühmade alla sobitada, kuid need olid siiski piisavad erinevad, et omaette rühm tekitada. Muu alla kuulusid nt *süda*, *dada* (uudissõna), *juhi koht*, *äripool*.

Ruum esines agendina viis korda, nt *tuba*, *võimla*, *jõusaal*, või ka lihtsalt *ruum*.

Maa-ala oli agendiks neli korda – *kalmistupind*, *uusmaa*, *krunt* ja *linn*. Kõige vähem, kolm korda, esines agendina periood – *pikk päev*, *täistööpäev* ja *periood*. Kuna need aga ühegi teise agenditüübi alla ei sobinud, siis jätsin need eraldi rühmaks.

Kaheksal korral lauses agenti ei esinenud. Nende lausete puhul oli verb *da*-infinitiivis, seda ühendverbiga *vastu seisma* (näide 47) või esines verb käändelises vormis määrusena (näide 48).

(47) *Et aga kiusatusele “asjast midagi ka ise arvata” on lootusetu vastu seista, siis edastame A&A lugejatele mõned “ kodukootud ” mõttekillud.* [AJA]

(48) *Liikudes ja seistes.* [AJA]

Kui vaadata agentide **elusust** ehk agentide jagunemist elus ja elutu vahel (kodeeritud kolm tasandit: *jah*, *ei*, *ei saa määrata*), siis arvestades inimeste hulka valimis, on elus agente rohkem, kokku 309 ja eluta 183. Kaheksas lauses, kus agent puudus ei saanud ka elusust määrata. Peale inimeste sai elus agentideks määrata veel elusorganismid ning kogumi rühma alt *perekonna*, *jõugu*, *ametnikkonna* ja *koguduse*. Muu alt sobis elus kategooriasse süda, see oli ka ainuke kehaosa, mis valimis esines.

Vertikaalsuse all (kodeeritud kolm tasandit: *jah*, *ei*, *ei saa määrata*), on lähtunud konkreetse agendi vertikaalsusest, see kas lause oli abstraktse või konkreetse tähendusega ei omanud sildi märgendamisel tähtsust. Vertikaalseid objekte oli kokku 335, neist 297 inimesed, peale nende esines vertikaalsust agendirühmades anum, ehitis, elusorganism, kogum, kuju, vertikaalne objekt. Mittevertikaalseid agente esines 156 korral.

Selle peatüki kokkuvõtteks võib võrdluseks tuua Schönefeldi (2006) töö, kus ta on uurinud kolme miljoni sõnaga BNC ajakirjanduskorpusest asendiverbide kasutust ja tõlkevasteid inglise, saksa ja vene keeles.

Täpsemalt uuris ta 8000 kollokatsiooni ja ka seda, kes on agent. Tulemuseks oli, et 59,6% on seisjaks inimene, järgnesid abstraktsed objektid 19,1%-ga. Schönefeld oli jaganud agendid kuueks, peale juba mainitute esinesid veel konkreetset objektid (8,8%), nt ehitised, sõidukid ja personifitseeritud ehk isikustatud objektid, nt *maailm*, *armee*. Kõige vähem esines elusorganisme, konkreetsemalt loomi (1,4%). Seega kui vaadata minu tulemusi, siis need on üsna sarnased, kuid agendid on jagatud kitsamatesse rühmadesse. Seda eelkõige seetõttu, et eesmärk polnud vaadata, kas objektid seisavad, vaid täpsemalt millised need objektid on. Kui mingit agenditüüpi esines üle viie korra (nt anum), siis on see juba eraldi märgendatud, sest see ei pruugi olla juhuslik ja mingid seismisele omased omadused peavad sel juhul olema täidetud.

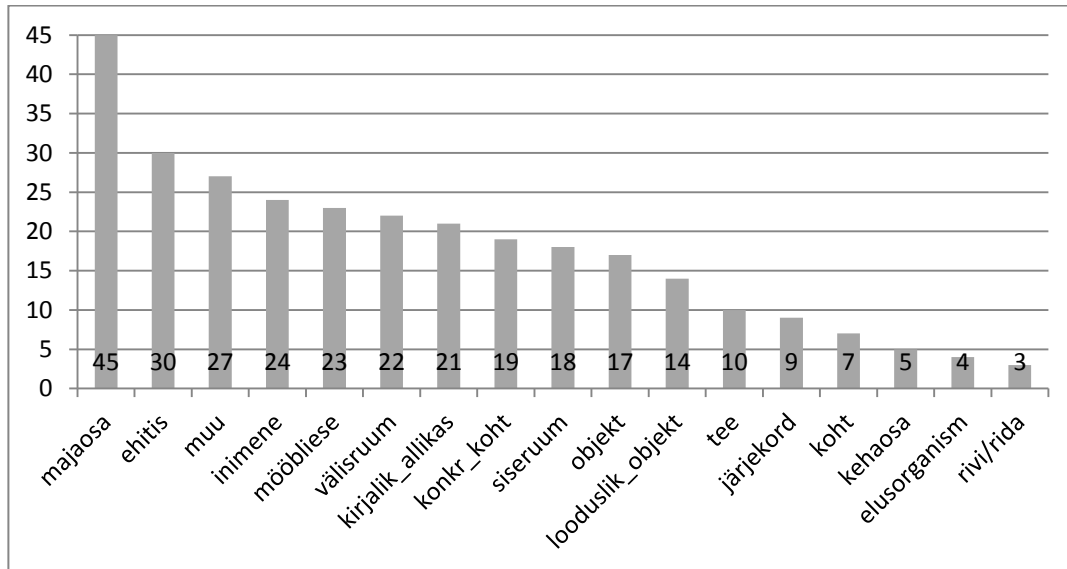
3.6. Taust⁹ ja sellega seotud ID-sildid

Selles alapeatükis tuleb juttu taustatüüpidest, mis esinesid ehk kus seisti. Samuti sellest, kas lauses esines kohamäärus ning mis kujul see oli – kaassõnafraas, nimisõnafraas, asemäärsõna. Täpsemalt märgendasin ka ära, mis käändes määrus esines või mis kaassõna kasutati. Ka on selles peatükis esitatud sagedusandmed viisi- ja ajamääruse esinemise kohta (vt ka lisa 3).

Kohamäärus esines valimi lausetes 297 korral 500st, seega u 60% lausetest oli välja toodud koht, kus seisti. Samuti nagu agentide puhul, jagasin ka taustad tüüpidesse. Kokku esines **17 taustatüüpi**, 203 korral kohamäärust lauses ei esinenud.

Joonisel 7 on esitatud, kuidas taustatüübid jagunesid. Kõige enam esines taustana majaosa, seda 45 korral. Uks esines taustana 14 korda ning aken 13. Peale nende seisti veel trepil (3), leti taga (3), seinä ääres (3), astmel, nurgas (2), terrassil, lävel (3), kuigi viimase võiks kokku lugeda ka uksega.

⁹ Töös on kasutusel Talmy käsitus ruumisemantikas kasutuses olevatest mõistetest taust (*ground*) ja kujund (*figure*). Taust on see, mille suhtes midagi liigub või asetseb (staatiline). (Talmy 2000: 311)

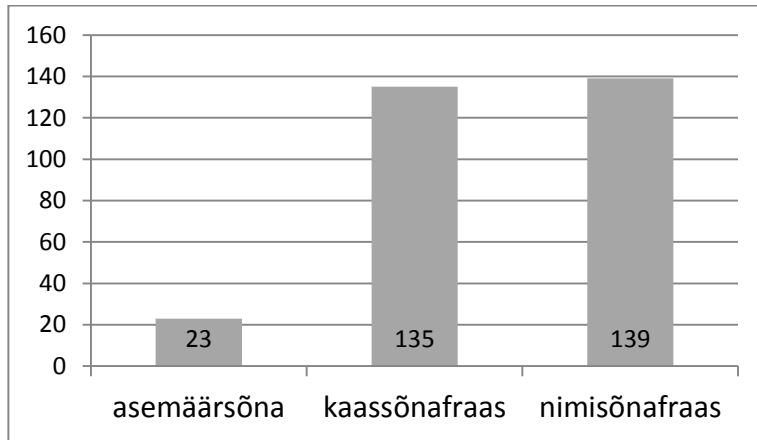


Joonis 7. Taustatüübid ja nende esinemise sagedus (N=297)

Ehitistest esines kõige sagedamini *maja*, seda seitse korda. Muu alla on loetud nt *kõrgemal, kerkokellale lähemal, teineteise kõrval* jne. Välisruumide all on nt *kõrb, aed, taevas*. Konkreetse koha all on silmas peetud asemäärsõnu nagu *seal, sealsamas, siin*.

Objekti all on mõeldud nt *maalil, pumba juures*. Loodusliku objekti korral on aga tegemist nt *kividel* või *vees* seismisega. Taustatüüp *koht* hõlmab üldisemaid paiku nagu nt *avariipaik* või *suitsetamiskoht*, elusorganism selliseid, kui taust on nt *puu all, põõsa varjus*. Kõige vähem esinenud *rivi/rida* hõlmab taustasid nagu *rivi, rida, tagarida*.

Nagu eelpool mainitud, siis 203 korral tausta ei esinenud, 297 korral oli lauses taust olemas ja nende lausete juures kasutati 135 korda kaassõna ja 139 korda käänet, 23 korda kasutati asemäärsõna. Joonisel 8 on toodud, kuidas kohamäärused korpuses jagunesid.



Joonis 8. Määruse tüübid korpuse lausetes (N=297)

Asemäärsõnadest esinesid *seal* (10), *sealsamas* (5), *siin* (3), *kusagil*, *selles*, *sinna* ja *sinnasamasse*.

Käänetest kasutati kõige rohkem adessiivi, seda 72 korda ja inessiivi 62 korda. Allatiiv esines vaid kolm ja illatiiv kaks korda.

Kõige rohkem kasutati kaassõnadest adpositsiooni *ees*, seda 31 korda. Peale selle *juures* (14), *keset* (15), *kõrval* (15) ja *taga* (10), *all* (9). Sihtkohamäärustest kasutati *ette*, *juurde*, *kõrvale*, *taha*, *äärde*. Kõik kaassõnad, mis korpuses leidsid on esitatud joonisel 9.

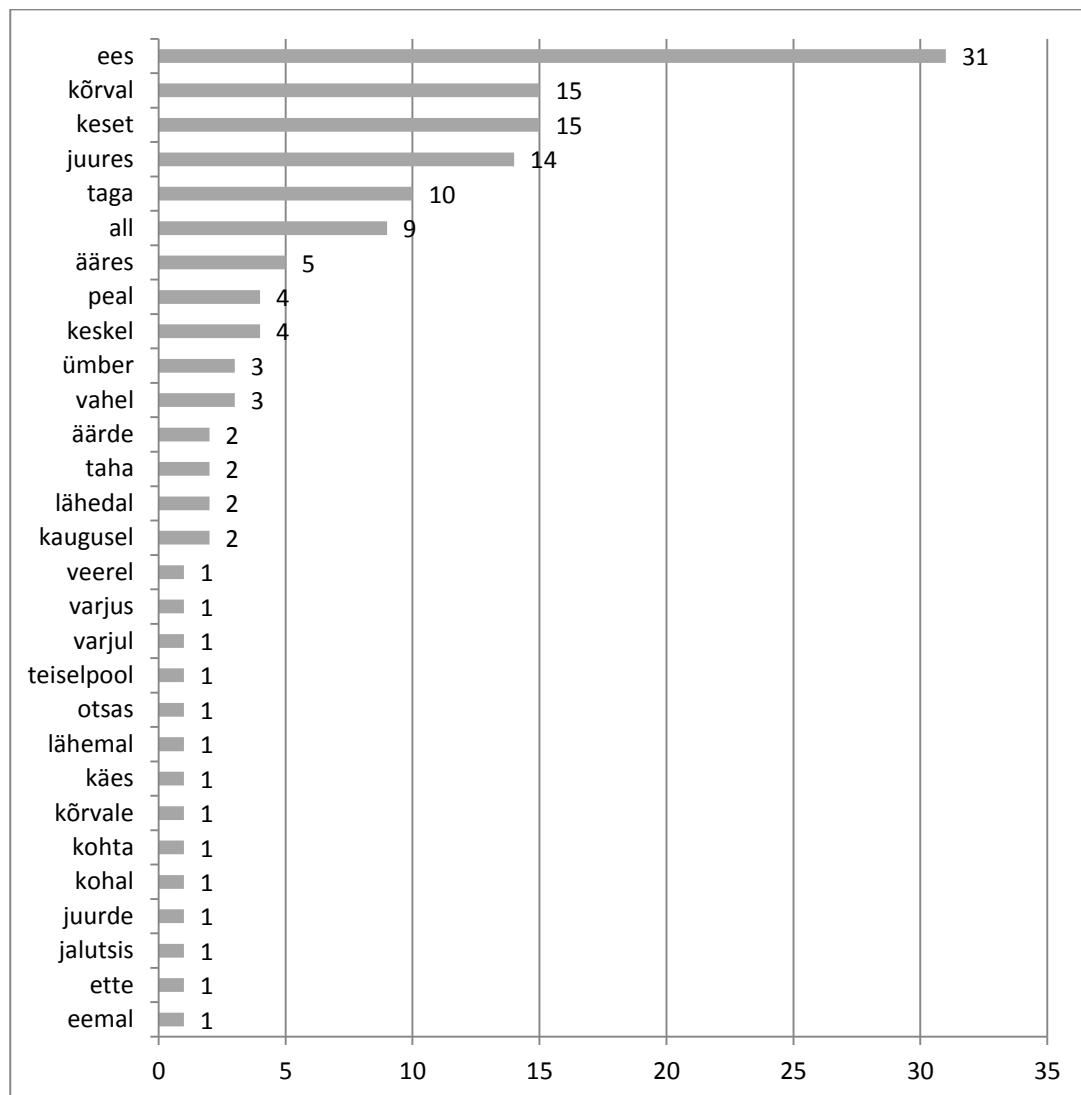
Ajamäärus esines 80 lauses ja **viisimäärus** 99 lauses. Mõlemad esinesid lõppanalüüsis binaarsete tunnustena (jah, ei), kuid töö esmaetapis märgendasin ka konkreetsed lekseemid.

Kõige sagedamini (6 korda) esines ajamäärusena *tükk aega*. Korduvalt kasutati ka *nüüd*, *hetke*, *endiselt*, *praegu*, *alati*, *korraaks*. Pikemaid perioode väljendati määrustega nagu *aastaid*, *hulk aega*, *kõik õhtud*, *lõputu aja*, *pikka aega*, *ikka* ja *ikka veel* (kõiki üks kord). Momentaansust väljendati seevastu adverbiaalidega *hetkeks*, *hetkel*, *järsku*, *mõne hetke*, *viivuvilkel*, *äkki*

Viisimäärustest esines kõige sagedamini *käed puusas*, *nõutult* ja *tühjalt*, igauht oli korpuses kasutatud neli korda. Rohkem kui korra esinesid veel *eraldi*, *halvatult*, *korralikult*, *kütmata*, *püsti*, *vaikides*, *süngelt*, *kasutuseta*. Kõige enam kasutati viisimäärust, kui agent oli inimene, 58 korda (nt *alasti*, *harkisjalu*, *jalul*, *kõikudes*, *käed rippu*, *käed*

ristis, raidkujuna). Veel kasutati seismise kirjeldamiseks võrdlust: *nagu liim, nagu soolasammas, nagu tulp, nagu üks pukk*. Viimastest on näha, et inimese seismist võrreldakse vertikaalse esemega.

Alati kasutati viisimäärust *ruumi* seismise juures, korduvalt ka *ehitise, ruumi* või *maa-ala* kirjeldamisel, seevastu mitte kordagi ei kasutatud seda *perioodi, sündmuse* ja *kuju* kirjeldamisel.



Joonis 9. Kaassõnad, mis korpuse lausetes esinesid ja nende esinemissagedus (N=135)

3.7. Morfoloogilised kategooriad

Morfoloogilistest kategooriatest märgendati *aeg, pööre, kõneviis, tegumood* ja *kõneliik*.

Ajakategooria on käitumisprofiilis jagatud kolmeks: minevik, olevik ja tulevik. Mineviku all on lihtminevik, täisminevik ja enneminevik. Mineviku otsustasin selliselt märgendada, sest osades *nud*-partitsiibiga lausetes ei olnud võimalik määrata, kas tegemist on enne- või täisminevikuga ning kuna pöördelõpud ja mineviku partitsiibid olid nagunii eraldi märgendatud, siis see ei tundunud ka oluline. Tuleviku all on märgendatud laused, kus oli tegemist eelseisva sündmusega, st kasutati ühendverbi *ees seisma*.

Laused jagunesid kolme aja vahel järgmiselt: minevikku esines kõige rohkem, 290 korda (58%), olevikku 196 korda (39%) ja tulevikku 14 korda (3%).

Sildi **pööre** all on märgendatud verbi pööre või käändeline vorm. Kõige enam esines singulari kolmandat pööret, kokku 244 korda, neist 138 minevikus. Singulari teist ja esimest isikut oli vastavalt 7 ja 29 korda. Pluurali kolmandat isikut esines 66 korda. Pluurali teist isikut 2 korda ja esimest 8 korda. Tabelis 2 on toodud, millised verbi lõppe kasutati ja kas sealjuures oli tegemist elus või elutu agendiga.

Tabel 2. Pöördelõpud ja agendi elusus

Verbi pööre	Elus	Eluta	Ei saa määrata	KOKKU
<i>da-inf</i>	19	2	5	26
<i>ma-inf</i>	36	18	1	55
<i>-mas</i>	4	0	0	4
<i>-nud</i>	12	22	0	34
pl1	8	0	0	8
pl2	2	0	0	2
pl3	35	31	0	66
sg1	28	1	0	29
sg2	7	0	0	7
sg3	137	107	0	244
<i>-tes</i>	20	1	1	22
<i>-tud</i>	1	0	0	1
<i>-vat</i>	0	1	0	1
Ei ole	0	0	1	1

4. Klasteranalüüs *seisma* verbi tähendustega

Kui peatükis 3.2. oli esitatud võimalik viis, kuidas *seisma* verbi tähendused võivad omavahel seotud olla ning toetunud korpusest saadud sagedusandmetele, siis selles peatükis on arvestatud nende „käitumisomadusi“ ning selle alusel tähendused rühmitatud. „Käitumisomaduste“ all pean eelkõige silmas seda, kui sarnased on parafraasid kasutus- põhiselt (kui paljud ID-sildi tasemed erinevate tähenduste puhul ühtivad ja mil määral).

Kasutusse on võetud märgendatud ID-sildid ja vaadatud, kuidas andmeanalüüsi abil suudab andmetöötlusprogramm tähendused rühmitada. Samas tuleb arvestada, et polüseemia uurimisel ja lausete märgendamisel peab uurija paratamatult toetuma enda intuitsioonile ja hinnangutele, kuid käitumisprofili meetodi juures on püütud seda minimaliseerida, seetõttu on arvesse võetud ka nt morfoloogilised ja süntaktilised sildid.

Kõigi ID-siltide märgendamise eesmärk oli saada nii palju erinevat infot *seisma* verbiga lausete kohta kui võimalik ja jagada koostatud parafraasid hierarhilise aglomeratiivse klasteranalüüsi abil rühmadesse. Järgimistes alapeatükkides tutvustan klasteranalüüsi läbiviimise kulgu ja tulemusi.

4.1. Andmetöötlus

Andmetöötluse jaoks tuli kõigepealt koostada koosinemistabel, kus oli kirjas kui mitu korda esines iga ID-sildi tase teatud parafrasiga. Kuna parafraase oli koos verbiühenditega kokku 23 ja ID-sildi tasemeid 117, siis oli tabelis väga palju 0-väärtusi. Nt parafrasiga *ajaliselt säilima* oligi võimalik iga ID-sildi taseme juures väärtus 1 või 0, kuna see esines valimis vaid ühe korra.

Tabelis 3 on toodud näide koosinemistabelist ja kolmest parafrasist: PÜSIMA (metafooriliselt), ASUMA ja KIRJAS OLEMA.

Tabel 3. (Osaline) koosinemistabel kahe ID-sildiga ja nende tasemete esinemine korpuses.

ID-silt	ID-sildi tase	Püsima (metaf)	Asuma	Kirjas olema	
Allikas	Ajakirjandus	3	8	10	} 10+13=23
	Ilukirjandus	12	27	13	
	Teaduskirjandus	0	1	0	
Üldtähendus	Abstraktne	15	0	23	} 23+0=23
	Konkreetne	0	36	0	

Esimeses tulbas on ID-silt, teises ID-sildi tase ja järgnevas on kirjas mitu korda mingi tase parafraasiga esines. Selliselt sai risttabeliga andmed iga parafraasi ja iga ID-sildi taseme kohta. Siinjuures peab vaatama, et iga sildi juures oleks parafraasi esindavate tasemete summa sama. Nt *kirjas olema* puhul peab see iga ID-sildiga olema 23.

Seejärel tuli koosinemistabelis iga parafraasi iga ID-sildi raames arvutada sagedusandmed – kui suure osa moodustas iga ID-sildi tase terve ID-sildi raames. Näiteks esines parafraas KIRJAS OLEMA 10 korda ajakirjandustekstis, 13 korda ilukirjanduses ja 0 korda teaduskirjanduses. Seega ühe ID-sildi (allikas) raames esines parafraasiga KIRJAS OLEMA 43% ($10/23=0,43$) ajakirjandustekstist pärit lauseid ja 57% ($13/23=0,57$) ilukirjandusest, teaduskirjandust oli 0%.

Selliselt sai võrdlusandmed, mil määral parafraasid üksteisest erinevad. Lähtudes sellest, et jagunemissageduse järgi saab vaadata ka funktsionaalset sarnasust, olid andmed sellisel kujul (kümnenmurdudena) ette valmistatud põhjalikumaks andmeanalüüsiks ja klasterdamiseks. Edasine analüüs toimus statistikaprogrammiga R.

Järgmise sammuna tuli otsustada, kuidas mõõta distantse parafraaside vahel. Nagu peatükis 2.3. mainitud, on selleks erinevaid viise, nt Eukleidese, Manhattani ja Canberra vahemaad. Selles klasteranalüüsis on kasutatud **Canberra distantsi**, kuna see sobib hästi,

kui andmehulgas on palju 0-väärtusi, sealjuures koheldakse neid analüüsis justkui puuduolevaid väärtusi.

Kui vahemaad mõõdetud, siis tuli parafraasid hierarhilise liitmismeetodi abil klasterdada. Seega kõigepealt ühendati omavahel kõige sarnasemad parafraasid ja seejärel liideti sammhaaval juurde sarnasuselt järgmised klastrid. Klastritevahelise distantsi mõõtmiseks kasutasin **Wardi meetodit**.

Tulemuseks on dendrogramm, millel on kõik parafraasid ja verbiühendid, mida märgendasin. Seega on sagedusandmetest moodustatud struktuuriline üksus, mis näitab, millised tähendused on sarnased või käituvad sarnaselt. Edasine analüüs ja klastrite tõlgendamine on aga uurija töö.

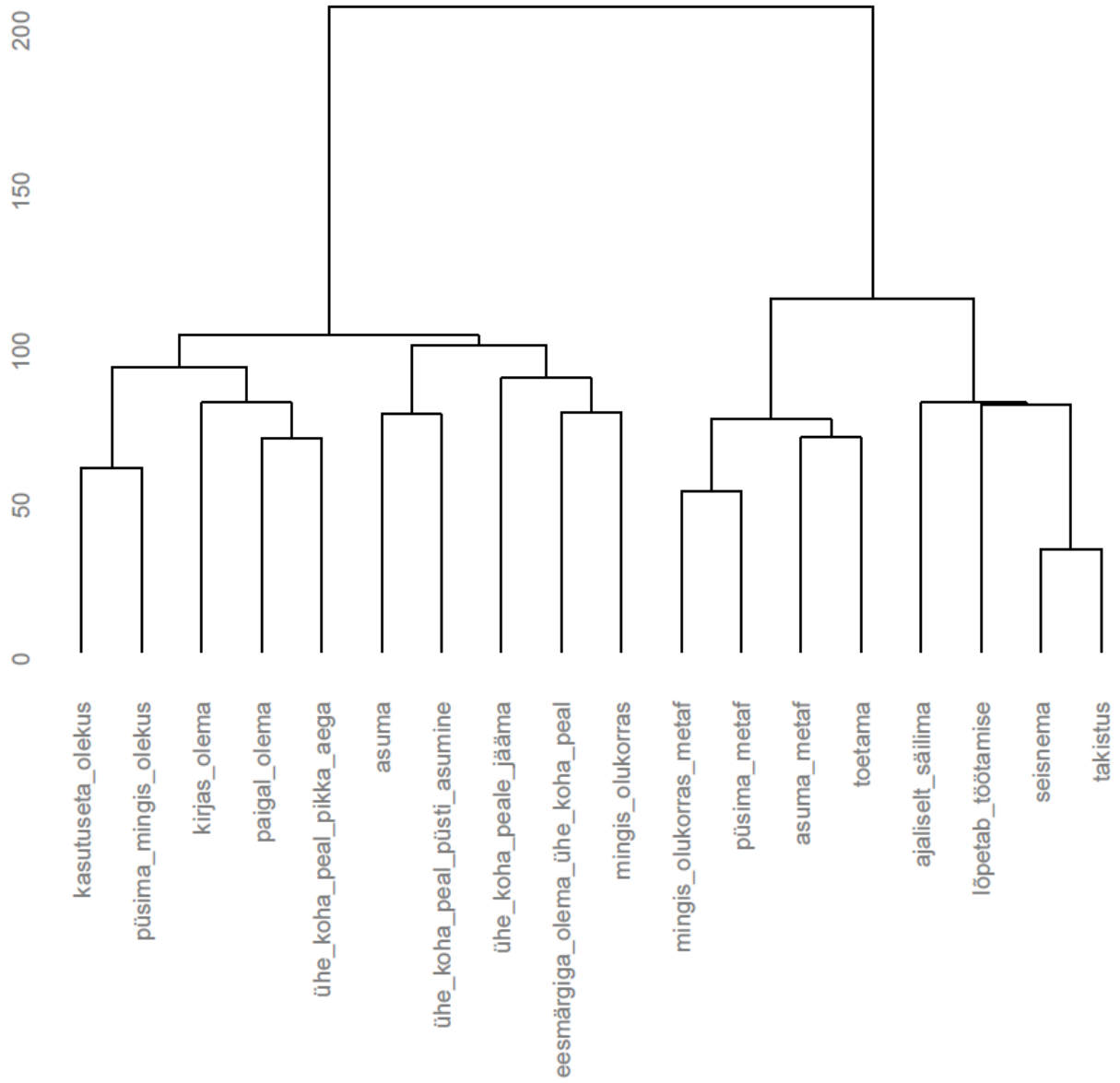
Klasterdamisel on ID-siltidest välja jäetud *Allikas*. See silt märgendati, et vaadata kuidas konkreetset ja abstraktsed tähendused tekstiliikide vahel jagunevad (vt ptk 3.1.), kuid laiemalt ei oma see tähtsust, sest töö eesmärk oli uurida verbi kasutust lausetes, mitte jagunemist tekstiliikide vahel.

4.2. Klasteranalüüsi tulemused ja tõlgendus

Joonisel 10 asuval dendrogrammil¹⁰ on esitatud ainult parafraasid ja nende jaotumine. Joonisel 12 (lk 62) on lisatud analüüsi ka ühendverbid.

Parafraaside omavaheliste seoste selgitamisel toon näiteid sagedusandmetest, mis profiilide võrdluses esinesid ja mis kirjeldadavate parafraaside juures võisid rolli mängida. Samas tuleb arvestada, et kõiki parameetreid ei suuda inimene (erinevalt andmetöötlusprogrammist) analüüsides nagunii arvesse võtta. Seetõttu on eri parafraaside juures toodud välja erinevad ID-sildi tasemete võrdlused, mis parajasti olulised tundusid.

¹⁰ Täpsemalt oli dendrogrammi tõlgendamisest ja lugemisest juttu alapeatükis 2.3.2.



Joonis 10. Dendrogramm kõigi parafraasidega, mis valimis esinesid.

Joonisel 10 on näha, et moodustunud on kaks suuremat klastrit. Ühte kuuluvad parafraasid KASUTUSETA OLEKUS, PÜSIMA MINGIS OLEKUS, KIRJAS OLEMA, PAIGAL OLEMA, ÜHE KOHA PEAL PIKKA AEGA, ASUMA, ÜHE KOHA PEAL PÜSTI ASUMA, ÜHE KOHA PEALE JÄÄMA, EESMÄRGIGA OLEMA ÜHE KOHA PEAL ja MINGIS OLUKORRAS (joonisel vasakul). Suurem kimp hõlmab seega enamjaolt konkreetseid tähendusi (KASUTUSETA OLEKUS ja PÜSIMA MINGIS OLEKUS on piiripealsed juhtumid). Ainus, mis ei EKSSi ega ka minu koostatud tähendusvõrgustiku järgi konkreetse tähenduse alla ei kuulu on KIRJAS OLEMA.

Teine suurem klaster (joonisel 10 paremal) sisaldab abstraktsemaid tähendusi ning pole nii homogeenne, kui esimene. See jaguneb omakorda kaheks: ühele poole jäävad MINGIS OLUKORRAS (metafooriliselt), PÜSIMA (metafooriliselt), ASUMA (metafooriliselt) ja TOETAMA. Teisele poole aga AJALISELT SÄILIMA, LÕPETAB TÖÖTAMISE, TAKISTUS ja SEISNEMA.

Seega jagunevad parafraasid ID-siltide järgi suures osas füüsilise ja mittefüüsilise ruumi vahel. Võrreldes peatükis 3.3. esitatud tähendusväljaga, on klasterdamise tulemus sarnane.

Vasaku poole peal joonisel 10, kus enamasti konkreetseid tähendused, on omavahel kõige sarnasemad parafraasid KASUTUSETA OLEKUS ja PÜSIMA MINGIS OLEKUS. Mõlema parafraasi juures domineerib seejuures abstraktne üldkasutus, kuigi PÜSIMA MINGIS OLEKUS puhul jaguneb kasutus peaaegu pooleks (53% lausetest abstraktset kasutust¹¹), KASUTUSETA OLEKU puhul on 80% lausetest abstraktse kasutusega. Tegemist on enamasti elutute agentidega, mis pole vertikaalsed. Kohamäärus esineb vaid vastavalt 39% ja 30% parafraasi alla kuuluvatest lausetest, kuid viisimäärus esineb seevastu 72% ja 71%. Siin on juba parafraaside nimetusest näha, et oluline on *olek*, seega mängib rolli just see *kuidas* seistakse ja seetõttu on need parafraasid ka teineteisele sarnased. Kui üldisem PÜSIMA MINGIS OLEKUS hõlmab igasugust olekut, nagu *üks seisab lukus* või *ristseliti, rattad seisavad teatud asendis* ja *värvid peal*, siis KASUTUSETA OLEK on kitsam määratlus, asjad seisavad *tühjalt, kasutuseta* või *niisama*. Ajamäärus esineb võrreldes teiste parafraasidega

¹¹ Kõik protsendid, mis selles peatükis on välja toodud on arvatatud ühe (kirjeldatava) ID-sildi raames, mis parafraasi alla kuulub.

samuti rohkem vastavalt 44% ja 35%. Ka on neid kahte parafraasi hea võrrelda, sest esinemissagedus korpuses on sarnane – 18 ja 17 korda.

Kui nende kahe parafraasi sarnasust arvesse võtta ja need ühe parafraasi all kokku panna, siis on näha, et nende esinemissagedus korpuses on küllalt suur – 35 korda. Seega on *seisma* verb kui mingit ajalist püsivust ja olekut näitav verb eesti keeles üsna sage.

Järgmisena on omavahel kõige sarnasemad PAIGAL OLEMA ja ÜHE KOHA PEAL PIKKA AEGA. Mõlemas on ligi 100% konkreetset üldkasutust. Kui PAIGAL OLEMA puhul on agendiks 68% lausetest sõidukid, siis teise parafraasiga domineerivad objektid, kujud. Seega on mõlema juures tegemist elutute agentidega. Kohamäärust on parafraasides vastavalt 72% ja 84%, seega ühendab neid tausta esinemine ja oluline on lokatiivsus. Kui PAIGAL OLEMA parafraasiga on taustatüübiks enamasti ehitised (36%), siis teise juures on põhiline välisruum (26%). Domineeriv verbi isik on mõlemal singulari kolmas.

Nende kahe parafraasiga on ühendatud KIRJAS OLEMA. See on huvitav, sest kirjas olemist arvatakse pigem piltlikuks kasutuseks (nt EKSS), kuigi mingi tekst, sõnad reaalselt seisavad tausta peal samamoodi nagu objektid või inimesed. Piltlikuks peetakse seda pigem seetõttu, et tekst iseenesest on midagi abstraktset ja mitte käega katsutavat.

Käitumisprofiilist tuleb aga välja, et kuigi üldtähenduses olen tekstid märkinud 100% abstraktseks, on teised sildid siiski piisavalt sarnased, et kasutus ühtib pigem konkreetse paigalolekuga. Sarnasused eelmistega on agendi elutus, vertikaalsuse puudumine, kohamääruse olemasolu (91,3%), adessiivi kasutus (30,4%), ka on enim kasutatud singulari kolmandat isikut. Selle kolmese klasteri juures on ehk kõige määravam kohamääruse kasutus.

Kolmene klaster on omakorda seotud parafraasidega KASUTUSETA OLEKUS ja PÜSIMA MINGIS OLEKUS.

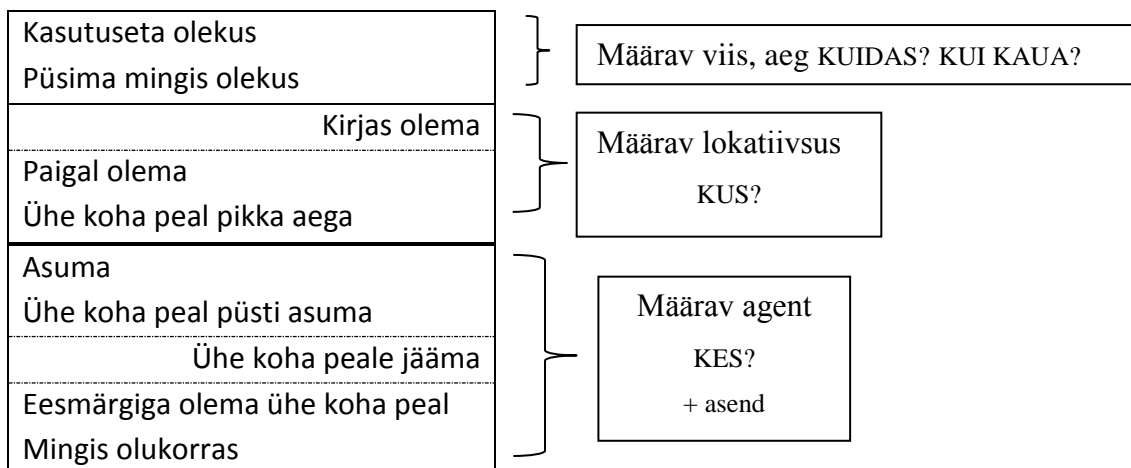
ASUMA ja ÜHE KOHA PEAL PÜSTI ASUMA parafraaside juures on tegemist 100% konkreetse kasutusega. PÜSTI ASUMISEL on agendiks inimesed (96%), ASUMISEL on kõige rohkem agendiks vertikaalsed objektid (25%) ja objektid (11%). Agendi vertikaalsus on mõlema puhul üle 70% lausetest. Kohamäärust esineb ASUMA parafraasiga 100% lausetes, ÜHE KOHA PEAL PÜSTI ASUMISEGA 80%.

EESMÄRGIGA ÜHE KOHA PEAL ja MINGIS OLUKORRAS parafraside juures on enamasti tegemist elus agentidega, inimestega (üle 95%). Esimesega on kohamäärust 83% lausetest, teisel vaid 25%. Ajamäärust esineb mõlemal u 23%. Ajavormidest domineerib minevik (üle 70% lausetest). Nende sarnasuse hindamisel tuleb arvestada, et parafras MINGIS OLUKORRAS esines vaid neli korda.

Viimase kahega on ühendatud ka parafras ÜHE KOHA PEALE JÄÄMA, mis jääb natukene kaugemale ASUMA ja ÜHE KOHA PEAL PÜSTI ASUMA parafrasist. Nii nagu eelmiste puhul on sellegi parafrasiga agendiks enamasti inimene (78%) ja üldtähendus konkreetne (88%). Kohamäärus esineb 35% lausetest, ajamäärust on 18%. Verb esineb enamasti minevikus.

Nende viie parafrasi (ASUMA, ÜHE KOHA PEAL PÜSTI ASUMA, EESMÄRGIGA ÜHE KOHA PEAL, MINGIS OLUKORRAS ja ÜHE KOHA PEALE JÄÄMA) ühiseks nimetajaks võib pidada agenti, enamasti on tegemist inimesega, määrused ei mängi sealjuures suuremat rolli.

Joonisel 11 on esitatud kokkuvõtlikult esimese suurema klatri jagunemine (nn füüsilise ruumi jagunemine) ja võimalik põhjus, mis kokku klasterdatud tähendusi ühendab.



Joonis 11. Suurema klatri jagunemine ja sagedusandmetel põhinev põhjus, miks tähendused võivad olla just nii kokku klasterdunud.

Abstraktsete tähenduste hulgas seevastu ei moodustu nii selget jagunemist. Võib eeldada, et piltlikud tähendused on keskmest kaugemale triivunud ja nende esmase tähendusega seostamine võib seega olla keerulisem.

Kõige esimesena on ühendatud parafrasid SEISNEMA ja TAKISTUS (need on üldse omavahel kasutuse poolest kõige sarnasemad). Kui nende profiile lähemalt vaadata, siis on seal väga palju sarnast. Arvesse tuleb võtta ka seda, et kumbagi parafraasi esines vähe (vastavalt 2 ja 3 korda), seega on võrdlusmaterjal napp. Nii tähendus kui ka agenditüüp oli abstraktne, tegemist oli seega elutute ja mittevertikaalsete agentidega, tausta ei esinenud üheski lauses, ajast kasutati olevikku ja kõneliik oli jaatav.

Teisena on kõige sarnasemad MINGIS OLUKORRAS (metaf) ja PÜSIMA (metaf). Mingis olukorras seisis enamasti inimene (73%), samas PÜSIMISE parafrasiga oli agent 40% lausetest abstraktne. Veel esinesid agendina kogum, objekt, sündmus. Kohamäärust, ajamäärust ei esinenud kummagi parafrasiga kordagi. Ajast kasutati enam olevikku.

Kahe eelnevaga on ühendatud ASUMA (metaf) ja TOETAMA. Agendiks on enamikul juhtudel inimene (50% ja 67%) ning peale selle abstraktne ja kogum. Taustatüüp esineb u 30% mõlema lausetest, ajamäärust esineb kummalgi 11%, ajast kasutatakse rohkem olevikku.

Seda neljast klastrit (MINGIS OLUKORRAS (metaf), PÜSIMA (metaf), ASUMA (metaf), TOETAMA) iseloomustab see, et verb on olevikus, üldtähendus abstraktne, agendiks abstraktne, kogum või inimene ja tausta puudumine või vähene esinemine.

Kaks parafraasi, AJALISELT SÄILIMA ja LÕPETAB TÖÖTAMISE jäävad võrdlemisi eraldiseisvaks, ka ei ole neid teineteisega kokku klasterdatud ehk siis ei ole nad omavahel piisavalt sarnased, kuid kõigi teistega on erinevus veelgi suurem.

Kui LÕPETAB TÖÖTAMISE on 100% konkreetne tähendus, siis AJALISELT SÄILIMA 100% abstraktne. Viimast on üldiselt raske analüüsida, sest seda esines vaid üks kord ja agendi näol oli tegemist elusorganismiga, lilledega. Pigem võiks ajalist säilimist seostada elutute entiteetidega, nagu toit (nt *seisab kauem värskes*). Kuna selle parafraasi juures oli iga ID-sildi taseme juures võimalik 100% või 0%, siis pole üllatav, et see küllalt eraldi jäi.

Parafraas LÕPETAB TÖÖTAMISE võis aga seetõttu ilma paariliseta jääda, et see käitub nagu metafoorne tähendus, kuigi üldtähendus on 100% konkreetne – midagi lõpetab liikumise, töötamise. Seega ei sobi see kumbagi rühma väga hästi. Agendiks on objektid ja sõidukid, seega elutud entiteedid. Samuti ei esine ei koha-, aja- ega viisimäärust, mis konkreetse poole peal paigale jäämise juures rolli mängisid.

Kui nüüd lisada viis verbiühendit, mis eelnevalt välja jäid, siis asetuvad tähendused veidi ümber, aga põhilised klastrid jäävad siiski samaks.

Joonisel 12 on lisatud verbiühendid *ees seisma*, *hea seisma*, *vastu seisma*, *eest seisma* ja *koos seisma* (joonisel 12 kastiga ümbritsetud). Kasutuse järgi paigutuvad need abstraktsete tähenduste vahele.

Parafraas MINGIS OLUKORRAS liigub konkreetse poole pealt piltlikusse, seega on see verbiühenditele sarnasem, kui *seisma* konkreetsematele tähendustele. Kokku liiguvad ÜHE KOHA PEALE JÄÄMA ja EESMÄRGIGA OLEMA ÜHE KOHA PEAL.

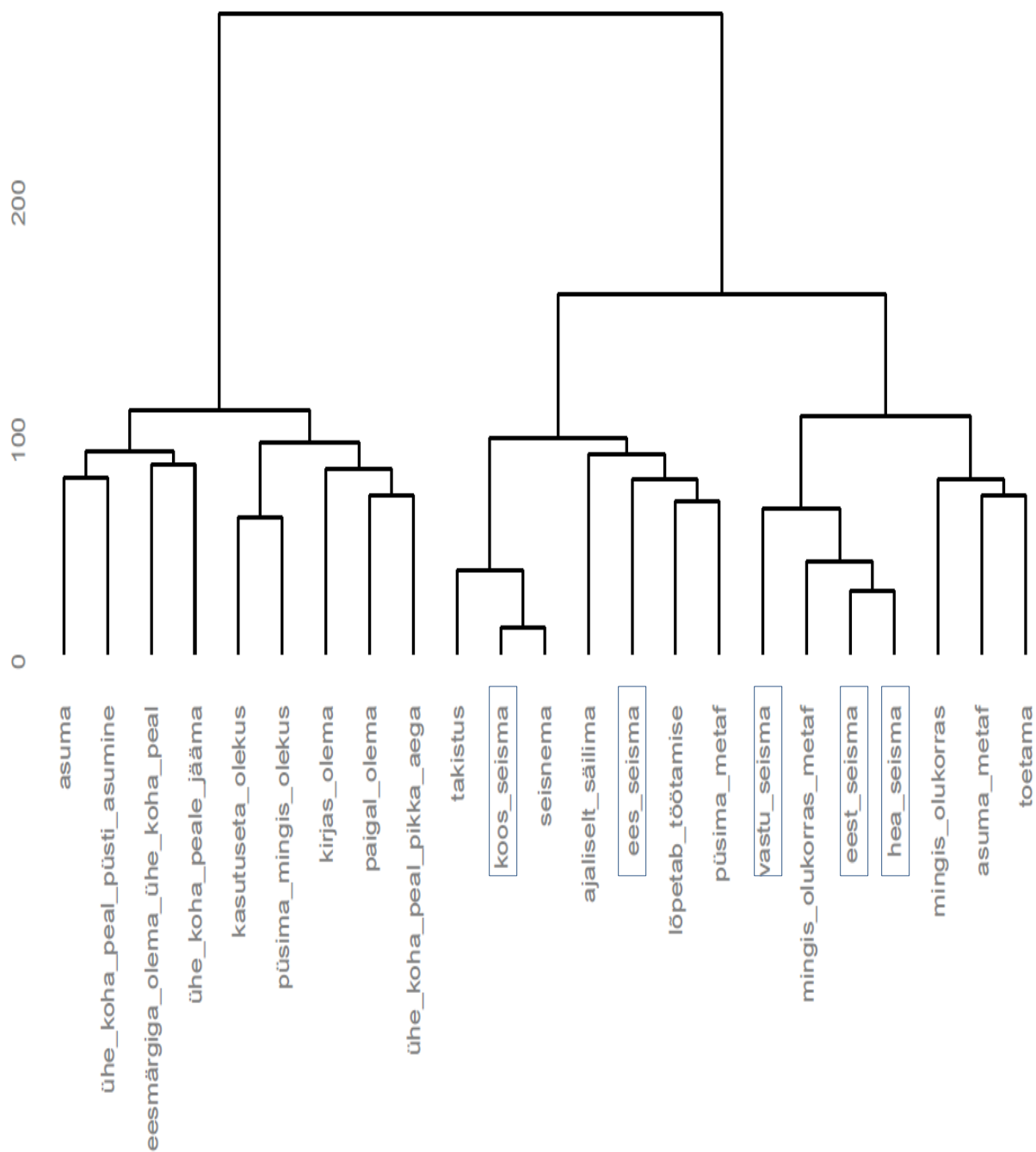
Kokku on klasterdunud *eest seisma* ja *hea seisma*, mis on ka tähenduse poolest sarnased. Samas ei ole need ühes klastris TOETAMA parafraasiga, kuhu need võiksid tähenduse poolest kuuluda.

Koos seisma on paigutunud SEISNEMA ja TAKISTUSE vahele. Seega on *koos seisma* ja SEISNEMA kasutuse järgi omavahel kõige sarnasemad.

LÕPETAB TÖÖTAMISE on kokku läinud metafoorilise püsimisega. Kõige eraldiseisvamateks kasutusteks on endiselt AJALISELT SÄILIMA ning sellele on lisandunud *ees seisma*.

Jooniselt 12 on näha, et ühendverbid *vastu seisma*, *eest seisma* ja *hea seisma* käituvad kõige sarnasemalt parafraasiga MINGIS OLUKORRAS (metafooriliselt) ning ühendverbid *ees seisma* ja *koos seisma* käituvad neist erinevalt. *Ees seisma* võrdlemisi eraldijääv paigutus võib tuleneda ka sellest, et selle ühendverbiga on ainsana kasutatud ajakategooriat *tulevik*.

Üks huvitav nähtus, mis profile võrreldes veel välja tuleb on see, et konkreetse seismise juures kasutatakse rohkem minevikuaega, samas metafoorse poole peal olevikku.



Joonis 12. Dendrogramm kõigi valimis esinenud parafraside ja verbiühenditega.

4.3. Võrdlus eelnevalt tehtud uurimustega

Selles peatükis on lühidalt välja toodud selle uurimistöö tulemused võrrelduna varasemate olulisemate töödega. Seejuures on arvestatud ka teiste keelte kohta tehtud uurimusi, lähtudes sellest, et inimesed on üldiselt sarnased ning tajuvad seismist ühtemoodi, see kuidas tähendus abstraktsemaks muutub ja grammatiseerub erineb aga keeleti.

Veismanni 2012. aastal tehtud uurimusest selgus, et ajakirjanduskeeles kasutatakse *seisma* verbi pooltel juhtudel abstraktses tähenduses (47% 205 lausest), sarnased tulemused on ka selles töös, ajakirjandustekstides on kasutatud 51% kõigist lausetest abstraktset tähendust ja ühendverbe. Teadustekstides on aga 70% abstraktset kasutust, seega mida formaalsema tekstiga on tegemist, seda kaugemale jääb asendiverb *seisma* esmasest tähendusest ehk jalgadel püsti seismisest ja ülekaalus on ülekantud tähendus ja ühendverbid.

See, mis on esmane või keskne tähendus ei ole tegelikult päris selge. Selles töös on võetud aluseks, et keskne tähendus on see, mis on korpuses kõige enam esindatud ning kasutusel kõige rohkemate ID-sildi tasemetega (inimese püsti seismine). Samuti on arvestatud seda, et asendiverb viitab inimese asendile ning seetõttu on seda kõige kergem defineerida just inimese seismise kaudu (Newmani 2002 prototüübid), ka on see esmane tähendus EKSSis.

Pajusalu (2001b) uuringust selgub, et sorteerimiskatses ei suutnud tudengid määrata verbi *seisma* keskset tähendust. Tema sõnul osutus põhitähenduse määramine informantide jaoks liiga raskeks või arusaamatuks ülesandeks. 41 analüüsitud vastuselehest oli vaid 11 lehel välja toodud sõna põhitähendus. 11st neli pidasid põhitähenduseks *paigal olema* ja *püsti olema* kombinatsiooni, mille näideteks olid valitud *Poiss seisab õues* ja *Koer seisab tagajalgadel* (EKSSis I rühm). Neid võib pidada põhitähenduse variantideks, kuid rohkem

kui korra toodi põhitähendusena välja laused *Buss seisab peatuses* ja *Auto seisab niisama, keegi ei kasuta seda*. Sellest lähtudes peaks Pajusalu arvates need tähendused olema EKSSis välja toodud põhitähendusele lähemal (praegu EKSSis V tähendusrühm). Pajusalu (2001b: 258)

Newman ja Rice (2004) on samuti teinud korpusanalüüsi, kus uurisid kolme põhiasendiverbi *seisma*, *istuma* ja *lebama*. Nende tööst selgub, et inglise keeles on põhiasendiverbid tihedalt seotud elususega, st elutuid entiteete seisis väga vähe. Peale selle on asendiverbid tugevalt seoses pikaajalise liikumatusega. Kuna eesti keeles on aga *seisma* verb seotud *jääma* verbi semantikaga, on seismine tugevamalt seotud just liikumise ja selle vastandusega, tegevuse muutusega. Nii on seismisel veel lisatähendused nagu AJALISELT SÄILIMA või HETKELINE PEATUMINE (*buss seisab peatuses*), samuti TÖÖTAMISE LÕPPEMINE (*kell jäi seisma*), mis näiteks inglise keeles ei esine, seega saab verbi kasutada enam ka elutute entiteetidega.

Selles töös oli parafras PAIGALE JÄÄMA korpuses püsti seismise järel kõige sagedasem, seega võib nõustuda Pajusalu (2001b) mõttekäiguga ning informantide vastustega paigalseisu kui esmase tähenduse lähedal asuvast tähendusest. Seejuures pole PAIGALE JÄÄMA ID-siltide taseme poolest püsti seismise järel teisel kohal. Kuna see on küllalt konkreetse kasutusega konstruktsioon (*jääma* verb enamasti kolmanda isiku preesensis või imperfektis ja *seisma* verb supiinis, agendiks inimesed, sõidukid), siis ei saa täielikult nõustuda Griesi (2006) väitega, et keskset tähendust saaks määrata ID-sildi tasemete arvu järgi. Samal seisukohal on Glynn (2012), kes arvab, et ID-sildi tasemete järgi võib sellest n-õ kesksest tähendusest saada hea ülevaate, see ei tähenda, et see peaks ilmtingimata prototüüpne olema.

Selle töö eesmärk ei olnud aga välja selgitada, mis on *seisma* verbi keskne tähendus, vaid uuritud on, kuidas eri tähendused omavahel võivad seotud olla. See on ka suurim erinevus varasemaselt tehtud töödest.

4.4. Järeldused

Sõna käitumisprofiili koostamine on võimalus, kuidas koguda ühe sõna kohta võimalikult palju infot ning seda arvestades vaadata, kuidas polüsemse sõna eri tähendused jagunevad ja kui tugevalt mis tähendused omavahel seotud on. Selleks on eelnevalt vaja põhjalikku tööd korpusega, et laused märgendada ja eri tähenduste vahel jagada.

Klasteranalüüsi tulemustest võib järeldada, et käitumisprofiil on võimekas meetod, kuidas leida suures andmehulgas mingit struktuuri. Kuigi selles töös on klastrid moodustunud üsna loogiliselt, st konkreetsemad tähendused moodustavad omaette suurema klatri ja abstraktsemad tähendused asetsevad neist palju kaugemal ning moodustavad eraldi-seisvaid klastreid, tuleb siiski arvestada, et multifaktoriaalses analüüsis võiks esineda igat kasutusjuhtu palju rohkem. Selle töö valimis esines aga mõnda tähendust ainult üks või kaks korda. Kuna polüsemia uurimisel ei saa kindlasti tähendusi välja jätta, siis on neid analüüsitud ja klasterdatud. Sellisel juhul ei pruugi aga saada kõige täpsemat pilti. Samas ainult sagedusandmeid vaadates on kõnekad ka need tähendused, mis vähem esinevad, st näeb kas ja kui palju need võivad keeles kasutuses olla.

500-lauseline valim ei pruugi eelnevat arvesse võttes olla piisav, kuid sobib magistritöö mahuga ning mingi esmase pildi saab verbi käitumisest siiski. Kindlasti võiks veel täpsemalt uurida abstraktseimaid *seisma* verbi kasutusi, sest seal on varieeruvus suurem ja neid esineb arvuliselt kokku ka rohkem, kui konkreetseid tähendusi. Seega on *seisma* verbi tähendusväli laiem, kui esialgu arvatud.

Eriti huvitavad on seejuures ühendverbid ja see, mis võib nende kasutust motiveerida. Siin võiks vaadata ühendverbide seotust suuna- ja ruumimetafooridega. Korpuses olid kõige sagedasemad ühendverbid *ees seisma* ja *taga seisma*. *Taga seisma* kasutus on omakorda jagunenud mitme tähenduse vahel.

Nagu eelpool mainitud, siis klasterdades jagunesid tähendused konkreetse (füüsilise ruumi) poole peale ja abstraktse poole peale. Esimeses olid tähendused, mis seostusid inimese seismise ja lokatiivsusega (nt PAIGAL OLEMA, KIRJAS OLEMA), ka olid selles klastris laused, mille seismine väljendas üldisemat olekut (KASUTUSETA, TEGEVUSETA OLEK) ja mille alla kuulusid nii konkreetse kui ka abstraktse üldtähendusega laused (*riided seisavad higisoolast püsti, maja seisab kütmata*). Teisele poole jäid abstraktsemad tähendused nagu AJALISELT SÄILIMA, TOETAMA ja PÜSIMA (metaf).

Korpuse lausetest lähtuvalt võib järeldada, et asendiverbi *seisma* tähendused on omavahel tihedalt seotud, siiski võib jagada laused kolme valdkonna vahel: ASEND (inimese seismine, vertikaalsus), LIKUMATUS (püsimine, hetkeline peatumine, säilimine = liikumisele vastandamine) ja LOKATIIVSUS (asetsemine, koht ja aeg). Need kolm on seismise juures kesksel kohal ning olenevalt tähendusest on olulisem üks või teine või kõik korraga.

Kõik valdkonnad korraga mängivad rolli näiteks inimese seismise juures. Seda võib pidada ka asendiverbi keskseks tähenduseks. Kuid kuna olulised on kõik kolm eelpool mainitud valdkonda, siis on inimese seismine kitsamalt piiritletud. Seevastu *seisma* tähendus vastanduses liikumisele on oluliselt laiemas kasutusega. Seda näitab ka liikumisega opositsioonis olevate abstraktsete tähenduste rohkus (PAIGAL OLEMA, PAIGALE JÄÄMA, LÕPETAB TÖÖTAMISE, PÜSIMA), samas ei ole nende juures vertikaalsus üldse oluline (nt *maailm seisab revolutsiooni lävel* = võime liikuda).

Liikumise vastanduse tajumine tundub seejuures oluline enamike lausete juures, sest *seisma* verbi kasutatakse koos eesti keele ühe tuumverbiga *jääma* (vt Tragel 2003). See on aga nii kinnistunud konstruktsioon (*paigale, seisma jääma*), et *seisma* verbi juures võib tajuda ka võimet liikuda, erinevalt nt teisest põhiasendiverbidest *istuma* ja *lebama*.

Samuti on oluline osa lokatiivsusel ja kuna tähendused on omavahel seotud, siis on kohati raske vahet teha, mis on millise tähenduse juures rohkem oluline.

Kui inimese seismine välja jätta, siis pole verbi kasutuse juures vertikaalsus määrav, kuigi see võib rolli mängida verbi valikul, st kui objekt sarnaneb rohkem inimesele, siis võidakse eelistada *seisma* verbi. Kui võtta valimi peale kokku kõik sõidukid, objektid,

vertikaalsed objektid, ehitised ja kujud, siis selgub, et vertikaalseid on neist 28, samas mittevertikaalseid 48. Seega ei mängi objektide seismisel vertikaalsus nii suurt rolli, kui võiks eeldada.

Sellest võib järeldada, et *seisma* asendiverbi kesksest tähendusest hargneb mitmeastmeline kontiinum paiknemisverbini, st kui esmase tähenduse juures on olulised lokatiivsus, asend ja liikumatus, siis verbi kasutusala laienedes nende kõigi osatähtsus mingil määral väheneb. Pigem asendab *seisma* verb sel juhul koopulat ja soovitakse väljendada tugevamat, püsivamat (ajas ja ruumis) olekut. Esmast tähendust on metafoorsetes valdkondades raskem tajuda ja seetõttu käitub *seisma* rohkem nagu paiknemisverb.

Täpsemalt võiks uurida katsega, kuidas katseisikud iga parafraasi esindavad näitelauseid jagaksid ning missugused omavahel sagedamini kokku paneksid, samuti seda millise tähenduse juures peavad nad oluliseks liikumisega vastandumist ja võimet liikuda.

Selle töö jätk võiks olla klastrite põhjalikum analüüs ning eri ID-siltide järgi klasterdamine, et vaadata, millised tähendused väiksemat hulka silte arvestades kokku võiksid liikuda. Samuti oleks huvitav analüüsida teisi põhiasendiverbe *istuma* ja *lebama* ning võrrelda nende kasutusmustreid *seisma* verbi omaga.

Kokkuvõte

Selles magistritöös uuriti eesti keele asendiverbi *seisma* polüseemiat. Asendiverbe *istuma*, *lebama* ja *seisma* peetakse põhiasendiverbideks (*cardinal posture verbs*) ja need viitavad esmajoones inimese keha asendile teatud positsioonis olles. Seismise korral on see jalgadele toetuv ja vertikaalne. Samal ajal on *seisma* verbil ka palju muid kasutusi ning selles magistritöös vaadati *seisma* verbi korpuses esinevaid tähendusi.

Varem on eesti keele *seisma* verbi kasutust uurinud R. Pajusalu (2001a, 2001b) ja Veismann (2012), kuid põhjalikku korpusanalüüsi pole seni tehtud. Selle asendiverbi semantikat oli huvitav vaadata just seetõttu, et *seisma* verbi kasutus on laienenud paljudesse metafoorsetesse valdkondadesse, EKSS annab verbile näiteks seitse põhitähendust hulga piltlike alltähendustega.

Töö teoreetiline taust on kognitiivne lingvistika, laiemalt kasutus põhine lingvistika ning meetod korpuslingvistika. Töös on kasutatud korpuslingvistide Stefan Th. Griesi ja Dagmar Divjaki välja töötatud korpuspõhist meetodit käitumisprofili (*Behavioral Profile*) ja andmeanalüüsis hierarhilist aglomeratiivset korpusanalüüsi.

Sõna käitumisprofili koostamiseks tuli esmalt leida korpusest lekseemi kõik kasutusjuhud ning seejärel koostada sobiva mahuga valim. Seejuures oli oluline kontekst ja kindlasti oli vaja tervet lauset, vahel ka seda ümbritsevaid lauseid. Järgmise sammuna märgendati süntaktilised, semantilised ja morfoloogilised variaablid (ID-sildid) ning nendest moodustus sõna profiili. Seejärel koostati koosinemistabel, kus kirjas mitu korda mingi variaabel mingi tähendusega esines. Koosinemistabeli põhjal arvutati sagedusandmed ja nii oli materjal ette valmistatud põhjalikumaks andmeanalüüsiks.

Selle töö aluseks oli 500-lauseline valim tasakaalus korpusest. 500 lauset jagati 23 tähenduse vahel, sealjuures moodustati 18 parafraasi, 5 verbiühendit jäeti parafraseerimata.

Käitumisprofili koostamisel märgendati 17 variaablit (ID-silti). Neist 16 kasutati lõplikus andmeanalüüsis (vt ptk 3.1.).

Töös leiti sõna *seisma* tähendustele järgmised parafrasid (esitatud tähestikulises järjekorras): AJALISELT SÄILIMA, ASUMA, ASUMA (metaf), EESMÄRGIGA OLEMA ÜHE KOHA PEAL, KASUTUSETA OLEKUS, KIRJAS OLEMA, LÕPETAB TÖÖTAMISE, MINGIS OLUKORRAS, MINGIS OLUKORRAS (metaf), PAIGAL OLEMA, PÜSIMA (metaf), PÜSIMA MINGIS OLEKUS, SEISNEMA, TAKISTUS, TOETAMA, ÜHE KOHA PEAL PIKKA AEGA, ÜHE KOHA PEAL PÜSTI ASUMA ja ÜHE KOHA PEALE JÄÄMA.

Verbiühendid, mis esinesid on *ees seisma, eest seisma, vastu seisma, koos seisma ja hea seisma*.

Klasteranalüüsi tulemusena moodustus kaks suuremat klastrit, kus ühele poole jäid pigem konkreetsed tähendused, näiteks ÜHE KOHA PEAL PÜSTI ASUMA (*inimese seismine*) ja ÜHE KOHA PEALE JÄÄMA, samuti oli seal PÜSIMA MINGIS OLEKUS (*ruumid seisavad tühjalt*), milles *seisma* verbi kasutatakse nii abstraktses kui ka konkreetse tähenduses.

Teise suure klasteri moodustasid abstraktsemad tähendused ja verbiühendid. Abstraktsema poole peale asetus parafras LÕPETAB TÖÖTAMISE, mis viitab peatumisele, olukorra muutumisele ja mille kasutust võiks pidada pigem konkreetseks. Samuti paigutus abstraktsema poole peale MINGIS OLUKORRAS (nt *silmitsi seisma kellegagi*), mis käitumisprofiilis oli märgendatud konkreetse üldtähendusega. Üldiselt oli klastrite jagunemine loogiline ja selgitatav. Seega on käitumisprofiil arvestatav meetod, millega suures andmehulgas struktuuri leida ja polüseemiat uurida.

Käitumisprofiil kui meetod annab võimaluse vaadata, milline võiks olla polüseemse sõna keskne tähendus ja millised tähendused võiksid teineteisele lähemal asetseda, tehes seda seejuures objektiivsematel alustel kui pelgalt intuitsioonil põhinev uurimus. Seejuures tuleks aga arvestada, et parimate tulemuste saamiseks võiks valim olla suurem kui selles töös. Seda arvesse võttes tasuks mõelda, kuidas oleks võimalik märgendamist rohkem automatiseerida. Kindlasti võiks käitumisprofili meetodit kasutada ka sünonüümsete sõnade uurimisel ja eesti keele konstruktsioonide võrdlemisel mõne teise keeltega.

Kirjandus

- Berez, Andrea L., Stefan Th. Gries* 2008. In Defense of Corpus-based Methods: A Behavioral Profile Analysis of Polysemous *get* in English. *International Journal of American Linguistics* 76 (1), lk 145–165.
- Bybee, Joan, Clay Beckner* 2009. Usage-based Theory. – *Handbook Linguistic Analysis*. H. Narrog and B. Heine (toim), lk 827–855. Oxford: Oxford University Press.
- Bybee, Joan* 2012. Usage-based Models in Linguistics: an Interview with Joan Bybee. <http://www.letras.ufrj.br/poslinguistica/revistalinguistica/index.php/volume-8-numero-1-junho-2012/usage-based-models-in-linguistics-an-interview-with-joan-bybee/> (vaadatud 13.05.2014)
- Brugman, Claudia* 1981. *Story of over*. – Magistritöö. Berkley Ülikool, California.
- Brugman, Claudia, George Lakoff* 2006. Cognitive Topology and Lexical Networks. – *Cognitive Linguistics: Basic Readings*. Dirk Geeraerts (toim), lk 109–139. Berlin/New York: Mouton de Gruyter.
- Divjak, Dagmar, Stefan Th. Gries* 2006. Ways of Trying in Russian: Clustering Behavioral Profiles. – *Corpus Linguistics and Linguistic Theory* 2, lk 23–60.
- Divjak, Dagmar* 2013. *Corpus linguistics: Principles and Methods*. Suuline ettekanne slaididega Lille'i kevadkoolis Prantsusmaal.
- Divjak, Dagmar, Nick Fieller (ilmumas)*. *Finding Structure in Linguistic Data*.
- EKSS = Eesti keele seletav sõnaraamat. IV. 2009. Tallinn: Eesti Keele Sihtasutus.
- Eslon, Pille* 2010. Keelekorpuste koostamisest, liikidest ja kasutamisest. http://www.tlu.ee/~jaanike/loengud/Mis_on_korpus.pdf (vaadatud 13.05.2014)

- Geeraerts, Dirk* 2006. Introduction: A Rough Guide to Cognitive Linguistics. – Cognitive Linguistics: Basic Readings. Dirk Geeraerts (toim), lk 1–28, Berlin: Mouton de Gruyter.
- Gibbs jt= Gibbs, Raymond W. jr., Dinara A. Beitel, Michael Harrington, Paul E. Sanders* 1994. Taking a Stand on the Meanings of *Stand*: Bodily Experience as Motivation for Polysemy. – Journal of Semantics 11, lk 231–251.
- Gibbs, Raymond W., Herbert L. Colston* 2006. The Cognitive Psychological Reality of Image Schemas and their Transformations. – Cognitive Linguistics: Basic Readings. Dirk Geeraerts (toim), lk 239–268. Berlin/New York: Mouton de Gruyter.
- Gilquin, Gaëtanelle, Stefan Th. Gries* 2009. Corpora and Experimental Methods: A state-of-the-art Review. – Corpus Linguistics and Linguistic Theory 5-1, lk 1–26. Walter de Gruyter.
- Glynn, Dylan* 2012. The Many Uses of Run: Corpus Methods and Socio-cognitive Semantics. –Corpus Methods in Cognitive Semantics. D. Glynn, J. Robinson (toim), lk 1–31. Amsterdam: John Benjamins.
- Goldberg, Adele* 1995. Constructions: A Construction Grammar Approach to Argument Structure. Chicago: University of Chicago Press.
- Gries, Stefan Th., Dagmar Divjak* 2009. Behavioral Profiles: a Corpus-based Approach to Cognitive Semantic Analysis. – New directions in Cognitive Linguistics. Vyvyan Evans ja Stephanie S. Pourcel (toim), lk 57–75. Amsterdam/Philadelphia: John Benjamins.
- Gries, Stefan Th., Dagmar Divjak* 2010. Quantitative Approaches in Usage-based Cognitive Semantics: Myths, Erroneous Assumptions, and a Proposal. – Quantitative Methods in Cognitive Semantics: Corpus-driven Approaches. Dylan Glynn, Kerstin Fischer (toim), lk 333–354. Berlin/New York: Mouton de Gruyter.

- Gries, Stefan Th.* 2006. Corpus-based Methods and Cognitive Semantics: The Many Senses of to Run. – Corpora in Cognitive Linguistics: Corpus-based Approaches to Syntax and Lexis, Stefan Th. Gries, Anatol Stefanowitsch (toim), lk 57–99. Berlin/New York: Mouton de Gruyter.
- Hopper, Paul* 1988. Emergent Grammar and the A Priori Grammar Postulate. – Linguistics in Context. Deborah Tannen (toim), lk 117–135. Ablex Publishing Corporation.
- Johnson, Mark* 1987. The Body in the Mind. The Bodily Basis of Meaning, Imagination, and Reason. Chicago: The University of Chicago Press.
- Kask, Kristiina* 2012. Asendiverbide *istuma*, *lebama*, *rippuma* ja *seisma* kasutus moodustuskatses. – Bakalaureusetöö. Tartu Ülikool.
- Kleimann, Maria* 2007. Paiknemisverbide semantika kognitiivsest vaatepunktist. – Bakalaureusetöö. Tartu Ülikool.
- Krillo, Kerli* 2010. Statistilise analüüsi teostamine Exceli ja SPSSi abil. – Konspekt. http://www.e-ope.ee/download/euni_repository/file/853/Statistika.pdf (vaadatud 13.05.2014)
- Kukk, Marek* 2011. Kontseptuaalne integratsioon ja seisma-verbi polüseemia. – Bakalaureusetöö. Tartu Ülikool.
- Lakoff George, Mark Johnson (1980)* 2011. Metafoorid, mille järgi me elame. Tallinn: TLÜ Kirjastus.
- Lance, G. N., W. T. Williams* 1967. Mixed-data classificatory programs. Agglomerative Systems. Australian Computer Journal 1, lk 15–20.
- Langacker, Ronald W.* 2000. Grammar and Conceptualization. Berlin: Mouton de Gruyter.
- Lemmens, Maarten* 2001. Computational Economy in Metaphor Analysis; <http://perso.univ-lille3.fr/~mlemmens/docspdf/MetaphorEconomy.pdf> (vaadatud 12.05.2014)

Lemmens, Maarten 2005. Motion and Location: Toward a Cognitive Typology.

<http://stl.recherche.univ-lille3.fr/sitespersonnels/lemmens/docspdf/MotionLocation.PDF>

[\(vaadatud 12.05.2014\)](#)

Lemmens, Maarten 2012. A typological View on Location. Loengukonspekt. 29.02.2012

McEnery Tony, Andrew Wilson 2006. Corpus Linguistics 2nd Edition. Edinburgh:
Edinburgh University Press.

McEnery, Tony, Andrew Hardie 2012. Corpus Linguistics. Method, Theory and Practice.
Cambridge University Press

Nerlich, Brigitte, Zazie Todd, Vimala Herman, David D. Clarke 2003. Polysemy: Flexible
Patterns of Meaning in Mind and Language. Berlin: Mouton de Gruyter.

Newman, John 2002. A cross-linguistic overview of the posture verbs 'Sit', 'Stand' and
'Lie'. – The Linguistics of Sitting, Standing and Lying. Koost. John Newman.
Amsterdam/Philadelphia: John Benjamins.

Newman, John, Sally Rice 2004. Patterns of Usage for English 'Sit', 'Stand' and 'Lie': A
Cognitively Inspired Exploration in Corpus Linguistics. – Cognitive Linguistics 15, lk
351–396.

Oosten, Jeanne van 1984. Sitting, Standing and Lying in Dutch: A Cognitive Approach to
the Distribution of the Verbs *Zitten*, *Staan*, and *Liggen*. Jeanne van Oosten, Johan
Snapper (toim.). – Dutch linguistics at Berkeley, lk 137–160.

- Pajusalu, Renate* 2001a. The polysemy of *seisma* 'to stand': multiple motivations for multiple meanings. – Papers in Estonian Cognitive Linguistics. Toim. Ilona Tragel. Tartu: Tartu Ülikooli Kirjastus, lk 170–191.
- Pajusalu, Renate* 2001b. Kas moos ja buss seisavad sarnaselt ehk väike katse verbiga *seisma*. – Keele kannul. Pühendusteos Mati Ereli 60. Sünnipäevaks. TÜ eesti keele õppetooli toimetised 17. Tartu, lk 222–231.
- Raukko, Jarno* 1999. An „Intersubjective“ Method for Cognitive-semantic research on Polysemy: The Case of *get*. – Cultural, Psychological and Typological Issues in Cognitive Linguistics. Masako K. Hiraga, Chris Sinha, Sherman Wilcox (toim), lk 87–106. Amsterdam, Philadelphia: John Benjamins.
- Raukko, Jarno* 2003. Polysemy as Flexible Meaning: Experiments with English *get* and Finnish *pitää*. – Polysemy. Flexible Patterns of Meaning in Mind and Language. Brigitte Nerlich jt (toim), lk 161–193.
- Ravin, Yael, Claudia Leacock* 2000. Polysemy: Theoretical and Computational Approaches. Oxford: Oxford University Press.
- Rosch, Eleanor* 1978. Principles of Categorization. – Cognition and Categorization. Hillsdale, E. Rosch, B. B. Lloyd (toim), lk 27–48. N.J: Lawrence Erlbaum.
- Talmy, Leonard* 2000. Toward a Cognitive Semantics. Vol 2. Typology and Process on Concept Structuring. Cambridge, MA: the MIT Press.
- Schönefeld, Doris* 2006. From conceptualization to linguistic expression: Where languages diversify. – Corpora in Cognitive Linguistics: Corpus-Based Approaches to Syntax and Lexis. Gries, Stefan Th., Anatol Stefanowitsch (toim), lk 297–344. Berlin /New York: Mouton de Gruyter.
- Soodla, Karin* 2003. Eesti keele verbid muutuma, minema, saama ja jääma muutumis-verbidena. – Bakalaureusetöö. Tartu Ülikool.

Tomasello, Michael 2003. *Constructing a Language*. Harvard University Press.

Tragel, Ilona 2003. *Eesti keele tuumverbid*. – Doktoritöö. Tartu: Tartu Ülikooli Kirjastus.

Tummers jt= Jose Tummers, Kris Heylen, Dirk Geeraerts 2005. *Usage-based Approaches in Cognitive Linguistics: A Technical State of the Art*. – *Corpus Linguistics and Linguistic Theory* vol. 1–2, lk 225–261.

Veismann, Ann 2004. *Sõna üle tähendustest*. *Keel ja Kirjandus* 10, lk 762–777.

Veismann, Ann 2012. *Korpusuuringud ja katsed semantikas sõna seisma näitel*. Suuline ettekanne slaididega Eesti Kognitiivse Keeleteaduse Ühingu I aastakonverentsil.

LISA 1. Parafraasid, nende esinemise arv korpuses ja näitelause.

Parafraas	Arv	Näide
ÜHE KOHA PEAL PÜSTI ASUMINE	191	<i>Ta seisis, käed ristis, hing tühi, vaatas maad enda jalgade ees.</i>
ÜHE KOHA PEALE JÄÄMA	40	<i>Siis jäi ta seisma ning avas peo</i>
ASUMA	36	<i>Kõrgematel kohtadel seisid mõned kuhjad.</i>
ASUMA (METAF)	26	<i>Maailm seisab põllumajandusliku revolutsiooni lävel.</i>
PAIGAL OLEMA	25	<i>Maja ees seisis mingi tattide auto.</i>
KIRJAS OLEMA	23	<i>Minu passis seisab nimi Neeme Truu.</i>
ÜHE KOHA PEAL PIKKA AEGA	19	<i>Riiulil seisis veel praegugi vana pruuniks roostetanud plekklehter.</i>
EESMÄRGIGA OLEMA ÜHE KOHA PEAL	18	<i>Said jupikese vorsti, kui järjekorras seisid.</i>
KASUTUSETA OLEKUS	18	<i>Ruumid on seisnud pikka aega pooltühjana.</i>
PÜSIMA MINGIS OLEKUS	17	<i>Nende riided seisid higisoolast püsti.</i>
EES SEISMA	15	<i>Põhjalikud muutused seisavad ees ka liidu regionaalpoliitika rahastamisel</i>
PÜSIMA (METAF)	15	<i>Vaikus seisis meie vahel.</i>
MINGIS OLUKORRAS (METAF)	11	<i>Sellest tulenevalt seisab praegune Eesti silmitsi sisemiste ohtudega</i>
TOETAMA	9	<i>See on natsionalism kitsamas mõttes, selle taga seisavad inimesed – rahvuslased.</i>
EEST SEISMA	7	<i>Ainult neid, kes ise suudavad enda eest seista, suudab ka kentuura aidata.</i>
VASTU SEISMA	7	<i>Anpilovi töölissalklased üritasid Kremlile jõuga vastu seista.</i>
LÕPETAB TÖÖTAMISE	6	<i>Vedru käib maha ja masinavärk jääbki seisma</i>
HEA SEISMA	5	<i>Seaduste järgi peab arst hea seisma oma kohustuste korrektse täitmise eest.</i>
MINGIS OLUKORRAS	4	<i>Ta seisis oma suureks ehmatuseks silm silma vastu suure oinaga</i>
TAKISTUS	3	<i>Maja hävitada, et ta ei seisaks enam elul ja arusaamisel põigiti ees.</i>
KOOS SEISMA	2	<i>Kuid, noormees, elu ei seisa koos ainult romantikast.</i>
SEISNEMA	2	<i>Probleem seisab ju selles, et kui oled kaotanud, siis tahad tagasi võita.</i>
AJALISELT SÄILIMA	1	<i>Ei need kaua seisa, aga või neid kauaks vaja ongi.</i>

LISA 2. Koosinemistabel parafraasi ja agenditüübi, elususe, vertikaalsusega.

ID-SILT	ID-SILDI TASE	ajaliselt säilima	asuma	asuma (metaf)	ees seisma	eesmärgiga olema ühe koha peal	eest seisma	hea seisma	kasutuseta olekus	kirjas olema	koos seisma	lõpetab töötamise	mingis olukorras	mingis olukorras (metaf)
AGENDI- TÜÜP	abstraktne	0	0	3	1	0	0	0	0	0	2	0	0	0
	anum	0	1	0	0	0	0	0	0	0	0	0	0	0
	ehitis	0	1	0	0	0	0	0	4	0	0	0	0	0
	ei ole	0	0	0	0	1	0	0	0	0	0	0	0	0
	elusorganism	1	3	0	0	0	0	0	0	0	0	0	0	0
	inimene	0	11	13	0	17	5	4	4	2	0	0	4	8
	institutsioon	0	0	1	1	0	2	1	0	0	0	0	0	2
	kogum	0	4	6	0	0	0	0	1	0	0	0	0	1
	kuju	0	0	0	0	0	0	0	0	0	0	0	0	0
	maa-ala	0	1	0	0	0	0	0	3	0	0	0	0	0
	muu	0	2	0	0	0	0	0	1	0	0	2	0	0
	objekt	0	4	0	0	0	0	0	0	0	0	2	0	0
	periood	0	0	0	3	0	0	0	0	0	0	0	0	0
	ruum	0	0	0	0	0	0	0	3	0	0	0	0	0
	sõiduk	0	0	0	0	0	0	0	0	0	0	2	0	0
	sündmus	0	0	1	10	0	0	0	1	0	0	0	0	0
	tekst	0	0	2	0	0	0	0	1	21	0	0	0	0
vert objekt	0	9	0	0	0	0	0	0	0	0	0	0	0	
ELUSUS	elus	1	15	14	0	17	5	4	4	2	0	1	4	8
	eluta	0	21	12	15	0	2	1	14	21	2	5	0	3
	ei saa määrata	0	0	0	0	1	0	0	0	0	0	0	0	0
VERTI- KAALSUS	jah	1	26	13	0	17	6	4	4	2	0	0	4	8
	ei	0	10	13	15	0	1	1	13	21	2	6	0	3
	ei saa määrata	0	0	0	0	1	0	0	1	0	0	0	0	0

LISA 2.

ID-SILT	ID-SILDI TASE	paigal olema	püsima (metaf)	püsima mingis olekus	seisnema	takistus	toetama	vastu seisma	ühe koha peal pikka aega	ühe koha peal püsti asumine	ühe koha peale jääma
AGENDI- TÜÜP	abstraktne	0	6	0	2	2	2	0	0	0	3
	anum	1	0	1	0	0	0	0	3	1	0
	ehitis	0	0	1	0	1	0	1	3	0	0
	ei ole	1	0	0	0	0	0	4	0	1	1
	elusorganism	0	0	1	0	0	0	0	0	2	0
	inimene	1	1	2	0	0	6	2	2	184	31
	institutsioon	0	1	0	0	0	0	0	0	0	0
	kogum	2	2	1	0	0	1	0	3	1	2
	kuju	0	0	0	0	0	0	0	5	1	0
	maa-ala	0	0	0	0	0	0	0	0	0	0
	muu	0	0	0	0	0	0	0	0	0	1
	objekt	1	2	8	0	0	0	0	2	0	0
	periood	0	0	0	0	0	0	0	0	0	0
	ruum	0	0	2	0	0	0	0	0	0	0
	sõiduk	17	0	0	0	0	0	0	0	0	1
	sündmus	1	2	0	0	0	0	0	0	0	0
	tekst	0	1	0	0	0	0	0	0	0	1
vert objekt	1	0	1	0	0	0	0	0	1	1	0
ELUSUS	elus	1	1	3	0	0	7	2	2	187	31
	eluta	23	14	14	2	3	2	1	17	3	8
	ei saa määrata	1	0	0	0	0	0	4	0	1	1
VERTI- KAALSUS	jah	2	1	5	0	0	6	2	13	190	31
	ei	22	14	12	2	3	3	1	6	0	8
	ei saa määrata	1	0	0	0	0	0	4	0	1	1

LISA 3. Koosinemistabel parafraasi aja-, koha- ja viisimäärusega ja (koha)määrusetüüp.

ID-SILT	ID-SILDI TASE	ajaliselt säilima	asuma	asuma (metaf)	ees seisma	eesmärgiga olema ühe koha peal	eest seisma	hea seisma	kasutuseta olekus	kirjas olema	koos seisma	lõpetab töötamise	mingis olukorras
KOHA-MÄÄRUS	jah	0	36	9	0	15	0	0	7	21	0	0	1
	ei	1	0	17	15	3	7	5	11	2	2	6	3
MÄÄRUSE-TÜÜP	asemäärsõna	0	4	0	0	0	0	0	2	2	0	0	0
	ei saa määrata	1	0	17	15	3	7	5	11	2	2	6	3
	kaassõnafraas	0	11	5	0	2	0	0	1	2	0	0	0
	nimisõnafraas	0	21	4	0	13	0	0	4	17	0	0	1
AJA-MÄÄRUS	jah	0	3	3	2	4	0	0	8	1	0	0	1
	ei	1	33	23	13	14	7	5	10	22	2	6	3
VIISI-MÄÄRUS	jah	0	2	4	0	2	0	0	13	1	0	0	1
	ei	1	34	22	15	16	7	5	5	22	2	6	3

ID-SILT	ID-SILDI TASE	mingis olukorras (metaf)	paigal olema	püsima (metaf)	püsima mingis olekus	seisnema	takistus	toetama	vastu seisma	ühe koha peal pikka aega	ühe koha peal püsti asumine	ühe koha peale jääma
KOHA-MÄÄRUS	jah	0	18	0	5	0	0	2	0	16	153	14
	ei	11	7	15	12	2	3	7	7	3	38	26
MÄÄRUSE-TÜÜP	asemäärsõna	0	1	0	1	0	0	0	0	1	9	3
	ei saa määrata	11	7	15	12	2	3	7	7	3	38	26
	kaassõnafraas	0	11	0	1	0	0	2	0	6	89	7
	nimisõnafraas	0	6	0	3	0	0	0	0	9	55	4
AJA-MÄÄRUS	jah	0	3	0	6	0	1	1	0	9	31	7
	ei	11	22	15	11	2	2	8	7	10	160	33
VIISI-MÄÄRUS	jah	3	3	4	12	0	1	0	0	0	51	2
	ei	8	22	11	5	2	2	9	7	19	140	38

Summary

The Polysemy of Estonian Posture Verb *seisma* 'Stand': Behavioral Profile and Cluster Analysis

The present MA thesis focuses on the Estonian polysemous posture verb *seisma* 'stand'. It is considered to be one of the cardinal posture verbs together with 'sit' and 'lie'. Although posture verbs refer to the posture of human beings their usage can be expanded to non-animate entities and to metaphoric uses.

The framework of the thesis is usage-based linguistics and the method used is corpus analysis, more precisely the *Behavioral Profile* method introduced by the corpus linguists Stefan Th. Gries and Dagmar Divjak in 2006.

This method is based on the assumption that distributional similarity also reflects functional similarity. To this end one needs to compile a representative sample of all the instances of a word lemma in a corpus and annotate different properties of the use of the word forms. These are referred to as ID-tags and comprise morphological, syntactic and semantic characteristics of the usage of the word in question.

The data used in the analysis is taken from the Estonian balanced corpus. The sample consists of 500 randomly selected sentences with the verb *seisma*. 17 different ID-tags were annotated, e.g. the sense of the word, agent, ground, tense, mood etc. These ID-tags were used in multifactorial hierarchical agglomerative cluster analysis to see how similar are the senses to each other and what kind of network is compiled based on the usage similarities. Cluster analysis provides a graphical image (dendrogram) of the semantic network a word may have.

In this research 23 different senses of *seisma* were identified and annotated in the sample and clustered by the multifactorial analysis.

As a result of clustering two clear cut clusters are formed on the basis of usage similarity. One cluster consists of senses mostly belonging to the physical domain like

standing of a human being or being located. The second major cluster involves abstract senses. For example, *to remain in some state* is used both in an abstract and concrete sense, but according to its behavioural profile it is more similar to the senses belonging to physical rather than the abstract domain.

In Estonian the meaning of the verb *seisma* has a strong component of motion. For example if in English the verb *stand* suggests prolonged inactivity (Newman, Rice 2004), in Estonian, it can be instead used to illustrate instantaneous activity (*A bus is standing in a bus stop*). It is also used together with the verb *jääma* 'stay', which is one of the core verbs of Estonian, as in *to stay stationary* (for a short period of time). This does not mean that the verb cannot be used to express prolonged inactivity as well. In that sense the semantic network of the Estonian *seisma* is wider than for example that of the English *stand*.

In the present data set a considerably larger number of abstract uses of the verb *seisma* were identified, e.g. *to preserve* (food, flowers) or *in some situation*. The richness of the abstract senses can be explained with the usage of adpositions used together with the verb.

Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks

Mina, Kristiina Kask

(sünnikuupäev 29.11.1988)

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) enda loodud teose

“*Seisma* verbi polüseemia: korpuspõhine käitumisprofiil ja klasteranalüüs”, mille juhendajad on Jane Klavan ja Ann Veismann

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 22.05.2014