

TARTU ÜLIKOOL  
HUMANITAARTEADUSTE JA KUNSTIDE VALDKOND  
EESTI JA ÜLDKEELETEADUSE INSTITUUT

Kertu Saul

EESTI KEELE LAUSEMALLIDE AUTOMAATNE TUVASTAMINE  
LIIGUTAMISVERBIDE NÄITEL

Magistritöö

Juhendajad kaasprofessor Kadri Muischnek ja vanemteadur Jelena Kallas

Tartu 2024

# SISUKORD

<b>SISSEJUHATUS.....</b>	<b>5</b>
<b>1. LAUSEMALLID TEOREETILISTES RAAMISTIKES.....</b>	<b>8</b>
1.1. Lausemall.....	8
1.2. Verbi laiendid.....	9
1.2.1. Seotud laiend.....	10
1.2.2. Vaba laiend.....	12
1.2.3. Argument.....	12
1.2.4. Adjunkt.....	14
1.2.5. Laiendi tüübi eristamine.....	14
1.3. Valents ja rektsioon.....	16
1.4. Lausemallid osana konstruktsioonigrammatikast ja konstruktikonidest.....	18
1.5. Semantilised klassid lausemallide vaatepunktist.....	20
1.6. Süntakiline märgendus Universal Dependencies raamistikus.....	22
<b>2. LAUSEMALLIDE KOOSTAMISE MEETODID.....</b>	<b>25</b>
2.1. Käsitsi koostamine.....	25
2.1.1. Lausemallide loendid ja sõnaraamatud.....	25
2.1.2. Elektroonilised ressursid.....	27
2.2. Automaatne tuvastamine.....	31
2.2.1. Uurimused eesti keele kohta.....	31
2.2.2. Uurimused teiste keelte kohta.....	32
<b>3. MATERJAL JA MEETOD.....</b>	<b>36</b>
3.1. Materjal.....	36
3.2. Meetod.....	40
3.2.1. Lausemallide käsitsi koostamine.....	40
3.2.2. Lausemallide automaatne tuvastamine.....	42
3.2.3. Hindamine.....	59
<b>4. TULEMUSED.....</b>	<b>63</b>
4.1. Liigutamisverbide lausemallide lauseliikmed.....	63
4.2. Verbiülesed lausemallid.....	66
<b>5. ANALÜÜS.....</b>	<b>75</b>
5.1. Kvaliteedi hindamine.....	75
5.2. Tuvastamata jäänud lauseliikmed ja lausemallid.....	79
5.2.1. Tuvastamata jäänud lauseliikmed.....	79
5.2.2. Tuvastamata jäänud lausemallid.....	83
5.3. Valesti tuvastatud lauseliikmed ja lausemallid.....	88

5.4. Uued lausemallid.....	95
<b>6. JÄRELDUSED.....</b>	<b>98</b>
6.1. Probleemid tuvastamata jäänud lauseliikmete ja -mallidega.....	98
6.2. Probleemid lausemallidest puuduvate lauseliikmetega.....	99
6.3. Probleemid valesi tuvastatud lauseliikmete ja -mallidega.....	100
6.4. Probleemid väljendverbidega.....	101
6.5. Probleemid meetodi filtriga.....	102
6.6. Probleemid märgendusvigadega.....	103
6.7. Probleemid käsitsi koostatud mallidega.....	104
<b>KOKKUVÕTE.....</b>	<b>105</b>
<b>LÜHENDITE JA MÄRGENDITE SELGITUSED.....</b>	<b>108</b>
<b>KIRJANDUS.....</b>	<b>111</b>
<b>AUTOMATIC DETECTION OF ESTONIAN ARGUMENT STRUCTURE CONSTRUCTIONS ON THE EXAMPLE OF CAUSED-MOTION VERBS. SUMMARY.....</b>	<b>125</b>
<b>LISAD.....</b>	<b>129</b>
Lisa 1. Käsitsi koostatud semantilised mallid.....	129
Lisa 2. Käsitsi koostatud semantiliste mallide grammatilised teisendused.....	131

## **Autorsuse kinnitus**

Kinnitan, et olen käesoleva lõputöö ise kirjutanud ning toonud korrektselt välja teiste autorite panuse. Töö on kirjutatud lähtudes Tartu Ülikooli eesti ja üldkeeleteaduse instituudi lõputöö nõuetest ning on kooskõlas heade akadeemiliste tavadega.

## SISSEJUHATUS

Lausemall on konstruktsioon, mis koosneb leksikaalselt väljendatud verbist ning sellega süntaktiliselt seotud abstraktselt väljendatud lauseliikmetest. Need lauseliikmed tulenevad selle verbi tähendusest ehk on selle seotud laiendid. (Rätsep 1978: 18) Verbist sõltuvalt on seotud laiendeid erinev arv ja need saavad esineda erinevates vormides (Erelt 2017a: 65; Karlsson 2002: 195). Varasemalt on eesti keele lausemallidega tegelenud Huno Rätsep oma raamatus “Eesti keele lihtlausete tüübid” (1978). Näites 1 on esitatud Huno Rätsepa koostatud lausemall, mis koosneb verbist *andma* ja selle neljast seotud laiendist: kolmest substantiivist ( $N^1$ ,  $N^2$ ,  $N^3$ ) ja ühest verbist ( $V^2$ ). Need saavad esineda ainult kindlates vormides:  $N^1$  nominatiivis,  $N^2$  nominatiivis, genitiivis või partitiivis,  $N^3$  allatiivis ja  $V^2$  *da*-infinitiivis.

(1)  $N^1$ +nom.  $V^1$   $N^2$  +ngp.  $N^3$  +all. ( $V^2$  +da).

$V^1$ =andma (*Peremees andis ühe toa külalistele kasutada.*) (Rätsep 1978: 161)

Nende lausemallide leidmiseks ei ole kasutatud andmepõhiseid meetodeid, vaid lähtutud ainult uurija enda keeletajust (Rätsep 1978: 6–8). Tänapäeva vahendid võimaldavad ag lausemalle leida automaatselt korpustest kogutud statistika alusel. Taolisi võimalusi on eesti keele jaoks kasutatud küll verbiga seotud lauseliikmete leidmiseks (Muischnek, Sahkai 2009; Orasmaa 2013), kuid kumbki neist ei seadnud eesmärgiks tuvastada terveid lausemalle, millele käesolev töö keskendub. Töö tulemust kasutatakse Eesti Keele Instituudi projekti PRG1978 “Uue aja sõnastik: grammatika ja keelepädevuse kirjeldamine integreeritud multifunktsionaalses leksikograafilises ressursis”<sup>1</sup> raames EKI ühendsõnastiku (Langemets jt 2021) grammatilise info esituse parandamiseks.

---

<sup>1</sup> <https://www.etis.ee/Portal/Projects/Display/66a57ac4-e0c6-40cd-b9df-148048ce9f95>

Kui Huno Rätsep võttis enda eesmärgiks kirjeldada kõiki lausemalle, mis eesti keeles võimalikud on (Rätsep 1978: 8), siis mina püüan oma töö praktilisest kaalutlusest tulenevalt leida just kõige sagedasemaid malle. Sellest oleks eriti kasu keeleõppijatele, kes näeksid sõnastikus vastavalt verbile, kuidas eesti keeles lauseid tavaliselt moodustatakse. Lausemalle on vaja ka eesti keele semantiliste rollide märgendaja väljatöötamiseks<sup>2</sup>. Semantiliste rollide määramine toetub süntaksile, kus verbi juurde kuuluvate laiendite ehk lausemalli osiste leidmine on selle esimene samm (Haverinen 2015: 913). Lausemallide leidmise osaline automatiseerimine hõlbustaks seetõttu semantiliste rollide märgendaja arendust.

Töö eesmärk on luua esmane töövoog lausemallide automaatseks tuvastamiseks ning saadud tulemuste põhjal välja selgitada kerkivad probleemid, et meetodit edasi arendada. Eesmärgist lähtudes on sõnastatud järgmised uurimisküsimused:

- Mis teoreetilist raamistikku lausemallide käsitlemiseks kasutada?
- Mis meetod sobib kõige paremini eesti keele lausemallide automaatseks tuvastamiseks?
- Mis kvaliteediga võimaldab meetod liigutamisverbide lausemalle automaatselt tuvastada?
- Mis on lausemallide automaatse tuvastamise peamised kitsaskohad?
- Mis on meetodi võimalikud edasiarendused?

Senisele kogemusele ja kirjandusele (Herbst 2014; Croft 2022) toetudes võib järeldada, et lausemallide automaatse tuvastamise peamiseks probleemideks kujuneb ilmselt seotud ja vaba laiendi eristamine, keeleliselt oluliste ebasagedaste lausemallide leidmine ja nende automaatse märgenduse vigadest eristamine.

---

<sup>2</sup> Vt nt <https://www.etis.ee/Portal/Projects/Display/66850b14-c97c-413b-9c08-4b9150728864>

Meetodi ja töövoogu töötan ma välja kõigi verbide lausemallide tuvastamiseks, kuid selle kvaliteedi hindamisel ja meetodi kitsaskohtade väljaselgitamisel võtan laialivalgumise ennetamiseks vaatluse alla ainult ühe semantilise grupi verbid: liigutamisverbid. Liigutamisverbidena käsitlen verbe, mille põhitähendus on liigutamine. Neile tuvastatakse lauseliikmete koosinemise sagedusele tugineva meetodi abil automaatselt lausemallid, mille lauseliikmed pannakse kirja morfoloogiliste ja süntaktiliste märgendite kombinatsiooniga. Andmeid kogutakse 15 miljoni sõna suurusest Tasakaalus korpusest<sup>3</sup>. Tulemuse kvaliteedi hindamiseks võrreldakse automaatselt tuvastatud lausemalle EDT korpuse<sup>4</sup> materjali alusel käsitsi koostatud liigutamisverbide mallidega, mille koostas Tartu Ülikooli EKTB75 projekti “Semantilise analüüsi baasvahendid“<sup>5</sup> raames. Võrdlusest saadud tulemuste põhjal pannakse kirja meetodi peamised kitsaskohad ning nende võimalikud lahendused.

Töö koosneb kuuest osast. Esimeses osas antakse ülevaade teoreetilisest taustast ja lausemallidega seotud mõistetest ning teises osas tööga sarnastest uurimustest. Kolmandas osas tutvustatakse materjali ja meetodit ning tuuakse välja nende piirangud. Neljandas osas esitatakse töö tulemused ja viiendas analüüsitakse neid, avades meetodi head küljed ja probleemkohad. Kuuendas osas võetakse analüüs kokku ning pakutakse probleemidele võimalikke lahendusi.

---

<sup>3</sup> <https://www.cl.ut.ee/korpused/grammatikakorpus/index.php?lang=et>

<sup>4</sup> [https://universaldependencies.org/treebanks/et\\_edt/index.html](https://universaldependencies.org/treebanks/et_edt/index.html)

<sup>5</sup> <https://www.etis.ee/Portal/Projects/Display/66850b14-c97c-413b-9c08-4b9150728864>

# 1. LAUSEMALLID TEOREETILISTES RAAMISTIKES

Lausemalle ja nendega seotud mõisteid iseloomustab suur variatiivsus eri teoreetiliste raamistike vahel. Töö üks eesmärkidest on leida lausemallide kirjeldamiseks kõige sobivam teoreetiline raamistik. Selle jaoks käsitlen siin peatükis erinevaid teooriad, võrdlen neid omavahel ning määratlen, kuhu sinse töö lausemallide käsitus nende teooriate sees paigutub.

## 1.1. Lausemall

Lause väljendab mingit sündmust, selle osalisi ning sellega seotud asjaolusid (EG 2023: 471). Süntaktiliselt väljendatakse neid lauseliikmete abil. Lauseliikmed on üksteisega alistus- või rinnastussuhetes. Alistussuhe koosneb mitmest moodustajast, kus alistavat kutsutakse põhjaks ning sellele alistuvaid laienditeks. (Erelt 2017a: 64) Lausemalli põhjaks on verb, sest see määrab kogu lause omadused (EG 2023: 472), selle hulgas ka teised lauses esinevad lauseliikmed ning nende vormi (Rätsep 1978: 11). Lausemalli osaks arvestataksegi selliseid verbile alistuvad laiendid, mille olemasolu või vorm tuleneb selle verbi tähendusest. Neid lauseliikmeid kutsutakse seotud laienditeks. Lausete üldistamiseks esitatakse seotud laiendeid lausemallis abstraktselt. Niisiis on lausemall konstruktsioon, mis koosneb leksikaalselt väljendatud verbist ning tema tähendusest sõltuvatest abstraktselt väljendatud lauseliikmetest. (Rätsep 1978: 15–18) Näites 2 on toodud lausemalli näitena Huno Rätsepa koostatud lausemall verbile *panema*, mis koosneb kolmest lauseliikmest: subjekti rollis nimisõnast nominatiivis, määruse rollis nimisõnast illatiivis ja objekti rollis nimisõnast, mis võib olla nii nominatiivis, genitiivis kui partitiivis.

(2)  $N^1 + \text{nom. V } N^2 + \text{ill. } N^3 + \text{ngp.}$

V = panema (*Parun pani ettevõttesse kogu raha.*) (Rätsep 1978: 159)

Kuna lausemalli liikmed sõltuvad verbi tähendusest, saab neid väljendada ka tähenduse tasandil semantiliste rollide kaudu (Goldberg 1995: 50; Hilpert 2014: 26–28). Semantilised rollid on suhted tegevuse ja selle osaliste vahel (Pajusalu 2009: 81). Näites 3 on esitatud verbiga *andma* seotud semantilised rollid.

(3) *andma*: AGENT, PATSIENT, SAAJA (Allerton 1982: 52)

Vahel esitatakse lausemallides koos nii süntaktiline info kui ka semantilised rollid (Herbst 2014: 182). Näites 4 on korraga esitatud süntaktilised rollid SCU ehk subjekt, PCU1 ehk objekt ja PCU2 ehk määrus, fraasiliik NP ehk nimisõnafraas ja semantilised rollid ANDJA, ANTAV ASI ja SAAJA (Herbst 2014: 183).

(4) [SCU: NP “ANDJA”] *andma* [PCU1: NP “ANTAV ASI”] [PCU2: NP “SAAJA”]

*Sa võid selle talle anda.*

Siin magistritöös käsitletakse lausemalle nii süntaktilise kui semantilise tasandi koosmõjuna kui ka ainult (morfo)süntaktiliselt.

## 1.2. Verbi laiendid

Lausemallide liikmeid defineeritakse olenevalt teoreetilisest käsitlusest ja keele tasandist erinevalt. Nende väljendamiseks kasutatakse mõisteid seotud laiend, komplement, vaba laiend, atribuut, argument, adjunkt, osalejaroll ja kõrvalroll. Eri allikad defineerivad neid mõisteid aga erinevalt. Järgnevas peatükis üritan anda neist ülevaate ning piiritleda, kuidas neid siin töös kasutatakse.

### 1.2.1. Seotud laiend

Seotud laiend on süntaksi tasandi mõiste (Erelt 2017a: 65). See on süntaktiliste lauseliikmete abil väljendatud laiend, mis tuleneb põhja tähendusest ehk selle esinemine on põhjast tingitud. Näites 5 on lause põhjaks verbivorm *luges*, mille seotud laiendid on subjekt *poiss* ja objekt *raamatut*, sest verb *lugema* eeldab oma tähendusest tulenevalt nii lugejat kui ka seda, mida loetakse. Eesti keeles on subjekt ja objekt alati seotud laiendid, kuid määrused saavad samuti seotud laiendid olla. Näiteks on näites 6 seotud laiend lisaks subjektile *Kultuuriministeerium* ka määrus *Tallinnas*, sest asumist ei saa väljendada ilma kohata. (EG 2023: 483)

(5) *Poiss luges raamatut.*

(6) *Kultuuriministeerium asub Tallinnas.* (EG 2023: 483)

Selle nähtuse väljendamiseks kasutatakse eri allikates mitmeid erinevaid mõisteid. Eesti grammatikakäsitlustes kutsutakse seda kõige sagedamini seotud laiendiks (Rätsep 1978: 16; EG 2023: 483; Erelt 2017a: 65). Saksa süntaksis domineerivad valentsiteoorias kutsutakse seda komplemendiks (Herbst 2014: 170), mille kaudu on see jõudnud ka eestikeelsetesse süntaksikäsitlustesse (EG 2023: 483; Erelt 2017a: 65; Karlsson 2002: 187). Sõltuvussüntaktilise automaatse analüüsi ehk parsimisega tegutsevas Universal Dependencies raamistikus (vt ptk 1.6) kutsutakse seda argumendiks (Nivre jt 2016: 1661) ja sõltuvusgrammatika teoorias aktandiks (Tesnière 2015: 97; Karlsson 2002: 187). Komplemendiks kutsutakse vahel ka neid seotud laiendeid, mis ei ole subjektid (Przepiórkowski, Patejuk 2018: 3837). Siinses töös selliseid eristusi ei tehta ning kasutatakse süntaktilise tasandi verbist sõltuva laiendi jaoks ainult terminit *seotud laiend*, kuna see on eesti keele senistes süntaksikäsitlustes kõige levinum.

Seotud laiendid võivad olla kõigil põhjana esinevatel lauseliikmetel: verbidel, nimisõnadel, omadussõnadel, kaassõnadel ja hulgasõnadel (EG 2023: 483; Herbst 2014:

168), kuid eelkõige siiski verbidel (EG 2023: 483). Selles töös keskendutakse verbide seotud laienditele, sest verb on lausemalli põhi (EG 2023: 472).

Seotud laiendid jaotuvad vabatahtlikeks, kontekstist sõltuvateks ja kohustuslikeks (Herbst 2014: 173), Rätsepa käsitluses fakultatiivseteks ja obligatoorseteks (Rätsep 1978: 16). Vabatahtlik ehk fakultatiivne seotud laiend saab jääda mainimata, kui ilma selleta on lause samuti tähenduslik (Herbst 2014: 173; Rätsep 1978: 16). Selline on näiteks verbi *pesema* objekt, mis saab vabalt lausest puududa, kui lause tähelepanu keskmes on verb ise (EG 2023: 618).

(7) Lähen (*ennast*) pesema.

Kontekstist sõltuv laiend on selline seotud laiend, mida on varem diskursuses mainitud, mistõttu ei pea see lauses avalduma (Herbst 2014: 173). Eesti keeles on näiteks verbe, mille puhul on objekti väljendamine kohustuslik, aga see saab lausest siiski vormiliselt puududa, kui see on konteksti põhjal tuletatav. Selline on näiteks verb *sisaldama*. (EG 2023: 617–618)

(8) Kas see toit sisaldab pähkleid? *Sisaldab küll.*

Seotud laiend on kohustuslik ehk obligatoorne, kui seda peab igal juhul valentsi kandjaga koos mainima (Herbst 2014: 173; Rätsep 1978: 17). Sellised on näiteks verbi *riputama* subjekt, objekt ja sihtkoht, sest nende mainimata jätmine jätab kontekstivaba *riputama*-verbiga lause ebagrammatiliseks (Rätsep 1978: 17).

(9) *Daam riputas mantli konksu otsa.*

\**Daam riputas*; \**Daam riputas mantli*; \**Daam riputas konksu otsa*; \**Riputas*;  
\**Riputas mantli*; \**Riputas mantli konksu otsa.*

Rätsep ei eristanud kontekstist sõltuvaid ja kohustuslikke laiendeid, sest ta käsitles oma teoses ainult kontekstivabu lauseid (Rätsep 1978: 71).

### 1.2.2. Vaba laiend

Vabad laiendid on laiendid, mis ei kuulu põhja tähendusstruktuuri (Erelt 2017a: 65). Nende olemasolu ega vorm ei sõltu verbist, vaid lause sisust. Seetõttu saavad need esineda pea iga verbiga. (Rätsep 1978: 15) Vabad laiendid väljendavad sündmuse asjaolusid, mistõttu on enamik määrusi tavaliselt vabad laiendid (EG 2023: 484–486). Kõige tüüpilisemad vabad laiendid on aja, koha ja viisimäärused (Karlsson 2002: 187). Vabade laiendite üheks tunnusjooneks peetakse seda, et nende eemaldamisel lausest ei muutu see ebagrammatiliseks (Herbst 2014: 170–171). Näiteks on näites 10 ajamäärus *eelmisel aastal* vaba laiend, sest see ei tulene lõpetamise tähendusest ning selle eemaldamisel jääb lause grammatiliseks (EG 2023: 484).

(10) Ta lõpetas *eelmisel aastal* ülikooli. (EG 2023: 484)

Vaba laiendit kutsutakse ka atribuudiks (Karlsson 2002: 185–187), adjunktiks (Herbst 2014: 170) ja *circumstandiks* (Tesnière 2015: 97). Eesti grammatikakäsitlustes on kõige laialdasemalt kasutusel mõiste vaba laiend (Rätsep 1978: 15; Erelt 2017a: 65; EG 2023: 483; Karlsson 2002: 187), mistõttu kasutatakse seda ka siinses töös.

### 1.2.3. Argument

Eesti grammatikakäsitluses loetakse semantilise tasandi mõisteteks argumenti ja adjunkti (Erelt 2017a: 69). “Eesti keele süntaksi” ja “Eesti grammatika” kohaselt on argumendid semantilise predikaadi seotud laiendid, mis avalduvad peamiselt tegevussubjekti ja tegevusobjekti rollis. Need võivad aga avalduda ka erinevate semantiliste rollidena. (Erelt 2017a: 69; EG 2023: 486–487) Renate Pajusalu õpikus “Sõna ja tähendus” (2009) käsitletakse argumente ainult semantiliste rollide kandjatena. Seal peetakse argumentideks kõiki predikaadi laiendeid, sealhulgas ka neid, mida

süntaksikäsitluses peetakse vabadeks laienditeks. Küll aga eristatakse selle põhjal semantilisi rolle osalejarollideks ja kõrvalrollideks (Pajusalu 2009: 81–82). Fred Karlssoni õpikus “Üldkeeleteadus” (2002: 182–187) kasutatakse argumenti aga valentsiga seoses kohustusliku laiendi tähenduses ning mainitakse, et taoline abstraktne lauseliikmete kirjeldus on viis süntaksit keelteüleselt kirjeldada.

Nii sõltuvusgrammatikas, konstruktsioonigrammatikas kui ka fraasistruktuuri-grammatikas defineeritakse argumenti kui predikaadi otsest alluvat, mis on semantilisel ja süntaktilisel tasandil vajalik selle tähenduse väljendamiseks (Tesnière 2015: 100; Croft 2001: 272; Barbu, Toivonen 2016: 1961). Neis raamistikutes jaguneb argument omakorda süntaktiliseks ja semantiliseks. Semantiline argument on predikaadi väljendatud tegevuse osaline. Neid saab väljendada läbi semantiliste rollide. (Ackema 2015: 247–248) Sellisena on defineeritud argumenti ka eesti keele semantiliste rollide märgendaja arendamise projektis (EKTB75 “Semantilise analüüsi baasvahendid“) kasutatav PropBanki raamistik (Palmer jt 2005). Süntaktiline argument on nende semantiliste rollide väljendamine süntaktiliste lauseliikmetena (Ackema 2015: 247–248). Konstruktsioonigrammatika seisukohalt kasutatakse selles tähenduses mõistet argumendifraas (Croft 2022: 175). Eesti keele terminoloogias vastab selline argumendikäsitlus seotud laiendile.

Semantilise ja süntaktilise tasandi mõistete selgeks eristamiseks kasutatakse selles töös mõistet argument seotud laiendi semantilise tasandi vastena, mida väljendatakse semantiliste rollide abil. Eesti grammatikakäsitlustes vastavad nendele kõige paremini Renate Pajusalu raamatus “Sõna ja tähendus” (2009: 82) kasutatud osalejarollid, mis on defineeritud kui rollid, mis lähtuvad predikaadi olemusest ning on vajalikud predikaadi kujutatud sündmuse väljendamiseks. Muudes grammatikateooriates vastab sellele kõige paremini semantilise argumendi mõiste, aga kuna eesti käsitlustes on süntaktilise argumendi jaoks juba kasutusel oma termin seotud laiend, kutsutakse selle semantilise tasandi mõistet siin töös lihtsalt argumendiks. Töös ei kasutata selle jaoks mõistet

osalejaroll vaid argument, et paremini ühilduda projektidega, kus töö tulemusi kasutatakse.

#### **1.2.4. Adjunkt**

Adjunkti mõiste ei ole eestikeelses grammatikakäsitluses seni kasutusel, kuid konstruktsioonigrammatikas, fraasistruktuurigrammatikas ja PropBanki raamistikus kasutatakse seda argumendi vastandina ehk vaba laiendi semantilise vastena (Croft 2001; Ackema 2015; Barbu, Toivonen 2016; Palmer jt 2005). Nende käsitluste kohaselt on adjunktid lause mittekohustulikud liikmed (Croft 2001: 272), mis iseloomustavad sündmuse asjaolusid. Nendeks on näiteks sündmuse toimumise aeg, koht, viis, põhjus jne. (Ackema 2015: 259) Eesti semantikakäsitluses kutsutakse neid sündmuse kõrvalrollideks (Pajusalu 2009: 87). Adjunkti ei kasutata aga alati semantilises tähenduses, vaid vahel vastandatakse seda ka valentsiteoorias komplemendile ehk seotud laiendile (Herbst 2007: 15; Herbst 2014: 170). Niisiis kasutatakse seda ka vahel vaba laiendi tähenduses.

Siinses töös kasutatakse adjunkti vaba laiendi semantilise tasandi vastena. Töös ei kasutata eesti terminoloogias defineeritud mõistet kõrvalroll, et paremini ühilduda nende projektidega, kus töö tulemusi kasutatakse.

#### **1.2.5. Laiendi tüübi eristamine**

Verbi tegevuse osaliste eristamine seotud ja vabadeks laienditeks ning argumentideks ja adjunktideks ei ole aga alati lihtne. Seda tehakse tavaliselt põhimõttel, et mida tähtsam lauseliige predikaadi tähenduse väljendamiseks on, seda seotum see on. Tähtsaimaid on tavaliselt need tegevuse osalised, kes tegevuse algatasid ja keda tegevus kõige rohkem mõjutas. Seetõttu on tüüpiliselt kõige seotum subjekt, tähtsuselt teine on objekt ja neile järgneb obliikva. (Croft 2022: 173) Semantilistest rollidest on argumendid tüüpiliselt tegijarollid (AGENT, AUTOR, INSTRUMENT), sündmuse vastuvõtja rollid (PATSIENT, KOGEJA, SAAJA) ja ruumirollid (TEEMA, LÄHTEKOHT,

SIHTKOHT). Adjunktid on tüüpiliselt KOHT, PÕHJUS, EESMÄRK, AEG, VIIS ja TEE. (Pajusalu 2009: 82–87)

Seotud ja vaba laiendit eritatakse tihti põhimõttel, et seotud laiendid on lause kohustuslikud liikmed ja vabad laiendid valikulised liikmed (EG 2023: 484). Siiski on selle abil seotud ja vaba laiendit keeruline kindlapiirilisel eristada, sest vahel võivad tüüpiliselt valikulised liikmed nagu kohamäärus (11) ja viisimäärus (12) olla ka kohustuslikud ning verbi väljendatud tegevusega tihedalt seotud osalised olla lausest välja jäetud (13) (Croft 2022: 173–174; EG 2023: 483–484).

(11) Lamp asetseb *laual*. – \*Lamp asetseb. (Rätsep 1978: 52)

(12) Ta käitub *halvasti*. – \*Ta käitub. (EG 2023: 484)

(13) Ta suitsetab *sigaretti* – Ta suitsetab. (EG 2023: 484)

Üks viis seotud ja vabade laiendite eristamiseks on toetuda nende grammatilisele markeeritusele. Seda tehakse tihti morfoloogilise markeerituse kaudu. (Croft 2022: 186) Morfoloogilise markeerituse teooria alusel on lauseliikmed seda tähtsamad, mida lihtsamalt nad on kodeeritud (Karlsson 2002: 143). See tähendab, et kui sõnal on morfoloogiliselt markeerimata vorm, on see suure tõenäosusega seotud laiend (Croft 2022: 186–188). Eesti keeles on markeerimata ainsus, nominatiiv, lühike sisseütlev ehk aditiiv, algvõrre, olevik, kindel kõneviis ja ainsuse kolmas isik. Vahel on markeerimata ka genitiiv ja partitiiv. (Karlsson 2002: 142) Selle kohaselt peaksid eesti keeles laiendid olema suure tõenäosusega seotud, kui need on nominatiivis, genitiivis, partitiivis või aditiivis. Siiski ei kehti see aditiivi puhul, sest see on lühike vorm illatiivist, mille moodustamise võimalikkus sõltub sõna morfoloogilistest tunnustest ja tähendusest (EKK 2020: 217).

Teine viis grammatilise markeerituse abil laiendite seotust tuvastada on sõnajärje abil (Croft 2022: 186). Keele tüüpiline sõnajärg näitab ära lauseliikmete hierarhia.

Hierarhias eespool olev element on suurema tõenäosusega ka seotud laiend. (Croft 2022: 191) Eesti keeles on täheldatud aga vastupidist: mida kaugemal lauseliige öeldisverbist on, seda tugevam on selle seos verbiga. Siiski ei saa eesti keele puhul seotud ja vabade laiendite eristamiseks sellest lähtuda, sest eesti keele sõnajärg on suhteliselt vaba ning oleneb pigem lause infostruktuurist kui laiendi seotusest. (Lindström 2017: 547–548)

Ei ole olemas raudkindlat testi seotud ja vabade laiendite eristamiseks. Seetõttu saab väita, et seotud ja vabad laiendid moodustavad pigem skaala kui on absoluutsed vastandid. (Herbst 2014: 171) Siinses töös määratakse seotud laienditeks lauseliikmed, mis on tihedalt seotud verbi tähendusega ja esinevad korpuses küllalt sagedasti. Kriteeriumitest rohkem peatükis 3.2.1.

### **1.3. Valents ja rektsioon**

Lausemalliga on tihedalt seotud mõisted valents ja rektsioon (Rätsep 1978: 61–68). Verbi valents väljendab seda, kui palju seotud laiendeid verbil on (Erelt 2017a: 65). Seotud laiendist saab mõelda kui valentsi lünga realisatsioonist (Herbst 2014: 168).

On olemas nullvalentsed, ühevalentsed, kahevalentsed, kolmevalentsed ja neljavalentsed verbid (Rätsep 1978: 242). Tüüpiliselt ilmastikunähtusi väljendavatel nullvalentsiga verbidel pole ühtegi seotud laiendit (14). Ühevalentsetel verbidel on ainult üks seotud laiend (tavaliselt subjekt) (15), kahevalentsetel kaks seotud laiendit (tavaliselt subjekt ja objekt) (16), kolmevalentsetel kolm seotud laiendit (tavaliselt subjekt, objekt ja obliikva) (17) (Karlsson 2002: 189) ja neljavalentsetel neli seotud laiendit (18) (Rätsep 1978: 242).

(14) Müristab.

(15) *Kõneleja* aevastas.

(16) *Peremees tappis sea.*

(17) *Ülemus pakkus talle paremat palka.* (Karlsson 2002: 189)

(18) *Ma muretsesin Peetrile Jürilt paremad konspektid.* (Rätsep 1978: 242)

Valentsi mõistega on tihedalt seotud transitiivsus. Transitiivsuse kaudu väljendatakse seda, kas verbi üks seotud laienditest on objekt või mitte. Verbid, mis ei nõua objekti, on intransitiivsed ning verbid, mis nõuavad, on transitiivsed. (Karlsson 2002: 188–189) Kahte objekti nõudvaid verbe nimetatakse ditransitiivseteks, aga neid on eesti keeles väga vähe (EG 2023: 618). Tüüpiliselt on ühevalentsed verbid intransitiivsed ning kahe- ja kolmevalentsed transitiivsed (Karlsson 2002: 188–189).

Eesti keeles saab paljusid transitiivseid verbe kasutada ka ilma objektita. Selline on näiteks verb *sööma* (19) (Metslang 2017a: 261). Transitiivsete verbidega, millel on objekti väljendamine kohustuslik, saab objekt lausest siiski vormiliselt puududa, kui see on konteksti põhjal tuletatav (20) (EG 2023: 618).

(19) *Sõin vaikselt köögis.* (Metslang 2017a: 261)

(20) "Kas sa ostsid värvi?" Naine noogutas. "Jah, ostsin küll," ütles ta. (ÜK 2021)

Rektsioon ehk sõltumine väljendab seda, mis vormi verb oma seotud laienditelt nõuab (Karlsson 2002: 195). Laiendi vormi määrab sel juhul verbi tähendus või selle leksikaalne kuju. Rektsioonist mõjutatud vormid on käandsõna kääne (21), käandsõnaga seotud kaassõna (22), verbi käändeline vorm (23) ja lause moodustaja sidend (24) (EG 2023: 484).

(21) *kohanes linnaeluga*

(22) *uhke poja üle*

(23) *tahab laulda*

(24) *Noored räägivad, et elu on ilus.* (EG 2023: 484)

Kuna sama tähendust saab edasi anda mitme vormiga, ei tingi verb tavaliselt mitte üht spetsiifilist vormi, vaid sarnaste tähendustega vormide klassi. Näiteks tingib verb *osalema* oma tähendusest tulenevalt selle, milles osaletakse, aga seda saab väljendada nii nimisõnalise määrusega adessiivi (25), nimisõnalise määrusega inessiivi (26), kaassõnaga (27) kui ka mäarsõnalise määrusega (28). (EG 2023: 485)

(25) Mari osales *talgutel*.

(26) Mari osales *arutelus*.

(27) Mari osales *veiniteo juures*.

(28) Mari osales *kõikjal*. (EG 2023: 485)

Vabu laiendeid ei käsitleta rektsioonidena, sest nende vorm ei ole mitte mõjutatud verbist, vaid laiendist endast ja konstruktsioonist, milles see esineb (EG 2023: 485).

#### **1.4. Lausemallid osana konstruktsioonigrammatikast ja konstruktikonidest**

Konstruktsioon on tähenduse ja vormi paar ehk kombinatsioon sellest, mida kõneleja väljendada tahab ja kuidas ta seda teeb (Croft 2022: 17–19). Konstruktsioonigrammatika alusel on vorm ja tähendus omavahel seotud (Boas 2013: 234), mistõttu väljendatakse eri tähendusi väljendavaid konstruktsioone tihti skemaatiliselt vormide kujul (Croft 2022: 3). Lausemall on üks konstruktsiooni liikidest (Rätsep 1978: 18). Seda kutsutakse ka argumentstruktuuri konstruktsiooniks (Goldberg 1995; Croft 2022: 45).

Konstruktsioonid erinevad oma skemaatilisuses. Need moodustavad kontiinumi, kus ühes otsas on täiesti leksikaalsed konstruktsioonid (sõnad, idioomid) ja teises otsas täielikult abstraktsed konstruktsioonid. Lausemallid on selles kontiinumis pigem

abstraktsemas otsas, sest need koosnevad leksikaalselt väljendatud verbist ja abstraktselt väljendatud laienditest. (Hoffmann 2013: 307–308)

Konstruksioone eristatakse ka selle järgi, kas need koosnevad ühest või mitmest osast. Ühest osast koosnevaid konstruktsioone kutsutakse atoomilisteks ning mitmest osast koosnevaid kompleksseteks. (Hoffmann 2013: 307–308) Lausemallid on peamiselt kompleksed, aga kuna esineb ka nullvalentsiga verbe (nt *müristama*, *sadama*) on lausemallid vahel ka atoomilised, sest koosnevad ainult verbist (Rätsep 1978: 77).

Komplekssete konstruktsioonide osad võivad olla ka teised konstruktsioonid ehk konstruktsioonid saavad üksteise sees esineda (Croft 2022: 27). Näiteks peetakse lauset omaette konstruktsiooniks. Lause koosneb tüüpiliselt verbina avalduvast predikaadist ja selle seotud laienditest. Seotud laiendit käsitletakse konstruktsioonigrammatikas samuti konstruktsioonina, seda kutsutakse argumentfraasiks. (Croft 2022: 171) Argumentfraas koosneb tüüpiliselt nimisõnast ja seda iseloomustavatest täiendfraasidest, mis on täienditest koosnevad konstruktsioonid (Croft 2022: 44). Lisaks saab konstruktsioonideks pidada ka üksikuid sõnu või lausa morfeeme (Croft 2001: 17). Niisiis saab ühe lause sees olla peidus mitmeid konstruktsioone, mis on üksteisest sõltuvad (Croft 2022: 44).

Konstruksiooni tähtsaim liige on pea ehk ülemus, mis on konstruktsiooni funktsiooni kõige paremini edasi andev element. See, mis sõna konstruktsiooni pea on, oleneb sellest, mida see konstruktsioon rõhutab. Näiteks fraasi *see puu, mis suri* pea on *puu*, sest fookuses on puu ise ja suremine annab selle kohta lihtsalt lisainfot, aga fraasi *puu suri ära* pea on *suri*, sest fraasi fookuses on suremise tegevus. Konstruktsiooni elemente, mis pole pead, kutsutakse alluvateks. Alluvad on peaga tähenduse kaudu seotud. (Croft 2022: 36–37) Lausemallide puhul on konstruktsiooni peaks verb (Croft 2022: 40).

Konstruksiooniline lausemallide käsitus erineb muudest käsitlustest peamiselt selle poolest, et kui varasemates teooriates määras verb üksi lause süntaktilise struktuuri ja semantilise funktsiooni, siis konstruksioonigrammatika teooria kohaselt eksisteerivad ka verbiülesed argumentstruktuuri konstruksioonid, mis ei sõltu mitte verbi tähendusest, vaid lause üleüldisest vormist, mis annab edasi mingit tähendust. See tähendab, et lausemallid ei eksisteeri mitte üksi vaakumis, vaid on osa suuremast konstruksioonide süsteemist, mis lausete struktuuri mõjutavad. (Boas 2013: 235–236; Goldberg 1995: 224) See annab seletuse nähtusele, kus verbi transitiivsus ja valents ei lähe alati kokku verbi reaalse kasutusega. Näiteks on *tapma* tüüpiline transitiivne kahevalentne verb, mis nõuab tavaliselt nii TAPJA kui ka TAPETAVA esitamist (29). Siiski saab seda kasutada ka nii, et objekti ehk TAPETAVAT ei spetsifitseerita (30). Seda seepärast, et see on mõjutatud teisest, väljendamata objekti konstruksioonist (*deprofiled object construction*). (Boas 2013: 239)

(29) Kiievi loomaaias tappis *tiiger* tema puuri sattunud *koristaja*. (EKI ühendsõnastik 2024)

(30) *Tiigrid* tapavad vaid öösiti. (Goldberg 2001: 506)

Argumentstruktuuri konstruksioone esitatakse samuti nagu lausemalle süntaktilise ja semantilise struktuuri vastavustena. Süntaktiliselt esitatakse neid eri täpsuse tasanditel, näiteks süntaktiliste rollidena, fraasiliikidena, osalausete või sekundaartarindite tüüpidenä. (Goldberg 2006: 20; Herbst 2014: 195).

## 1.5. Semantilised klassid lausemallide vaatepunktist

Magistritöös käsitletakse ühe semantilise klassi, liigutamise seotud verbe. Verbe jagatakse semantilistesse klassidesse nende tähenduse ja lausemallide alusel. See tähendab, et sama semantilise klassi verbidel kipub olema ka sarnane argumentstruktuur. (Tragel, Taremaa 2019: 212) Taolise semantilise liigitamise kohta on

üks põhjapanevamaid teoseid Beth Levini raamat “English Word Classes and Alternations” (1993). Eesti keele jaoks selline põhjalik ülevaade puudub, kuid eesti grammatikakäsitlustes kasutatakse verbide semantilisse klassi liigitamiseks Leviniga sarnast põhimõtet ehk lähtutakse samuti verbi lausemallist. Lausemallist lähtumine on ka üks viis polüseemsete verbide semantiliste klasside eristamiseks. (Tragel, Taremaa 2019: 212) Protsess toimib aga ka vastupidiselt: verbide semantilisi klasse kasutatakse ka eesti keele süntaksi, sõnamoodustuse ja morfoloogia seletamiseks, mis viitab samuti sellele, et lause semantiline ja süntaktiline tasand on omavahel tihedalt seotud. (Tragel, Taremaa 2019: 208–209)

Verbe liigitatakse semantilistesse klassidesse erinevalt. Üks laialt levinud meetod on määrata verb semantilisse klassi selle põhitähenduse järgi, toetudes kas uurija enda või lingvistilise katse raames suurema keelekogukonna intuitsioonile. Teine variant on lähtuda verbi kasutuskontekstist ehk toetuda korpusele ja kasutada statistilisi meetodeid. (Tragel, Taremaa 2019: 210–211)

Kuigi eesti keele jaoks ei ole ülevaatlikku semantiliste klasside uuringut nagu Beth Levini oma (1993), on semantiliste klasside ja nende lausemallide kohta tehtud väiksemahulisemaid uuringuid. Liikumisega seotud verbe, sealhulgas liigutamisverbe, on uurinud Pajusalu jt (2013) artiklis “The encoding of motion event in Estonian”. Uuritavad verbid saadi eesti Wordneti<sup>6</sup> kaudu verbide *liikuma* ja *liigutama* alammõisteid ekstraheerides (Pajusalu jt 2013: 48–49). Artikkel toob nende verbidega välja kaheksa semantilist rolli, millega liikumisega seotud verbid peamiselt esinevad. Need on kolm üldisemat rolli: AGENT (liikuja), PÕHJUS (liikumise algataja) ja OBJEKT (liigutatav), ning viis ruumirolli: LÄHTEKOHT, SIHTKOHT, TEEKOND, ASUKOHT ja KADUMINE. Artiklis tuuakse välja ka see, kuidas neid rolle eesti keeles morfoloogiliselt ja süntaktiliselt väljendatakse. (Pajusalu jt 2013: 44) Analüüs näitas, et ruumirollide väljendamiseks kasutatakse kõige rohkem kolmemõõtmelisi kohakäandeid

---

<sup>6</sup> <https://www.cl.ut.ee/ressursid/teksaurus/>

ning kõige sagedamini väljendatakse ruumirollidest SIHTKOHTA (Pajusalu jt 2013: 64).

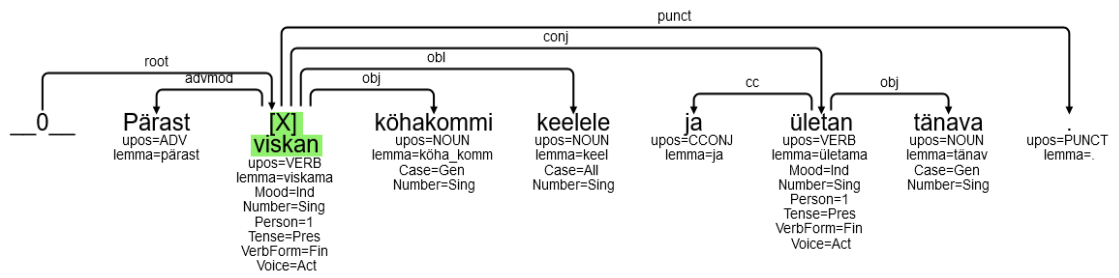
## 1.6. Süntaktiline märgendus Universal Dependencies raamistikus

Siinses magistritöös kasutatakse materjalina Universal Dependencies raamistikus annoteeritud korpusi. Universal Dependencies (edaspidi UD) on teksti keelteüleseks morfoloogiliseks ja sõltuvussüntaktiliseks annoteerimiseks mõeldud raamistik. UD süsteem kasutab oma teoreetilise raamistikuna sõltuvusgrammatikat, kus osalused, fraasid ja neis sisalduvad sõnad on üksteisega sõltuvussuhete kaudu hierarhiliselt seotud. (Marneffe jt 2021: 255–257) Sõltuvussuhted avalduvad märgenditena esitatud süntaktiliste funktsioonidena: objekt ehk *obj*, käändsõnaline subjekt ehk *nsubj* jne (Marneffe jt 2021: 265–266). Fraaside ja (osa)lausete tähtsaimat liiget ehk tuumsõna kutsutakse peaks ning sellest sõltuvaid liikmeid alluvateks (Marneffe jt 2021: 257). Eesti keeles kutsutakse tuumsõna ka ülemuseks (Muischnek, Müürisep 2017: 124).

Lause hakkab hargnema selle kõige tähtsamast elemendist ehk juurtipust, mis on sündmust väljendav predikaat. Selle alluvad saavad olla nii käändsõnafraasid, määrsõnafraasid kui ka kõrvallused ja lauselühendid. Kuna kõigil lause osalistel peavad olema ülemused, määratakse ülemused ka kirjavahemärkidele ja muudele mitte-keelelistele elementidele. Ülemus-alluv suhetena esitatakse ka selliste sõnaühendite sisemisi suhteid, kus sõltuvussuhet tegelikult ei ole. Nii näiteks on verbi liitvormi “ülemuseks” selle täistähenduslik komponent: nt fraasi *ei oleks pidanud tegema* ülemus on supiivorm *tegema*, millele alluvad kõik teised perifrastilised verbivormi osad. Sõltuvussuhete näitamiseks on juurtipp otseselt ühendatud aga ainult osalause ja fraaside peasõnaga. Selliseid lauseliikmeid kutsutakse ka pea otsesteks alluvateks. (Marneffe jt 2021: 257; Marneffe jt 2014: 4586) Igal otsesel alluval saavad olla ka enda alluvad ning neil alluvatel oma alluvad, millest kokku moodustub hargnev puustruktuur (Marneffe jt 2021: 257).

Lisaks süntaksile on iga sõna kohta toodud ka selle morfoloogiline info ja muu edasise automaatanalüüsi jaoks vajalik lisainfo, näiteks kas sõna on võõrsõna või ortograafiaviga (Marneffe jt 2021: 259). Eesti keeles kuulub morfoloogilise info alla sõna lemma, sõnaliik ja morfoloogilised kategooriad. Morfoloogilistest kategooriatest kasutatakse eesti keele UD süsteemis märgendamiseks märgendeid nii käandsõnade kui verbide iseloomustamiseks. Käandsõnade jaoks kasutatakse eesti keeles märgendeid käände, arvu, omadussõnade võrdlusastme, arvsõna liigi ja asesõna tüübi näitamiseks. Kaassõnadel on eraldi märgend kaassõna tüübi näitamiseks. Verbidel märgendatakse arvu, isikut, aega, tegumoodi, kõneviisi, eitust, verbi eitava liitvormi osa ja verbivormi (finiitne, da-infinitiiv, supiin, partitsiip, konverb). (Muischnek, Müürisep 2022) Lausemallide määramiseks on neist olulised kääne, verbivorm, kõneviis ja tegumood, sest verb määrab oma laiendite käände ja verbivormi (Rätsep 1978: 40) ning verbivorm, kõneviis ja tegumood mõjutavad lausemallis teiste lauseliikmete esinemist (Erelt jt 2017: 252–253; Metslang 2017b: 529).

Joonisel 1 on kujutatud UD süsteemis märgendatud süntaksipuu. Seal on sõltuvussuhted näidatud joontega, kus joon algab peast ning noole tipp lõpeb selle otsese alluva juures. Selle puu juurtipp on verbivorm *viskan*, mille süntaktilise funktsiooni (*root*) näitamiseks luuakse UD süsteemis sellele fiktiivne pea (*\_0\_*). (Marneffe jt 2021: 265) Juurtipu juurest lähevad jooned otse sõnade *pärast*, *kõhakommi*, *keelele*, rindlause peasõna *ületan* ja lauselõpupunkti juurde. Need sõnad on juurtipu otsesed alluvad. Iga joone peal on kirjas selle alluva süntaktilise funktsiooni märgend. Näiteks on sõnu *viskan* ja *kõhakommi* ühendava joone peal näha märgendit *obj*, mis näitab, et *kõhakommi* on verbivormi *viskan* objekt. Iga sõna juures on märgendite abil kirjas ka selle morfoloogiline info. Näiteks on sõna *keelele* juures kirjas, et selle sõnaliik on nimisõna (*upos=NOUN*), selle lemma on *keel* (*lemma=keel*), see on allatiivses käändes (*Case=All*) ja ainsuses (*Number=Sing*). (Muischnek, Müürisep 2022)



Joonis 1. Tegusõna *viskama* süntaksipuu EDT korpusest

## **2. LAUSEMALLIDE KOOSTAMISE MEETODID**

Lausemalle saab koostada nii käsitsi kui ka automaatselt. Selles peatükis tuleb esiteks juttu eri tüüpi käsitsi koostatud verbi süntaktilist ja semantilist käitumist kirjeldatavatest leksikaalsetest ja leksikograafilistest ressursidest. Teiseks annan ülevaate meetoditest, mis võimaldavad verbi süntaktilist ja semantilist käitumist automaatselt tuvastada.

### **2.1. Käsitsi koostamine**

Selles peatükis kirjeldan konkreetseid projekte, mille eesmärk oli kas lausemallide või verbivalentsi/verbirektsiooni kirjeldus.

#### **2.1.1. Lausemallide loendid ja sõnaraamatud**

Eesti keele kõige põhjalikum lausemallide käsitus on Huno Rätsepa kirjutatud raamatus “Eesti keele lihtlauset tüübid”, mis põhineb tema 1974. aastal kaitstud samanimelisel doktoritööl (Rätsep 1978: 3). Teos koosneb põhjalikust lausemallide teoreetilisest käsitlusest, lausemallide loendist ning eesti keele lihtlauset struktuuri kohta tehtavatest järeldustest. Lausemallide käsitluses toetub teos osaliselt generatiivse grammatika teooriale, määrates töö eesmärgiks kirjeldada, missuguse struktuuriga lihtlaused on eesti keeles võimalikud ja grammatilised. Sellest tulenevalt on teoses kasutatud materjali koostanud autor ise, lähtudes seejuures enda kui emakeelse kõneleja keelepädevusest. Seetõttu jäi teosest kõrvale aga lausetegelik kasutus. (Rätsep 1978: 4–8)

Rätsep esitab lausemalle morfoloogiliste vormide, sõnaliikide ja nende substitutsiooniklasside kombinatsioonina. Ta hoidub lausemallide lauseliikmete tasemel esitamisest, sest lauseliikmed on tema sõnul lausemallide esitamiseks liiga abstraktsed. Seetõttu kasutab ta näiteks subjekti asemel sõnaliigist ja käändest koosnevat vormiklassi *N+nom* ehk nimisõna nominatiivis. (Rätsep 1978: 18) Selliste lauseelementide jaoks, mis kannavad sama tähendust, aga saavad esineda väga erinevates vormides, kasutab

Rätsep substitutsiooniklasse. Näiteks selle asemel, et panna verbi *tulema* LÄHTEKOHT mallis kirja kõigi selles esinevate vormiklassidena (*N+el.*, *N+abl.*, *N+gen. tagant*, *N+gen. alt*, *N+gen.+pl. vahelt*, adverbid *altpoolt*, *sealt*), koondab ta need ühe substitutsiooniklassi alla: ekstralokaalne direksionaal ehk kust-suund. (Rätsep 1978: 43–44) Siinses töös kasutatakse substitutsiooniklasside asemel semantilisi rolle.

Huno Rätsep koostas lausemallid mahupiirangute tõttu ainult liht- ja liitverbidele, väljend- ja ühendverbid jäid teosest välja. Kokku on teoses fiktseeritud lausemallid üle 6000 verbile, kuid polüseemiat arvestades on lausemallid tuvastatud kokku üle 12000 verbi tähendusele. Need jaotuvad teoses esitatud 380 verbiülese elementaarlause malli alla. Töös on niisiis autori sõnul kirjeldatud ära enamik eesti keele lihtlausetest ning sealhulgas ka mõned laused väljaspool seda piiri. (Rätsep 1978: 212–215)

Huno Rätsepa uurimus on eesti keeles ainuke, mille fookuses on terved laused. Traditsiooniliselt kirjeldatakse verbi süntaktilist käitumist rektsioonimallidena. Näiteks võib tuua Raili Pooli sõnastiku “Eesti keele verbirektsioone” (1996). Erinevalt Rätsepa kirjeldusest on selle eesmärk abistada eesti keele õppijaid ja õpetajaid keeleõppes ja -õpetamises, mistõttu keskendub see ainult rektsioonide endi esitusele. Raamatus on 318 verbi (sealhulgas ühend- ja väljendverbi) rektsioonid ja neid ilmestavad näitelauseid koos vene, inglise ja soome tõlgetega. Rektsioonide juures on kirjas nii verbi seotud kui vabad laiendid, kuid vabad on pandud sulgudesse. (Pool 1996: 5–6) Tabelis 1 on toodud verbi *vedama* esitus selles raamatus (Pool 1996: 72).

Tabel 1. Verbi *vedama* reksiooniesitus raamatus “Eesti keele verbireksioone” (Pool 1996)

Verb	Reksioon	Näited
vedama/vedada veab	kellel? (+ milles?)	<b>Reinul</b> ei vea ( <b>kaardimängus</b> ) kunagi. <i>Rein has no luck in cards.</i> <i>Reinillä on aina huono onni korttipelissä.</i> <i>Рейну никогда не везет в карточной игре.</i>
vedama/vedada veab	kellel? (+ millega? kellega?)	<b>Meil</b> vedas ( <b>ilmaga</b> ). <i>We were lucky to have good weather.</i> <i>Meillä oli hyvä tuuri sään suhteen.</i> <i>Нам повезло с погодой.</i>

Inglise keele jaoks on korpuse põhjal käsitsi koostatud “Inglise valentsisõnaraamat” (Herbst jt 2004), kus on antud info õppijatele keerulise valentsiga predikaatide kasutuse kohta. Iga predikaadi kohta on tähenduste kaupa kirjas predikaadi lausemallid ning selles võimalikud esinevad seotud laiendid, nende semantilised rollid ja kollokatsioonid.

### 2.1.2. Elektroonilised ressursid

Eelmises peatükis kirjeldatud “Inglise valentsisõnaraamatu” (Herbst jt 2004) põhjal on tehtud ka **the Erlangen Valency PatternBank** (Herbst, Uhrig 2009), mis koondab kõiki “Inglise valentsi sõnaraamatus” esitatud lausemalle verbide üleselt. Nende kahe ressursi peamine vahe on see, et kui sõnaraamat lähtub verbist ja näitab selle all lausemalle, siis the Erlangen Valency Patternbank lähtub lausemallist ja näitab selle all verbe. (Herbst, Uhrig 2019: 173–174)

Üks kõige laialdasemalt kasutatavaid lausemallide esitamise raamistikke on **FrameNeti** projekt, mis põhineb freimisemantika teoorial. Freimisemantika teooria kohaselt on tegevustel tegelikult palju rohkem osalisi kui seotud laiendeid. Näiteks on *kirjutama* kahevalentne verb kahe argumendiga: kirjutaja ja see, mida kirjutatakse. Kirjutamisprotsessi osalised saavad aga olla ka kirjutusvahend, adressaat, kirjutamiseks vajaminev materjal, kaaskirjutaja, kirjutamise põhjus, kirjutamisega seotud tunded, kirjutamise asukoht ja toimumisaeg. (Croft 2022: 177) Freimisemantika seisukohalt tuleks verbi tähendusi avada just verbi kujutatud tegevuse kõigi osaliste kaudu (Fillmore 1982). Freimisemantikast kasutatakse tihti konstruktsioonigrammatikaga käsikäes, sest see annab võimaluse arvestada konstruktsioonide semantikaga (Borin, Lyngfelt 2025 (ilmumas)).

FrameNeti projekti eesmärk on koostada freimisemantikast põhinev freimide andmebaas. See koondab freimide alla freimi tähendusest tuleneva tegevuse osalisi ehk freimi elemente ja seda tegevust väljendavaid sõnu ehk leksikaalseid üksusi. Tüüpiliselt on leksikaalseks üksuseks verb. (Ruppenhofer jt 2016: 7–8) Freimi elemente väljendatakse semantiliste rollide abil, mis jaotuvad tuumelementideks ehk argumentideks ja perifeerseteks elementideks ehk adjunktideks (Fillmore 2007: 133; Ruppenhofer jt 2016: 22). Tabelis 2 on toodud näiteks freimi *liigutamine* (*Cause\_motion*) kirje.

Tabel 2. **Freim *liigutamine* FrameNetis**<sup>7</sup>

Freim	Liigutamine
<b>Definitsioon</b>	AGENT paneb TEEMA liikuma LÄHTEKOHAST TEEKONDA SIHTKOHANI. Freimi erinevad liikmed rõhutavad trajektoori erineval määral ja freim jätab tavaliselt osa LÄHTEKOHAST, TEEKONNAST ja/või SIHTKOHAST väljendamata. Liikumise lõpetamine ei ole nõutav, kuigi selle freimiga annoteeritud üksikud laused võivad SIHTKOHTA rõhutada.

<sup>7</sup> [https://framenet2.icsi.berkeley.edu/fnReports/data/frameIndex.xml?frame=Cause\\_motion](https://framenet2.icsi.berkeley.edu/fnReports/data/frameIndex.xml?frame=Cause_motion)

<b>Tuumelemendid</b>	AGENT, ALA, JÕUD, SIHTKOHT, ALGUSPUNKT, TEEKOND, TULEMUS, LÄHTEKOHT, TEEMA
<b>Perifeersed elemendid</b>	MÄÄR, SEISUND, KAUGUS, KESTVUS, SELGITUS, PIDE, INSTRUMENT, VIIS, PÕHJUS, ASUKOHT, ALAMPIIRKOND, AEG
<b>Leksikaalsed üksused</b>	<i>ligi_tõmbama.v, viskama.v, tirima.v, liigutama.v, lükkama.v</i> (..)
<b>Freimidevahelised suhted</b>	Pärib millelt: Transitiiivne_tegevus Pärandab millele: Sujuva_liikumise_põhjustamine, Tulistamine, Edasi_andmine Kasutatakse milles: Toomine, Väljundamine, Kokku_kogumine, Söömine; Alamfreimid: Asetamine, Eemaldamine; On kausatiivne vorm: Liikumine
<b>Näitelauseid</b>	[ <i>Ta</i> ]AGENT <i>viskas</i> [ <i>oma kingad</i> ]TEEMA [ <i>kuivatisse</i> ]SIHTKOHT

Freimid ja nende rollid on koostatud käsitsi korpusenäidete abil. Seega on FrameNet käsitsi koostatud ressurss, mis koondab verbide lausemalle ehk nende semantilist ja süntaktilist valentsi semantiliste rollide abil väljendatud verbiülestest konstruktsioonide alla. (Ruppenhofer jt 2016: 7–8)

FrameNetiga sarnane süsteem on vene keele jaoks koostatud **FrameBank**, mis põhineb samuti freimisemantikal ning koostati käsitsi korpusenäiteid annoteerides. Sellel on FrameNetiga võrreldes aga kaks olulist erinevust. Esimene on see, et kui FrameNet lähtub freimist ja lahterdab selle alla kuuluvaid sõnu, siis FrameBank lähtub verbist ning lisab freimid selle eri tähendustele. Teine oluline erinevus on see, et kui FrameNet keskendub semantilisele informatsioonile, on FrameBankis välja toodud ka freimi elementide morfoloogilised ja süntaktilised omadused: süntaktiline funktsioon, sõnaliik, freimielementi täitva sõna kääne või selle juurde kuuluv kaassõna. (Lyashevskaya, Kashkin 2015: 350–351)

Eesti keele jaoks on samuti FrameNeti taolist freimidel põhinevat andmebaasi koostatud, küll aga väga piiratud mahus. Freimisemantilist lähenemist katsetati morfoloogiliselt ja süntaktiliselt märgendatud korpusega, mis koosnes 370 lausest Huno Rätsepa raamatust “Eesti keele lihtlausete tüübid” (1978). Sealt võeti välja liikumis- ja liigutamisverbe käsitlevad laused ja määrati neile sobiv freim. Laused märgendati seejärel käsitsi ära, määrates igale lauseliikmele sellega kokku langev freimielement. (Müürisep jt 2008)

Üks lausemalle elektrooniliselt esitav ressurs on veel **VerbNet**, mis jaotab verbid suurematesse semantilistesse klassidesse. Iga klassi jaoks määratakse süntaksi tasandil seal esinevad seotud laiendid ning semantika tasandil argumentidele lähtuvad piirangud. Semantilised klassid on omavahel hierarhilises struktuuris, kus üks klass võib olla teise alamliik või üleliik. (Kipper-Schuler 2005) On arutatud ka eesti keele VerbNeti loomist (Jentson 2013), kuid see projekt ei saanud teoks.

Lausemallide kirjeldamisega tegeleb ka **PropBanki** projekt, kus ei üldista lausemalle suurematele semantilistele klassidele, vaid lausemallid koostatakse igale verbile eraldi. Vastupidiselt VerbNeti ja FrameNeti lähenemistele ei täpsusta see mitte semantilisi rolle, vaid võtab need kokku suuremateks ja üldisemateks. Neid väljendatakse mitte nimedega, vaid nummerdatakse: Arg0 kuni Arg5 ning ArgM. (Palmer jt 2005) Arg0 tähistab PROTO-AGENTI ehk agendilaadseid semantilisi rolle, Arg1 PROTO-PATSIENTI ehk patsiendilaadseid semantilisi rolle. Arg2 kuni Arg5 alla kuuluvad erinevad semantilised rollid, mis sõltuvad tavaliselt verbi tähendusest. Arg2 on tavaliselt INSTRUMENT, KASUSAAJA või ATRIBUUT, Arg3 LÄHTEKOHT, KASUSAAJA, ATRIBUUT, Arg4 SIHTKOHT ja ArgM tähistab tavaliselt vabu määrusi. Üldistustega koos teeb PropBank aga ka täpsustusi. Iga verbi jaoks on inimloetavuse huvides üles kirjutatud, mida iga argument selle verbi juures märgendab. Näiteks pannakse verbi *nõustuma* puhul eraldi faili kirja, et Arg0 on *nõustuja*, Arg1 *ettepanek* ja Arg2 *see, kellega nõustutakse*. (Jurafsky, Martin 2023: 5–6)

## 2.2. Automaatne tuvastamine

Peale lausemallide käsitsi koostamise saab püüda neid ka automaatselt tuvastada, mis võtab küll vähem töömahtu, kuid on see-eest keerulisem. Järgnevalt tutvustatakse varasemaid uurimusi, mis tegelevad lausemallide või nendega seotud keeleliste nähtuste automaatse tuvastamisega.

### 2.2.1. Uurimused eesti keele kohta

Eesti keele jaoks on seni proovitud automaatselt tuvastada verbide seotud laiendeid ja nende realisatsioone (Orasmaa 2013; Muischnek, Sahkai 2009). Selleks kasutati automaatselt morfoloogiliselt märgendatud eesti keele korpusi: Tasakaalus korpust (Muischnek, Sahkai 2009: 22) ja Koondkorpuse ilukirjanduse alamosa (Orasmaa 2013: 587). Koondkorpusele oli lisaks morfoloogilisele infole märgendatud juurde ka osalausete piirid (Orasmaa 2013: 583). Kuigi kasutatavad korpused olid sarnased, olid konstruktsioonide automaatse tuvastamise meetodid neil uurimustel erinevad.

Kadri Muischnek ja Heete Sahkai (2009) tuvastasid konstruktsioone kollokatsioonide baasilt. Korpusest otsiti verbiga *minema* esinevaid trigramme nii grammatiliste kategooriate kui ka nende realisatsioonidena. Need pandi t-skoori abil esilduvuse ja produktiivsuse põhjal järjekorda. Leiti, et verbi *minema* laiendite kõige esilduvumad grammatilised kategooriad on ma-tegevusnimi, ainsuse aditiiv, illatiiv, allatiiv ja translatiiv. Tulemusi kontrolliti erinevate assotsiatsiooniskooride abil, mis kinnitasid leitud. Grammatiliste kategooriatena esitatud konstruktsioonide realisatsioonid vaadati ka käsitsi üle ning leiti, et meetod oli tuvastanud ka varem kirjeldamata konstruktsioone. (Muischnek, Sahkai 2009: 22–23) Uuringu tulemusena leiti, et kollokatsioonide otsimise arendamine oleks üks võimalus konstruktsioonide automaatseks tuvastamiseks. (Muischnek, Sahkai 2009: 27)

Siim Orasmaa (2013) üritas tuvastada verbi seotud laiendeid, mis on semantilistes käänetes nimisõnafraasid. Selle jaoks otsis ta korpusest automaatselt kindlas kõneviisis ja aktiivis verbidega osalauseid ning leidis, mis käänetes sõnad selle osalause piirides esinesid. Iga käände kohta pandi vastavalt sellega koosinenud verbile kirja käände sagedus, jättes välja idiomatilised väljendid. Saadud sagedusi võrreldi käände esinemisega kogu korpuses sõltumata verbist ning käänded pandi sageduste alusel ritta. Kui kääne oli sagedusejärjestuses verbiülesest järjestusest eespool, arvestati see verbi potentsiaalseks seotud laiendiks. Saadud tulemuste kontrollimiseks võrreldi neid eesti süntaksianalüsaatori jaoks käsitsi koostatud valentsisõnastikuga. Selle abil leiti, et meetodi täpsus on keskmiselt 72%, aga täheldati, et see ei sobi väga hästi verbidele, mis korpuses vähe esinevad. (Orasmaa 2013: 584–589)

### **2.2.2. Uurimused teiste keelte kohta**

Kuna magistritöö valmib projekti PRG1978 “Uue aja sõnastik: grammatika ja keelepädevuse kirjeldamine integreeritud multifunktsionaalses leksikograafilises ressursis” raames, mis tegeleb ka konstruktikoni loomisega, uuritakse järgnevas peatükis ka konstruktikonide loomist. Konstruktikonides kirjeldatakse verbide süntaktilist ja semantilist käitumist. Neid on loodud vähemalt seitsme keele jaoks: inglise, rootsi, saksa, vene, jaapani, brasiilia portugali ja ungari (Lyngfelt jt 2018). Enamike konstruktikone ehitavate keelte jaoks on juba olemas konstruktsioone kirjeldav käsitsi koostatud ressurss. Inglise, saksa, jaapani, rootsi ja brasiilia portugali keele jaoks on olemas verbide semantilist ja süntaktilist valentsi kirjeldav FrameNet (Ruppenhofer jt 2016; Ziem jt 2019; Ohara jt 2004; Dannélls jt 2021; Torrent jt 2014), inglise keele jaoks on lisaks sellele olemas ka valentsisõnaraamat (Herbst jt 2004). Vene keele jaoks FrameNetti ei ole, aga neil on selle sarnane ressurss FrameBank (Lyashevskaya, Kashkin 2015).

Ainus keel, millele luuakse konstruktikoni ja nagu eesti keelegi jaoks pole varasemat konstruktsioone koondavat ressursi, on ungari keel. Radikaalse konstruktsiooni-

grammatika teooriale (Croft 2001) toetudes peab ungari konstruktikoni projekt konstruktsioonideks ka sõnu ja morfeeme, mida see üritab sõnastikust automaatselt tuvastada ja kasutajale kuvada. Näiteks kui sõnastikust puudub liitsõna kirje, luuakse liitsõna lahku ja kuvatakse selle osade kirjed või kui sõnastikust puudub mõne tuletise kirje, tuvastatakse selle tüvi ja morfeem ning kasutajale kuvatakse nende kirjed. (Sass 2023: 535-540)

Enamikes konstruktikoni loovates keeltes on põhjana juba mingi käsitsi koostatud ressurss aga olemas, mistõttu ei keskendu enamik neist konstruktsioonide automaatselt tuvastamisele, vaid olemasolevate ressursside konstruktikoniks pööramisele (Fillmore jt 2012; Ohara 2018; Lyngfelt jt 2018; Torrent jt 2014; Ziem jt 2019; Janda jt 2018; Perek, Patten 2019). Konstruktikonide täiendamine käib tavaliselt käsitsi, näiteks õpikuid analüüsides või korpustest käsitsi konstruktsioone otsides (Janda jt 2020: 166) ja korpuse materjali annoteerides (Ziem jt 2019: 75).

Siiski on mõni projekt proovinud konstruktikone ka automaatse tuvastuse teel täiendada. Seda tehti näiteks rootsi keele jaoks. Nende prioriteet oli leida konstruktsioonid, mis koosnevad leksikaalsest osast ja grammatilisest osast. Selle jaoks otsiti automaatse morfosüntaktilise märgendusega korpusest *n*-gramme, mille osad kodeeriti sõnade, sõnaliikide ja fraaside tasandile. Fraaside leidmiseks muudeti automaatselt märgendatud sõltuvusstruktuurid fraasistruktuurideks. Peale kõigi võimalike kombinatsioonide tuvastamist arvutati nende esinemise tõenäosuse leidmiseks iga konstruktsiooni suhteline sagedust kõigi konstruktsioonide lõikes. Seejärel leiti PMI algoritmiga (*Pointwise mutual information*), kas kahe sõna koosinemine on tähenduslik või juhuslik. Tuvastatud konstruktsioonid anti pärast manuaalsele hindamisele. (Forsberg jt 2014: 118–122)

Kuigi inglise keele jaoks eksisteerib mitmeid käsitsi koostatud materjali põhjal loodud konstruktikone (Fillmore jt 2012; Perek, Patten 2019), on katsetatud ka automaatselt

masinõppeliste meetoditega uue loomist. Selle jaoks võeti ette korpus, lisati selle sõnadele automaatselt algoritmide abil süntaktilised ja semantilised piirangud ning otsiti nendevahelisi seoseid. Meetodi peamine probleem seisnes selles, et see on puhtalt fraasipõhine, mistõttu ei olnud selle meetodiga võimalik fraase suuremateks üksusteks kombineerida. (Dunn 2023) Masinõpet pole inglise keele jaoks kasutatud ainult laiemalt konstruktsioonide otsimiseks vaid ka ainult lausemallide automaatseks tuvastamiseks. Uuringus kasutati selle jaoks juhendamata masinõpet, mis tähendab, et sisendina anti mudelile ette ainult sõnade sõnaliigid, mille põhjal otsis see ise nende vahelisi mustreid ilma, et sellele oleks näidetena lausemalle ette antud. (Lippincott jt 2012: 427)

Konstruktsioonide automaatsel tuvastamisel toetutakse peamiselt korpustele ja sealt leitavatele kollokatsioonidele. Korpusest saab kätte statistika iga üksiku lauseliikme või realisatsiooni sageduse, lauseliikmete või sõnade koosinemise sageduse ja nende omavahelise seose tugevuse kohta. Tüüpiliselt kasutatakse lauseliikmete koosinemise leidmiseks n-gramme või aknameetodit. N-grammi meetodiga arvestatakse kollokatsioonideks ainult üksteise järel tulevad sõnad, aknameetodiga vaadatakse aga sõna ümbritsevat laiemat konteksti. Pärast kollokatsioonide tuvastamist arvutatakse, kui tugevalt selle osad üksteisega seotud on. (Simon 2023: 14) Seose tugevuse leidmiseks kasutatakse vastavalt ülesandele erinevaid statistikuid: MI-skoor on hea haruldaste kollokatsioonide, nt idioomide leidmiseks ja t-skoor, log-likelihood ja  $\chi^2$  on head sagedaste mustrite leidmiseks (Evert 2009: 1229-1230). Kollokatsioonidele põhinedes ei pruugi olla tuvastatud ühend aga alati konstruktsioon, sest üksteise läheduses esinevad sõnad ei pruugi olla alati süntaktiliselt seotud. (Simon 2023: 15)

Kollokatsioonidest lähtub ka kollostruktuuriline analüüs, mis on sõnade ja grammatiliste struktuuride seose uurimiseks mõeldud statistiliste meetodite kogumik. Peamiselt kasutatakse selle raamistiku meetodeid verbidega seotud konstruktsioonide uurimiseks, mille hulka kuulub ka verbi argumentstruktuur. (Stefanowitsch 2013: 290–291) Selle töö vaatepunktist on kollostruktuurilise analüüsi meetoditest kõige

olulisem lihtne kollekseemi analüüs (Stefanowitsch, Gries 2003), mis keskendub konstruktsioonide koosesinemisele ehk uurib seda, kui tihedalt on kaks konstruktsiooni omavahel seotud. Tavaliselt on üks neist konstruktsioonidest leksikaalne (sõna) ja ning teine grammatiline (lausemall). Niisiis saab öelda, et selle meetodiga leitakse automaatselt, kui hästi sõna sobitub konstruktsiooni. Seda meetodit saaks kasutada leitud verbiülestes lausemallides esinevate verbide automaatseks tuvastamiseks. (Stefanowitsch 2013: 291–295)

Peale puhtale kollokatsioonidele toetumise on automaatselt süntaksi kaudu otsitud ka konstruktsioonide osi, kasutades selle jaoks morfoloogiliselt ja süntaktiliselt märgendatud korpusi. Seda tehti näiteks baski keele verbide seotud ja vabade laiendite eristamise jaoks, sest nagu eesti keelgi on baski keel vaba sõnajärgjega. See tähendab, et kõrvuti olevad sõnad ei pruugi suure tõenäosusega süntaktiliselt seotud olla, mistõttu ei pruugi laiendite lähestikulise koosesinemise baasil otsimine nii häid tulemusi anda. Eesti keelega sarnaselt on baski keel ka morfoloogiliselt rikas keel, mistõttu abstraheriti laiendid käänete ja kaassõnade tasandile. Vabade ja seotud laiendite eristamiseks kasutati varasemalt mainitud MI-skoori ja ka Fisheri täpset testi, mis näitab, kui suure tõenäolisusega saaks eksperimendi kordamisel samad tulemused. Tulemusi hinnati käsitsi ning leiti, et need olid väga head: nii täpsus kui ka saagis olid üle 90%. Materjali vähesuse tõttu ei proovitud aga tuvastada terveid malle, vaid ainult verbide seotud laiendeid. (Aldezabal jt 2002: 42–47)

### 3. MATERJAL JA MEETOD

Järgnevas sisuosas tutvustatakse magistritöö tegemiseks kasutatud materjali ja kuidas seda materjali rakendati. Jutuks tulevad ka nii materjali kui meetodiga seotud probleemid ja piirangud.

#### 3.1. Materjal

Töö eesmärgiks on töötada välja esmane töövoog lausemallide automaatseks tuvastamiseks ning kirjeldada ära lausemallide automaatse tuvastusega tekkivad probleemid, et meetodit nende põhjal edasi arendada. Töövoog koosneb neljast sammust: tuvastamiseks vajalike andmete kogumine, nende põhjal lausemallide automaatne tuvastamine, käsitsi võrdleva materjali koostamine ja nende abil automaatsete lausemallide kvaliteedi hindamine.

Lausemallide automaatseks tuvastamiseks kasutan ma andmeid Tasakaalus korpusest<sup>8</sup>. Tasakaalus korpus on Eesti keele koondkorpuse allosa, mis sisaldab võrdses koguses tekste aja-, ilu- ja teaduskirjandusest: igapäevaste 5 miljonit sõna. See on märgendatud automaatselt nii morfoloogiliselt kui süntaktiliselt. Morfoloogiliseks märgendamiseks kasutati eesti keele töötamiseks mõeldud Pythoni (Rossum, Drake 2009) teekide kogu EstNLTK<sup>9</sup> versiooni 1.6.9 (Laur jt 2020; Koppel, Kallas 2022: 208) morfoloogilist analüsaatorit ja ühestajat Vabamorf<sup>10</sup>. Süntaktiliseks märgendamiseks kasutati mudelina eesti keele peal treenitud Stanzat (Qi jt 2020). Mina kasutan Tasakaalus korpuse 2023. aastal märgendatud versiooni, kus õige ülemuse ja lauseliikme märgendi on saanud

---

<sup>8</sup> <http://www.cl.ut.ee/korpused/grammatikakorpus/index.php?lang=et>

<sup>9</sup> <https://github.com/estnlTK>

<sup>10</sup> <https://github.com/Filosoft/vabamorf>

ümardatult 85% sõnadest<sup>11</sup>. Morfoloogilise märgenduse kvaliteet on analüüside mitmesust arvesse võttes viimase hindamise seisuga ümardatult 86% (Saul 2022: 36).

Otsustasin oma töö jaoks kasutada Tasakaalus korpust mitmel põhjusel. Esiteks on seal võrdselt esindatud kõik kirjakeele peamised žanrid ning need žanrid on samad, mis lausemallide käsitsi koostamiseks kasutataval korpusel. See tagab, et võrreldavad mallid põhinevad sarnasel materjalil. Teiseks on Tasakaalus korpus 15 miljoni sõnaga piisavalt suur, et selle põhjal saaks lausemalle üldistada, ning piisavalt väike, et selle põhjal andmete kogumine nõuaks võrdlemisi vähe arvutusvõimsust. Lisaks tähendab korpuse suhteliselt väike suurus, et andmed on ka käsitsi vaatlusel hoomatavad, mis teeb võimalike andmekogumisel tekkinud vigade leidmise lihtsamaks. Tulevikus on plaanis kasutada andmete kogumiseks ka suuremaid korpuseid, sest need annaksid võimaluse leida ka sellistele verbidele lausemallid, mida Tasakaalus korpuses ei esine või esineb üldistuste tegemiseks liiga vähe. Väike korpus on aga hea viis lausemallide tuvastamise esmaseks katsetamiseks, et probleemid enne suurte korpuste kasutamist tuvastada ja ära siluda.

Valisin meetodi kvaliteedi kontrollimiseks välja ühe semantilise klassi verbid: liigutamisverbid. Ühe semantilise klassi piiridesse jäämisel on kaks eelist. Esiteks aitab see takistada valimi laialivalgumist ning teiseks jagavad ühe semantilise klassi verbid tihti ka lausemalli (Tragel, Taremaa 2019: 212), mistõttu annab sama semantilise klassi verbide kasutamine veel ühe tahu lausemallide kvaliteedi kontrolliks. Varasemas liigutamisverbe käsitlevas töös määrati liigutamisverbideks Wordnetist leitud *liigutama* alammõisted, mida oli kokku 123 (Pajusalu jt 2013: 49–50). Mahupiirangutest tulenevalt lähtuti siinses töös verbide semantilisse klassi liigitamisel aga uurija enda sisetundest, kus liigutamisverbide hulka liigitati ainult need verbid, mille põhitähendus on liigutamisega seotud. Siiski peab analüüsis arvestama ka sellega, et üks ja sama verb

---

<sup>11</sup>[https://github.com/estnltk/estnltk/blob/main/tutorials/nlp\\_pipeline/C\\_syntax/03\\_syntactic\\_analysis\\_with\\_stanza.ipynb](https://github.com/estnltk/estnltk/blob/main/tutorials/nlp_pipeline/C_syntax/03_syntactic_analysis_with_stanza.ipynb)

võib polüseemsuse tõttu kuuluda mitmesse eri semantilisse klassi (Tragel, Taremaa 2019: 213).

Valikuks osutusid spetsiifiliselt liigutamisverbid, sest olles eesti keele semantiliste rollide märgendaja arendamise projektis (EKTB75 “Semantilise analüüsi baasvahendid“) mitme erineva semantilise klassi verbe märgendanud, märkasin, et liigutamisverbidega esinesid kõik tuvastusega seotud suuremad probleemid, mistõttu otsustasin nende kaardistamiseks liigutamisverbide peale jääda. Liigutamisverbidest võtsin uurimise alla ainult need, mis esinesid projektis kasutatavas korpuses vähemalt kümme korda. Piir jäi kümne peale, sest vähemate näidetega oleks lausemalli väga keeruline määrata ning projekti ajalise piirangu tõttu ei uuritud seal verbe alla selle sageduse. Seetõttu koosneb minu valim 28 verbist: *liigutama, tõstma, viskama, seadma, paigutama, asetama, pistma, poetama, laduma, toppima, sättima, torkama, paiskama, loopima, riputama, tirima, raputama, langetama, sõidutama, vedama, heitma, lükkama, tõukama, keerama, tõmbama, viima, tooma, panema*.

Liigutamisverbidega lausete korpusest väljavõtmiseks, lausemallide automaatseks tuvastamiseks ja lausemalli põhjal näitelauseste otsimiseks kirjutasin ma valmis programmid. Programmeerimiskeelena kasutasin Pythoni versiooni 3.6.10 (Rossum, Drake 2009). Koodi kirjutasin Jupyter Notebooki keskkonnas, mis võimaldab koodi loetavuse parandamiseks selle vahele ka tekstielemente lisada (Kluyver jt 2016). Kood on saadaval Githubis, sellele saab ligi järgnevalt lingilt: [https://github.com/wertepure/Lausemallide\\_automaatne\\_tuvastamine](https://github.com/wertepure/Lausemallide_automaatne_tuvastamine).

Pärast lausemallide automaatse tuvastamise valmimist peab saadud lausemallide kvaliteeti ka hindama. Selleks kasutan enda käsitsi loodud lausemalle. Neist semantilisi on 49 ja grammatilisi 92 (vt lähemalt ptk 5.1). Mallide koostamiseks on kasutatud materjali UD korpuse EDT alamkorpusest (Muischnek jt 2014).

EDT korpus<sup>12</sup> on käsitsi märgendatud korpus, milles sisaldub iga sõne kohta nii morfoloogiline kui ka süntaktiline info. Korpus koosneb 437769 sõnest ning 30972 lausest, millest moodustab suurima osa ajakirjandus 266806 sõne ja 18690 lausega. Suuruselt järgmine on teaduskirjandus 94022 sõne ja 5483 lausega ning väikseim on ilukirjandus 67744 sõne ja 5522 lausega. Lisaks on korpuses väike osa HamleDT 3.0 puudepangast, mis võtab enda alla 9200 sõne ning 1277 lauset. Selle andmestiku seas sisalduvad ka näitelauseid Huno Rätsepa raamatust “Eesti keele lihtlausete tüübid” (1978) (Bick jt 2004). UD korpuses on ka EWT alamkorpus, mis koosneb uuest meediast, foorumitest ning kommentaariumitest (Muischnek jt 2019: 23). Selle laused on aga žanrist tulenevalt keeruliste struktuuridega, mille tõttu jäid need magistritöö raamistikust välja.

---

<sup>12</sup> [https://universaldependencies.org/treebanks/et\\_edt/index.html](https://universaldependencies.org/treebanks/et_edt/index.html)

## 3.2. Meetod

Järgnevas peatükis seletatakse nii lausemallide käsitsi koostamise kui ka automaatse tuvastamise jaoks kasutatud meetodit ning tuuakse välja nende kitsaskohad ja materjalist tulenevad piirangud.

### 3.2.1. Lausemallide käsitsi koostamine

Ma alustasin magistritöö tegemist lausemallide käsitsi tuvastamisest, sest enne meetodi välja arendamist on hea tutvuda reaalse materjaliga, mida see meetod analüüsima hakkab. Lausemallid koostati osana Tartu Ülikooli EKT75 projektist “Semantilise analüüsi baasvahendid“. Selle raames võtsin EDT korpusest enda loodud programmi abil välja kõik laused, kus minu valimi 28 verbi esinevad. Kokku oli selliseid lauseid 1774. Seejärel vaatasin ma iga lause käsitsi üle ning koostas tabeli, kus panin iga verbi (sealhulgas sama peasõnaga ühend- ja väljendverbi) kohta kirja kõik tema otseste alluvatena esinenud semantilised rollid ja neile vastavad morfosüntaktilised märgendid.

Eristasin oma andmestikus argumente ehk lausemalli osi adjunktidest lauseliikmetest, mille saab lause põhistruktuuri muutmata välja jätta (Karlsson 2002: 187). Argumentideks määrati laiendid, mis tulenesid verbi tähendusest ning esinesid lausetes sageli. Esitasin lauseliikmeid PropBanki märgenduskeemi kaudu (Palmer jt 2005), lisades neile selgitusena spetsiifilisemad semantilised rollid. Kuidas üks käsitsi koostatud mall lõpuks välja nägi on toodud näites 31, kus iga semantilise rolli jaoks on esimese osana välja toodud sõna süntaktiline roll ja teise osana kääne. Kaassõnafraaside puhul lisandub neile ka kaassõna. Lühendite seletusteks vt peatükki “Lühendite ja märgendite selgitused”.

(31) lükkama

Arg0 (AGENT) = nsubj\_Nom

Arg1 (PATIENT) = obj\_Nom, obj\_Gen, obj\_Par

Arg2 (SIHTKOHT) = obl\_All, obl\_Adit, obl\_Ill, obl\_Gen+*alla*, obl\_Gen+*taha*

Lauseliikme lausemalli osaks määramisel toetutakse tavaliselt selle seotusele verbi tähendusega (Erelt 2017a: 64–65), selle kohustuslikkusele (EG 2023: 484) ja grammatilisele markeeritusele, mina otsustasin aga lisaks sellele arvestada ka sagedust. Seda seepärast, et lõin selle andmestiku osana “Semantilise analüüsi baasvahendid“ projektist. Projektis lähtuti PropBanki põhimõttest mitte eristada argumente adjunktidest lingvistiliste tavade kohaselt, vaid panna märgenduse järjepidevuseks kirja kõik rollid, mis verbiga küllalt sageli esinevad (Palmer jt 2005: 76). Sageduse abil ei püüta argumente adjunktidest selgepiirilisel eristada, vaid näidata, kummal pool argument-adjunkti skaalal need verbi suhtes asetsevad. Verbiga sagedamini esinevaid laiendeid loetakse märgendamise jaoks pigem argumenditaolisteks ja vähem sagedasi pigem adjunktlaadseteks. (Haverinen jt 2015: 910) Otsustamiseks, mis on küllalt sage ja mis mitte, ei lähtunud aga kindlast sageduse piirmäärast, vaid emakeelse kõneleja sisetundest. Kuna märgenduse hõlbustamiseks arvestati eesti semantilise märgendaja arendamise projektis sagedusega, otsustasin ka mina võtta selle üheks lausemalli osade määramise tunnuseks.

Argumentide ja adjunktide eristamise keerukus on aga üks peamistest põhjustest, miks käsitsi koostatud lausemalle ei saa pidada lõplikult õigeks. Peale selle on käsitsi koostamisel aga ka teine kitsaskoht: **materjali vähesus**. Nimelt otsustati varasemalt nimetatud projektis leida lausemallid kõigile verbidele, mis esinevad EDT korpuses vähemalt kümme korda. See tähendab, et on verbe, mille jaoks oli käsitsi lausemalli tuvastamiseks toetuda vaid kümnele näitele. Jääb aga küsimus, kas see on küllalt suur materjal argumentide adjunktidest eristamiseks. Näiteks eksisteerib semantilisi rolle, mis võivadki esineda vähe, aga olla siiski verbi tähendusest tingitud. Sellised on liigutamisverbide juures tüüpiliselt nt LÄHTEKOHT ja SAAJA. Materjali vähesuse tõttu võib juhtuda, et ühtegi nende rollidega näidet ei sattunud valimisse, mistõttu jäid

need argumentidena tuvastamata. Kuna EDT on suurim eesti keele jaoks olemasoleva käsitsi süntaktiliselt märgendatud korpus, ei ole valimi suurendamine võimalik.

Kuigi EDT korpus on käsitsi märgendatud, pole selles sisalduv info sada protsenti õige. Lauseid ja nende märgendust üle vaadates leidsin ma **mitmeid vigu, mida käsitsi märgendatud korpuses üldse esinema ei peaks**. Vigu oli mitut tüüpi: sõnale oli määratud vale ülemus; sõna oli saanud vale süntaktilise märgendi (näiteks nimisõnalise määruse asemel mäarsõnaline määrus); ühendverbi osis oli määratud tavaliseks määruseks, kuigi sellel peaks olema täiesti oma märgend, või sõna oli saanud vale käände märgendi (nt illatiivis sõna on nominatiivi märgendiga). Need olid vead, mille mina tuvastasin, aga on ka võimalik, et minu käsitsi koostatud mallidesse sattus veel vigu, mis mul kahe silma vahele jäid. See faktor võis käsitsi koostatud lausemallide kvaliteeti mõjutada.

### 3.2.2. Lausemallide automaatne tuvastamine

Rööpselt lausemallide käsitsi koostamisega hakkasin välja arendama nende automaatse tuvastamise meetodit. Kuna see magistritöö on seotud ka EKI projektiga<sup>13</sup>, mille üks eesmärkidest on töötada välja meetodid, mis võimaldavad lisada sõnastikku eri tüüpi konstruktsioone, sh verbide argumentsstruktuuri konstruktsioone, siis esmalt analüüsisin just eri keelte konstruktionide koostamispõhimõtteid (vt ptk 2.2.2). Nende kohta käivaid uurimusi lugedes avastasin, et enamikes keeltes, kus üritatakse konstruktion luua, ei püüta lausemalle ega muid konstruktsioone automaatselt tuvastada, sest nende keelte jaoks on neid sisaldav käsitsi tehtud ressurss juba olemas (Fillmore jt 2012; Ohara 2018; Torrent jt 2014; Janda jt 2018; Ziem jt 2019; Perek, Patten 2019). Eesti keele jaoks selline ressurss aga puudub. Samas olukorras oli ainult ungari keel, kuid nemad keskenduvad lausemallist väiksemate konstruktsioonide tuvastamisele (Sass 2023). Teised uuringud, kus konstruktsioonide automaatse tuvastusega tegeleti, põhinesid aga suuresti eesti keelest väga erineva struktuuri ja teoreetiliste tavadega

---

<sup>13</sup> <https://www.etis.ee/Portal/Projects/Display/66a57ac4-e0c6-40cd-b9df-148048ce9f95>

keeltel nagu inglise ja rootsi keel, mistõttu ei saanud ma nende meetodeid enda uurimuseks kasutada (Forsberg jt 2014; Dunn 2023; Lippincott jt 2012). Nii eesti keele kui ka teiste sarnaste keelte kohta tehtud sarnased uurimused ei üritanud aga tuvastada terviklikke lausemalle, vaid ainult verbide seotud laiendeid ja nende taga peituvaid realisatsioone (Muischnek, Sahkai 2009; Orasmaa 2013; Aldezabal jt 2002). Neil põhjustel otsustasin ma luua eesti keele lausemallide tuvastamiseks täiesti uue meetodi.

Tuginedes oma kogemustele lausemallide käsitsi koostamisel otsustasin rajada meetodi bigrammide sagedusel põhinevale tuvastusele. Valik osutus selle lähenemise kasuks, sest see tundus mulle kõige lihtsam viis arvestada lausete keerulistest struktuuridest tuleneva müraga. Nimelt jäi lausemalle käsitsi koostades silma, et päriselt kasutuses olevad korpuselaused järgivad täpset lausemalli struktuuri üpris harva. Põhjuseid selleks on mitmeid: lausetes võib esineda lõpmatu hulk vabu laiendeid, eesti keeles saavad olla elliptilised nii lauseliikmed kui ka muutetunnused ning lausemallid lõikuvad teiste konstruktsioonidega, mis vastavalt nende struktuuri muudavad. Lisaks sellele saab sama funktsiooniga lauseliikmeid grammatiliselt mitmel moel väljendada, mis hajutab lause struktuuride sagedust veelgi. Peale nende grammatilistest väljendusvõimalustest tulevate probleemide on 15% lauseliikmetest ka valesti märgendatud (vt ptk 3.1). Märgendusvigade tõttu võib lausete tervikstruktuure otsides juhtuda see, et kui lause koosneb neljast lauseliikmest ja üks neist on saanud vale märgendi, on selle tulemusel kogu lause struktuur vale. Neist probleemidest tuleb selles peatükis hiljem põhjalikumalt juttu (vt lk 51–59). Kõigil nendel põhjustel ei tundunud aga otstarbekas vaadata lausemallide leidmiseks lausete tervikstruktuuri, vaid läheneda lauseliikmetele kahekaupa. Sel viisil saab leida sageli koos esinevaid lauseliikmeid ilma, et lausete struktuurilised erinevused, lauseliikmete eri grammatilised väljendusviisid ja märgendusvead tulemusi liialt mõjutama hakkaksid.

Meetodi esimene samm on tuvastada iga verbi jaoks kõik selle otsesed alluvad ja koostada nende sagedusest andmestik. Kuna magistritöö valmis koostöös EKiga,

koostas andmestiku minu väljaarendatud meetodi põhjal EKI vanemtarkvaraarendaja Katrin Tsepelina.

Kõigepealt leitakse iga verbi jaoks lauses selle otsesed alluvad. Iga otsese alluva kohta pannakse esiteks andmestikku kirja see, kui mitu korda see just selle verbi otsese alluvana esines. Teiseks vaadatakse, mis teiste selle verbi otseste alluvatega see lauseliige koos esines ning pannakse kirja nende koosinemise sagedus. Seda tehakse aga üldistamise eesmärgil märgendite tasandil ja kahekaupa ehk bigrammidena. Iga lauseliige pannakse kirja nelja märgendi kombinatsioonina: süntaktiline märgend ehk *deprel*, käändsõnadel käändemärgend ehk *Case*, pöördõnadel verbivorm ehk *Verbform* ning kaassõnaühenditel kaassõna. Kui lauseliikmel puudub mõni neist märgenditest, pannakse selle asemel tühi tekstiväärtus. Andmeid ei kogutud aga kõigi verbi otseste alluvate jaoks, vaid ainult nende, mis saavad lausemallis esineda ning sobivad üldistuste tegemiseks. Need on toodud tabelis 3.

Tabel 3. Lausemallide automaatseks tuvastuseks otsitavad lauseliikmed

Süntaktiline märgend	Seletus <sup>14</sup>	Näide <sup>15</sup>
nsubj	käändsõnaline subjekt	<i>Hansapank seda ei uuri.</i>
obj	käändsõnaline objekt	<i>Eesti on jõudnud arengufaasi, kus <u>isiksust</u> peedistatakse igal tasandil.</i>
obl	nimisõnaline määrus ehk obliikva	<i>Juba <u>reedel</u> andis ta kahe <u>tunniga</u> ettevõtte üle.</i>
xcomp	ahelverbi infiniitsed osad, da-infiniitsed objektid, da-infiniitsed verbid otstarbelause	<i>Olen lapsepõlvest peale harjunud seda <u>vaatama</u> ; Selline hiigelhoius pidanuks</i>

<sup>14</sup> seletused pärinevad UD dokumentatsioonist:  
[https://github.com/EstSyntax/EstUD/blob/master/UD\\_dokumentatsioon\\_oktoobris\\_2022.pdf](https://github.com/EstSyntax/EstUD/blob/master/UD_dokumentatsioon_oktoobris_2022.pdf);  
<https://universaldependencies.org/u/dep/>

<sup>15</sup> näited pärinevad EDT korpusest: [https://universaldependencies.org/treebanks/et\\_edt/index.html](https://universaldependencies.org/treebanks/et_edt/index.html)

	öeldisena, translatiivsed predikatiivadverbiaalid, seotud laiendina toimivad otstarbemäärused	<i>Hansapanga <u>valvsaks</u> tegema.</i>
advcl*	määrusena toimiv infiniitarid	<i>Vene-Tšetšeeni sõja tõttu jäi see siiski <u>tegemata</u>.</i>
advmod	määrsõnaline määrus	<i>Seadusest on <u>alati</u> kinni peetud.</i>
root	predikaat, mitmesõnalise öeldise puhul infiniitne komponent	<i>Pere ei <u>võta</u> seda seika siiski päris tõsiselt.</i>
case	kaassõna	<i>Peastaapi jõuab ta <u>pärast</u> kella kahte.</i>
compound:prt	ühendverbi afiksaaladverbiline osis	<i>Ajakirjanikud mõtlevad <u>välja</u> valesid jne.</i>
ccomp	komplementlause peasõna ja da-infinitiivne öeldistäide	<i>"<u>Tegutsesime</u> nende võimaluste piirides, mis meile on antud," ütles ta.</i>

Märgend *advcl* on märgitud tärniga, sest selle kolmest erinevast funktsioonist - määruskõrvallause, määrusena toimiv infiniitarid ja võrdlustarind - jäetakse sisse ainult määrusena toimivad infiniitaridid. Seda põhjusel, et need saavad olla lausemalli osad (Rätsep 1987: 226–229) ning need on vormi tõttu teistest kergesti eristatavad. Määruslikud kõrvallused aga ei saa seotud laienditena esineda (Rätsep 1987: 58–60), mistõttu jäid need välja. Võrdlustarind saab küll lausemallis esineda, aga väga harva, eesti keele varasemas lausemallide käsitluses ainult verbi *olema* puhul (Rätsep 1978: 145).

Objekti ja subjekti puhul ei arvestatud sageduse kokku lugemisel käänat, sest nende kääne oleneb lause aspektist ning tegevussubjekti ja -objekti rollist ehk pole

lausemallide lauseliikmete eristamisel oluline (Erelt 2017a: 77). Meetodi väljatöötamisel kaalutleti ka allatiivi, illatiivi ja aditiivi koondamist üheks suureks sihtkäändeks, sest need kõik väljendavad peamiselt SIHTKOHTA (EG 2023: 150). Otsustasin need siiski eraldi jätta. Allatiivi ei koondatud illatiivi ja aditiiviga, sest vastupidiselt sisekohakäänetele väljendab allatiiv lisaks SIHTKOHALE ka SAAJAT (EG 2023: 155), mille arvestasin lausemallide käsitsi koostamisel mõne verbi argumendiks. Kuigi aditiiv on lühike vorm illatiivist, jäid need kaks käänet eraldi, sest paljude sõnade puhul väljendab aditiiv alati kohta ja illatiiv sõltuvust. See ei kehti aga kõigi sõnade puhul (EKK 2020: 216). Siiski tähendab see, et nende käänete eristamine võib teatud verbide puhul, mis emmas-kummas tähenduses sõnu võtavad, oluliseks osutada. Samuti on verbe, mis nõuavad kas ainult illatiivis või ainult aditiivis laiendeid. Näiteks nõuavad Tiit-Rein Viitso (1976: 152) sõnul alati illatiivis määrust *segunema* ja *suhtuma* ning alati aditiivis määrust *kaduma*, *kukkuma*, *minema* ja *tulema*. Lisaks annab illatiivi ja aditiivi eristamine võimaluse leida, kas verbiga moodustatakse mõni tihti esinev väljendverb, mille osis on ainult ühes neist vormidest (nt *mängu panema* või *riidesse panema*).

Kõik teised verbi otsesed alluvad otsustasin ma välja filtreerida, sest neid kas ei arvestata traditsiooniliselt lausemallide osadeks, need vajaksid eraldi käsitlemist või need esinevad suhteliselt harva ja väga spetsiifilistes funktsioonides. Need lauseliikmed ja neile vastavad märgendid on toodud tabelis 4.

Tabel 4. Lausemallide automaatselt tuvastusest välja jäetud lauseliikmed

Süntakiline märgend	Seletus	Näide
csubj	infiniitne või osalauseline subjekt	<i>Tundub, et see on fotograafi teadlik tahe.</i>
advcl*	määruskõrvallause peasõna või võrdlustarind	<i>Raha liikus samamoodi nagu narkodollarid Ameerikas</i>
discourse	hüüundid ja diskursuspartiklid	<i>Mmm... kurat, ma vaatan seda.</i>

aux	abiverb ja mitmesõnalise verbivormi komponent	<i>Me <u>ei</u> anna mingeid kommentaare.</i>
cop	koopula	<i>Eesti Pank tunnistab: meie käed <u>on</u> liiga lühikesed!</i>
mark	alistussidendid osalause algul või küsisõnad küsilause algul	<i>Talle on jäänud mulje, <u>et</u> ühel hetkel hakatakse nõudma liiga lihtsaid lahendusi.</i>
nmod	nimisõnaline täiend	<i>Nemad peavad olema veendunud ja tuvastama <u>raha</u> allika.</i>
appos	lisand	<i>27. aprilli hommikul läks AS <u>Liviko</u> müügijuht <u>Arno</u> Voika tööle.</i>
nummod	arvsõnaline laiend või kvantor	<i>Juuli lõpus <u>1997</u> läks Cantradest mitmes osas teele kokku ligi <u>50 miljonit</u> dollarit ehk üle <u>700 miljoni</u> krooni.</i>
acl	täiendkõrvallause peasõna või täiendina toimiv infiniitaring	<i>See oli rohkem kui <u>10 protsenti</u> kogu Hansapanka <u>paigutatud</u> rahast.</i>
amod	omadussõnaline täiend	<i>Iga üheksas kroon tuli <u>salapäraselt</u> isikutelt.</i>
det	määratleja	<i><u>Selline</u> hiigelhoius pidanuks Hansapanga valvsaks tegema.</i>
conj	rinnastatud element	<i>Nemad peavad olema veendunud ja <u>tuvastama</u> raha allika.</i>
cc	rinnastussidend	<i>Suurem osa rahast tuli välismaalt <u>ja</u> anonüümselt.</i>
fixed	grammatiseerunud väljend	<i>Nii <u>et</u> suusakeskuse tegelased muutusid närviliseks.</i>
flat	mitmesõnalised nimed, võõrkeelsed väljendid	<i>Esmalt liikus see raha New <u>Yorgis</u> asuvasse Bankers <u>Trusti</u> ja sealt edasi Hansapanka.</i>
list	süntaktilise struktuurita sõnajärjendid	<i>Väitekirja avalik kaitsmine toimub 11. juunil 2004. a. algusega kell 11.00 Eesti Põllumajandusülikooli Loomakasvatuseinstituudi teadusnõukogu saalis (Tartu, <u>Kreutzvaldi</u> 1)</i>

parataxis	nõrgalt seotud lauseosa, kõrvuasend (saatelause otsekõne vahel)	" Ei, " <u>vastab</u> Kuusmann ning lisab: " Peamine vastutus lasub pangal endal.
orphan	verbi ellipsi tõttu sõltuvusstruktuuris loomuliku ülemuseta jäänud lauseelemendid	Tähendab ... mina konkreetselt vaatan ETVd väga vähe, <u>aeg-ajalt</u> AKd.
goeswith	valesti lahku kirjutatud sõnad	Nüüd mõtleb ta hoopis väikse tapamaja peale maksumusega 50 <u>000</u> --60 <u>000</u> krooni.
compound	mitmesõnalised arvud	Zürichist tuleb üle <u>700</u> miljoni
punct	kirjavahemärgid	Iga üheksas kroon tuli salapäraselt isikutelt.
dep	kui muud relatsiooni ei osata panna	"Paljude inimeste ettekujutus on, et <u>stepp</u> , see on aeroobikasuund - steppaeroobika!"
acl:relcl	relatiivkõrvallause peasõna	See on Šveitsi suurima panga UBSi haru, mis on <u>tuntud</u> numbrikontode poolest.
cc:preconj	lahksidendi esikomponent	Perel endal on valida <u>kas</u> Brüssel või mõni kinnisvara-arendusega tegelev firma.
csubj:cop	koopulalause infiniitne või osalauseline subjekt	Selleks, et uurijatel oleks lihtsam rahapesu vastu <u>võidelda</u> .
nsubj:cop	koopulalause käändsõnaline subjekt	Eesti Pank tunnistab: meie <u>käed</u> on liiga lühikesed!

Kuigi suuremat osa neist lauseliikmetest ei arvestata lausemallide osaks, võib rinnastatud element, koopulalause subjekt ja infiniitne või osalauseline subjekt esineda ka seotud laienditena (Rätsep 1978: 59; Rätsep 1978: 130; Rätsep 1978: 202). Infiniitne ja osalauseline subjekt jäi automaatselt analüüsist välja, sest see on subjekti erijuht, mis võib ka predikaadita esineda ning millel puuduvad subjekti põhitunnused (Erelt jt 2017: 241). Koopulalausest seotud lauseliikmeid jäid välja, sest koopulalused, milleks UD puudepankades on ainult *olema*-verbiga laused, vajavad eraldi käsitlust, mis selle magistritöö raamidesse ei mahu. Sõnad märgendiga *conj* jäid välja, sest verbi otseste alluvatena on selle märgendiga tavaliselt rinnastatud lause peasõnad, aga rinnastatud

lauseid arvestatakse ainult mõne väljendverbi (*kätte võtma, jalad selga võtma*) lausemalli osaks (Rätsep 1978: 59–60).

Lisaks alluvate filtreerimisele ei kogu ma statistikat kõigis vormides esinevate liigutamisverbide jaoks. Välja filtreeritakse sellistes vormides verbid, mille morfoloogiline vorm mõjutab seotud laiendite esinemise võimalust ja sellised, mis esinevad lauses perifrastilise verbi osana või sisestatud konstruktsioonina. Need on loetletud tabelis 5.

Tabel 5. Välja filtreeritavad verbivormid

Märgend	Vorm	Näide
sup	supiin ehk <i>ma</i> -infiniitiv	<i>elama</i>
sup ps el	supiin elatiivis ehk <i>mast</i> -vorm	<i>elamast</i>
sup ps abes	supiin abessiivis ehk <i>mata</i> -vorm	<i>elamata</i>
sup ps tr	supiin translatiivis ehk <i>maks</i> -vorm	<i>elamaks</i>
sup ps in	supiin inessiivis ehk <i>mas</i> -vorm	<i>elamas</i>
partic pres ps	<i>v</i> -kesksõna	<i>elav</i>
inf	<i>da</i> -infiniitiv	<i>elada</i>
ger	gerundiiv ehk <i>des</i> -vorm	<i>elades</i>
imper	imperatiiv ehk käskiv kõneviis	<i>ela/elagu jne</i>
-	kui verbil on abiverb	<i>on <u>elanud</u>, ei <u>ela</u>, saab <u>elada</u> jne</i>
syn_rel=aux	kui verb on abiverb	<i><u>on</u> elanud, <u>võib</u> elada</i>
imps	umbisikuline tegumood	<i>elatakse</i>

Umbisikuline tegumood, *da*-infinitiiv ja käskiv kõneviis jäid välja seetõttu, et selles vormis öeldisverbi puhul ei esine lauses tavaliselt subjekti (Erelt jt 2017: 252–253; Metslang 2017b: 529). *Ma*-infinitiivi tarindid jäid välja, sest need esinevad ainult verbi laiendite, ahelverbide osade või perfifrastiliste verbivormide koosseisus, mistõttu ei moodusta nad üksi lausemalle (Erelt 2017e: 790–805). *V*-partitsiip jäi välja, sest see ei esine lauses predikaadina (Erelt 2017e: 820). *Des*-vorm jäi välja, sest see esineb tavaliselt adverbiaalina ning kui see ka esineb sõltumatult, on see leksikaliseerunud ehk ei võta muid laiendeid peale subjekti (Erelt 2017e: 808–813). Abiverbid jäid välja, sest neil ei ole oma lausemalli - lausemalli määrab põhiverb (Erelt 2017b: 93). Kuna eelnevalt oli plaan leida ka lausemalli lauseliikmete kõige tüüpilisem järjekord, jäid abiverbiga põhiverbid analüüsist välja, sest mitmesõnaline verb muudab objekti asukohta lauses (Lindström 2017: 562). Kokkuvõttes koostati materjal isikulise tegumoe kindlas, tingivas ja kaudes kõneviisis, oleviku ja lihtmineviku ja jaatava kõne vormides verbidele.

Filtri läbinud lauseliikmete põhjal koostatakse järgmise sammuna sagedusel põhinev andmestik, kus filtrid läbinud lauseliikmete ja nende omavaheliste paaride sagedused loetakse verbide kaupa kokku. Sel puhul eristatakse samade märgenditega lauseliikmetega arvestamiseks kahte tüüpi sagedust: *count\_all* ja *count*. Kui verbi alluvateks on kaks samade märgenditega sõna, loetakse *count\_all* juures selle sageduseks kaks, *count* juures aga üks. See eristus on tehtud seepärast, et kaks identset lauseliiget ei tohiks seotud laiendina verbi alluvana korraga esineda, aga süntaktilise märgenduse piirangute tõttu saavad samade märgenditega olla ka kaks erinevat lauseliiget. Kuna algselt oli plaanis leida ka lausemalli liikmete kõige tüüpilisem järjekord, pandi andmestikku kirja ka see, kumb lauseliige enne teist esines ning kas verb esines enne esimest lauseliiget. Joonisel 2 on kujutatud näide selle meetodi põhjal koostatud andmestikust. Joonisel kujutatava tabeli päiste ja märgendite selgitused on leitavad peatükis “Lühendite ja märgendite selgitused”.

id	verb	verb_compound	deprel1	case1	verbform1	obl_case1	deprel2	case2	verbform2	obl_case2	count <sup>1</sup>	count_all	deprel1_before	deprel2_before	verb_before
Filter	viskama	Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter
442...	viskama		obj	*			obj	*			907	985	81	81	882
188...	viskama		nsubj	*			nsubj	*			620	630	10	10	189
687...	viskama		nsubj	*			obj	*			520	571	530	41	143
687...	viskama		obj	*			nsubj	*			520	571	41	530	505
188...	viskama		advmod	<puudub>			advmod	<puudub>			330	400	80	80	270
113...	viskama		obl	all			obl	all			258	260	2	2	235
100...	viskama		advmod	<puudub>			obj	*			224	283	193	90	178
100...	viskama		obj	*			advmod	<puudub>			224	283	90	193	216

## Joonis 2. Lausemallide automaatseks tuvastamiseks koostatud andmestiku väljavõte tegusõna *viskama* näitel.

Järgmise sammuna pannakse lauseliikmete paaridest kokku suuremad lausemallid. Kõigepealt arvutatakse selle jaoks välja iga üksiku lauseliikme ja lauseliikmete paari suhteline sagedus. See näitab verbipõhiselt, mitmes protsendis selle verbi lauses lauseliige või lauseliikmete paar esines. Lausemalli kandidaatideks arvestatakse ainult need lauseliikmed või paarid, mis esinevad üle 5% lauses. Lävend on sellel piiril seepärast, et 5% on minu esialgse hinnangul piisav, et tuvastada ka olulisi kuid harvemaid lausemalle ilma, et liiga palju märgendusvigu ja vabu laiendeid lausemalli sisse jääks. Lävendi ületanud lauseliikmete paarid kombineeritakse suuremateks potentsiaalseteks lausemallideks põhimõttel, et kui lauseliikmed esinevad kõik paaridena üksteisega koos üle 5% lauses, esinevad nad ka kolmikute ja nelikutena koos. Abstraheerituna saab seda seletada nii, et kui lävendi ületasid paarid a+b, a+c ja b+c, arvestatakse lausemalli kandidaadiks ka a+b+c. Kuna selle tulemusel tekib palju lausemalli kandidaate, mis üksteist hõlmavad, otsitakse õigete lausemallide saamiseks iga lauseliikme jaoks üles pikim mall, milles see lauseliige esineb. Neist duplikaatide eemaldamisel saadakse kätte lõplikud lausemallid. Kõige eelnimetatu automaatseks tegemiseks koostas ma Pythoni keeles koodi Jupyter Notebook keskkonnas.

Lausemallide automaatsel tuvastamisel on mitmeid probleeme, mille kõigi lahendamine pole kahjuks võimalik. Üks probleemide allikas on see, et üht **sama semantilise funktsiooniga lauseliiget on süntaktiliselt võimalik väljendada väga erinevatel viisidel** (Rätsep 1978: 43; Pajusalu 2009: 82; EG 2023: 485; Croft 2022: 175).

Lausemalle käsitsi koostades sain ma näiteks verbi *asetama* 24 lause põhjal kirja SIHTKOHA väljendamiseks 7 erinevat märgendite kombinatsiooni: *obl\_All* (ööhämärase rajale), *obl\_Adit* (teise inimese nahka), *obl\_Ill* (ridadesse), *obl\_Gen + case vahele* (vaesuse ja laiskuse vahele), *obl\_Gen + case peale* (vormi peale), *obl\_Gen + case äärde* (seina äärde) ja *obl\_Par + case väljaspoole* (väljaspoole koostööd). SIHTKOHA roll pole selles osas ainulaadne, korpuselauses esineb pea iga semantiline roll samamoodi mitmetel väga erinevatel viisidel.

Veel üks märgenditega seotud kitsaskoht tuleneb sellest, et **mitut erinevat semantilist rolli saab väljendada samade grammatiliste vahenditega** (Pajusalu 2009: 82; Croft 2022: 175; Fillmore 2007: 145). Näiteks saab nimisõnaline määrus allatiivis (*obl\_All*) tähistada nii SIHTKOHTA kui ka SAAJAT, mis on mitmete liigutamisverbide puhul mõlemad lausemalli eraldi osad. Selline on näites 32 toodud lause EDT korpusest, kus nii SIHTKOHT *pintsakule* kui ka SAAJA *mulle* on mõlemad allatiivis nimisõnalised määrused.

(32) "Huvitav, miks *mulle* korporatsiooni värvilist paela *pintsakule* ei pandud,"  
märkab ta ühtäkki, kui pildistamise ajal esmaspäeval oma Rotalia  
korporatsiooni mütsi kohendab.

Kuigi süntaksi tasandil pole neid võimalik sama märgendi tõttu eristada ning semantiliste rollidega märgendatud korpuseid eesti keele jaoks veel ei eksisteeri, peaks minu meetod andma ka automaatsel tuvastamisel vähemalt aimu, mis verbide juures selline olukord võis tekkida. Seda sellepärast, et ma kasutan kahte erinevat sageduse statistikut. Võttes näite 2 seda illustreerimaks, siis ühes sageduse arvestuses (*count*) loen ma kokku, kas nimisõnaline määrus allatiivis (*obl\_All*) üldse selles lauses selle verbi alluvana esines ning teises sageduses (*count\_all*) loen kokku, mitu korda nimisõnaline määrus allatiivis (*obl\_All*) samas lauses kindla verbi alluvana kokku esines. Niisiis oleks näite 2 seisukohast allatiivis nimisõnalise määruse (*obl\_All*) *count* 1 ja *count\_all*

2. Nende kahe statistiku vahe näitab seega, kas selle ühe märgendi taga võib peituda kaks erinevat semantilist rolli. Kui selle sagedus, kas märgend verbi alluvana üldse esines (*count*), on näiteks poole väiksem kui selle sagedus, mitu korda märgend tegelikult verbi alluvana esines (*count\_all*) näeme, et tegelikult esines see märgend verbi alluvana iga kord mitu korda. Selle tulemusel saab järeldada, et seal võib ühe märgendi all olla tegemist kahes erinevas semantilises rollis esineva laiendiga.

Siiski võib nende kahe sageduse erinevus tuleneda ka **märgendusvigadest**. Nagu varasemalt peatükis 3.1 mainitud, on automaatse märgenduse korrektsus nii süntaktiliselt kui morfoloogiliselt umbes 85%<sup>16</sup> (Saul 2022: 36). See aga tähendab, et 15% sõnadel on valed märgendid. Seetõttu esineb automaatselt märgendatud Tasakaalus korpuses ka selliseid juhtumeid, kus ühes lauses esineb samade märgenditega lauseliige mitu korda märgendusvea tõttu. Näites 33 on toodud lause Tasakaalus korpusest, kus *viskama* subjektiks on loetud nii *väravad* kui ka *Mihhail*, kuigi tegelikult peaks üks olema objekt ja teine subjekt.

(33) Eesti meeskonna *väravad* viskasid *Mihhail* Kozlov (2), Andrei Markov, Dmitri Raskidajev, Eduard Valiullin ja Boriss Dobrov.

Olukorda raskendavad veel **vabad laiendid**, sest need võivad verbi alluvana täiesti grammatiliselt mitu korda esineda ja seda ka täpselt sama märgendiga (Kroeger 2005: 76). Näiteks on näites 34 esitatud lause Tasakaalus korpusest, kus *viskama* otsesed alluvad on korraga kaks määrsõnalist määrust (*advmod*): *keskpaiku* ja *järjest*.

(34) Poolaja *keskpaiku* viskas Jesper Parve neli punkti *järjest* ja viis Ehitustööriista juhtima 23:21.

---

<sup>16</sup>[https://github.com/estnltk/estnltk/blob/main/tutorials/nlp\\_pipeline/C\\_syntax/03\\_syntactic\\_analysis\\_with\\_stanza.ipynb](https://github.com/estnltk/estnltk/blob/main/tutorials/nlp_pipeline/C_syntax/03_syntactic_analysis_with_stanza.ipynb)

Vabad laiendid pole küll lausemalli osad, aga traditsiooniliselt vaba laiendit esindavaid lauseliikmeid nagu määrsõnaline määrus tundub ka vale andmestikust täiesti välja jätta, sest need võivad vahel olla ka seotud laiendi rollis (EG 2023: 484). Näites 35 on lause EDT korpusest, kus määrsõnaline määrus *kuhugi* (märgendiga *advmod*) on hoopis *panema* sihtkoht ehk selle seotud laiend.

(35) "Mul on kolm lehma, kelle piima pole *kuhugi* panna," ütleb Paslepa küla Allika talu perenaine Leili Raidla.

Samuti on verbe, mille jaoks määrsõnalised määrused on pea alati seotud laiendid. Selline on näiteks verb *käituma*, mida ei saa väljendada ilma käitumise viisi mainimata (EG 2023: 484). Näites 36 on selleks seotud laiendiks määrsõnaline määrus *salapäraselt* ning nagu näha, teeks selle lausest välja jätmine lause ebagrammatiliseks.

(36) Viimati käitus Swedbank *salapäraselt* enne aastavahetust, kui tõstis osalust Hansapangas nelja protsendi võrra.

Kuigi siinses töös uurin ma meetodi kvaliteeti ainult liigutamisverbide vaatepunktist, siis meetodi arendamisel lähtusin ma Eesti Keele Instituudi vajadustest parandada kõigi verbide rektsiooniesitust EKI ühendsõnastikus (Langemets jt 2021), mistõttu otsustasin ma selliseid lauseliikmeid mitte välja arvata. See aga süvendab liigutamisverbide puhul automaatse tuvastamise probleeme vabade ja seotud laiendite eristamisega.

Lausemallide tuvastamist raskendab veel verbide **polüseemia**. Eesti keele jaoks puudub hetkel verbide tähendust eristav ressurss, mistõttu ei ole automaatselt võimalik lauseid tähenduse järgi jaotada, et malle tähenduste kaupa tuvastada. Verbi eri tähendustel saavad aga olla erinevad lausemallid (Karlsson 2002: 189). Näiteks on verb *lugema* ühes tähenduses kahevalentne ehk kahe seotud laiendiga: subjekt ja objekt (37), teises

tähenduses aga kolmevalentne ehk kolme seotud laiendiga: subjekt, objekt ja translatiivis obliikva (38):

(37) *Kaarel loeb raamatut.*

(38) *Kaarel loeb Joonast oma sõbraks.* (Karlsson 2002: 189)

Sellistel puhkudel võib polüsemse sõna tähendusspetsiifiline lausemallide tuvastus võimalik olla, aga see teeb lauseliikmete esinemise statistika veel vähem usaldusväärseks, sest suhteline sagedus arvutatakse verbi tähenduste ülese sageduse põhjal. Seetõttu võib ka mõne perifeersema tähenduse lausemall siiski kaduma minna, sest sõna esines selles tähenduses liiga vähe, et selle unikaalne lausemall mahuks statistikasse.

Lausemallide automaatset tuvastamist takistab ka see, et **lausemallid lõikuvad teiste konstruktsioonidega**, mis lause struktuuri mõjutavad (Hilpert 2014: 28–29). Näiteks tõstab muutmiskonstruktsioon verbi valentsi, lisades verbi laiendite hulka translatiivse predikatiivadverbiaali ning intransitiivsetele verbidele objekti (Erelt 2017c: 292). Võrdle nt verbi *magama* oma tüüpilises intransitiivses kasutuses (39) ja muutmiskonstruktsioonis (40):

(39) Täna öösel magasin lõpuks korralikult. (EKI ühendsõnastik 2024)

(40) Ta magas oma kaela kangeks. (Erelt 2017c: 292)

Automaatsel tuvastamisel on väga keeruline selliste konstruktsioonide lõikumisega arvestada, mistõttu loetakse näites 40 verbi otsesteks alluvateks ehk lausemalli osade kandidaatideks ka objekt *kaela* ja translatiivne predikatiivadverbiaal *kangeks*. Kui selliseid juhte esineb mingil põhjusel palju, võivad need ka pärast andmete töötlust lausemalli osaks jääda.

Automaatselt leitud sageduse põhjal seotud laiendite tuvastamisel peab arvestama ka sellega, et eesti keeles saab lauseliikmeid lausest välja jätta, kui neid saab tuletada kontekstist, teise lauseliikme grammatilisest infost või ümbritsevast situatsioonist (vt ptk 1.2.1 ja 1.3). Seda kutsutakse **ellipsiks**. Elliptilised võivad olla nii subjekt (41), objekt (42), määrus (43) kui ka verb ise (44). (Erelt 2017d: 590–597)

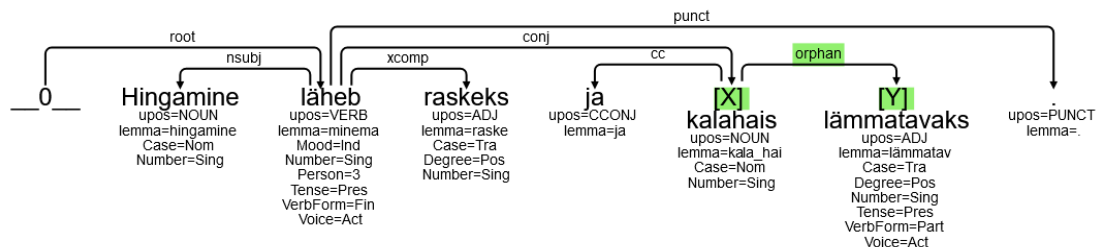
(41) Jüripäeva hommikul ärkas kahe lapsega üksi kodus olnud *Inga Sepp* sünnitusvalu peale ning Ø [=Inga Sepp] hakkas end haiglasse minekuks sättima. (EDT)

(42) Korjasin oma *riided* võõraste hulgast kokku ja toppisin Ø [=riided] kotti. (EDT)

(43) *Festivali korralduskomiteele* esitati osalemistaotlus ja Ø [=festivali korralduskomiteele] saadeti koori tutvustav plaat. (etTenTen<sup>17</sup>)

(44) Jüri ja Mari *sõitsid* Võrru, Juhan Ø [=sõitis] Narva ja mina Ø [=sõitsin] Tallinnasse. (Erelt 2017d: 599)

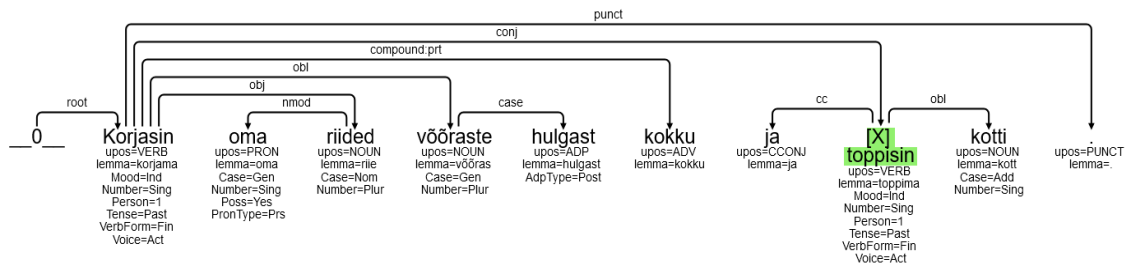
Neid on keeruline automaatselt tuvastada, sest UD süsteemis märgendatakse ellipse ainult juhul, kui rinnastatud on kaks lauset ning ühes neist on peasõna elliptiline. Sellistel juhtudel ühendatakse esimese rinnastatud osalause pea teise elliptilise pea esimese seotud laiendiga rinnastatud lauseliikmeid tähistava *conj* märgendi kaudu. Elliptilisele peasõnale alluvad lauseliikmed on omakorda selle seotud laiendiga seotud elliptilisust tähistava *orphan* märgendi kaudu.



### Joonis 3. Ellipsi esitamine UD süsteemis

<sup>17</sup> <https://www.keeleveeb.ee/dict/corpus/ettenten/about.html>

Peamiselt saabki UD süsteemis elliptilisi lauseliikmeid leida just rinnastatud liikmeid ühendava märgendi *conj* kaudu. Joonisel 4 esitatud lauses peaks verbivormi *toppisin* objekti leidmiseks liikuma mööda *conj* märgendit verbivormi *korjasin* juurde ja seal edasi *obj* märgendi kaudu sõna *riided* juurde.



Joonis 4. Rinnastusel tekkinud ellipsi esitamine UD süsteemis

Automaatselt pole aga võimalik tuvastada, missugune lauseliige elliptiliseks jäi, mistõttu oleks ainus võimalus nende leidmiseks kõigi esimese rinnastusega ühendatud lause peasõna laiendite omistamine teisele. See tekitaks aga rohkem kahju kui kasu, sest alati ei ole elliptilise lauseliikme peasõna ja sellega rinnastatud osalause peasõna samade seotud laienditega ning kui ongi, võib elliptiline lauseliige olla oma lähtevormist süntaktiliselt funktsioonilt ja vormilt erinev (Erelt 2017d: 597). Näites 45 on toodud EDT korpusest üks selline juhtum, kus koopulalause predikaadil *kehaehitusega* on sõna *Oleinikov* subjekti rollis, kuid verbivormi *lükatakse* puhul, mille alluvate seast on see ellipsi tulemusel välja jäetud, on tegemist hoopis objektiga.

(45) Nii *Oleinikov* kui Terry Hailey on kerge kehaehitusega ja lükatakse oma kohalt ära, seetõttu läks palju palle kaduma, " sõnas klubi peatreener Allan Dorbek.

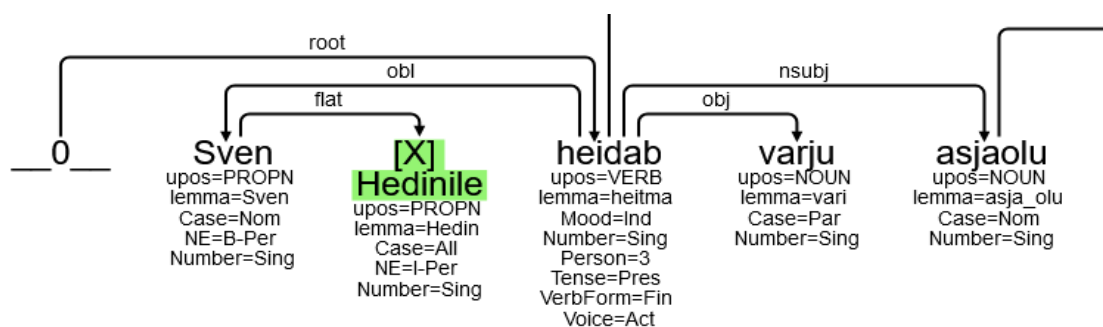
Rinnastus ei tekita aga probleeme ainult lauseliikme ellipsi puhul, vaid ka **vormilise ellipsi** puhul. Nimelt kannab terminatiivis, essiivis, abessiivis ja komitatiivis rinnastatud noomenifraaside puhul tüüpiliselt käänat ainult rinnastuse viimane liige. UD süsteemis on rinnastatud liikmete puhul aga otsene ühendus lause peaga esimesel liikmel ehk

liikmel, mis on genitiivis. Näite 46 puhul näiteks on verbivormide *tõsta* ja *langetada* otsene alluv ainult *ravimite*, mitte aga sõna *lihaskõõgastitega*, mis kannab lausemalli seisukohalt ainsana õiget käändeinfot.

(46) Põiesisest rõhku saab tõsta kolinergiliste *ravimite*, beeta-adrenergiliste agonistide ning langetada antikolinergiliste ravimite, beeta-adrenoblokaatorite ning *lihaskõõgastitega*. (EDT)

Selle näite teeb veel eriti keeruliseks see, et rinnastusi on kaks: rinnastatud on nii laused kui ka neis esinevad noomenifraasid. See tähendab, et verbivormi *tõsta* puhul peaks õige käände leidmiseks liikuma mitte ainult kahe rinnastust tähistava *conj* märgendi kaudu, vaid ka verbivormi *langetada* ja vormi *ravimit* ühendava *obl* märgendi kaudu. Niisiis oleks õiget vormi automaatselt väga keeruline leida.

Õige käände ei lähe aga kaduma ainult rinnastuse pärast, vaid ka muude UD süsteemi omapärade tõttu. Üks neist on **pärisnimede esitus**, kus mitmesõnaliste nimede puhul on pea otsene alluv mitmesõnalise nime esimene element. Teine element on esimese elemendiga ühendatud *flat* märgendi kaudu.



Joonis 5. Liitnimede esitamine UD süsteemis

Nagu jooniselt 5 näha, on mitmesõnaliste nimede puhul nimisõnalist määrust tähistava märgendiga *obl* eesnimi *Sven* verbi otsene alluv ning perekonnanimi on sellega

ühendatud *flat* märgendi kaudu. Probleem tuleneb sellest, et just perekonnanime juures on verbi rektsioonist tulenev kääne: allatiiv. Hetkel aga tuvastatakse verbi otsese alluvana obliikva nominatiivis, mitte väljendverbi *varju heitma* rektsioonist tulenev obliikva allatiivis.

Probleeme tekitavad ka **väljendverbid**, mis automaatselt tuvastatavate verbide hulgast välja jäävad, sest vastupidiselt ühendverbidele puudub neil UD süsteemis enda märgend. Väljendverbi käändsõnaline osis saab märgendi vastavalt oma süntaktilisele funktsioonile ja väljendverbi ei kohelda ühe tervikliku predikaadina. See tähendab aga seda, et lihtverbi alluvate sageduse hulka võib sattuda ka väljendverbidest tulenev sagedus. Õnneks kipub väljendverbi osaks olev käändsõna järgima lihtverbi malli (Rätsep 1978: 21), mis tähendab, et vähemalt määrusliku seotud laiendi tuvastamisel ei tohiks väljendverbid takistuseks saada. Siiski võib probleemiks kujuneda see, et väljendverbidel võivad olla lihtverbist erinevad seotud laiendid (Rätsep 1978: 23), mis aga märgendusüsteemi piirangutest tulenevalt loetakse lihtverbi laiendite hulka. See tähendab, et kui lihtverbist erineva lausemalliga väljendverb esineb küllalt sagedasti, võib selle mall sattuda ekslikult automaatselt tuvastatud lihtverbi malliks. See ei juhtu aga ühendverbide puhul, mis filtreeritakse vastavalt nende osist märkivale märgendile (*compound:prt*) lihtverbide seast välja.

### 3.2.3. Hindamine

Hindan oma välja töötatud meetodit mahupiirangute tõttu liigutamisverbide näitel. Liigutamisverbide automaatselt tuvastatud lausemallide kvaliteedi hindamiseks kasutan ma enda EDT korpuse põhjal käsitsi koostatud andmestikku. Hindamise teeb aga keeruliseks see, et käsitsi koostatud andmestik erineb vormiliselt automaatselt tuvastatud lausemallidest.

Lähtusin lausemallide koostamisel korpusest ja statistikast, mis tähendab, et ma ei pannud kirja kõiki malle, mis eesti keeles võiksid esineda, vaid ainult need, mis

materjalis esinesid ja sagedasti kasutatakse. Minu koostatud lausemallides on lauseliikmed esitatud semantiliste rollidena, sest andmestik koostati projekti “Semantilise analüüsi baasvahendid“ raames. Andmestikus olevate semantiliste rollide alla panin ma kirja kõik variandid, kuidas see morfosüntaktiliselt avaldus, kaasa arvatud kaassõnaühenditena. Lisaks sellele panin ma projekti vajadustest lähtuvalt ühe lausemalli alla kõik semantilised rollid, mis verbi ühe tähenduse all küllalt tihti esinesid, isegi siis, kui mõned neist esinesid üksteisega harva koos (47).

(47) viskama

AGENT = nsubj\_Nom

PATSIENT = obj\_Nom, obj\_Gen, obj\_Par

SIHTKOHT = obl\_All, obl\_Adit, obl\_Ill, obl\_Gen+ette, obl\_Gen+peale, obl\_Gen+vahele, obl\_Gen+taha, obl\_Gen+alla, obl\_Gen+pihta

Lõplikul hindamisel ma aga kõiki enda koostatud lausemalle ei arvesta. Kuigi ma koostas in käsitsi mallid ka ühend- ja väljendverbidele, jäävad need mahupiirangute tõttu magistritööst välja. See tähendab, et kokku leitakse lausemallid 28 lihtverbile.

Automaatselt tuvastatud lausemallid erinevad käsitsi koostatutest ühe peamise aspekti poolest: kuna praeguse seisuga ei ole eesti keele jaoks semantiliste rollide tuvastamise hõlbustamiseks loodud ressursi, puudub automaatsel mallidel süntaksit koondav semantiline roll. Lausemall on selle asemel esitatud ainult süntaktiliste ja morfoloogiliste märgendite abil (48). Iga lauseliige on kirja pandud kombinatsioonina neljast elemendist: süntaktiline roll, kääne, kaassõna ja verbivorm. Kui lauseliikmel saab olla kääne, aga seda ei arvestata, pannakse käände asemele tärn. Vastavalt lauseliikmele võivad osad elemendid ka tühjaks jääda. Tühja elementi esitatakse kahe ülakomaga, mis tähistab Pythoni programmeerimiskeeles tühja tekstielementi.

(48) lükkama ('nsubj', '\*', ", ") ('obj', '\*', ", ") ('obl', 'all', ", ")

Erinevuste vähendamiseks tuleb lausemallide esituses teha teatavaid teisendusi. Kuna käsitsi koostatud mallid on esitatud semantiliselt, mille alla on koondatud ka nende grammatiline avaldumine, tuleb semantilised mallid automaatsete mallidega võrdlemiseks morfosüntaktilisteks mallideks lahku lüüa. Käsitsi koostades panin ma iga semantilise rolli alla kirja kõik variandid, kuidas see korpuses grammatiliselt avaldus, aga mõningad neist olid väga harvad. Seetõttu lõön ma käsitsi koostatud mallid lahku ainult iga argumendi kõige tüüpilisemate grammatiliste realisatsioonide pealt. Näiteks lähevad rolli SIHTKOHT puhul eraldi mallidesse määrused sihikäändetes allatiiv, illatiiv ja aditiiv, aga mitte selles rollis ainult paar korda avaldunud määrsõnaline määrus ja erinevad kaassõnaühendid. Kui automaatselt tuvastati mõni kaassõnaühendiga mall, ei lähe see samal põhjusel kvaliteedi hindamisel arvesse. Selle tulemusel tekkis 49 semantilisest mallist 92 grammatilist malli. Nii semantilised kui ka grammatilised mallid on leitavad töö lisades.

Meetodi kitsaskohtade paremaks kaardistamiseks hindan ma automaatse tuvastuse kvaliteeti nii lauseliikmetele kui ka lausemallidele. Lauseliikmete kvaliteedi hindamisel kontrollin, kui paljud käsitsi koostatud grammatilistele kujudele viidud mallide lauseliikmed esinevad ka automaatselt tuvastatud mallides ning kui paljud tuvastati valesti. Lausemallide põhisel hindamisel kontrollin, kas lausemall on tervikuna automaatselt tuvastatud. Peale automaatse tuvastuse vigade leian ma ka käsitsi koostamisel esinenud vead, kus käsitsi koostatud materjalist oli verbi seotud laiend materjali vähesuse tõttu puudu jäänud. Mall või lauseliige arvestatakse käsitsi valesti veaks arvestatute hulka, kui see käsitsi koostatud materjal ei esinenud, aga Tasakaalus korpusse kontrolli tulemusel on näha, et tegemist on siiski õige lausemalli või verbi seotud laiendiga.

Kvaliteedi hindamiseks kasutan ma saagist ja täpsust. Saagis kirjeldab seda, kui paljud käsitsi koostatud materjalid on olemas lauseliikmed ja lausemallid automaatselt tuvastati

ning täpsus seda, kui paljud automaatselt tuvastatutest õiged olid. Saagise ja täpsuse arvutamiseks jaotatakse lauseliikmed ja lausemallid kolme erinevasse klassi:

- a) *True positive* (TP) - lausemall või lauseliige on käsitsi koostatud materjalis ja tuvastati ka automaatselt.
- b) *False positive* (FP) - automaatselt tuvastati lausemall või lauseliige, mis käsitsi koostatud materjalis ei esine.
- c) *False negative* (FN) - käsitsi koostatud materjalis olev lausemall või lauseliige jäi automaatselt tuvastamata. (Faaß jt 2010: 805)

Täpsuse ja saagise arvutamiseks kasutatakse järgmisi valemeid (Olson, Delen 2008: 138):

$$täpsus = \frac{TP}{TP + FP}$$

$$saagis = \frac{TP}{TP + FN}$$

Siinkohal peaks aga ära märkima, et kuna käsitsi koostatud ja automaatselt tuvastatud mallid olid loodud erinevatel põhimõtetel ja on algselt ka erineva kujuga, on arvatud statistikud lõpuks siiski üldistused ega anna täiesti täpset ettekujutust lausemallide kvaliteedist. Need on esiteks mõeldud analüüsi abistavate vahenditena, mis aitavad meetodi puudujääke tuvastada.

## 4. TULEMUSED

Kokku tuvastas meetod 28 verbi peale 107 verbikeskset lausemalli, mille keskmeks on leksikaalne verb. Verbideüleseid unikaalseid lausemalle, mis saavad esineda nii ühel kui ka mitmel verbil leiti 41. 107 lausemallist 25 on kahe lauseliikmega, 55 kolme lauseliikmega ja 27 nelja lauseliikmega. Verbiülestest lausemallidest on 15 kahe liikmega, 16 kolme lauseliikmega ja 10 nelja liikmega. Keskmiselt leiti iga verbi kohta 4 lausemalli, kuid kõige sagedasem lausemallide arv oli 5. Kõige vähem oli lausemalle verbil *heitma*, millel tuvastati ainult üks lausemall ja kõige rohkem oli lausemalle verbil *sõidutama*, millele leiti seitse lausemalli.

### 4.1. Liigutamisverbide lausemallide lauseliikmed

Liigutamisverbide mallides esines 21 erinevat lauseliiget, millest üheksa esinesid ainult ühe verbi lausemallides ja kolm kõigi verbide lausemallides.

Tabel 6. Automaatselt tuvastatud lausemallides esinevad lauseliikmed

Lauseliige	Lauseliige märgenditena	Mitme verbiga lauseliige esineb	Verbid, millega lauseliige esineb
käändsõnaline objekt	('obj', '*', ", ")	28	<i>riputama, tirima, raputama, torkama, sättima, toppima, laduma, poetama, asetama, sõidutama, paigutama, lükkama, keerama, viskama, tõmbama, loopima, paiskama, pistma, langetama, vedama, tõukama, seadma, tõstma, liigutama, viima, tooma, panema, heitma</i>
käändsõnaline subjekt	('nsubj', '*', ", ")	28	<i>riputama, tirima, raputama, torkama, sättima, toppima, laduma, poetama, asetama,</i>

			<i>sõidutama, paigutama, lükkama, keerama, viskama, tõmbama, loopima, paiskama, pistma, langetama, vedama, tõukama, seadma, tõstma, liigutama, viima, tooma, panema, heitma</i>
määrsõnaline määrus	('advmod', '<puudub>', ", ")	28	<i>riputama, tirima, raputama, torkama, sättima, toppima, laduma, poetama, asetama, sõidutama, paigutama, lükkama, keerama, viskama, tõmbama, loopima, paiskama, pistma, langetama, vedama, tõukama, seadma, tõstma, liigutama, viima, tooma, panema, heitma</i>
nimisõnaline määrus ehk obliikva allatiivis	('obl', 'all', ", ")	26	<i>tirima, loopima, paiskama, poetama, pistma, langetama, paigutama, vedama, lükkama, tõukama, keerama, tõmbama, tõstma, riputama, torkama, sättima, toppima, laduma, asetama, sõidutama, heitma, seadma, viskama, viima, tooma, panema</i>
nimisõnaline määrus ehk obliikva aditiivis	('obl', 'adit', ", ")	21	<i>tirima, torkama, sättima, poetama, asetama, viskama, loopima, laduma, pistma, paigutama, lükkama, tõukama, seadma, tõmbama, viima, tooma, panema, riputama, paiskama, toppima, sõidutama</i>
nimisõnaline määrus ehk obliikva illatiivis	('obl', 'ill', ", ")	10	<i>tirima, torkama, sõidutama, riputama, paiskama, toppima, paigutama, viima, tooma, pistma</i>
nimisõnaline määrus ehk obliikva adessiivis	('obl', 'ad', ", ")	9	<i>tirima, paiskama, laduma, viima, tooma, loopima, langetama, sõidutama, vedama</i>

nimisõnaline määrus ehk obliikva komitatiivis	('obl', 'kom', ", ")	6	<i>lökkama, loopima, vedama, tõukama, tõmbama, sõidutama</i>
ma-infinitiivis laiend	('xcomp', 'ill', ", 'ma')	5	<i>riputama, pistma, tõukama, sättima, panema</i>
nimisõnaline määrus ehk obliikva elatiivis	('obl', 'el', ", ")	5	<i>raputama, tõmbama, tooma, tirima, sõidutama</i>
nimisõnaline määrus ehk obliikva inessiivis	('obl', 'in', ", ")	3	<i>torkama, paiskama, liigutama</i>
nimisõnaline määrus ehk obliikva genitiivis kaassõnaga <i>alla</i>	('obl', 'gen', 'alla', ")	2	<i>toppima, seadma</i>
nimisõnaline määrus ehk obliikva genitiivis kaassõnaga <i>külge</i>	('obl', 'gen', 'külge', ")	1	<i>riputama</i>
komplementlause või da-infinitiivne öeldistäide	('ccomp', '<puudub>', ", ")	1	<i>torkama</i>
nimisõnaline määrus ehk obliikva genitiivis kaassõnaga <i>ette</i>	('obl', 'gen', 'ette', ")	1	<i>asetama</i>
nimisõnaline määrus ehk obliikva translatiivis	('obl', 'tr', ", ")	1	<i>paigutama</i>
nimisõnaline määrus ehk obliikva genitiivis kaassõnaga <i>poole</i>	('obl', 'gen', 'poole', ")	1	<i>keerama</i>
nimisõnaline määrus ehk obliikva terminatiivis	('obl', 'term', ", ")	1	<i>viima</i>
nimisõnaline määrus ehk obliikva genitiivis kaassõnaga <i>juurde</i>	('obl', 'gen', 'juurde', ")	1	<i>sõidutama</i>
translatiivne öeldistäitemäärus e predikatiivadverbiaal või otstarbemäärus	('xcomp', 'tr', ", ")	1	<i>seadma</i>
des-lauselühend	('advcl', '<puudub>', ", 'des')	1	<i>tirima</i>

## 4.2. Verbiülesed lausemallid

Nagu teoorias mainitud, jagavad sama semantilise klassi verbid tihti ka lausemalle (vt ptk 1.5). See tendents kehtib ka liigutamiverbide puhul. Minu tuvastatud 41 verbiülesest mallist olid 25 unikaalsed ühele verbile ja 16 malliks kahele või enamale verbile. Neist kõige laialdasema kasutusega olid korraga kaks lausemalli, mis mõlemad olid lausemalliks ühteteistkümnele verbile. Neist esimene koosneb subjektist, objektist ja määrõnalisest määrusest (49), mis tuvastati mallina verbidele *loopima*, *raputama*, *poetama*, *langetama*, *paigutama*, *lukkama*, *tõukama*, *keerama*, *tõmbama*, *tõstma* ja *liigutama*. Teine koosneb subjektist, objektist ja nimisõnalisest määrusest allatiivis (50) ning tuvastati mallina verbidele *loopima*, *paiskama*, *poetama*, *langetama*, *paigutama*, *vedama*, *lukkama*, *tõukama*, *keerama*, *tõmbama* ja *tõstma*. Näitelauses on kaldkirjas märgitud ära lausemallis esinevad lauseliikmed. Kõik selle ja järgnevate peatükkide näitelaused pärinevad Tasakaalus korpusest, kui ei ole just viidatud teisiti. Sõltuvus-süntaktilisest esitusest tulenevalt (vt ptk 1.6) on näitelauses märgitud ära ainult fraasi peasõna.

(49) ('nsubj', '\*', ", "), ('obj', '\*', ", "), ('advmod', '<puudub>', ", ")

Teinekord oli tal tahtmine, et *keegi teda tugevasti raputaks!*

(50) ('nsubj', '\*', ", "), ('obj', '\*', ", "), ('obl', 'all', ", ")

Tuhandete hobujõudude rammust värisev *alus* keerab *end merele*.

Kolmas kõige suuremale arvule verbidele tuvastatud lausemall oli nende kahe malli kombinatsioon (51), mis oli lausemalliks kümnele verbile: *sättima*, *laduma*, *asetama*, *sõidutama*, *heitma*, *seadma*, *viskama*, *viima*, *tooma* ja *panema*. Selle malli subjektita variant (52) esines samuti, kuid ainult *tõukama* verbil.

(51) ('nsubj', '\*', ", "), ('obj', '\*', ", "), ('obl', 'all', ", "), ('advmod', '<puudub>', ", ")

Ülempreester oli ainus, kes heitis käed rinnale risti ja pomises lõua alla loitsu.

(52) ('obj', '\*', ", "), ('obl', 'all', ", "), ('advmod', '<puudub>', ", ")

Eva sööstab köögist koridori ja tõukab *Bronka jõuliselt põrandale*.

Kõik neis mallides esinevad lauseliikmed osutusid liigutamisverbide seas ka kõige sagedasemateks lausemalli osadeks. Neist tuvastati subjekt, objekt ja määrsõnaline määrus kõigi verbide lausemallide liikmeks. Nimisõnaline määrus allatiivis ei esinenud vaid kahe verbi lausemallides: *raputama* ja *liigutama*.

Nagu allatiivis määruse sagedusest näha, olid liigutamisverbide seas ootuspäraselt sagedased SIHTKOHTA väljendavate lauseliikmetega mallid. Kokku oli 107 verbikesksest lausemallist 28 realisatsiooniga “subjekt, objekt, SIHTKOHT” ehk “keegi liigutab midagi kuhugi”, kus SIHTKOHT esines nimisõnalise määrusena allatiivis, illatiivis, aditiivis, terminatiivis ja nimisõnalise määrusena genitiivis koos kaassõnaga. 21 lausemallis oli nendele kolmele lauseliikmele lisandunud mõni teine lauseliige: määrsõnaline määrus, obliikva allatiivis või obliikva elatiivis, mis esindavad tavaliselt VIISI, SAAJA või LÄHTEKOHA rolli. Lisaks lisanditega mallidele oli ka 15 malli, kus oli küll väljendatud objekt ja SIHTKOHT, aga polnud subjekti. Tuvastati ka üks, kus oli SIHTKOHT, kuid polnud ei subjekti ega objekti. Kokku oli SIHTKOHT väljendatud 107 verbikesksest mallist 65-s ning 41 verbiülesest mallist 20-s.

Kuigi kõige sagedasem oli SIHTKOHTA väljendada allatiivis nimisõnalise määrusega, olid selle jaoks sagedased ka nimisõnalised määrused aditiivis ja illatiivis. Aditiivis nimisõnaline määrus oli liigutamisverbide lausemallide neljas kõige sagedasem lauseliige. Lausemall subjekti, objekti ja aditiivis nimisõnalise määrusega (53) tuvastati üheksale verbile: *loopima*, *laduma*, *paigutama*, *lökkama*, *seadma*, *tõmbama*, *viima*, *tooma* ja *panema*. Lausemall, kus subjektile, objektile ja aditiivis määrusele lisandus veel määrsõnaline määrus (54), esines veel kuue verbiga: *riputama*, *paiskama*, *torkama*,

*toppima, pistma, sõidutama*. Lausemall, kus määrsõnalise määruse asemel lisandus neile tüüpiliselt SAAJA rolli iseloomustav allatiivis määrus (55), tuvastati kolmele verbile: *torkama, toppima* ja *riputama*. Ainult objektist ja aditiivsest määrusest koosnev lausemall (56) tuvastati veel viiele verbile: *tirima, sättima, poetama, asetama* ja *viskama*. Ülejäänud neli lausemallid, mis sisaldasid nimisõnalist määrust aditiivis, leiti ainult ühele verbile: *pistma* (57), *tõukama* (58), *sõidutama* (74), *torkama* (85).

(53) ('nsubj', '\*', ", "), ('obj', '\*', ", "), ('obl', 'adit', ", ")

*Ta ladus need virna ning kandis tabelitesse.*

(54) ('nsubj', '\*', ", "), ('obj', '\*', ", "), ('obl', 'adit', ", "), ('advmod', '<puudub>', ", ")

*Ma tahaksin, et ta sõidutaks mu nüüd koju.*

(55) ('nsubj', '\*', ", "), ('obj', '\*', ", "), ('obl', 'adit', ", "), ('obl', 'all', ", ")

*"Ta toppis selle mulle põue ja pistis taskutesse," lisab Liina.*

(56) ('obj', '\*', ", "), ('obl', 'adit', ", ")

*Mänd lööb salamahti enesele naba kohal risti ette ja poetab veidra pesukompsu nurka.*

(57) ('obj', '\*', ", "), ('obl', 'adit', ", "), ('obl', 'all', ", ")

*"Oota siin, ma tulen kohe tagasi," ütlesin ma ja pistsin Tirtsule kaisujänku sülle.*

(58) ('obj', '\*', ", "), ('obl', 'adit', ", "), ('advmod', '<puudub>', ", ")

*Näiteks Tõnu Trubetskyst, kes ta üldse kirjandusse tõukas, või Hando Runnelist, kes oma mõttesügavusega võlus.*

Illatiivi esines liigutamisverbide lausemallides aditiiviga võrreldes üle poole vähem: kui aditiiv esines 21 verbi lausemallides, siis illatiiv ainult kümne. See tuleneb ilmselt

sellest, et paljudel sõnadel kasutatakse aditiivi koha ja illatiivi sõltuvuse väljendamiseks (EKK 2020: 216) ning liigutamisel on tähendusest tulenevalt levinum koha väljendamine. Siiski on illatiivse määruse kui tüüpilise sihtkäändega määruse lausemallid liigutamisverbide seas üpris laialt levinud. Subjekti, objekti ja illatiivis määrusega lausemall (59) esineb kuue verbiga: *riputama*, *paiskama*, *toppima*, *paigutama*, *viima* ja *tooma*. Mitmele liigutamisverbile tuvastati ka lausemall, mis koosnes ainult objektist ja illatiivis määrusest (60). See leiti kolmele verbile: *tirima*, *torkama*, *sõidutama*. Lisaks neile tuvastati veel ainult üks illatiivis nimisõnalist määrust sisaldav lausemall, kus lisaks subjektile, objektile ja illatiivis määrusele on ka määrsõnaline määrus *advmod* (61). See esines aga ainult ühe verbiga: *pistma*.

(59) ('nsubj', '\*', ", "), ('obj', '\*', ", "), ('obl', 'ill', ", ")

*Ülikoolid toovad ajud Tartusse, ent vaid Tartu ise saab hoolitseda nende siia kinnistumise eest.*

(60) ('obj', '\*', ", "), ('obl', 'ill', ", ")

“Eks tulevikus maksan,” ütles ta ja torkas *kviitungi taskusse*.

(61) ('nsubj', '\*', ", "), ('obj', '\*', ", "), ('obl', 'ill', ", "), ('advmod', '<puudub>', ", ")

*Ta pistis pardi niikauaks pükstesse, kui särki ja pluusi selga pani.*

Liigutamisverbide seas olid sagedased ka mallid, mis sisaldasid nimisõnalist määrust komitatiivis. Neist suurima kattuvusega oli lausemall, mille lauseliikmeteks on subjekt, objekt ja nimisõnaline määrus komitatiivis (62). See lausemall tuvastati neljale verbile: *loopima*, *vedama*, *tõukama* ja *tõmbama*. Lisaks neile esines kaks lausemalli ka ühe verbiga: verbi *lükka* puhul objekti ja komitatiivse määrusega lausemall (63) ja verbi *sõidutama* puhul subjekti, objekti, komitatiivse määruse ja määrsõnalise määrusega lausemall (64). Niisiis oli see lauseliige esindatud kokku kuue verbi kolmes lausemallis.

(62) ('nsubj', '\*', "", ""), ('obj', '\*', "", ""), ('obl', 'kom', "", "")

*Nad loopisid mind kividega.*

(63) ('obj', '\*', "", ""), ('obl', 'kom', "", "")

*Lükkasin läikivat tera jalaga, mängitasin teda käes, silitasin tema siledat pead.*

(64) ('nsubj', '\*', "", ""), ('obj', '\*', "", ""), ('obl', 'kom', "", ""), ('advmod', '<puudub>', "", "")

*“Nii ma sõidutasin neid oma autoga,” räägib ta.*

Lausemallides tulid esile ka lauseliikmed, mis olid asukohakäändetes ehk inessiivis ja adessiivis. Nimisõnaline määrus adessiivis esines kokku üheksa ja nimisõnaline määrus inessiivis kolme verbi lausemallides. Mallides, kus on nimisõnaline määrus adessiivis ja mis hõlmavad mitut verbi, varieerub subjekti ja objekti lausemalli liikmeks tuvastamine. Nii subjekti, objekti kui adessiivis nimisõnalise määrusega lausemalli (65) kui ka ilma objektita lausemalli (66) jagavad kolm verbi, vastavalt *loopima*, *langetama*, *sõidutama* ja *paiskama*, *viima*, *tooma*. Objektita lausemalli realisatsioonid olid objektita märgendusvigade tõttu või tähendustes, mis polnud otseselt liigutamisega seotud. Ilma subjektita lausemall (67) leiti kahele verbile: *tirima* ja *laduma*. Üks lausemall on adessiiviga veel (68), kuid see on unikaalne ainult ühele verbile: *vedama*.

(65) ('nsubj', '\*', "", ""), ('obj', '\*', "", ""), ('obl', 'ad', "", "")

Lepistiku tsirkuse *artistid* aga loobivad *jalgadel kuubikuid* ja silindreid ning lõpus koguni inimesi.

(66) ('nsubj', '\*', "", ""), ('obl', 'ad', "", "")

“Tesaitetemaltlöögi!” paiskas *teine kiirukõnel*.

(67) ('obj', '\*', "", ""), ('obl', 'ad', "", "")

Nad löid naise pikali, tirisid *tal käekoti* õlalt, mehele virutati rusikahoop näkku.

(68) ('nsubj', '\*', ", "), ('obj', '\*', ", "), ('obl', 'ad', ", "), ('advmod', '<puudub>', ", ")

*Seekord vedas meid koduloo kõrgnivool kohalik guru hr Olavi Vallimäe.*

Nimisõnaline määrus inessiivis esineb lauseliikmena ainult kahes lausemallis: ühes koos subjekti ja objektiga (69) ja teises ainult subjektiga (70). Esimene neist on lausemalliks kahele verbile: *paiskama* ja *liigutama*, teine aga ainult ühele verbile: *torkama*. Liigutamise tähenduses *torkama* selles mallis aga ei esinenud, vaid vajas liigutamise väljendamiseks ka objekti.

(69) ('nsubj', '\*', ", "), ('obj', '\*', ", "), ('obl', 'in', ", ")

*Instituudis paiskas ta segamini kogu seal aastakümneid kestnud oleskelu.*

(70) ('nsubj', '\*', ", "), ('obl', 'in', ", ")

Samal hetkel torkas *miski südames* valusasti, ja uni oli kui peoga pühitud.

Kohakäänetest oli veel esindatud lähtekääne elatiiv, kuid lausemallides ei esinenud kordagi lauseliiget teises lähtekäändes ablatiivis. Nimisõnaline määrus elatiivis esines aga neljas lausemallis, millest üks esines kahe verbiga (71) ja teised kolm ühe verbiga (72–74). Need olid vastavalt *raputama* ja *tõmbama* (71), *tooma* (72), *tirima* (73) ja *sõidutama* (74). Niisiis esines elatiivis nimisõnaline määrus kokku viie verbi neljas lausemallis.

(71) ('obj', '\*', ", "), ('obl', 'el', ", ")

*Tõmbas särgi seljast, jäi palja ülakehaga peegli ette.*

(72) ('nsubj', '\*', ", "), ('obj', '\*', ", "), ('obl', 'el', ", ")

*Janis tõi suled loomaaiast, loomadest ja lindudest sai alati tuge, jõudu ja energiat.*

(73) ('nsubj', '\*', "", ""), ('obj', '\*', "", ""), ('obl', 'el', "", ""), ('advmod', '<puudub>', "", "")

*Mendijupats tiris kabuurist lagedale "Makarovi" püstoli ja suunas relva minu poole.*

(74) ('nsubj', '\*', "", ""), ('obj', '\*', "", ""), ('obl', 'adit', "", ""), ('obl', 'el', "", "")

Seepeale sõidutas *pereisa Pärnust* 30 kilomeetri kaugusele *koju lastetohtri*, kes vaatas imiku üle ja leidis , et temaga on kõik korras.

Liigutamisverbide kohta oli üllatavalt sage lauseliige ma-infinitiivne laiend, mis esines viie verbi neljas lausemallis. Lausemallid on üldiselt väga sarnased ja erinevad subjekti, objekti ja määrsõnalise määruse esinemise poolest. Subjekti, objekti, ma-infinitiivse laiendi ja määrsõnalise määrusega lausemall (75) tuvastati verbidele *sättima* ja *panema*, ilma subjekti ja määrsõnalise määruseta mall (76) verbile *riputama*, ilma objektita lausemall (77) verbile *pistma* ja määrsõnalise määruseta mall (78) verbile *tõukama*.

(75) ('nsubj', '\*', "", ""), ('obj', '\*', "", ""), ('xcomp', 'ill', "", 'ma'),

('advmod', '<puudub>', "", "")

*Tüdruk ilmselt sätib end magama.*

(76) ('obj', '\*', "", ""), ('xcomp', 'ill', "", 'ma')

Väljas, riputab *pesu kuivama* ja vahib pilvi.

(77) ('nsubj', '\*', "", ""), ('xcomp', 'ill', "", 'ma'), ('advmod', '<puudub>', "", "")

Mitte keegi ei märganud ega tundnud vähimatki muret, igatahes mitte sõjaväelased, kuni *Pannküla äkki jooksmas* pisti.

(78) ('nsubj', '\*', "", ""), ('obj', '\*', "", ""), ('xcomp', 'ill', "", 'ma')

Kõik see tõukab *inimest* mässuideooloogiat omaks *võtma*.

Teiste lauseliikmetega lausemallidest ei olnud ükski lausemall mitmel verbil ühine, küll aga esines veel üks lauseliige mitme verbi lausemallides: nimisõnaline määrus genitiivis koos kaassõnaga *alla*. See esines kahe verbi erinevates mallides: verbi *seadma* mallis subjekti ja objektiga (79) ja verbi *toppima* mallis objektiga (80).

(79) *seadma* ('nsubj', '\*', "", ""), ('obj', '\*', "", ""), ('obl', 'gen', 'alla', "")

Vahekohtu teatav *ebajärjekindlus* seab aga taolise *järelduse kahtluse alla*.

(80) *toppima* ('obj', '\*', "", ""), ('obl', 'gen', 'alla', "")

Põlvitas mu kõrval ja nuttis ning toppis *käsi mu külje alla*.

Kaassõnadest esines mallides veel *külge* (81), *ette* (82), *poole* (83) ja *juurde* (84), igauks eri verbidega omas mallis. Tüüpiliselt on need mallid ilma subjektita.

(81) *riputama* ('obj', '\*', "", ""), ('obl', 'gen', 'külge', "")

Otsisin kummutist välja ühe tädi voodilinadest ja riputasin *selle kardinakonksude külge*.

(82) *asetama* ('obj', '\*', "", ""), ('obl', 'gen', 'ette', "")

Üks neist astub aeglaselt minu juurde ning asetab mu parema *näo ette* veidra õhus rippuva *kuulikese*.

(83) *keerama* ('obj', '\*', "", ""), ('obl', 'gen', 'poole', "")

Keerasin *näo Kalle poole*.

(84) *sõidutama* ('nsubj', '\*', "", ""), ('obj', '\*', "", ""), ('obl', 'gen', 'juurde', "")

Nad istusid siin autosse ja *Kristjan* sõidutas *teda* oma *maja juurde*.

Lauseliikmetest esines liigutamisverbidele tuvastatud lausemallide osana veel verbiga *torkama* komplementlause, mis osutus kõigis näidetes valesti komplementlauseks märgendatud kõrvallausealiseks subjektiks (85), verbiga *paigutama* nimisõnaline määrus translatiivis (86) ja verbiga *viima* terminatiivis (87), verbiga *seadma* translatiivne öeldistäitemäärus (88) ja verbiga *tirima* määrusena toimiv infiniittarind des-vormis (89).

(85) torkama ('obl', 'adit', "", ""), ('ccomp', '<puudub>', "", "")

Ometi torkab *silma*, et Venemaa suhtes *valitseb* vaikus.

(86) paigutama ('obj', '\*', "", ""), ('obl', 'tr', "", "")

Tulemuseks võttis linn laenu ja paigutas 102 miljonit *krooni* ASi Anne Soojus omandis seisva pooleli oleva Luunja katlamaja *viimiseks* turba-puiduküttele.

(87) viima ('nsubj', '\*', "", ""), ('obl', 'term', "", "")

Igasuguse kontrolli ja sanktsiooni rakendamise võimaluse *puudumine* viib *anarhiani*.

(88) seadma ('nsubj', '\*', "", ""), ('obj', '\*', "", ""), ('xcomp', 'tr', "", "")

Mehe käed värisesid samuti , kui *ta helitugevuse talutavaks* seadis.

(89) tirima ('nsubj', '\*', "", ""), ('obj', '\*', "", ""), ('advmod', '<puudub>', "", ""), ('advcl', '<puudub>', "", 'des')

*Arvatavasti* *gospodi pomilui pomisedes* tiris *mees surnukeha* üle läve, vinnas seina külge seotud hobusele, päästis ratsutid ja andis loomale malakat.

## 5. ANALÜÜS

Järgnevas peatükis hinnatakse saadud tulemuste kvaliteeti, tuuakse välja nii tuvastamata jäänud kui ka valesti tuvastatud lausemallid ning nende vigade tekkimise põhjused.

### 5.1. Kvaliteedi hindamine

Nagu tulemustest näha, kasutan meetodi kvaliteedi hindamiseks verbe, mille põhitähendus on liigutamine. Need verbid on aga tihti polüseemsed, mistõttu esinevad need ka tähendustes, mis pole liigutamisega seotud. Nende tähenduste mallid võivad liigutamise omadega kokku minna, kuid ei pruugi, mis tähendab, et neil tähendustel võivad olla ka täiesti omad mallid. Kuigi need pole otseselt liigutamisega seotud, hindan ma analüüsi osas ka nende mallide õigsust, sest käsitsi koostati mallid kõigi verbi tähenduste jaoks ning automaatselt pole tähendusi võimalik üksteisest eristada.

Hindasin kvaliteeti käsitsi koostatud materjali põhjal eraldi nii lauseliikmete kui lausemallide jaoks. Käsitsi koostatud materjal on leitav lisades. Kvaliteedi hindamiseks panin iga verbi kohta kirja, kui paljud selle automaatselt tuvastatud lausemallid ja neis esinevad lauseliikmed esinesid ka käsitsi koostatud materjalis (TP), kui paljud käsitsi koostatutest jäid automaatsel tuvastamata (FN) ning kui paljud automaatselt tuvastatutest käsitsi koostatud materjalis ei esinenud (FP). Lisaks sellele panin ma iga verbi juurde kirja ka selle, kui paljud lauseliikmed ja lausemallid olid küll õiged, aga lahterdusid valesti tuvastatute alla sellepärast, et need olid materjali vähesuse tõttu käsitsi koostatud mallide seast välja jäänud. Need tulemused on toodud tabelis 7.

Tabel 7. Lauseliikmete automaatse tuvastamise kvaliteet verbide kaupa

Verb	TP	FN	FP	Õige, aga käsitsi koostatutest puudu
<i>riputama</i>	5	-	2	1
<i>tirima</i>	5	-	4	1
<i>loopima</i>	4	1	3	1
<i>paiskama</i>	4	-	4	1
<i>raputama</i>	3	1	1	-
<i>torkama</i>	5	-	3	1
<i>sättima</i>	4	1	2	1
<i>toppima</i>	5	-	1	-
<i>laduma</i>	4	1	2	-
<i>poetama</i>	4	1	1	-
<i>pistma</i>	5	-	2	1
<i>asetama</i>	4	1	1	-
<i>langetama</i>	3	-	2	-
<i>sõidutama</i>	5	-	4	1
<i>paigutama</i>	5	-	2	-
<i>vedama</i>	5	2	1	-
<i>heitma</i>	3	2	1	-
<i>lukkama</i>	4	1	2	-
<i>tõukama</i>	4	1	3	1
<i>seadma</i>	5	1	1	-
<i>keerama</i>	4	3	-	-
<i>viskama</i>	4	1	1	-
<i>tõmbama</i>	5	2	1	-

<i>tõstma</i>	3	3	1	-
<i>liigutama</i>	2	-	2	1
<i>viima</i>	6	2	2	-
<i>tooma</i>	6	3	2	-
<i>panema</i>	5	1	1	-
<b>kokku</b>	<b>121</b>	<b>28</b>	<b>52</b>	<b>10</b>

Nende andmete põhjal arvutati lauseliikmete automaatse tuvastuse saagiseks 0,812 ja täpsuseks 0,699. Kui lahutada valesti tuvastatud lauseliikmetest maha need, mis olid tegelikult õiged, aga tuvastati valesti ainult käsitsi koostatud mallide puudujääkide tõttu, oli lauseliikmete automaatse tuvastamise täpsus 0,742. Kokkuvõttes tähendab see, et automaatselt jäi tuvastamata 19% lauseliikmetest ja tuvastatud lauseliikmetest oli õigesti tuvastatuid ümardatult 74%. Järgnevalt vaadatakse, kuidas olid tulemused lausemallide lõikes.

Tabel 8. Lausemallide automaatse tuvastamise kvaliteet verbide kaupa

Verb	TP	FN	FP	Õige, aga käsitsi koostatutest puudu
<i>riputama</i>	1	1	3	1
<i>tirima</i>	-	2	6	-
<i>loopima</i>	2	1	3	1
<i>paiskama</i>	1	1	4	1
<i>raputama</i>	-	2	2	-
<i>torkama</i>	1	1	4	-
<i>sättima</i>	-	2	3	-
<i>toppima</i>	1	2	2	1
<i>laduma</i>	1	2	2	-

<i>poetama</i>	1	2	2	-
<i>pistma</i>	-	3	4	-
<i>asetama</i>	-	3	2	-
<i>langetama</i>	1	-	2	-
<i>sõidutama</i>	1	1	5	-
<i>paigutama</i>	3	-	2	-
<i>vedama</i>	1	3	2	-
<i>heitma</i>	-	3	1	-
<i>lükama</i>	2	1	2	-
<i>tõukama</i>	1	2	4	1
<i>seadma</i>	2	2	1	-
<i>keerama</i>	2	3	-	-
<i>viskama</i>	-	3	2	-
<i>tõmbama</i>	3	2	2	-
<i>tõstma</i>	1	4	1	-
<i>liigutama</i>	-	1	2	1
<i>viima</i>	2	5	3	1
<i>tooma</i>	3	4	2	-
<i>panema</i>	1	5	2	-
<b>kokku</b>	<b>31</b>	<b>61</b>	<b>70</b>	<b>7</b>

Lausemallide automaatse tuvastamise saagis oli 0,337 ja täpsus 0,307. Käsitsi koostatud mallide puudujääke arvesse võttes oli lausemallide automaatse tuvastamise täpsus 0,330. See tähendab, et tuvastamata jäi 66% käsitsi koostatud materjalis olevatest lausemallidest ja kõigist automaatselt tuvastatud lausemallidest oli õigesti tuvastatud 33%. Automaatse tuvastuse puudujääkide põhjuseid avatakse järgmises peatükis.

## 5.2. Tuvastamata jäänud lauseliikmed ja lausemallid

Selles peatükis uuritakse lauseliikmeid ja lausemalle, mis esinesid käsitsi koostatud materjalis kuid jäid automaatselt tuvastamata. Esimesena vaadatakse, miks jäid tuvastamata üksikud lauseliikmeid ning teisena seda, miks jäid tuvastamata terved lausemallid.

### 5.2.1. Tuvastamata jäänud lauseliikmed

Iga tuvastamata jäänud lauseliikme jaoks tuuakse välja verbid, mille lauseliikmena see tuvastamata jäi, mitu neid verbe oli ning mitme verbi lausemallides lauseliige käsitsi koostatud materjalis pigem seotud laiendi rollis esines.

Tabel 9. Automaatselt tuvastatud lausemallidest välja jäänud lauseliikmed

Tuvastamata jäänud lauseliige	Verbid, mille lauseliige tuvastamata jäi	Mitmel verbil lauseliige tuvastamata jäi	Mitme verbi seotud laiend lauseliige on
(‘obl’, ‘ill’, ‘’, ‘’)	<i>loopima, sättima, laduma, asetama, vedama, heitma, lükkama, tõukama, seadma, keerama, viskama, tõmbama, tõstma, panema</i>	14	20
(‘xcomp’, ‘ill’, ‘’, ‘ma’)	<i>vedama, heitma, tõmbama, viima</i>	4	6
(‘obl’, ‘adit’, ‘’, ‘’)	<i>heitma, keerama, tõstma</i>	3	23
(‘obl’, ‘abl’, ‘’, ‘’)	<i>viima, tooma</i>	2	2
(‘xcomp’, ‘tr’, ‘’, ‘’)	<i>tooma</i>	1	2
(‘obl’, ‘tr’, ‘’, ‘’)	<i>keerama</i>	1	1
(‘obl’, ‘es’, ‘’, ‘’)	<i>tooma</i>	1	1
(‘obl’, ‘all’, ‘’, ‘’)	<i>raputama</i>	1	22
(‘obl’, ‘term’, ‘’, ‘’)	<i>tõstma</i>	1	2

Tuvastamata jäid nimisõnaline määrus illatiivis, aditiivis, ablatiivis, essiivis, allatiivis ja terminatiivis, ma-infinitiivis laiend ja öeldistaitemäärus translatiivis. Kõige sagedamini jäi tuvastamata nimisõnaline määrus illatiivis, mis moodustas kõigist tuvastamata jäänud lauseliikmetest 50% ning ei tuvastatud 70% verbides, milles see käsitsi koostatud mallis seotud laiendina esines. Korpusematerjali uurides tundus, et nimisõnaline määrus illatiivis jäi lausemallidest välja seetõttu, et aditiivi eelistavad sõnad kivinevad üldiselt püsiühenditeks tihemini kui illatiivi eelistavad sõnad, mis tähendab, et paljud sagedased sõnad kasutavad just aditiivi (Kio 2006: 126). Põhjus võis seisneda ka selles, et paljude sõnade puhul kasutatakse aditiivi alati kohta ja illatiivi sõltuvuse väljendamiseks (EKK 2020: 216). Aditiivi eelistamise väidet toetab ka see, et 14 verbil, kus illatiivis määrus tuvastamata jäi, oli kümnes tuvastatud nimisõnaline määrus aditiivis. Aditiivis määrus jäi tuvastamata ainult kolmel verbil ehk 13% selle seotud laiendiga verbidest, mis on 53% parem tulemus kui illatiivse nimisõnalise määruse oma.

Sageduselt teine tuvastamata jäänud lauseliige oli *ma*-infinitiivne laiend, mis jäi tuvastamata kahel kolmandikul seda nõudvatest verbidest. Korpust vaadates jäi silma, et nende puhul on tegemist küll lingvistiliselt oluliste, aga üpris perifeerse kasutusega mallidega (90–93). Ilmselt jäid need tuvastamata just sellel töös rakendatava statistikal põhineva meetodi tõttu, mis tuvastab ainult kõige sagedasemaid malle. Halliga on mallides märgitud laiendit, mis jäid seotud ja vaba laiendi piirimaile, mistõttu ei võetud neid kvaliteedi hindamisel arvesse.

(90) ('nsubj', '\*', ' ', ' '), ('obj', '\*', ' ', ' '), ('xcomp', 'ill', ' ', 'ma')

Lapsed aitavad kokku korjata suurema prügi, *mille autod kohe minema* veavad.

(91) ('nsubj', '\*', ' ', ' '), ('xcomp', 'ill', ' ', 'ma'), ('obl', 'adit', ' ', ' ')/  
('obl', 'all', ' ', ' ')/('obl', 'ill', ' ', ' ')

Väljast *tulnu* aga heitis *narile puhkama*.

(92) ('nsubj', '\*', ' ', ' '), ('obj', '\*', ' ', ' '), ('xcomp', 'ill', ' ', ' ', 'ma')

*Eesav tõmbas uue tiku põlema ja lähendas jälle pealaele.*

(93) ('nsubj', '\*', ' ', ' '), ('obj', '\*', ' ', ' '), ('xcomp', 'ill', ' ', ' '),

('obl', 'adit', ' ', ' ')/('obl', 'all', ' ', ' ')/('obl', 'ill', ' ', ' ')

"Tahad, *ma* viin *sind* *Volgale* paadiga sõitma?"

Tüüpiliselt LÄHTEKOHA rolli väljendav nimisõnaline määrus ablatiivis jäi tuvastamata mõlema kahe verbi jaoks, mis seda nõudsid. Verbi *viima* puhul pole see üllatav, sest nagu *ma*-infiniitvise laiendigagi on selle LÄHTEKOHTA nõudev mall väga spetsiifilise tähenduse väljendamiseks ja seega ka vähe kasutatav.

(94) ('nsubj', '\*', ' ', ' '), ('obl', 'ill', ' ', ' '), ('obl', 'abl', ' ', ' ')

*Tänavalt* viib *üks* otse *maaliateljeesse*.

*Tooma* puhul on tulemus üllatavam, sest LÄHTEKOHA väljendamine tundus intuiitiivselt olevat selle verbi jaoks üpris tüüpiline.

(95) ('nsubj', '\*', ' ', ' '), ('obj', '\*', ' ', ' '), ('obl', 'abl', ' ', ' ')

Vahepeal tõi *Salih turult* pätsi *leiba*, selle oli ta vahetanud granaadi vastu.

Siiski on *tooma* seotud laiendiks automaatselt tuvastatud nimisõnaline määrus elatiivis (72), mis iseloomustab tavaliselt samuti LÄHTEKOHTA. Niisiis on ablatiivis määrus välja jäänud ilmselt elatiivi eelistamise tõttu.

Viis seotud laiendit jäi tuvastamata ainult ühel verbil. Neist esines käsitsi koostatud mallides kõige rohkem nimisõnaline määrus allatiivis, mis jäi tuvastamata ainult verbi *raputama* puhul. Sellele verbile ei tuvastatud ka ühtegi muud tüüpilist SIHTKOHA rolli

väljendavat laiendit. Selle põhjuseks võib olla see, et verbil *raputama* on tüüpiliselt SIHTKOHT ainult siis, kui raputatakse mõnda loendamatu eset, nii et ainult väga spetsiifilises tähenduses (96). Nimisõnaline määrus allatiivis võib esineda ka SAAJA rollis, aga ka see on raputamise jaoks väga spetsiifiline tähendus (97).

(96) Geša võttis taskust peeglikese, pani selle teleriesisele klaaslauale ja raputas  
*peeglile valget pulbrit.*

(97) "SSSA TULINE KURRAT!" vannub ta ning raputab *Lauritsale* rusikat.

Tähenduse spetsiifilisuse tõttu jäi tuvastamata ka verbi *keerama* seotud laiend nimisõnaline määrus translatiivis. See esineb ainult siis, kui keerama tähendab “mingit nuppu keerates seisundimuutust põhjustama”.

(98) Seda riistapuud (mille nimi oli teleekraan) sai küll *vaiksemaks* keerata, aga võimatu oli seda täiesti välja lülitada.

Tuvastamata jäi ka verbi *tõstma* seotud laiendina esinev terminatiivis obliikva. Põhjus tuleneb ilmselt jällegi verbi polüseemiast ja korpuse mahust, sest nimisõnaline määrus terminatiivis esineb ainult selles *tõstma* tähenduses, kus suurendatakse mingisugust hulka või määra või viiakse midagi kõrgemale tasemele.

(99) Languse peatamiseks tõstis Kanada keskpank intressimäära viie *protsendini*.

Translatiivne öeldistäitemäärus oli seotud laiend *tooma* ja *seadma* mallides, kuid jäi *tooma* puhul tuvastamata. Siiski on minu hinnangul näiteid vaadates intuitiivselt tegemist üpris tüüpilise *tooma* kasutusjuhtumiga (100–102). Algandmeid uurides tuli aga välja, et translatiivne öeldistäitemäärus esines vaid 2,5% *tooma* lausetes.

- (100) Ta tõi *vabanduseks* selle, et nii väga oli ta soovinud õnnistada raportit millegi huvitavaga.
- (101) Tõi *põhjuseks* tervislikku seisundit või pagan teab mida veel.
- (102) Aga ta oleks meeleldi selle *ettekäändeks* toonud, et minema hakata.

Verbil *tooma* jäi tuvastamata ka nimisõnaline määrus essiivis, mis on verbi *tooma* seotud laiend translatiivse öeldistaitemäärusega samas tähenduses.

- (103) *Näitena* tõi Reinsalu lubaduse, kus allakirjutanud kohustavad rakendama õiglast palka.
- (104) Ühe lahkumise *põhjuse*na tõi ta ka võimuvahetuse, kuid ta lisis, et preemia ja ametist lahkumise vahel ei ole mingit seost.

Võimalik, et need lauseliikmed jäid seotud laienditena tuvastamata, sest meetod käsitleb korpuses väga sagedasti esinevaid ja harva esinevaid verbe täpselt samamoodi, kuigi sagedasi verbe kasutatakse tüüpiliselt rohkemates tähendustes ja seega ka suurema hulga mallidega, mistõttu vajaksid need eraldi lähenemisi.

### **5.2.2. Tuvastamata jäänud lausemallid**

Lausemallide tuvastamise madala saagise ja täpsuse taga on mitu erinevat põhjust. Hindamisel ei tooda välja kõiki tuvastamata jäänud malle, vaid ainult vead ja nende tekkimise põhjused. Vigu ja põhjuseid ei panda kirja lausemallide kaupa, sest malle on kokku 92 tükki ja kuna verbiti on nende tuvastamata jäämise põhjused erinevad, pole mõtet neid ka hindamise jaoks verbiülesteks mallideks koondada. Seetõttu olen töös toonud välja ainult vead, mille tõttu mall tuvastamata jäi ja nende vigade tekkimise põhjused. Need on esitatud tabelis 10.

Tabel 10. Tuvastamata jäänud lausemallide põhjused

Tuvastamata jäämise põhjus	Vea liik	Vea tekkimise põhjus	Verbid	Kokku
(‘advmod’, ‘<puudub>, ‘, ‘)	üleliigne lauseliige	lauseliikmed esinesid üksteisega liiga sagedasti koos	<i>paiskama, raputama, sättima, toppima, laduma, pistma (2x), asetama, heitma, seadma, viskama, liigutama, viima, tooma, panema (2x), tõstma</i>	17
(‘obl’, ‘ill’, ‘, ‘)	tuvastamata lauseliige	lauseliige ei esinenud verbiga küllalt sagedasti koos	<i>loopima, sättima, laduma, asetama, vedama, lükkama, tõukama, seadma, keerama, viskama, tõmbama, tõstma, panema</i>	13
(‘nsubj’, ‘*, ‘, ‘)	puuduv lauseliige	lauseliikmed ei esinenud üksteisega küllalt sagedasti koos	<i>riputama, tirima, raputama, torkama, poetama, asetama, viskama</i>	7
(‘obl’, ‘adit’, ‘, ‘)	tuvastamata lauseliige	lauseliige ei esinenud verbiga küllalt sagedasti koos	<i>heitma, keerama, tõstma</i>	3
(‘xcomp’, ‘ill’, ‘, ‘ma’)	tuvastamata lauseliige	lauseliige ei esinenud verbiga küllalt sagedasti koos	<i>heitma, tõmbama, vedama, viima</i>	4
(‘nsubj’, ‘*, ‘, ‘) ja (‘advmod’, ‘<puudub>, ‘, ‘)	puuduv lauseliige ja üleliigne lauseliige	üks lauseliige ei esinenud teistega küllalt sagedasti koos ja teine esines teistega liiga sagedasti koos	<i>tirima, tõukama</i>	2
(‘obl’, ‘abl’, ‘, ‘)	tuvastamata lauseliige	lauseliige ei esinenud verbiga küllalt sagedasti koos	<i>viima (2x)</i>	2
(‘obl’, ‘abl’, ‘, ‘)	puuduv lauseliige	lauseliikmed ei esinenud üksteisega küllalt sagedasti koos	<i>tooma</i>	1
(‘obj’, ‘*, ‘, ‘)	puuduv lauseliige	lauseliikmed ei esinenud üksteisega	<i>viima</i>	1

		küllalt sagedasti koos		
(‘obl’, ‘es’, ‘’, ‘’)	puuduv lauseliige	lauseliikmed ei esinenud üksteisega küllalt sagedasti koos	<i>tooma</i>	1
(‘xcomp’, ‘tr’, ‘’, ‘’)	puuduv lauseliige	lauseliikmed ei esinenud üksteisega küllalt sagedasti koos	<i>tooma</i>	1
(‘obj’, ‘*’, ‘’, ‘’) ja (‘advmod’, ‘<puudub>’, ‘’, ‘’)	puuduv lauseliige ja üleliigne lauseliige	üks lauseliige ei esinenud teistega küllalt sagedasti koos ja teine esines teistega liiga sagedasti koos	<i>pistma</i>	1
(‘obl’, ‘el’, ‘’, ‘’) ja (‘advmod’, ‘<puudub>’, ‘’, ‘’)	puuduv lauseliige ja üleliigne lauseliige	üks lauseliige ei esinenud teistega küllalt sagedasti koos ja teine esines teistega liiga sagedasti koos	<i>sõidutama</i>	1
(‘obl’, ‘adit’, ‘’, ‘’) ja (‘obl’, ‘all’, ‘’, ‘’)	lausemallist koos puudu	lauseliikmed ei esinenud üksteisega küllalt sagedasti koos + meetodi filter	<i>panema</i>	1
(‘obl’, ‘ill’, ‘’, ‘’) ja (‘obl’, ‘all’, ‘’, ‘’)	lausemallist koos puudu	lauseliikmed ei esinenud üksteisega küllalt sagedasti koos + meetodi filter	<i>panema</i>	1
(‘obl’, ‘ad’, ‘’, ‘’) ja (‘obl’, ‘kom’, ‘’, ‘’)	lausemallist koos puudu	lauseliikmed ei esinenud üksteisega küllalt sagedasti koos + meetodi filter	<i>vedama</i>	1
(‘ccomp’, ‘<puudub>’, ‘’, ‘’)	tuvastamata lauseliige	lauseliige ei esinenud verbiga küllalt sagedasti koos	<i>poetama</i>	1
(‘obl’, ‘tr’, ‘’, ‘’)	tuvastamata lauseliige	lauseliige ei esinenud verbiga küllalt sagedasti koos	<i>keerama</i>	1
(‘obl’, ‘adit’, ‘’, ‘’)	üleliigne lauseliige	lauseliikmed esinesid üksteisega liiga sagedasti koos	<i>toppima</i>	1
(‘nsubj’, ‘*’, ‘’, ‘’) ja (‘obj’, ‘*’, ‘’, ‘’)	kogu lausemall oli tuvastamata	meetodi filter	<i>tõstma</i>	1

92 verbikesksest käsitsi koostatud mallist jäi tuvastamata 61 ehk 66% mallides. Põhjused nende jaoks olid erinevad. Kõige sagedasem viga oli see, et tuvastatud lausemallis oli üleliigne määrsõnaline määrus, mis moodustas kõigist tuvastamata jäämist tekitanud vigadest 28%. Kui määrsõnaline määrus automaatsetest mallidest eemaldada, oleks tuvastamata jäänud malle olnud 18% võrra vähem ehk 48%. Teine kõige suurem probleemide allikas oli tuvastamata jäänud nimisõnaline määrus illatiivis, mis moodustas vigadest 21%. Kolmel korral oli malli tuvastamata jäämise põhjustanud veaks ka tuvastamata jäänud nimisõnaline määrus aditiivis, nii et kokku moodustas sihtkäändes määruse tuvastamata jäämine vigadest 26%. Kolmas kõige sagedasem malli tuvastamata jäämise põhjustanud viga oli subjekti mallist puudumine, mis moodustas vigadest 11%. Kokku moodustasid need kolm probleemi tuvastamata jäämise vigadest 66% ehk nende parandamisel paraneks lausemallide tuvastus kahe kolmandiku võrra. Selle tulemusel jääks praeguse 66% asemel tuvastamata ainult 23% lausemallidest.

Tuvastamata jäämise põhjuseid saab koondada ka suuremate vealiikide ja nende tekkepõhjuste alla. Kokku oli neid kuus eri liiki:

- lauseliige oli lausemallis üleliigne, sest lauseliikmed esinesid üksteisega liiga sagedasti koos,
- lauseliige jäi tuvastamata, sest see ei esinenud vähemalt 5% verbi lausetes,
- lauseliige oli lausemallist puudu, sest see ei esinenud teiste lauseliikmetega vähemalt 5% verbi lausetes koos,
- üks lauseliige oli mallis üleliigne, sest see esines teistega liiga sagedasti koos ja teine oli lausemallist puudu, sest see ei esinenud teistega küllalt sagedasti koos,
- kaks lauseliiget olid lausemallist koos puudu, sest need ei esinenud üksteisega küllalt sagedasti koos ja kui esineksidki, jääks see või üks teine tuvastatud lausemall meetodi filtri tõttu tuvastamata,
- kogu mall jäi tuvastamata, sest meetod võtab ainult pikima lausemalli.

Toodud põhjustest oli kõige sagedasem see, et lauseliige ei esinenud verbiga küllalt sagedasti koos, mistõttu jäi see seotud laiendina täielikult tuvastamata. See oli 40% vigade põhjuseks. Sellest oli rohkem juttu eelmises peatükis 5.2.1. Järgmine sagedasim põhjus oli see, et lauseliikmed esinesid üksteisega liiga sagedasti koos, mis tekitas malli üleliigse lauseliikme. Neist oli 16 juhul 17-st tegemist määrsõnalise määrusega. Kokku moodustas see klass vigadest 29%. Kuna üleliigsus viitab pigem valesti tuvastamisele kui tuvastamata jäämisele, tuleb sellest pikemalt juttu peatükis 5.3.

Kolmas kõige sagedasem viga oli see, et lauseliige oli mallist puudu, sest see ei esinenud teiste lauseliikmetega koos vähemalt 5% verbi lausetes. See moodustas tuvastamata jäämise põhjustanud vigadest 19%. Selle alla jaotusid ka probleemid subjekti ja objekti mallist puudumisega. Subjekt ja objekt jäävad mallist välja peamiselt seetõttu, et need saavad eesti keeles olla lauses väljajätelised (vt ptk 6.2), mis hakkab mõjutama ka nende esinemise sagedust teiste lauseliikmetega. Kuigi objekti kuulumine verbi seotud laiendite hulka võib verbilt verbile paljuski erineda (Metslang 2017a: 261–262), peaks subjekt olema peaaegu kõikide verbide seotud laiend, erandiks on ainult nullvalentsed verbid (Karlsson 2002: 181). Objekti väljajätte tõttu jäi tuvastamata verbi *viima* lausemall subjekti, objekti ja terminatiivis obliikvaga (105). Selle asemel tuvastati see mall ilma objektita (106).

(105) ('nsubj', '\*', '', ''), ('obj', '\*', '', ''), ('obl', 'term', '', '')

Nende sõpruse aga lõpetas armastuskolmnurk Harrisoni naisega, *mis* viis *Claptoni* ka *narkoprobleemideni*.

(106) ('nsubj', '\*', '', ''), ('obl', 'term', '', '')

Alles Uus-Meremaal saavutatud *tulemused* viisid *selleni*, et õpikeskuste kavandamine sai alata ka Soomes.

Oli ka neli malli, mis jäid tuvastamata, sest üks lauseliige oli mallist puudu ja teine - määrsõnaline määrus - oli üleliigne. See juhtub, kui üks lauseliige ei esine teistega vähemalt 5% lausetes, aga teine esineb liiga sagedasti. Need moodustasid malli tuvastamata jäämise põhjustanud vigadest 7%. Niisiis moodustavad liigsed ja puuduvad lauseliikmed kokku 55% malli tuvastamata jäämist põhjustavatest vigadest.

Tuvastamata jäänud 61 mallist jäi kolm leidmata kahe probleemi kombinatsiooni tõttu: lauseliikmete koosinemine ei ületanud 5% künnist, aga kui see ületaks, jääks kas see või mõni teine lausemall meetodi filtri tõttu tuvastamata. See juhtub, sest hetkel võtab meetod kõigist genereeritustest mallidest iga lauseliikme jaoks kõige pikema lausemalli. Ainult sellel põhjusel jäi tuvastamata üks mall: verbi *tõstma* mall subjekti ja objektiga, sest *tõstma* jaoks oli olemas ka mall subjekti, objekti ja määrsõnalise määrusega, mis sisaldab mõlemat lauseliiget ja on esimesest mallist pikem. Sama probleem oleks juhtunud ka *panema* ja *vedama* mallidega, kui nende kaks lauseliiget oleks küllalt sagedasti koos esinenud: üks uus mall oleks saadud tuvastatud, aga teised oleks selle tulemusel tuvastamata jäänud.

### 5.3. Valesti tuvastatud lauseliikmed ja lausemallid

Siin peatükis uuritakse lauseliikmeid ja lausemalle, mis tuvastati automaatselt, aga käsitsi koostatud materjalis ei esinenud. Iga sellise juhtumi kohta tuuakse välja valesti tuvastatud lauseliige, mis verbidele see valesti tuvastati, mitu neid verbe oli ja mitu neist vigadest loeti vigadeks käsitsi koostatud materjali puuduste tõttu ja olid tegelikult õigesti tuvastatud.

Tabel 11. Valesti tuvastatud lauseliikmed

Valesti tuvastatud lauseliige	Verbid, mille lauseliige valesti tuvastati	Mitmel verbil lauseliige valesti tuvastati	Õige, aga käsitsi koostatutest puudu
-------------------------------	--	--	--------------------------------------

(‘advmod’, ‘<puudub>’, ‘’, ‘’)	<i>riputama, tirima, loopima, paiskama, raputama, torkama, sättima, toppima, laduma, poetama, pistma, asetama, langetama, sõidutama, paigutama, vedama, heitma, lükkama, tõukama, seadma, viskama, liigutama, viima, tooma, panema, tõstma</i>	26	-
(‘obl’, ‘ad’, ‘’, ‘’)	<i>tirima, loopima, paiskama, laduma, langetama, sõidutama, viima, tooma</i>	8	-
(‘obl’, ‘kom’, ‘’, ‘’)	<i>loopima, sõidutama, lükkama, tõukama, tõmbama</i>	5	1
(‘obl’, ‘ill’, ‘’, ‘’)	<i>riputama, tirima, paiskama, sõidutama</i>	4	4
(‘obl’, ‘in’, ‘’, ‘’)	<i>paiskama, torkama, liigutama</i>	3	1
(‘obl’, ‘adit’, ‘’, ‘’)	<i>sättima</i>	1	1
(‘obl’, ‘all’, ‘’, ‘’)	<i>pistma</i>	1	1
(‘obl’, ‘tr’, ‘’, ‘’)	<i>paigutama</i>	1	-
(‘advcl’, ‘<puudub>’, ‘’, ‘des’)	<i>tirima</i>	1	-
(‘ccomp’, ‘<puudub>’, ‘’, ‘’)	<i>torkama</i>	1	-
(‘xcomp’, ‘ill’, ‘’, ‘ma’)	<i>tõukama</i>	1	1

Kõige rohkem tuvastati automaatselt seotud laiendiks valesti mäarsõnaline määrus (*advmod*), mis moodustas 52 valesti tuvastatud lauseliikmest 50% ning oli valesti tuvastatud 28 verbist 26-l. See juhtus ilmselt seetõttu, et mäarsõnalist määrust saab kasutada kõigi asjaolumääruste (koht, aeg, viis, hulk, seisund) ja ka mõõndusmääruse väljendamiseks (Veismann jt 2020: 304). Asjaolumäärused on ühed kõige tüüpilisemad vabad laiendid (Karlsson 2002: 187) ning kuna mäarsõnaline määrus esineb kõigi asjaolumääruse liikidena, on sel väga lai kasutusala. Mäarsõnalised määrused on

nimisõnalistega võrreldes ka problemaatilisemad seetõttu, et vastupidiselt nimisõnalistele määrustele puudub määrsõnalistel kääne, mis aitaks nende sagedust hajutada.

Tüüpiliselt on vaba laiend ka **nimisõnaline määrus adessiivis**, mis oli vigadest määrsõnalise määruse järel teisel kohal, moodustades valesti lausemalli osaks tuvastatud lauseliikmetest 15%. Nimisõnaline määrus adessiivis tuvastati neljale lausemallile (65–68), mis hõlmasid üheksat verbi, kuid kõigil neist peale *vedama* oli adessiivis määrus igas kontekstis vaba laiend. *Vedama* puhul on adessiivis obliikva seotud laiend ka ainult juhul, kui lauses pole objekti (107), mis tuvastatud mallis aga oli (108). Niisiis oli selle malli kontekstis tegemist siiski ka *vedama* vaba laiendiga ehk kõik adessiivis obliikvaga tuvastatud mallid olid valed.

(107) ('obl', 'ad', ' ', ' '), ('obl', 'kom', ' ', ' ')

Samas pidi ta tunnistama, et *sõbrataril* vedas *armsamaga* üsna hästi.

(108) ('nsubj', '\*', ' ', ' '), ('obj', '\*', ' ', ' '), ('obl', 'ad', ' ', ' ')

Seekord vedas *meid* koduloo *kõrgnivool* kohalik *guru* hr Olavi Vallimäe.

Asukohakäänetes määrustest oli tuvastatud *paiskama*, *liigutama* ja *torkama* lausemallide (69–70) liikmeks ka **nimisõnaline määrus inessiivis**, mis osutus siiski nende verbide vabaks laiendiks. *Liigutama* puhul tundus see mõnes tähenduses olevat aga ka seotud laiendi funktsioonis, sellest pikemalt peatükis 5.4. Vaba laiendina esines nimisõnaline määrus inessiivis koha-, aja- ja viisimäärusena ning märgendusveana.

(109) Samal hetkel torkas miski *südames* valusasti, ja uni oli kui peoga pühitud.

(110) Ainuüksi Õismäe trollide asendamine bussidega paiskab *aastas* õhku suure hulga heitmeid.

(111) "See oli Tarmo, kes selle lause *ägedas toonis* hiiepuude alla paiskas.

- (112) Kuradi baptistid paiskab *Põldroos* ise talle omase otsekohesuse ja ehedusega tõe välja.

Verbile *torkama* seotud laiendiks tuvastatud inessiivis määrus osutus aga hoopis väljendverbi *silma torkama* seotud laiendiks.

- (113) Rootsi kuninga kui kirikupea sekkumine Eestimaa piiskopkonna valitsemisse torkab kõige eredamalt silma *personalipoliitikas*.
- (114) *Haljastuses* torkasid silma palmid ja Norfolki männid.
- (115) *Ühes asjas* torkab Keskerakond küll positiivselt silma.
- (116) *Lausestuses* torkab silma osasihitisega liialdamine (panni Josepit nende jure), sest Hornungi grammatikast alates oli akusatiivina kui sihitise käändena esitatud praeguse partitiivi vorme.

Huvitaval kombel ei tuvastatud malli liikmeks aga obliikvat aditiivis, mis peaks tähistama väljendverbi osist *silma*. Tasakaalus korpust uurides jäi mulje, et see tuleneb kahest probleemist: korpuse märgendusvigadest ja väljendverbide lihtverbidega koos arvestamisest. Väljendverbi osis *silma* esines verbiga *torkama* kokku 208 korda, neist 150 korda obliikvana aditiivis, 55 korda objektina ja 3 korda nimisõnalise täiendina. Objektina oli *silma* märgendatud õigesti kolm korda: kaks korda osana väljendist *silma peast (välja) torkama* ja üks kord millegi torkamise otseses tähenduses. Nimisõnalise täiendi märgend oli aga kõigil kordadel valesti pandud.

- (117) Koolmeistri tütar torkab rehkenduspulgaga igal tähel *silma peast* ja veerib ja veerib.
- (118) Sest kui sa veel mu õlga puudutad, torkan sul pastakaga *silma peast* välja.
- (119) "Poiste sõnade järgi olid nad algul vaid kahe näpuga *silma* torganud, kuid siis oli neil lapsikust mängust isu täis saanud ning nad otsustasid kiisu puu külge siduda ja kastreerida.

Kokku oli neljandik *silma torkama* juhtudest saanud vale märgenduse, mille tulemusel selle sagedus hajuma hakkas. Väljendverbi ja selle laiendite sagedus hajub aga veelgi, sest praegu arvestab meetod väljendverbe lihtverbidega kokku. See tähendab, et kui laiend esineb ainult väljendverbiga, arvestatakse selle sageduse arvutamisel ka lihtverbi lauseid, kus see laiend seotud ei ole.

Valesti tuvastati ka **nimisõnaline määrus komitatiivis**, mis esines kuue verbi lausemallides (62–64). Neist viie verbi – *vedama, tõukama, tõmbama, lükkama, sõidutama* – puhul oli komitatiivis määrus käsitsi koostatud materjalis vabaks laiendiks määratud, kuid kõrge sageduse tõttu võib see pigem paigutada ka aga vaba ja seotud laiendi piirimaile.

- (120) Kaubandusettevõtte vedavat neid *bussiga* kohale Ida-Virumaalt.
- (121) Bronka tõukab *jalaga* väikest kappi.
- (122) Ta tõmbas *nimetissõrmega* üle oma raamaturea selgade.
- (123) Leena lükkas *käega* juuksed silmilt ja hingas sügavalt sisse.
- (124) Isa teid just vist palju oma *masinaga* ei sõiduta?

Verbi *loopima* puhul oli INSTRUMENTI rollis komitatiivse määruse näol tegemist siiski kindlalt seotud laiendiga, sest selle väljajätt muudaks lause tähendust (125). Kuigi minu käsitsi koostatud mallide seast oli see puudu, oli seda malli oli varasemalt kirjeldanud Huno Rätsep (1978: 188), aga tema malli oli lisatud ka translatiivis omadussõna (126).

- (125) Poisid loopisid teda *kaigastega*, lootes spontaanset sõimu kuulda.
- (126) N1+nom. V N2+ngp. N3+kom. A+tr  
V=loopima (*Poisid loopisid koera kividega vigaseks.*)

Verbide *tõmbama* ja *vedama* puhul oli samuti juhtumeid, kus komitatiivis nimisõnalise määruse eemaldamisel muutub lause arusaamatuks, aga sel juhul oli tegemist hoopis väljendverbide *ninaga õhku tõmbama* ja *ninaga õhku vedama* osisega *ninaga*.

(127) Tommi tõmbas *ninaga* õhku ja tagus sabaga vastu vanaema kõrtsikut.

(128) Tamara vedas *ninaga* õhku ja küsis: "Mis magusroog sealt küll tulemas on?"

Kokkuvõtteks saab öelda, et kuuest nimisõnalist määrust komitatiivis sisaldavast verbikesksest lausemallist oli ainult üks kindlalt õigesti tuvastatud, kaks osutusid väljendverbideks ja ülejäänutel on õigsuse puhul ruumi vaidluseks.

Valesti tuvastati veel verbi *paigutama* mall objekti ja **translatiivis obliikvaga** (86). See oli käsitsi koostades jäänud *paigutama* seotud laiendite seast välja ning korpusest kontrollides avaldus, et translatiivis obliikva on tõesti *paigutama* vaba laiend, mis esineb aja- (129) ja põhjusemäärusena (130).

(129) Patarei alaealiste kasvatajale pean olema eriti tänulik, et ta paigutas mind kaheks *nädalaks* vattide kambrisse.

(130) Infosüsteemide *rajamiseks* paigutas ESS hoonesse 15 miljonit krooni.

Vaba laiend sattus ka verbi *tirima* malli (89), kus malli liikmeks arvestatud **des-vormis infiniittarind** osutus vabaks viisi- (131) ja tagajärjemääruseks (132).

(131) Potapov, *hoides* ühe käega raudses haardes Kaarna kukalt ning pigistades teisega suud kinni, tiris teda nüüd armutult keset põõsaid seisva järelvanker-suvila poole, mille kõrval seisis sinine "Renault".

(132) Paraku tirivad reklaami- ja kampaaniameistrid solvumise pantvangi ka erakondi, *vähendades* nii Keskerakonna kui ka Reformierakonna väljavaateid valimistel.

Käsitsi koostatud lausemallidest oli vähese materjali tõttu seotud laiendina puudu üheksa lauseliiget, millest neli olid nimisõnalised määrused illatiivis, üks nimisõnaline määrus komitatiivis, üks nimisõnaline määrus inessiivis, üks nimisõnaline määrus aditiivis, üks nimisõnaline määrus allatiivis ja üks ma-infinitiivne laiend. Nendest ainult kahte – nimisõnalist määrust komitatiivis ja inessiivis – ei arvestatud iga kord veaks ainult käsitsi koostatud materjali puuduste tõttu.

Seotud laienditest, mis käsitsi koostamise puuduste tõttu veaks loeti, oli kõige sagedasem **illatiivne nimisõnaline määrus** (133–136). Kokku oli see valesti veaks loetud neli korda: verbide *riputama*, *tirima*, *paiskama* ja *sõidutama* mallides. Tasakaalus korpust uurides tuli välja, et tegelikult on see kõigi nende verbide puhul üks SIHTKOHA rolli grammatilisi realisatsioone ehk nende verbide seotud laiend. Kuna SIHTKOHT oli siiski semantiliste käsitsi koostatud mallide osa, ei peeta illatiivse määrusega tuvastatud malle uuteks varem kirjeldamata lausemallideks.

- (133) Kiirelt riputas ta oma dubljonka *nagisse* ja istus Viktoria vastu laua taha.
- (134) Nad tirisid oma vigase sõbra *autosse* ja sõitsid ära.
- (135) Saatuse tõuge mind *haiglasse* paiskas.
- (136) Politsei sõidutas bussireisijad tagasi *Tartusse*.

Käsitsi koostatud mallidest oli välja jäänud ka verbile *tõukama* automaatselt tuvastatud mall, mis koosnes subjektist, objektist ja **ma-infinitivsest laiendist** (78). Kuigi käsitsi koostates ma seda ei tuvastanud, on seda malli varasemalt kirjeldanud Huno Rätsep (1978: 197) ning Tasakaalus korpust uurides jõudsin ma temaga sama järelduseni. Tegemist on *tõukama* malliga, sest *ma-infinitiivse* laiendi välja jätmine muudaks lause arusaamatuks (137–138).

(137) Ilu imetlemine saab paheks, kui tõukab sind ilusat *omandama*.

(138) "Meie sotsiaalne sallimatus üksi ei tapa kedagi, ei juuri välja ühtegi arusaama, kuid ta tõukab inimesi neid *peitma* või *hoiduma* aktiivsest tegevusest nende levitamisel."

Käsitsi ei olnud ma määranud verbi *torkama* seotud laiendiks **komplementlauset**, kuid automaatselt tuvastasin ma verbile *torkama* aditiivis obliikva ja komplementlausega malli (85). Tasakaalus korpust silmitsedes avaldus, et komplementlause on *torkama* seotud laiend siis, kui *torkama* on tähenduses 'midagi kriitilist ütleva' (139). Sellist malli, kus see oleks ilma aditiivis määruseta, ma automaatselt aga ei tuvastanud.

(139) Opu torkaks kindlasti küüniliselt, et looduslik valik *paneb* inimesed oma kohtadele vastavalt võimetele.

Väljendverbi malliks osutus ka verbile *seadma* tuvastatud mall subjekti, objekti ja kaassõnaga *alla* ühend nimisõnalise määrusega (79), mille realisatsioon oli pea alati väljendverb *kahtluse alla seadma*. Vahel tundus see avalduvat siiski ka *seadma* enda mallina, esinedes ühendites *surve alla seadma*, *küsimuse alla seadma*, *vaatluse alla seadma*, *löögi alla seadma* ja *maksukoorma alla seadma*.

(140) Väike eksimus selles, ja *kuulaja* seab lause lingvistilise *tähenduse kahtluse alla*.

Ülejäänud iga kord ainult käsitsi materjali puuduste tõttu veaks loetud lauseliikmed esinesid mallides, mis ei olnud varasemalt kirjeldatud ei minu käsitsi koostatud materjalis ega Huno Rätsepa raamatus "Eesti keele lihtlausete tüübid" (1978) ega olnud ka pelgalt väljendverbid. Nendest pikemalt järgmises peatükis.

## 5.4. Uued lausemallid

Verbile *liigutama* tuvastati automaatselt kaks lausemalli, millest kummaski ei ole väljendatud liigutamisverbidele tüüpilised SIHTKOHT ja LÄHTEKOHT, vaid hoopis tüüpiliselt ASUKOHTA või KOGEJAT märkiv obliikva inessiivis ja tüüpiliselt vaba laiendina esinev määrsõnaline määrus.

(141) ('nsubj', '\*', ", "), ('obj', '\*', ", "), ('obl', 'in', ", ")

Mingi ähmane *aimdus* liigutas *end* *Luiges*.

(142) ('nsubj', '\*', ", "), ('obj', '\*', ", "), ('advmod', '<puudub>', ", ")

*Ta* liigutas *end kiiresti* ja jõudis paari minutiga tee otsa kiriku ette.

Nagu varasemalt valesti tuvastatud lauseliikmete peatükis räägitud, on *advmod* märgendiga määrsõnaline määrus liigutamisverbide puhul tüüpiliselt siiski vaba laiend ning tavaliselt on seda ka nimisõnaline määrus inessiivis. Vahel tundub nimisõnaline määrus inessiivis olevat aga *liigutama* puhul ka seotud laiendi moodi, kuigi varasemas enda koostatud materjalis seda ei esinenud. Lauseid, mis ei olnud Huno Rätsepa koostatud ehk olid ehedad kasutusnäited, oli aga EDT korpuses ainult 20, mistõttu võisid need valimi väikese suuruse tõttu välja jääda. Huno Rätsepa koostatud lausemallides esineb *liigutama* puhul küll inessiivis määrus, aga substitutsiooniklassis *Di* SIHTKOHA näitajana.

(143) N<sup>1</sup>+nom. V N<sup>2</sup> +part. ((De) Di)

V=liigutama (*Peeter liigutab kätt soovitud suunas.*) (Rätsep 1978: 126)

Inessiivis määrus tundub olevat siiski *liigutama* seotud laiend, aga ainult selles tähenduses, mis tunnete viitab. Sellistes lausetes ei kujuta inessiivis obliikva mitte SIHTKOHTA, vaid hoopis tunnete ASUKOHTA ehk ei liigitu Rätsepa koostatud lausemallide alla. Tunnete viitavates *liigutama* lausetes muutub lause inessiivis obliikva välja jätmisel arusaamatuks, mistõttu arvestan seda seotud laiendiks.

- (144) Aga aborditegemise otsus liigutas *temas* midagi.
- (145) Muusikapalade hulgas oli tihti lemmikpalasid, mis *hinges* imelisi keeli liigutasid.
- (146) Mingi ähmane aimdus liigutas end *Luiges*.
- (147) Ainult harva, oma voodi juures askeldavat naist vaadates, tundis ta, kuidas *südamepõhjas* liigutas end soojussäde.
- (148) Väike okas liigutas end Tiia *sisemuses* ja torkas talle...kas just südamesse, aga sinna lähedale küll.

Verbile *pistma* tuvastati automaatselt mall, mis koosnes objektist, nimisõnalisest määrusest aditiivis ja nimisõnalisest määrusest allatiivis. Korpust uurides avastasin, et allatiivis obliikva on selles mallis SAAJA rollis. Käsitsi ei olnud ei mina ega Huno Rätsep (1978) *pistma* verbile SAAJA rolli tuvastanud. Siiski on näitest näha, et see on lause mõistmiseks oluline info ehk seotud laiend.

- (149) ('obj', '\*', ", "), ('obl', 'adit', ", "), ('obl', 'all', ", ")  
Ma ei olnud jõudnud veel sadat meetritki sõita, kui üks õdedest *mulle käe püksi* pistis.

## 6. JÄRELDUSED

Järgnevas peatükis võetakse kokku lausemallide automaatse tuvastusega seotud probleemid ning pakutakse neile võimalikke lahendusi, mille toel meetodit edasi arendada.

### 6.1. Probleemid tuvastamata jäänud lauseliikmete ja -mallidega

Kõige sagedasem põhjus, miks õiged lausemallid tuvastamata jäid, oli see, et meetod ei tuvastanud verbile õigeid seotud laiendeid. Liigutamisverbide puhul oli kõige suurem probleem illatiivis määruse tuvastamata jäämisega. See juhtus ilmselt seetõttu, et aditiivis sõnad kivinevad püsiühenditeks illatiivsetega võrreldes tihemini (Kio 2006: 126) ja paljud sõnad on koha tähenduses pigem aditiivis ja sõltuvusmäärustena pigem illatiivis (EKK 2020: 216). Jätsin algselt illatiivi ja aditiivi eraldi, sest see võiks verbidel, millel on seotud laiendid nii kohamäärused kui ka sõltuvusmäärused, aidata määruse liiki tuvastada ning on verbe, mis tahavad kas ainult illatiivis või ainult aditiivi laiendit (Viitso 1976: 152). Liigutamisverbide puhul oli nende eristamine aga üks mallide tuvastamata jäämise sagedasemaid põhjuseid. Nende kahe käände kokku arvestamine parandaks liigutamisverbide juures kindlasti tulemusi, aga võib mõne teise verbiklassi puhul neid ka halvemaks muuta. Seetõttu tuleks enne nende käänete koondamist põhjalikumalt uurida, kui paljudel verbidel võiks koha- ja sõltuvusmääruste eristamine oluline olla ning siis otsustada, kas see teeks rohkem kahju või kasu. Teise variandina võib püsiühenditest tekkivast sageduserinevusest mööda pääsemiseks kaaluda aditiivis esinevate määruste sagedusse lugemist ainult sõna esimesel esinemisjuhul.

Peale illatiivi ja aditiivi eristamise oli üks seotud laiendite tuvastamata jäämise põhjuseid ka verbide polüseemia. Nimelt olid samal verbil eri tähendustes tihti ka erinevad mallid. Kui tähendus oli aga väga spetsiifiline, ei leitud neid malle üles, sest

nende sagedus oli sagedastema tähendustega võrreldes liiga madal. Kahjuks pole praeguste vahenditega verbide tähenduste automaatne eristamine võimalik, mistõttu ei ole see töö kirjutamise seisuga üks meetodi parandamise võimalusi. Teine võimalus on lävendit alandada, näiteks viielt protsendilt kahe protsendi peale. Sel juhul võib valesid analüüse küll juurde tulla, aga saadav kasu perifeersete mallide tuvastamisel võib selle üle kaaluda.

Kõrge sagedusega verbide puhul jäid tuvastamata ka mõned seotud laiendid, mis intuiitiivselt tundusid olevat laialt levinud. See võis juhtuda seepärast, et sagedastel verbidel on tüüpiliselt rohkem tähendusi, mistõttu on need polüseemsusest tekkivate probleemide suhtes eriti tundlikud. Hetkel ei käsitle meetod rohkete tähendustega verbe erinevalt kui väheste tähendustega verbe, kuid tundub, et tegelikult võiks nende eristamine tulemust parandada. Niisiis oleks võimalik edasiarendus jaotada verbid kas sageduse või näiteks EKI ühendsõnastikus (Langemets jt 2021) toodud tähenduste arvu alusel eraldi klassidesse, millel kõigil võiksid olla erinevad lävendid.

## **6.2. Probleemid lausemallidest puuduvate lauseliikmetega**

Üks meetodi põhilisi probleeme oli see, et 20 malli 41st ei eristanud üksteisest muu kui see, kas subjekt oli lausemalli liikmeks arvestatud. Subjekti mallist puudumine oli ka mallide tuvastamata jäämise kolmas kõige sagedasem põhjus. Selle parandamiseks on kaks võimalust. Esimene võimalus on subjekti automaatselt kõigi peale nullvalentsete verbide seotud laiendiks lugeda ehk selle esinemist üldse mitte tuvastada. Selle peamine puudus on see, et ei saa kindel olla, kas eesti keeleteaduses on kõik subjektita esinevad verbid ära kirjeldatud ning kas see nimekiri ei saa täieneda. Lisaks tekitaks vigu jällegi polüseemia, sest ka nullvalensed verbid saavad polüseemia tõttu subjektiga esineda. Teine võimalus on kasutada UD süsteemis arendatavat *enhanced dependencies* esitust. See lisab elliptiliste lauseliikmete asemele tühjad kohatäite väärtused, mis on otseses ühenduses nii elliptilise lauseliikme varem esinenud vaste kui ka elliptilise lauseliikme

laienditega. Nende kaudu on ellipsi tõttu välja jäänud lauseliikmeid ja selle laiendeid oluliselt lihtsam automaatselt tuvastada. (Nivre jt 2020) Niisiis saaks selle abil arvestada sagedusse ka varem tehnilise keerukuse tõttu tuvastamata jäänud lauseliikmed. Selle jaoks tuleks *enhanced dependencies* märgenduskiht korpustele automaatselt lisada. Parserit, mis seda eesti keele jaoks teeks, hetkel aga tehtud ei ole.

### 6.3. Probleemid valesti tuvastatud lauseliikmete ja -mallidega

Üks meetodi peamisi probleeme oli see, et automaatselt määrati lausemalli osaks pigem vaba laiendina esinev keelend. See oli teine kõige sagedasem põhjus, miks õiged mallid tuvastamata jäid. 17 korrast 16-l oli üleliigseks lauseliikmeks määrsõnaline määrus ning valesti lausemalli liikmeks tuvastatud laienditest oli 49% juhtudest tegemist samuti just määrsõnalise määrusega. Järelikult juba ainult määrsõnalise määruse tuvastamise parandamine viiks lauseliikmete tuvastamise saagise 82 protsendilt 92 peale ja täpsuse 75 protsendilt 88 protsendi peale. Üks variant oleks määrsõnaline määrus kõigist mallidest automaatselt välja jätta, aga on ka verbe, mil see on seotud laiend. Liigutamisverbidel olid need näiteks *keerama* ja *tõmbama*, mille puhul on see seotud laiend seisundimäärusena.

(150) Ta keeras end *selili*.

(151) Villem tõmbas mõlemad püssikuked *vinna*.

Teine variant vabu laiendeid seotutest paremini eristada on neile määruse liigi lisamine. See aitaks sama süntaktilise rolli ja ka sama käändega määruseid paremini eristada, mille tulemusel hajuks vabade laiendite sagedus loodetavasti küllalt, et neid ei arvestataks enam nii tihti seotud laiendite hulka. See annaks ka võimaluse eristada sama süntaktilise rolli ja käändega, aga erinevate semantiliste rollidega argumente üksteisest. Hetkel puudub määruse liigi märgendamiseks usaldusväärne andmestik, kuid üks meetodi edasiarendusi saaks kindlasti olla sellise andmestiku koostamine ja määruse

liigi märgendaja välja arendamine. Veel üks variant on kasutada vabade ja seotud laiendite eristamiseks statistilisi meetodeid. Näiteks on varasemalt sarnase struktuuriga keele seotud ja vabade laiendite eristamiseks kasutatud MI-skoori ja ka Fisheri täpset testi, mis andsid ka väga häid tulemusi (Aldezabal jt 2002: 42–47). See nõuaks määruse liigi märgendamise väljaarendamisest oluliselt vähem ressursi, kuid see meetod võib olla ka ebatäpsem.

#### 6.4. Probleemid väljendverbidega

Korpusnäiteid uurides avastasin, et kolm lihtverbile tuvastatud malli osutusid hoopis väljendverbide mallideks. Väljendverbidel võivad olla lihtverbidest erinevad mallid, mistõttu võib nende seotud laiendite sagedus lihtverbi omi mõjutama hakkama ning selle tulemusel tuvastatakse nende seotud laiendid ekslikult lihtverbide omadeks.

Lahendus võiks olla väljendverbide tuvastamine leksikoni põhiselt. Selle jaoks tuleks hankida näiteks eesti keele verbikesksete püsiühendite andmebaasist<sup>18</sup> või EKI Ühendsõnastiku andmebaasist (Langemets jt 2021) väljendverbide loend ning lugeda väljendverbide esinemiseks see, kui väljendverbi peasõna otsene alluv on väljendverbi osisega sama sõna. Ka see pole aga ilma probleemideta, sest väljendverbi moodustavad sõnad saavad koos esineda nii väljendverbina (152) kui ka otseses tähenduses (153). Lisaks sellele võib väljendverbi osis olla väljendatud ka liitsõnana, mitte lihtsõnana (154).

(152) Ajal, mil neonatsid kuulutuste kaudu mõttekaaslasi otsivad ja stalinistid niikuinii *pead tõstavad*, rikastab meie Issanda loomaaeda veel mitu tublit ristirüütlit. (EDT)

(153) Kaelkirjak silmitses meid pikalt, lõuad rütmiliselt akaatsialehti jahvatamas, siis *tõstis* jälle *pea* ja jätkas lehtede napsimist. (EDT)

---

<sup>18</sup> <https://cl.ut.ee/ressursid/pysiyhendid/>

(154) *Heidame* selle raamatu põhjal *kiirpilgu* naisekuju muutumisele meie sajandil.

(EDT)

Niisiis oleks väljendverbide eraldamine üks võimalik meetodi edasiarendusi, aga kahjuks pole see ilma probleemideta. Siiski usun, et väljendverbide sellisel viisil lihtverbidest eristamine tooks rohkem kasu kui kahju, sest korpus uurides tundusid otseses tähenduses ja liitsõnana väljendatud väljendverbi osised olevat siiski harv nähtus.

## 6.5. Probleemid meetodi filtriga

Hetkel genereerib meetod väga palju üksteist hõlmavaid lausemalli kandidaate, mistõttu arvestatakse neist lausemalliks ainult iga lauseliikme kõige pikem mall. See tähendab aga, et kui lausemall koosneb kahest lauseliikmest, mis mõlemad esinevad ka pikemates mallides, jääb lühem mall lausemalliks tuvastamata. Selles töös oli selliseid malle ainult üks, kuid oli ka kolm tükki, mis oleks jäänud meetodi teiste puuduste kõrvaldamisel sellel põhjusel tuvastamata. Hetkel ei näe ma lahendust, mis selle probleemi täielikult kõrvaldaks, aga üks võimalus oleks lugeda lausemalliks mitte iga lauseliikme pikim lausemalli kandidaat, vaid iga lauseliikme paari kõige pikem lausemalli kandidaat. See lahendaks filtrist tuleva probleemi näiteks verbi *vedama* lausemalli jaoks, mis koosneb nimisõnalistest määrustest adessiivis ja komitatiivis (155). Praegu tuvastati need kahte eraldi malli: nimisõnaline määrus komitatiivis koos subjekti ja objektiga (156) ja nimisõnaline määrus adessiivis koos subjekti, objektiga ja määrsõnalise määrusega (157).

(155) ('obl', 'ad', '', ''), ('obl', 'kom', '', '')

(156) ('nsubj', '\*', '', ''), ('obj', '\*', '', ''), ('obl', 'kom', '', '')

(157) ('nsubj', '\*', '', ''), ('obj', '\*', '', ''), ('obl', 'ad', '', ''), ('advmod', '<puudub>', '', '')

Hetkel jääks meetodi filtri tõttu esimene mall (155) tuvastamata, sest teised kaks (156–157) sisaldavad selle lauseliikmeid ning on sellest pikemad. Kui filtriga määrata aga lausemalliks iga kahe lauseliikme kombinatsiooni kohta kõige pikem mall, kus see esineb, tuvastaks meetod praeguse kahe asemel kõik kolm malli. See ei suudaks siiski lahendada olukorda, kus sama lausemalli paar esineb ühe verbi puhul kahes eri pikkuses lausemallis. Selline on näiteks verb *tõstma*, kus ühes mallis on koos subjekt, objekt ja SIHTKOHT (158) ja teises ainult subjekt ja objekt (159).

(158) ('nsubj', '\*', ' ', ' '), ('obj', '\*', ' ', ' '), ('obl', 'all', ' ', ' ')

*Ta tõstis Ilvese kätele ja kõndis väravast välja.*

(159) ('nsubj', '\*', ' ', ' '), ('obj', '\*', ' ', ' ')

“Tänane *esitus* tõstab *enesekindlust*,” ütles ManU kaitsja Rio Ferdinand.

Sel juhul tuvastatakse ainult esimene mall ka varem pakutud lahendusega. Siiski olen arvamusel, et kõigi lausemallikandidaatide lausemalliks arvestamine teeks rohkem kahju kui kasu, mistõttu jääb see probleem hetkel lahenduseta.

## 6.6. Probleemid märgendusvigadega

Korpusest näiteid uurides jäi silma, et automaatselt süntaktilisest märgendusest tulenevalt esineb korpuses ka üpris palju märgendusvigu. Mallide tuvastamisel hakkas see selgelt segama *torkama* malle tuvastades, kus automaatselt oli verbile *torkama* tuvastatud mall subjekti ja inessiivis obliikvaga, mis osutus hoopis väljendverbi *silma torkama* malliks. Mallist oli aga puudu väljendverbi osist *silma* märkiv lauseliige nimisõnaline määrus aditiivis. Statistikat uurides tuli välja, et väljendverbi osis jäi mallist tuvastamata just märgendusvigade tõttu. Hetkel on automaatse süntaktilise

analüsaatori täpsus 85%<sup>19</sup>, aga see on pidevas arenduses ehk on lootust, et see muutub ajaga paremaks. Minu teada puudub muu viis analüsaatori vigadest mööda pääsemiseks. Näeksin lahendust ainult subjekti ja objekti puhul, kus on variant jätta analüüsist välja need verbid, mis on oma otsesteks alluvateks saanud korraga kaks subjekti või kaks objekti, sest need vead on ilmselged. Teiste lauseliikmetega seda probleemi nii lahendada ei saa, sest samade morfosüntaktiliste märgenditega lauseliige võib esindada kahte erinevat argumenti, nii et vigade otsimisel ei saa lähtuda reeglist, et sama seotud laiend ei saa verbi alluvana rohkem kui üks kord esineda (Kroeger 2005: 76). Seetõttu näen ma märgendusvigade probleemi ainukese lahendusena analüsaatori kvaliteedi paranemise ootamist.

## **6.7. Probleemid käsitsi koostatud mallidega**

Üheks tuvastuse kvaliteedi probleemiks osutus ka käsitsi koostatud materjali puudujäägid, mille tõttu loeti mitmed automaatselt tuvastatud lausemallid valedel alustel vigaseks. Peamiselt tulenes probleem sellest, et käsitsi koostatud materjali oli liiga vähe, et kõik verbi tähendused ja seotud laiendid valimisse jõuaksid. Seetõttu tuleks tulevikus luua käsitsi koostatud materjal mõne suurema korpuse abil või võtta valimisse ainult need verbid, mis korpuses kõrge sagedusega esinevad.

---

<sup>19</sup>[https://github.com/estnltk/estnltk/blob/main/tutorials/nlp\\_pipeline/C\\_syntax/03\\_syntactic\\_analysis\\_with\\_stanza.ipynb](https://github.com/estnltk/estnltk/blob/main/tutorials/nlp_pipeline/C_syntax/03_syntactic_analysis_with_stanza.ipynb)

## KOKKUVÕTE

Magistritöö eesmärk oli luua töövoog eesti keele lausemallide automaatseks tuvastamiseks ning leida selle esmased probleemid, millele tuginedes meetodit parandada. Töö tulemusi kasutatakse Eesti Keele Instituudi PRG 1978 projektis “Uue aja sõnastik: grammatika ja keelepädevuse kirjeldamine integreeritud multifunktsionaalses leksikograafilises ressursis”, mille üks eesmärkidest on parandada EKI ühendsõnastiku (Langemets jt 2021) grammatilise info esitust. Töö eesmärkidest tulenevalt otsustasin lausemallide käsitlemisel toetuda eesti keele varasematele grammatikaraditsioonidele ja konstruktsioonigrammatikale.

Alustasin tööd automaatse tuvastamise kvaliteedi hindamiseks võrdleva materjali koostamisega, mis valmis Tartu Ülikooli EKT75 projekti “Semantilise analüüsi baasvahendid“ raames. Otsustasin hinnata meetodit liigutamisverbide peal, sest neis esinesid enamik lausemallide tuvastusega seotud peamisi probleeme. Lõplik valim koosnes 28 liigutamisverbist. Lausemallide käsitsi koostamiseks kasutasin materjali EDT korpusest, mis on käsitsi märgendatud nii morfoloogiliselt kui süntaktiliselt. Kokku koostas 49 malli, kus mallide liikmeteks olid semantiliste rollide alla koondatuna esitatud kõik variandid, kuidas see roll korpuses morfosüntaktiliselt esines. Automaatselt tuvastatud mallidega võrdlemiseks lõin need 49 semantilist malli lahku 92 grammatiliseks malliks, kus kõikide semantiliste rollide kõige tüüpilisemad morfosüntaktilised realisatsioonid olid eraldi välja toodud.

Järgmise etapina töötasin välja lausemalle automaatselt tuvastava meetodi, mis põhineb verbi otseste alluvate koosinemise sagedusel. Meetodi esimene samm on andmestiku kogumine, mida tegi minu antud juhiste põhjal EKI vanemtarkvaraarendaja Katrin Tsepelina. Iga verbi jaoks leiti kõigepealt nende otsesed alluvad ning loeti kokku nende esinemise sagedus. Otseseid alluvaid otsiti ainult nendes vormides verbidele, mis ei muuda oluliselt lause struktuuri. Lausemalli võimalikeks liikmeteks arvestati ainult

need laiendid, mis saavad lausemallis esineda ning sobivad üldistuste tegemiseks. Nende leidmiseks rakendati andmete kogumisel mitmeid filtreid. Seejärel vaadati iga otsese alluva juures, milliste teiste otseste alluvatega see koos esines. Et paremini arvestada lausete keerulise struktuuriga, pandi nende koosinemise sagedus kirja kahekaupa ehk bigrammidena. Nendest lauseliikmete paaridest pani minu kirjutatud programm kokku pikemad verbipõhised lausemallid. Lausemallidesse sattumiseks pidi iga lauseliige või lauseliikmete paar esinema vähemalt viies protsendis selle verbi lauses. Bigrammid pandi kokku põhimõttel, et kui lävendi ületasid bigrammid a+b, b+c ja a+c, esineb ka järjend a+b+c. Lõpuks jäeti neist mallidest alles ainult iga lauseliikme kõige pikem mall.

Meetodi tulemusel leidsin 28 liigutamisverbile kokku 107 verbikeskset ja 41 verbiülest lausemalli. Verbipõhiste lausemallide keskmeks on leksikaalne verb ja verbiülesed mallid on mallid, mis võivad esineda nii ühel kui ka mitmel verbil. Nende tuvastamise kvaliteeti hindasin nii lauseliikmete kui lausemallide lõikes, kasutades selleks täpsuse ja saagise statistikuid. Käsitsi koostatud materjalis esinevatest lauseliikmetest jäi tuvastamata 19% ning õigesti tuvastatuid lauseliikmeid oli ümardatult 74%. Kõige sagedamini jäi tuvastamata nimisõnaline määrus illatiivis, mis moodustas lauseliikme tuvastamata jäämise juhtudest 50%. Valesti seotud laiendiks tuvastatud lauseliikmetest moodustas enamiku määrsõnaline määrus, mis oli vigase lausemalli tuvastuse põhjuseks samuti 50% juhtudest. Käsitsi koostatud materjalis olevatest lausemallidest jäi tuvastamata 66% ning kõigist automaatselt tuvastatud lausemallidest oli õigesti tuvastatuid 33%. Kõige sagedasemad lausemallide tuvastamata jäämise põhjused olid vigaselt tuvastatud määrsõnaline määrus, tuvastamata jäänud nimisõnaline määrus illatiivis ja mallist puuduv subjekt, mis kokku moodustasid kõigist vigadest 66%. Meetodi hea küljena märgiksin ära, et see tuvastas kahele verbile varem kirjeldamata lausemallid. Need olid verbi *liigutama* mall subjekti, objekti ja inesiivis obliikvaga, mis avaldus ainult tunnetega seotud metafoorses tähenduses kasutusmustriga “STIIMUL liigutab midagi KOGEJAS” ning *pistma* mall objekti, SIHTKOHTA märkiva nimisõnalise määrusega aditiivis ja SAAJAT märkiva nimisõnalise määrusega allatiivis.

Nagu tulemustest avaldub, oli meetodil mitmeid kitsaskohti. Peamised kitsaskohad olid:

- verbi seotud laiend jääb tuvastamata;
- tuvastatud seotud laiend puudub mõnest lausemallist, mille osa ta tegelikult on;
- verbi seotud laiendina on tuvastatud lauseliige, mis tegelikult seda pole;
- väljendverb koos talle omase lausemalliga on tuvastatud lihtverbi lausemallina;
- meetodi filtrist tulenevad piirangud;
- automaatse märgenduse vead;
- meetodi hindamiseks kasutatud võrdlusmaterjal põhineb liiga väikese korpuse andmetel.

Nende kõige esilduvamad näited olid illatiivis nimisõnalise määruse tuvastamata jäämine, subjekti mallidest välja jäämine, liigse määrsõnalise määruse lausemalli liikmeks tuvastamine, lihtverbi *torkama* malliks tuvastatud väljendverbi *silma torkama* mall, filtri tõttu välja jäänud ainult subjektist ja objektist koosnev mall, *silma torkama* mallis märgendusvigade tõttu tuvastamata jäänud *silma* ja käsitsi koostatud materjali puuduste tõttu valeks loetud 9 õiget lausemalli.

Neist probleemidest lähtuvalt oleks meetodi võimalikud edasiarendused järgmised:

- tuvastamaks rohkem seotud laiendeid, alandada lävendit praeguselt 5% tasemelt allapoole;
- käsitleda erineva sagedusega verbe erinevalt, sest sagedasemaid mõjutab polüseemsus üldiselt rohkem;
- tulemaks paremini toime subjekti sagedase ellipsiga, lisada lausemallidesse automaatselt subjekt, v.a. nullvalentsete verbide puhul, või kasutada UD täiustatud sõltuvuste (*enhanced dependencies*) märgenduskihti;
- liigitada määruslikke laiendeid semantilistesse alaliikidesse;

- kasutada seotud ja vabade laiendite eristamiseks statistilisi meetodeid;
- tuvastada väljendverbid leksikonipõhiselt;
- rakendada meetodi filtrit mitte üksiklauseliikmetele, vaid lauseliikmete paaridele;
- eemaldada admebaasipäringu tulemusest selged märgendusvead, näiteks kahe subjektiga laused;
- koostada võrdlev materjal ainult nende verbide baasil, mis korpusel kõrgel sagedusega esinevad.

Sain magistratöös vastused sissejuhatuses püstitatud küsimustele, mille tulemusel astuti otsustavaid samme EKI ühendsõnatiku grammatikaesituse parandamise suunas. Seetõttu sai autori hinnangul magistratöös püstitatud eesmärk täidetud.

## LÜHENDITE JA MÄRGENDITE SELGITUSED

Tabel 12. Töö autori kasutatavad lühendid ja märgendid

Lühend/Märgend	Selgitus
verb_compound	ühendverbi osis
deprel1	bigrammi esimese lauseliikme süntaktilise funktsiooni märgend
deprel2	bigrammi teise lauseliikme süntaktilise funktsiooni märgend
case1	bigrammi esimese lauseliikme käändemärgend
case2	bigrammi teise lauseliikme käändemärgend
verbform1	bigrammi esimese lauseliikme verbivorm sufiksina
verbform2	bigrammi teise lauseliikme verbivorm sufiksina
obl_case1	bigrammi esimese lauseliikme kaassõna
obl_case2	bigrammi teise lauseliikme kaassõna
deprel1_before	kui palju esimene lauseliige enne teist esineb
deprel2_before	kui palju teine lauseliige enne esimest esineb
verb_before	kui palju esimene lauseliige enne verbi esines
nsubj	nimisõnaline subjekt ehk nimisõnaline alus
obj	objekt ehk sihitis
obl	obliikva ehk nimisõnaline määrus
advmod	määrsõnaline määrus
nom	nominatiiv ehk nimetav kääne

gen	genitiiv ehk omastav kääne
par	partitiiv ehk osastav kääne
all	allatiiv ehk alaleütlev kääne
adit	aditiiv ehk lühike sisseütlev kääne
ill	illatiiv ehk sisseütlev kääne
<puudub>	lauseliikmel ei saa kääned olla ja seda pole ka märgendatud
*	nominatiiv, genitiiv või partitiiv ehk nimetav, omastav või osastav kääne
Arg0	Propbanki raamistiku argument 0 ehk PROTO-AGENT
Arg1	Propbanki raamistiku argument 1 ehk PROTO-PATSIENT
Arg2	Propbanki raamistiku argument 2

Tabel 13. Huno Rätsepa (1978: 40-45) kasutatavad lühendid

Lühend	Selgitus
nom	nominatiiv ehk nimetav kääne
ngp	nominatiiv, genitiiv või partitiiv ehk nimetav, omastav või osastav kääne
part	partitiiv ehk osastav kääne
all	allatiiv ehk alaleütlev kääne
ill	illatiiv ehk sisseütlev kääne
el	elatiiv ehk seestütlev kääne
abl	ablatiiv ehk alaltütlev kääne
tr	translatiiv ehk saav kääne

kom	komitatiiv ehk kaasaütlev kääne
N	substantiiv ehk nimisõna
V	verb ehk tegusõna
A	adjektiiv ehk omadussõna
da	<i>da</i> -infitiiv
De	ekstralokaalne direksionaal ( <i>kust</i> -suund)
Di	intralokaalne direksionaal ( <i>kuhu</i> -suund)
pl	mitmus

## KIRJANDUS

**Ackema, Peter 2015.** Arguments and Adjuncts. – Syntax – Theory and Analysis: An International Handbook 1. Ed. by Tibor Kiss, Artemis Alexiadou. Berlin: De Gruyter Mouton, 246–273.

**Aldezabal jt = Aldezabal, Izaskun, Maxux Aranzabe, Koldo Gojenola, Kepa Sarasola, Aitziber Atutxa 2002.** Learning argument/adjunct distinction for Basque. – Proceedings of the ACL-02 workshop on Unsupervised lexical acquisition 9. Stroudsburg: Association for Computational Linguistics, 42–50; <https://doi.org/10.3115/1118627.1118633>. Vaadatud 26.05.2024.

**Allerton, David J. 1982.** Valency and the English Verb. London/New York: Academic Press.

**Barbu, Roxana Maria, Ida Toivonen 2016.** Event participants and linguistic arguments. – Proceedings of the 38th Annual Meeting of the Cognitive Science Society. Ed. by Anna Papafragou, Daniel J. Grodner, Daniel Mirman, John Trueswell. Philadelphia: Cognitive Science Society, 1961–1966.

**Bick jt = Bick, Eckhard, Heli Uiho, Kaili Müürisep 2004.** Arborest – a Growing Treebank of Estonian. – Nordic Language Technology. Ed. by Henrik Holmboe. Copenhagen: Museum Tusulanums Forlag, 125–142.

**Boas, Hans C. 2013.** Cognitive Construction Grammar. – The Oxford Handbook of Constructional Grammar. Ed. by Thomas Hoffmann, Graeme Trousdale. Oxford: Oxford University Press, 233–252.

**Borin, Lars, Benjamin Lyngfelt 2025 (ilmumas).** Framenets and constructiCons. – The Cambridge Handbook of Construction Grammar. Cambridge: Cambridge University Press; [https://www.academia.edu/95301779/Framenets\\_and\\_constructiCons](https://www.academia.edu/95301779/Framenets_and_constructiCons). Vaadatud 26.05.2024.

**Croft, William 2001.** Radical Construction Grammar: Syntactic Theory in Typological Perspective. Oxford: Oxford University Press.

**Croft, William 2022.** Morphosyntax: Constructions of the World's Languages. Cambridge: Cambridge University Press.

**Dannélls jt = Dannélls, Dana, Lars Borin, Markus Forsberg, Karin Friberg Heppin, Maria Toporowska Gronostaj 2021.** Swedish Framenet. – The Swedish FrameNet++: Harmonization, integration, method development and practical language technology applications. Ed. by Dana Dannélls, Lars Borin, Karin Friberg Heppin. Amsterdam: John Benjamins Publishing Company, 37–66; <https://doi.org/10.1075/nlp.14.02dan>. Vaadatud 26.05.2024.

**Dunn, Jonathan 2023.** Exploring the Constructicon: Linguistic Analysis of a Computational CxG. – Proceedings of the First International Workshop on Construction Grammars and NLP. Washington, D.C.: Association for Computational Linguistics, 1–11; <https://arxiv.org/pdf/2301.12642.pdf>. Vaadatud 26.05.2024.

**EG 2023 = Eesti grammatika 2023.** Toim. Helle Metslang. Koost. Helle Metslang, Mati Ereht, Külli Habicht, Tiit Hennoste, Reet Kasik, Pire Teras, Annika Viht, Eva Liina Asu, Liina Lindström, Pärtel Lippus, Renate Pajusalu, Helen Plado, Andriela Rääbis, Ann Veismann. Tartu: Tartu Ülikooli Kirjastus.

**EKI ühendsõnastik 2024.** Eesti Keele Instituut, Sõnaveeb 2024; <https://sonaveeb.ee>.  
Vaadatud 26.05.2024.

**EKK = Erelt, Mati, Tiiu Erelt, Kristiina Ross 2020.** Eesti keele käsiraamat.  
Uuendatud väljaanne. Tallinn: Eesti Keele Sihtasutus.

**Erelt, Mati 2017a.** Sissejuhatus süntaksisse. – Eesti keele süntaks. Eesti keele varamu III. Toim. Mati Erelt, Helle Metslang. Tartu: Tartu Ülikooli Kirjastus, 53–89.

**Erelt, Mati 2017b.** Öeldis. – Eesti keele süntaks. Eesti keele varamu III. Toim. Mati Erelt, Helle Metslang. Tartu: Tartu Ülikooli Kirjastus, 93–239.

**Erelt, Mati 2017c.** Öeldistäitemäärus. – Eesti keele süntaks. Eesti keele varamu III. Toim. Mati Erelt, Helle Metslang. Tartu: Tartu Ülikooli Kirjastus, 289–299.

**Erelt, Mati 2017d.** Ellips. – Eesti keele süntaks. Eesti keele varamu III. Toim. Mati Erelt, Helle Metslang. Tartu: Tartu Ülikooli Kirjastus, 590–602.

**Erelt, Mati 2017e.** Sekundaartarindiga laused. – Eesti keele süntaks. Eesti keele varamu III. Toim. Mati Erelt, Helle Metslang. Tartu: Tartu Ülikooli Kirjastus, 756–840.

**Erelt jt = Erelt, Mati, Metslang, Helle, Plado, Helen 2017.** Alus. – Eesti keele süntaks. Eesti keele varamu III. Toim. Mati Erelt, Helle Metslang. Tartu: Tartu Ülikooli Kirjastus, 240–257.

**Faaß jt = Faaß, Gertrud, Ulrich Heid, Helmut Schmid 2010.** Design and Application of a Gold Standard for Morphological Analysis: SMOR as an Example of Morphological Evaluation. – Proceedings of the Seventh International Conference on Language Resources and Evaluation. Valletta: European Language Resources

Association, 803–810;  
[http://www.lrec-conf.org/proceedings/lrec2010/pdf/409\\_Paper.pdf](http://www.lrec-conf.org/proceedings/lrec2010/pdf/409_Paper.pdf). Vaadatud  
27.05.2024.

**Fillmore, Charles J. 1982.** Frame semantics. – *Linguistics in the Morning Calm*. Seoul: Hanshin Publishing Company, 111–137.

**Fillmore, Charles J. 2007.** Valency issues in FrameNet. – *Valency: Theoretical, Descriptive and Cognitive Issues*. Ed. by Thomas Herbst, Katrin Götz-Votteler. Berlin: De Gruyter Mouton, 129–160.

**Fillmore jt = Fillmore, Charles J., Russell Lee-Goldman, Russell Rhodes 2012.** The FrameNet Constructicon. – *Sign-based Construction Grammar*. Ed. by Hans C. Boas, Ivan A. Sag. Chicago: CSLI Publications.

**Forsberg jt = Forsberg, Markus, Richard Johansson, Linnéa Bäckström, Lars Borin, Benjamin Lyngfelt, Joel Olofsson, Julia Prentice 2014.** From construction candidates to constructicon entries: An experiment using semi-automatic methods for identifying constructions in corpora. – *Constructions and Frames* 6 (1). Ed. by Lars Borin, Gerard de Melo, Karin Friberg Heppin, Tiago T. Torrent, 114–135; <https://doi.org/10.1075/cf.6.1.07for>. Vaadatud 27.05.2024

**Goldberg, Adele 1995.** *Constructions: a construction grammar approach to argument structure*. Chicago: University of Chicago Press.

**Goldberg, Adele 2001.** Patient arguments of causative verbs can be omitted: the role of information structure in argument distribution. – *Language Sciences* 23, 503–524.

**Goldberg, Adele 2006.** *Constructions at Work: The nature of generalization in language.* Oxford: Oxford University Press.

**Haverinen jt = Haverinen, Katri, Jenna Kanerva, Samuel Kohonen, Anna Missilä, Stina Ojala, Timo Viljanen, Veronika Laippala, Filip Ginter 2015.** *The Finnish Proposition Bank.* – *Language Resources and Evaluation* 49, 907–926; <https://doi.org/10.1007/s10579-015-9310-y>. Vaadatud 26.05.2024.

**Herbst jt = Herbst, Thomas, David Heath, Ian F. Roe, Dieter Götz 2004.** *A Valency Dictionary of English: A Corpus-Based Analysis of the Complementation Patterns of English Verbs, Nouns and Adjectives.* Berlin: De Gruyter Mouton; <https://doi.org/10.1515/9783110892581>. Vaadatud 26.05.2024.

**Herbst, Thomas 2007.** *Valency complements or valency patterns? – Valency: Theoretical, Descriptive and Cognitive Issues.* Ed. by Thomas Herbst, Katrin Götz-Votteler. Berlin: De Gruyter Mouton, 15–35.

**Herbst, Thomas 2014.** *The valency approach to argument structure constructions.* – *Constructions Collocations Patterns.* Ed. by Thomas Herbst, Hans-Jörg Schmid, Susen Faulhaber, Berlin: De Gruyter Mouton, 167–216; <https://doi.org/10.1515/9783110356854.167>. Vaadatud 26.05.2024.

**Herbst, Thomas, Peter Uhrig 2009.** *Erlangen Valency Patternbank. A corpus-based research tool for work on valency and argument structure constructions.*

**Herbst, Thomas, Uhrig, Peter 2019.** *Towards a valency and argument structure constructicon of English: Turning the valency patternbank into a constructicon.* – *Lexicographica* 35, 87–104; <https://doi.org/10.1515/lex-2019-0006>. Vaadatud 26.05.2024.

**Hilpert, Martin 2014.** Argument structure constructions. – Construction Grammar and its Application to English. Edinburgh: Edinburgh University Press, 25–49.

**Hoffmann, Thomas 2013.** Abstract Phrasal and Clausal Constructions. – The Oxford Handbook of Constructional Grammar. Ed. by Thomas Hoffmann, Graeme Trousdale. Oxford: Oxford University Press, 307–328.

**Janda jt = Janda, Laura, Olga Lyashevskaya, Tore Nessel, Ekaterina Rakhilina, Francis M. Tyers 2018.** A Constructicon for Russian: Filling in the Gaps. – Constructicography: Constructicon development across languages. Ed. by Benjamin Lyngfelt, Lars Borin, Kyoko Ohara, Tiago T. Torrent. Amsterdam: John Benjamins Publishing Company, 165–181; <https://doi.org/10.1075/cal.22.06jan>. Vaadatud 26.05.2024.

**Janda jt = Janda, Laura, Anna Endresen, Valentina Zhukova, Daria Mordashova, Ekaterina Rakhilina 2020.** How to build a constructicon in five years: The Russian example. – Belgian Journal of Linguistics 34, 162–175; <https://doi.org/10.1075/bjl.00043.jan>. Vaadatud 26.05.2024.

**Jentson, Indrek 2013.** Eesti VerbNet'i loomise võimalikkusest. – Eesti Rakenduslingvistika Ühingu Aastaraamat 9, 75–83; <http://dx.doi.org/10.5128/ERYa9.05>. Vaadatud 26.05.2024.

**Jurafsky, Daniel, James H. Martin 2023.** Semantic Role Labeling. – Speech and Language Processing. Draft of January 7, 2023; <https://web.stanford.edu/~jurafsky/slp3/24.pdf>. Vaadatud 17.02.2023.

**Karlsson, Fred 2002.** Üldkeeleteadus. Tõlkinud ja kohandanud Renate Pajusalu, Jüri Valge, Ilona Tragel. Tallinn: Eesti Keele Sihtasutus.

**Kio, Kati 2006.** Sisseütleva käände kasutus eesti kirjakeeles. Magistritöö. Tartu: Tartu Ülikool.

**Kipper-Schuler, Karin 2005.** VerbNet: A broad-coverage, comprehensive verb lexicon. Pennsylvania: University of Pennsylvania.

**Kluyver jt = Kluyver, Thomas, Benjamin Ragan-Kelley, Fernando Perez, Brian Granger, Matthias Bussonnier, Jonathan Frederic, Kyle Kelley, Jessica Hamrick, Jason Grout, Sylvain Corlay, Paul Ivanov, Damian Avila, Safia Abdalla, Carol Willing 2016.** Jupyter Notebooks – a publishing format for reproducible computational workflows. – Positioning and Power in Academic Publishing: Players, Agents and Agendas. Ed. by Fernando Loizides, Birgit Schmidt. Amsterdam: IOS Press BV, 87–90.

**Kroeger, Paul 2005.** Analyzing Grammar: An Introduction. Cambridge: Cambridge University Press; <https://doi.org/10.1017/CBO9780511801679>. Vaadatud 26.05.2024.

**Koppel, Kristina, Jelena Kallas 2022.** Eesti keele ühendkorpuste sari 2013–2021: mahukaim eestikeelsete digitekstide kogu. – Eesti Rakenduslingvistika Ühingu aastaraamat 18, 207–228; <http://dx.doi.org/10.5128/ERYa18.12>. Vaadatud 26.05.2024.

**Langemets jt = Langemets, Margit, Kristina Koppel, Jelena Kallas, Arvi Tavast 2021.** Sõnastikukogust keeleportaalsiks. – Keel Ja Kirjandus 64 (8–9), 755–770; <https://doi.org/10.54013/kk764a6>. Vaadatud 26.05.2024.

**Laur jt = Laur, Sven, Siim Orasmaa, Dage Särg, Paul Tammo 2020.** EstNLTK 1.6: Remastered Estonian NLP Pipeline. – Proceedings of The 12th Language Resources

and Evaluation Conference. Marseille: European Language Resources Association, 7154–7162; <https://www.aclweb.org/anthology/2020.lrec-1.884>. Vaadatud 26.05.2024.

**Levin, Beth 1993.** English Word Classes and Alternations. Chicago: University of Chicago Press.

**Lindström, Liina 2017.** Lause infostruktuur ja sõnajärg. – Eesti keele süntaks. Eesti keele varamu III. Toim. Mati Ereht, Helle Metslang. Tartu: Tartu Ülikooli Kirjastus, 537–565.

**Lippincott jt = Lippincott, Thomas, Diarmuid Ó Séaghdha, Anna Korhonen 2012.** Learning Syntactic Verb Frames Using Graphical Models. – Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics. Jeju Island: Association for Computational Linguistics, 420–429.

**Lyashevskaya, Olga, Egor Kashkin 2015.** FrameBank: A Database of Russian Lexical Constructions. – Analysis of Images, Social Networks and Texts, AIST 2015. Ed. by Mikhail Khachay, Natalia Konstantinova, Alexander Panchenko, Dmitry Ignatov, Valeri Labunets. Cham: Springer, 350–360; [https://doi.org/10.1007/978-3-319-26123-2\\_34](https://doi.org/10.1007/978-3-319-26123-2_34). Vaadatud 26.05.2024.

**Lyngfelt jt = Lyngfelt, Benjamin, Linnéa Bäckström, Lars Borin, Anna Ehrlemark, Rudolf Rystedt 2018.** Constructicography at work: Theory meets practice in the Swedish constructicon. – Constructicography: Constructicon development across languages. Ed. by Benjamin Lyngfelt, Lars Borin, Kyoko Ohara, Tiago T. Torrent. Amsterdam: John Benjamins Publishing Company, 41–106.

**Marneffe jt = Marneffe, Marie-Catherine de, Timothy Dozat, Natalia Silveira, Katri Haverinen, Filip Ginter, Joakim Nivre, Christopher D. Manning 2014.**

Universal Stanford dependencies: A cross-linguistic typology. – Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14), Reykjavik: European Language Resources Association (ELRA), 4585–4592; <https://aclanthology.org/L14-1045/>. Vaadatud 26.04.2024.

**Marneffe jt = Marneffe, Marie-Catherine de, Christopher D. Manning, Joakim Nivre, Daniel Zeman 2021.** Universal Dependencies. – Computational Linguistics 47 (2), 255–308; [https://doi.org/10.1162/coli\\_a\\_00402](https://doi.org/10.1162/coli_a_00402). Vaadatud 26.04.2024.

**Metslang, Helle 2017a.** Sihitis. – Eesti keele süntaks. Eesti keele varamu III. Toim. Mati Ereht, Helle Metslang. Tartu: Tartu Ülikooli Kirjastus, 258–277.

**Metslang, Helle 2017b.** Kommunikatiivsed lausetüübid. – Eesti keele süntaks. Eesti keele varamu III. Toim. Mati Ereht, Helle Metslang. Tartu: Tartu Ülikooli Kirjastus, 515–536.

**Muischnek, Kadri, Heete Sahkai 2009.** Using collocation-finding methods to extract constructions and to estimate their productivity. – Workshop on extracting and using constructions in NLP. Ed. by Magnus Sahlgren, Ola Knutsson, 22–27; [http://www.robos.org/sections/research/constructions\\_workshop.pdf](http://www.robos.org/sections/research/constructions_workshop.pdf). Vaadatud 02.05.2024.

**Muischnek jt = Muischnek, Kadri, Kaili Müürisep, Tiina Puolakainen, Eleri Aedmaa, Riin Kirt, Dage Särg 2014.** Estonian Dependency Treebank and its annotation scheme. – Proceedings of the 13th Workshop on Treebanks and Linguistic Theories, 285–291; <http://tlt13.sfs.uni-tuebingen.de/tlt13-proceedings.pdf>. Vaadatud 17.02.2023.

**Muischnek jt = Muischnek, Kadri, Kaili Müürisep, Dage Särg 2019.** CG Roots of UD Treebank of Estonian Web Language. – Proceedings of the NoDaLiDa 2019 Workshop on Constraint Grammar – Methods, Tools and Applications. Linköping: Linköping University Electronic Press, 23–26; <https://ep.liu.se/ecp/168/006/ecp19168006.pdf>. Vaadatud 24.04.2024.

**Muischnek, Kadri, Kaili Müürisep 2017.** Eesti keele sõltuvuspuude pank ja selle keeleteoreetilised lähted. – Emakeele Seltsi aastaraamat 62, 122–145; <http://dx.doi.org/10.3176/esa62.04>. Vaadatud 26.05.2024.

**Muischnek, Kadri, Kaili Müürisep 2022.** Eesti keele universaalsõltuvuste (Universal Dependencies, UD) eesti keele puudepank ja selle märgenduskeem; [https://github.com/EstSyntax/EstUD/blob/master/UD\\_dokumentatsioon\\_oktoobris\\_2022.pdf](https://github.com/EstSyntax/EstUD/blob/master/UD_dokumentatsioon_oktoobris_2022.pdf). Vaadatud 26.04.2024

**Müürisep jt = Müürisep, Kaili, Heili Orav, Haldur Ōim, Kadri Vider, Neeme Kahusk, Piia Taremaa 2008.** From Syntax Trees in Estonian to Frame Semantics. – Proceedings of The Third Baltic Conference on Human Language Technologies. Ed. by František Čermák. Vilnius: Vytautas Magnus University, Institute of the Lithuanian Language, 211–218.

**Nivre jt = Nivre, Joakim, Marie-Catherine de Marneffe, Filip Ginter, Yoav Goldberg, Jan Hajič, Christopher D. Manning, Ryan McDonald, Slav Petrov, Sampo Pyysalo, Natalia Silveira, Reut Tsarfaty, Daniel Zeman 2016.** Universal Dependencies v1: A multilingual treebank collection. – Proceedings of the Tenth International Conference on Language Resources and Evaluation, LREC 2016. Ed. by Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Marko Grobelnik, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk, Stelios Piperidis. Portorož: European Language Resources Association (ELRA), 1659–1666.

**Nivre jt = Joakim Nivre, Marie-Catherine de Marneffe, Filip Ginter, Jan Hajič, Christopher D. Manning, Sampo Pyysalo, Sebastian Schuster, Francis Tyers, and Daniel Zeman 2020.** Universal Dependencies v2: An Evergrowing Multilingual Treebank Collection. – Proceedings of the Twelfth Language Resources and Evaluation Conference. Marseille: European Language Resources Association (ELRA), 4034–4043.

**Ohara jt = Ohara, Kyoko, Seiko Yamaguchi Fujii, Toshio Ohori, Ryoko Suzuki, Hiroaki Saito, Shun Ishizaki 2004.** The Japanese FrameNet Project: An introduction. – Proceedings of the Satellite Workshop “Building Lexical Resources from Semantically Annotated Corpora”, 9–11.

**Ohara, Kyoko 2018.** Relations between frames and constructions: A proposal from the Japanese FrameNet constructicon. – Constructicography: Constructicon development across languages. Ed. by Benjamin Lyngfelt, Lars Borin, Kyoko Ohara, Tiago T. Torrent. Amsterdam: John Benjamins Publishing Company, 141–165.

**Olson, David L., Dursun Delen 2008.** Advanced Data Mining Techniques. Berlin: Springer.

**Orasmaa, Siim 2013.** Verb Subcategorisation Acquisition for Estonian Based on Morphological Information. – Text, Speech, and Dialogue 2013. Lecture Notes in Computer Science 8082. Ed. by Ivan Habernal, Vaclav Matoušek. Berlin: Springer, 583–590; [https://doi.org/10.1007/978-3-642-40585-3\\_73](https://doi.org/10.1007/978-3-642-40585-3_73). Vaadatud 26.05.2024.

**Pajusalu, Renate 2009.** Sõna ja tähendus. Tallinn: Eesti Keele Sihtasutus.

**Pajusalu jt = Pajusalu, Renate, Neeme Kahusk, Heili Orav, Ann Veismann, Kadri Vider, Haldur Õim 2013.** The encoding of motion events in Estonian. – Motion Encoding in Language and Space. Ed. by Mila Vulchanova, Emile van der Zee. Oxford: Oxford University Press, 44–66; <https://doi.org/10.1093/acprof:oso/9780199661213.003.0003>. Vaadatud 26.05.2024.

**Palmer jt = Palmer, Martha, Dan Gildea, Paul Kingsbury 2005.** The Proposition Bank: An Annotated Corpus of Semantic Roles. – Computational Linguistics Journal 31 (1), 71–106; <https://aclanthology.org/J05-1004.pdf>. Vaadatud 26.04.2024.

**Perek, Florent, Amanda L. Patten 2019.** Towards an English Constructicon using patterns and frames. – International Journal of Corpus Linguistics 24 (3), 354–384; <https://doi.org/10.1075/ijcl.00016.per>. Vaadatud 26.04.2024.

**Pool, Raili 1996.** Eesti keele verbireksioone. Tartu: Tartu Ülikooli Kirjastus.

**Przepiórkowski, Adam, Agnieszka Patejuk 2018.** Arguments and Adjuncts in Universal Dependencies. – Proceedings of the 27th International Conference on Computational Linguistics. Ed. by Emily M. Bender, Leon Derczynski, Pierre Isabelle. Santa Fe: Association for Computational Linguistics, 3837–3852.

**Qi jt = Qi, Peng, Yuhao Zhang, Yuhui Zhang, Jason Bolton, Christopher D. Manning 2020.** Stanza: A Python Natural Language Processing Toolkit for Many Human Languages. – Association for Computational Linguistics (ACL) System Demonstrations; <https://doi.org/10.48550/arXiv.2003.07082>. Vaadatud 26.05.2024.

**Rossum, Guido van, Fred L. Drake 2009.** Python 3 Reference Manual. Scotts Valley: CreateSpace.

**Ruppenhofer jt = Ruppenhofer, Josef, Michael Ellsworth, Miriam Petruck, Christopher Johnson, Jan Scheffczyk 2016.** FrameNet II: Extended Theory and Practice. Mannheim: Institut für Deutsche Sprache.

**Rätsep, Huno 1978.** Eesti keele lihtlauseste tüübid. Eesti NSV Teaduste Akadeemia Emakeele Seltsi toimetised nr. 12. Tallinn: Valgus.

**Sass, Bálint 2023.** From a dictionary towards the Hungarian Constructicon. – Electronic lexicography in the 21st century (eLex 2023): Invisible Lexicography. Proceedings of the eLex 2023 conference. Brno: Lexical Computing, 534–544; <https://elex.link/elex2023/wp-content/uploads/105.pdf>. Vaadatud 26.05.2024.

**Saul, Kertu 2022.** EstNLTK morfoloogilise analüsaatori ja ühestaja kvaliteedi hindamine. Bakalaureusetöö. Tartu: Tartu Ülikool.

**Simon, Gábor 2023.** Constructions, Collocations, and Patterns: Alternative Ways of Construction Identification in a Usage-based, Corpus-driven Theoretical Framework. – Proceedings of the First International Workshop on Construction Grammars and NLP (CxGs+NLP, GURT/SyntaxFest 2023). Ed. by Claire Bonial, Harish Tayyar Madabushi. Washington, D.C.: Association for Computational Linguistics, 12–30; <https://aclanthology.org/2023.cxgsnlp-1.2.pdf>. Vaadatud 26.05.2024.

**Stefanowitsch, Anatol, Stefan Th. Gries 2003.** Collostructions: Investigating the interaction between words and constructions. – International Journal of Corpus Linguistics 8 (2), 209–243; <https://doi.org/10.1075/ijcl.8.2.03ste>. Vaadatud 26.05.2024.

**Stefanowitsch, Anatol 2013.** Collostructional Analysis. – The Oxford Handbook of Constructional Grammar. Ed. by Thomas Hoffmann, Graeme Trousdale. Oxford: Oxford University Press, 290–306.

**Ziem jt = Ziem, Alexander, Johanna Flick, Phillip Sandkühler 2019.** The German Constructicon Project: Framework, methodology, resources. – *Lexicographica* 35, 15–40; <https://doi.org/10.1515/lex-2019-0003>. Vaadatud 26.05.2024.

**Tesnière, Lucien 2015.** Elements of Structural Syntax. Translated by Timothy Osborne, Sylvain Kahane. Amsterdam: John Benjamins Publishing Company.

**Tragel, Ilona, Piia Taremaa 2019.** Eesti keele verbide semantilisest liigitamisest. – *Keel ja Kirjandus* 3, 207–222; <https://doi.org/10.54013/kk736a4>. Vaadatud 26.05.2024.

**Torrent jt = Torrent, Timponi Tiago, Ludmila Meireles Lage, Thais Fernandes Sampaio, Tatiane da Silva Tavares, Ely Edison da Silva Matos 2014.** Revisiting border conflicts between FrameNet and Construction Grammar: Annotation policies for the Brazilian Portuguese Constructicon. – *Constructions and Frames* 6 (1), 34–51; <https://doi.org/10.1075/cf.6.1.03tor>. Vaadatud 26.05.2024.

**Veismann jt = Veismann, Ann, Mati Erelt, Helle Metslang 2020.** Määrus. – Eesti keele süntaks. Eesti keele varamu III. Toim. Mati Erelt, Helle Metslang. Tartu: Tartu Ülikooli Kirjastus, 300–375.

**Viitso, Tiit-Rein 1976.** Eesti muutkondade süsteemist. – *Keel ja Kirjandus* 3, 148–162.

**ÜK 2021 = Eesti keele ühendkorpus 2021;**  
<https://doi.org/10.15155/3-00-0000-0000-0000-08D17L>. Vaadatud 26.05.2024.

## **AUTOMATIC DETECTION OF ESTONIAN ARGUMENT STRUCTURE CONSTRUCTIONS ON THE EXAMPLE OF CAUSED-MOTION VERBS. SUMMARY**

The aim of the master's thesis was to create a workflow for the automatic detection of Estonian argument structure constructions and to find its primary problems to improve the method. The results will be used in the PRG1978 project of the Institute of the Estonian Language "Expanding the scope of a multi-purpose lexicographic resource to grammar and L2 competence", one of the goals of which is to improve the presentation of grammatical information in the EKI Combined Dictionary (Langemets et al. 2021).

I started work by compiling comparative material for evaluating the quality of automatic detection, which was prepared as part of the EKTB75 project "Basic tools for semantic analysis" of the University of Tartu. I chose to evaluate the method on caused-motion verbs because most of the primary problems related to detecting argument structure construction occurred in them. There were 28 caused-motion verbs in the sample. In order to manually compile argument structure constructions I used material from the manually morphologically and syntactically annotated EDT corpus. In total, I compiled 49 argument structure constructions, where arguments were presented as semantic roles that included all variants of how they appeared morphosyntactically in the corpus. In order to compare with the automatically detected constructions, I converted the 49 semantically represented construction to 92 grammatical ones based on their most typical morphosyntactic realizations.

In parallel with the analysis of caused-motion verbs, I familiarized myself with the theoretical literature on argument structure constructions. Argument structure constructions and their related concepts are defined very differently in different theoretical frameworks, which is why I thoroughly described basic concepts and

different approaches to identifying constructions. Due to the goals of the thesis, I decided to base my work mostly on the traditional approach to syntax analysis in Estonian linguistics and Construction Grammar framework.

Next I developed a method that automatically detects argument structure constructions based on the co-occurrence frequency of a verb's direct dependents. The first step of the method is collecting a dataset, which was done by Katrin Tsepelina, EKI's senior software developer, based on my instructions. First the method searched for each verb's direct dependents and counted their frequency. Only verbs in forms which do not significantly change the structure of the sentence were used. As the method didn't use all verbs, it also didn't count all dependents, as only those that can be part of an argument structure construction and are suitable for making generalizations were considered possible arguments. To find them, several filters were applied during data collection. Then each direct dependent's co-occurrence with another direct dependent was counted. In order to better take into account the complex structure of the sentences, the frequency of their co-occurrence was counted in pairs aka bigrams. Longer verb-based argument structures were formed from these pairs using a program I wrote. To be included in the construction, each dependent or pair of dependents had to appear in at least 5% of sentences with that verb. The bigrams were put together on the principle that if the bigrams  $a+b$ ,  $b+c$  and  $a+c$  crossed the threshold, the sequence  $a+b+c$  also occurs. Finally, only the longest construction of each dependent was retained from these templates.

As a result of the method, I found a total of 107 partially schematic and 41 fully schematic argument structure constructions for 28 verbs of movement. Partially schematic constructions include verbs while fully schematic are relevant for one or more verbs. I evaluated the quality of their detection both in terms of dependents and in terms of constructions using precision and recall. 19% of dependents in the manually compiled material were not identified and 74% of them were identified correctly. The

most frequently unidentified was the oblique in the illative, which accounted for 50% of unidentified dependents. Most of the incorrectly identified dependents were adverbial modifiers, which was the reason a construction remained unidentified in 50% of cases. Out of all argument structure constructions, 66% remained undetected and 33% were detected correctly. The most frequent reasons for not identifying argument structure constructions were an incorrectly identified adverbial modifier, an unidentified oblique in illative and a missing subject, which together accounted for 66% of all errors. Despite the low precision and recall the method also identified previously undescribed argument structures for two verbs: *liigutama* 'to move/to touch' and *pistma* 'to put'. For *liigutama* the method found an argument structure construction with subject, object and oblique in inessive case, which manifested itself only when the meaning of the verb *liigutama* was related to feelings. For *pistma* it found a construction with an object, an oblique in additive case representing a GOAL and one in allative representing a RECIPIENT.

As the results show, the method had several problems. The main problems were being unable to detect certain arguments, leaving some detected arguments out of constructions, detecting some adjuncts as arguments, phraseological verbs intersecting with non-phraseological verbs, limitations caused by the method's filter and not using enough material for compiling manual constructions. The most prominent examples of these were the oblique in illative being unidentified, subject being left out of the constructions, adverbial modifier being wrongly identified as an argument, a phraseological verb's argument structure construction being identified for a non-phraseological verb, a construction consisting of only a subject and object remaining unidentified due to the method's filter, a phraseological verb's nominal part remaining unidentified and 9 correct argument structure constructions deemed incorrect due to shortcomings in manually compiled constructions.

Based on these issues, possible improvements to the method would be to lower the 5% threshold; treating verbs separately based on their frequency; adding a default subject to the constructions (except for zero-valent verbs) or adding an UD enhanced dependencies layer to the corpus; developing a way to automatically annotate adverbs with their types or using statistical methods to distinguish arguments from adjuncts; using a lexicon to identify phraseological verbs; applying the method filter to bigrams instead of individual dependents; omitting sentences with two subject or object tags and basing comparative material on larger data.

In my master's thesis, I got the answers to the questions raised in the introduction, as a result of which decisive steps were taken towards improving the automatic detection of argument structure construction of Estonian verbs and the improvement of the presentation of grammatical information in the EKI Combined Dictionary.

## LISAD

### Lisa 1. Käsitsi koostatud semantilised mallid

Halliga on tähistatud rollid, mis jäid argumendi ja adjunkti piirimaale, mistõttu ei võetud neid automaatse tuvastuse kvaliteedi hindamisel arvesse.

Verb	Argument1	Argument2	Argument3	Argument4	Argument5
<i>riputama</i>	AGENT	PATSIENT	SIHTKOHT	SAAJA	
<i>riputama</i>	AGENT	PATSIENT	TULEMUS		
<i>tirima</i>	AGENT	PATSIENT	SIHTKOHT	LÄHTEKOHT	
<i>loopima</i>	AGENT	PATSIENT	SIHTKOHT		
<i>paiskama</i>	AGENT	PATSIENT	SIHTKOHT		
<i>raputama</i>	AGENT	PATSIENT	LÄHTEKOHT		
<i>raputama</i>		PATSIENT	SIHTKOHT		
<i>torkama</i>	AGENT	PATSIENT	SIHTKOHT	SAAJA	
<i>sättima</i>	AGENT	PATSIENT	SIHTKOHT	SAAJA	
<i>sättima</i>	AGENT	PATSIENT	TULEMUS		
<i>toppima</i>	AGENT	PATSIENT	SIHTKOHT		
<i>laduma</i>	AGENT	PATSIENT	SIHTKOHT	SAAJA	
<i>poetama</i>	AGENT	PATSIENT	SIHTKOHT	SAAJA	
<i>poetama</i>	AGENT	OTSEKÕNE			
<i>pistma</i>	AGENT	PATSIENT	SIHTKOHT		

<i>pistma</i>	AGENT	PATSIENT	TULEMUS		
<i>asetama</i>	AGENT	PATSIENT	SIHTKOHT		
<i>langetama</i>	AGENT	PATSIENT	SIHTKOHT		
<i>sõidutama</i>	AGENT	PATSIENT	SIHTKOHT	LÄHTEKOHT	
<i>paigutama</i>	AGENT	PATSIENT	SIHTKOHT		
<i>vedama</i>	AGENT	PATSIENT	SIHTKOHT		
<i>vedama</i>	KOGEJA	MILLEGA VEAB			
<i>vedama</i>	AGENT	PATSIENT	TULEMUS		
<i>heitma</i>	AGENT	PATSIENT	SIHTKOHT	LÄHTEKOHT	INSTRUMENT
<i>heitma</i>	AGENT	PATSIENT	SIHTKOHT	OTSTARVE	
<i>lukkama</i>	AGENT	PATSIENT	SIHTKOHT		
<i>tõukama</i>	AGENT	PATSIENT	SIHTKOHT		
<i>seadma</i>	AGENT	PATSIENT	SIHTKOHT		
<i>seadma</i>	AGENT	PATSIENT	EESMÄRK		
<i>keerama</i>	AGENT	PATSIENT	SIHTKOHT		
<i>keerama</i>	AGENT	PATSIENT	SEISUND		
<i>viskama</i>	AGENT	PATSIENT	SIHTKOHT		
<i>tõmbama</i>	AGENT	PATSIENT	SIHTKOHT	LÄHTEKOHT	
<i>tõmbama</i>	AGENT	PATSIENT	SAAJA		

<i>tõmbama</i>	AGENT	PATSIENT	SEISUND		
<i>tõstma</i>	AGENT	PATSIENT	SIHTKOHT		
<i>tõstma</i>	AGENT	PATSIENT	HULK		
<i>tõstma</i>	AGENT	PATSIENT			
<i>liigutama</i>	AGENT	PATSIENT			
<i>viima</i>	AGENT	PATSIENT	SIHTKOHT	LÄHTEKOHT	SAAJA
<i>viima</i>	AGENT	SIHTKOHT	LÄHTEKOHT		
<i>tooma</i>	AGENT	PATSIENT	SIHTKOHT	LÄHTEKOHT	SAAJA
<i>tooma</i>	AGENT	PATSIENT	LÄHTEKOHT		
<i>tooma</i>	AGENT	PATSIENT	SAAJA		
<i>tooma</i>	AGENT	PATSIENT	MILLENA/MILL EKS		
<i>panema</i>	AGENT	PATSIENT	SIHTKOHT	SAAJA	
<i>panema</i>	AGENT	PATSIENT	SIHTKOHT		
<i>panema</i>	AGENT	PATSIENT	SEISUND		
<i>panema</i>	AGENT	PATSIENT	TEGEVUS		

## **Lisa 2. Käsitsi koostatud semantiliste mallide grammatilised teisendused**

Halliga on tähistatud rollid, mis jäid seotud ja vaba laiendi piirimaale, mistõttu ei võetud neid automaatse tuvastuse kvaliteedi hindamisel arvesse.

Verb	Lauseliige1	Lauseliige2	Lauseliige3	Lauseliige4	Lauseliige5
<i>riputama</i>	(‘nsubj’, ‘*’, ‘’, ‘’)	(‘obj’, ‘*’, ‘’, ‘’)	(‘obl’, ‘adit’, ‘’, ‘’)	(‘obl’, ‘all’, ‘’, ‘’)	
<i>riputama</i>	(‘nsubj’, ‘*’, ‘’, ‘’)	(‘obj’, ‘*’, ‘’, ‘’)	(‘xcomp’, ‘ill’, ‘’, ‘ma’)		
<i>tirima</i>	(‘nsubj’, ‘*’, ‘’, ‘’)	(‘obj’, ‘*’, ‘’, ‘’)	(‘obl’, ‘adit’, ‘’, ‘’)	(‘obl’, ‘el’, ‘’, ‘’)	
<i>tirima</i>	(‘nsubj’, ‘*’, ‘’, ‘’)	(‘obj’, ‘*’, ‘’, ‘’)	(‘obl’, ‘all’, ‘’, ‘’)		
<i>loopima</i>	(‘nsubj’, ‘*’, ‘’, ‘’)	(‘obj’, ‘*’, ‘’, ‘’)	(‘obl’, ‘adit’, ‘’, ‘’)		
<i>loopima</i>	(‘nsubj’, ‘*’, ‘’, ‘’)	(‘obj’, ‘*’, ‘’, ‘’)	(‘obl’, ‘ill’, ‘’, ‘’)		
<i>loopima</i>	(‘nsubj’, ‘*’, ‘’, ‘’)	(‘obj’, ‘*’, ‘’, ‘’)	(‘obl’, ‘all’, ‘’, ‘’)		
<i>paiskama</i>	(‘nsubj’, ‘*’, ‘’, ‘’)	(‘obj’, ‘*’, ‘’, ‘’)	(‘obl’, ‘adit’, ‘’, ‘’)		
<i>paiskama</i>	(‘nsubj’, ‘*’, ‘’, ‘’)	(‘obj’, ‘*’, ‘’, ‘’)	(‘obl’, ‘all’, ‘’, ‘’)		
<i>raputama</i>	(‘nsubj’, ‘*’, ‘’, ‘’)	(‘obj’, ‘*’, ‘’, ‘’)	(‘obl’, ‘all’, ‘’, ‘’)		
<i>raputama</i>	(‘nsubj’, ‘*’, ‘’, ‘’)	(‘obj’, ‘*’, ‘’, ‘’)	(‘obl’, ‘el’, ‘’, ‘’)		
<i>torkama</i>	(‘nsubj’, ‘*’, ‘’, ‘’)	(‘obj’, ‘*’, ‘’, ‘’)	(‘obl’, ‘adit’, ‘’, ‘’)	(‘obl’, ‘all’, ‘’, ‘’)	
<i>torkama</i>	(‘nsubj’, ‘*’, ‘’, ‘’)	(‘obj’, ‘*’, ‘’, ‘’)	(‘obl’, ‘ill’, ‘’, ‘’)	(‘obl’, ‘all’, ‘’, ‘’)	

			‘)	‘)	
<i>sättima</i>	(‘nsubj’, ‘*’, ‘’, ‘’)	(‘obj’, ‘*’, ‘’, ‘’)	(‘obl’, ‘ill’, ‘’, ‘’)	(‘obl’, ‘all’, ‘’, ‘’)	
<i>sättima</i>	(‘nsubj’, ‘*’, ‘’, ‘’)	(‘obj’, ‘*’, ‘’, ‘’)	(‘xcomp’, ‘ill’, ‘’, ‘ma’)		
<i>toppima</i>	(‘nsubj’, ‘*’, ‘’, ‘’)	(‘obj’, ‘*’, ‘’, ‘’)	(‘obl’, ‘adit’, ‘’, ‘’)		
<i>toppima</i>	(‘nsubj’, ‘*’, ‘’, ‘’)	(‘obj’, ‘*’, ‘’, ‘’)	(‘obl’, ‘ill’, ‘’, ‘’)		
<i>toppima</i>	(‘nsubj’, ‘*’, ‘’, ‘’)	(‘obj’, ‘*’, ‘’, ‘’)	(‘obl’, ‘all’, ‘’, ‘’)		
<i>laduma</i>	(‘nsubj’, ‘*’, ‘’, ‘’)	(‘obj’, ‘*’, ‘’, ‘’)	(‘obl’, ‘adit’, ‘’, ‘’)	(‘obl’, ‘all’, ‘’, ‘’)	
<i>laduma</i>	(‘nsubj’, ‘*’, ‘’, ‘’)	(‘obj’, ‘*’, ‘’, ‘’)	(‘obl’, ‘ill’, ‘’, ‘’)	(‘obl’, ‘all’, ‘’, ‘’)	
<i>laduma</i>	(‘nsubj’, ‘*’, ‘’, ‘’)	(‘obj’, ‘*’, ‘’, ‘’)	(‘obl’, ‘all’, ‘’, ‘’)	(‘obl’, ‘all’, ‘’, ‘’)	
<i>poetama</i>	(‘nsubj’, ‘*’, ‘’, ‘’)	(‘obj’, ‘*’, ‘’, ‘’)	(‘obl’, ‘adit’, ‘’, ‘’)	(‘obl’, ‘all’, ‘’, ‘’)	
<i>poetama</i>	(‘nsubj’, ‘*’, ‘’, ‘’)	(‘obj’, ‘*’, ‘’, ‘’)	(‘obl’, ‘all’, ‘’, ‘’)	(‘obl’, ‘all’, ‘’, ‘’)	
<i>poetama</i>	(‘nsubj’, ‘*’, ‘’, ‘’)	(‘ccomp’, ‘<puudub>’, ‘’, ‘’)			
<i>pistma</i>	(‘nsubj’, ‘*’, ‘’, ‘’)	(‘obj’, ‘*’, ‘’, ‘’)	(‘obl’, ‘adit’, ‘’, ‘’)		

<i>pistma</i>	(‘nsubj’, ‘*’, ‘’, ‘’)	(‘obj’, ‘*’, ‘’, ‘’)	(‘obl’, ‘ill’, ‘’, ‘’)		
<i>pistma</i>	(‘nsubj’, ‘*’, ‘’, ‘’)	(‘obj’, ‘*’, ‘’, ‘’)	(‘xcomp’, ‘ill’, ‘’, ‘ma’)		
<i>asetama</i>	(‘nsubj’, ‘*’, ‘’, ‘’)	(‘obj’, ‘*’, ‘’, ‘’)	(‘obl’, ‘adit’, ‘’, ‘’)		
<i>asetama</i>	(‘nsubj’, ‘*’, ‘’, ‘’)	(‘obj’, ‘*’, ‘’, ‘’)	(‘obl’, ‘ill’, ‘’, ‘’)		
<i>asetama</i>	(‘nsubj’, ‘*’, ‘’, ‘’)	(‘obj’, ‘*’, ‘’, ‘’)	(‘obl’, ‘all’, ‘’, ‘’)		
<i>langetama</i>	(‘nsubj’, ‘*’, ‘’, ‘’)	(‘obj’, ‘*’, ‘’, ‘’)	(‘obl’, ‘all’, ‘’, ‘’)		
<i>sõidutama</i>	(‘nsubj’, ‘*’, ‘’, ‘’)	(‘obj’, ‘*’, ‘’, ‘’)	(‘obl’, ‘adit’, ‘’, ‘’)	(‘obl’, ‘el’, ‘’, ‘’)	
<i>sõidutama</i>	(‘nsubj’, ‘*’, ‘’, ‘’)	(‘obj’, ‘*’, ‘’, ‘’)	(‘obl’, ‘all’, ‘’, ‘’)	(‘obl’, ‘el’, ‘’, ‘’)	
<i>paigutama</i>	(‘nsubj’, ‘*’, ‘’, ‘’)	(‘obj’, ‘*’, ‘’, ‘’)	(‘obl’, ‘adit’, ‘’, ‘’)		
<i>paigutama</i>	(‘nsubj’, ‘*’, ‘’, ‘’)	(‘obj’, ‘*’, ‘’, ‘’)	(‘obl’, ‘ill’, ‘’, ‘’)		
<i>paigutama</i>	(‘nsubj’, ‘*’, ‘’, ‘’)	(‘obj’, ‘*’, ‘’, ‘’)	(‘obl’, ‘all’, ‘’, ‘’)		
<i>vedama</i>	(‘nsubj’, ‘*’, ‘’, ‘’)	(‘obj’, ‘*’, ‘’, ‘’)	(‘obl’, ‘ill’, ‘’, ‘’)		
<i>vedama</i>	(‘nsubj’, ‘*’, ‘’, ‘’)	(‘obj’, ‘*’, ‘’, ‘’)	(‘obl’, ‘all’, ‘’, ‘’)		

<i>vedama</i>	(‘obl’, ‘ad’, ‘’, ‘’)	(‘obl’, ‘kom’, ‘’, ‘’)			
<i>vedama</i>	(‘nsubj’, ‘*’, ‘’, ‘’)	(‘obj’, ‘*’, ‘’, ‘’)	(‘xcomp’, ‘ill’, ‘’, ‘ma’)		
<i>heitma</i>	(‘nsubj’, ‘*’, ‘’, ‘’)	(‘obj’, ‘*’, ‘’, ‘’)	(‘obl’, ‘adit’, ‘’, ‘’)	(‘obl’, ‘el’, ‘’, ‘’)	(‘obl’, ‘kom’, ‘’, ‘’)
<i>heitma</i>	(‘nsubj’, ‘*’, ‘’, ‘’)	(‘obj’, ‘*’, ‘’, ‘’)	(‘obl’, ‘all’, ‘’, ‘’)	(‘obl’, ‘el’, ‘’, ‘’)	(‘obl’, ‘kom’, ‘’, ‘’)
<i>heitma</i>	(‘nsubj’, ‘*’, ‘’, ‘’)	(‘xcomp’, ‘ill’, ‘’, ‘ma’)	(‘obl’, ‘adit’, ‘’, ‘’)/(‘obl’, ‘all’, ‘’, ‘’)/(‘obl’, ‘ill’, ‘’, ‘’)		
<i>lukkama</i>	(‘nsubj’, ‘*’, ‘’, ‘’)	(‘obj’, ‘*’, ‘’, ‘’)	(‘obl’, ‘adit’, ‘’, ‘’)		
<i>lukkama</i>	(‘nsubj’, ‘*’, ‘’, ‘’)	(‘obj’, ‘*’, ‘’, ‘’)	(‘obl’, ‘ill’, ‘’, ‘’)		
<i>lukkama</i>	(‘nsubj’, ‘*’, ‘’, ‘’)	(‘obj’, ‘*’, ‘’, ‘’)	(‘obl’, ‘all’, ‘’, ‘’)		
<i>tõukama</i>	(‘nsubj’, ‘*’, ‘’, ‘’)	(‘obj’, ‘*’, ‘’, ‘’)	(‘obl’, ‘adit’, ‘’, ‘’)		
<i>tõukama</i>	(‘nsubj’, ‘*’, ‘’, ‘’)	(‘obj’, ‘*’, ‘’, ‘’)	(‘obl’, ‘ill’, ‘’, ‘’)		
<i>tõukama</i>	(‘nsubj’, ‘*’, ‘’, ‘’)	(‘obj’, ‘*’, ‘’, ‘’)	(‘obl’, ‘all’, ‘’, ‘’)		
<i>seadma</i>	(‘nsubj’, ‘*’, ‘’, ‘’)	(‘obj’, ‘*’, ‘’, ‘’)	(‘obl’, ‘adit’, ‘’, ‘’)		

<i>seadma</i>	(‘nsubj’, ‘*’, ‘’, ‘’)	(‘obj’, ‘*’, ‘’, ‘’)	(‘obl’, ‘ill’, ‘’, ‘’)		
<i>seadma</i>	(‘nsubj’, ‘*’, ‘’, ‘’)	(‘obj’, ‘*’, ‘’, ‘’)	(‘obl’, ‘all’, ‘’, ‘’)		
<i>seadma</i>	(‘nsubj’, ‘*’, ‘’, ‘’)	(‘obj’, ‘*’, ‘’, ‘’)	(‘xcomp’, ‘tr’, ‘’, ‘’)		
<i>keerama</i>	(‘nsubj’, ‘*’, ‘’, ‘’)	(‘obj’, ‘*’, ‘’, ‘’)	(‘obl’, ‘adit’, ‘’, ‘’)		
<i>keerama</i>	(‘nsubj’, ‘*’, ‘’, ‘’)	(‘obj’, ‘*’, ‘’, ‘’)	(‘obl’, ‘ill’, ‘’, ‘’)		
<i>keerama</i>	(‘nsubj’, ‘*’, ‘’, ‘’)	(‘obj’, ‘*’, ‘’, ‘’)	(‘obl’, ‘all’, ‘’, ‘’)		
<i>keerama</i>	(‘nsubj’, ‘*’, ‘’, ‘’)	(‘obj’, ‘*’, ‘’, ‘’)	(‘obl’, ‘tr’, ‘’, ‘’)		
<i>keerama</i>	(‘nsubj’, ‘*’, ‘’, ‘’)	(‘obj’, ‘*’, ‘’, ‘’)	(‘advmod’, ‘<puudub>’, ‘’, ‘’)		
<i>viskama</i>	(‘nsubj’, ‘*’, ‘’, ‘’)	(‘obj’, ‘*’, ‘’, ‘’)	(‘obl’, ‘adit’, ‘’, ‘’)		
<i>viskama</i>	(‘nsubj’, ‘*’, ‘’, ‘’)	(‘obj’, ‘*’, ‘’, ‘’)	(‘obl’, ‘ill’, ‘’, ‘’)		
<i>viskama</i>	(‘nsubj’, ‘*’, ‘’, ‘’)	(‘obj’, ‘*’, ‘’, ‘’)	(‘obl’, ‘all’, ‘’, ‘’)		
<i>tõmbama</i>	(‘nsubj’, ‘*’, ‘’, ‘’)	(‘obj’, ‘*’, ‘’, ‘’)	(‘obl’, ‘adit’, ‘’, ‘’)	(‘obl’, ‘el’, ‘’, ‘’) / (‘obl’, ‘abl’, ‘’, ‘’)	
<i>tõmbama</i>	(‘nsubj’, ‘*’, ‘’, ‘’)	(‘obj’, ‘*’, ‘’, ‘’)	(‘obl’, ‘ill’, ‘’, ‘’)	(‘obl’, ‘el’, ‘’, ‘’)	

			‘)	‘)/ (‘obl’, ‘abl’, ‘, ‘)	
<i>tõmbama</i>	(‘nsubj’, ‘*’, ‘, ‘)	(‘obj’, ‘*’, ‘, ‘)	(‘obl’, ‘all’, ‘, ‘)		
<i>tõmbama</i>	(‘nsubj’, ‘*’, ‘, ‘)	(‘obj’, ‘*’, ‘, ‘)	(‘xcomp’, ‘ill’, ‘, ‘, ‘ma’)		
<i>tõmbama</i>	(‘nsubj’, ‘*’, ‘, ‘)	(‘obj’, ‘*’, ‘, ‘)	(‘advmod’, ‘<puudub>’, ‘, ‘)		
<i>tõstma</i>	(‘nsubj’, ‘*’, ‘, ‘)	(‘obj’, ‘*’, ‘, ‘)	(‘obl’, ‘adit’, ‘, ‘)		
<i>tõstma</i>	(‘nsubj’, ‘*’, ‘, ‘)	(‘obj’, ‘*’, ‘, ‘)	(‘obl’, ‘ill’, ‘, ‘)		
<i>tõstma</i>	(‘nsubj’, ‘*’, ‘, ‘)	(‘obj’, ‘*’, ‘, ‘)	(‘obl’, ‘all’, ‘, ‘)		
<i>tõstma</i>	(‘nsubj’, ‘*’, ‘, ‘)	(‘obj’, ‘*’, ‘, ‘)	(‘obl’, ‘term’, ‘, ‘)		
<i>tõstma</i>	(‘nsubj’, ‘*’, ‘, ‘)	(‘obj’, ‘*’, ‘, ‘)			
<i>liigutama</i>	(‘nsubj’, ‘*’, ‘, ‘)	(‘obj’, ‘*’, ‘, ‘)			
<i>viima</i>	(‘nsubj’, ‘*’, ‘, ‘)	(‘obj’, ‘*’, ‘, ‘)	(‘obl’, ‘adit’, ‘, ‘)	(‘obl’, ‘el’, ‘, ‘)/ (‘obl’, ‘abl’, ‘, ‘)	(‘obl’, ‘all’, ‘, ‘)
<i>viima</i>	(‘nsubj’, ‘*’, ‘, ‘)	(‘obj’, ‘*’, ‘, ‘)	(‘obl’, ‘ill’, ‘, ‘)	(‘obl’, ‘el’, ‘, ‘)/ (‘obl’, ‘abl’, ‘, ‘)	(‘obl’, ‘all’, ‘, ‘)
<i>viima</i>	(‘nsubj’, ‘*’, ‘, ‘)	(‘obj’, ‘*’, ‘, ‘)	(‘obl’, ‘all’, ‘, ‘)	(‘obl’, ‘el’, ‘, ‘)	(‘obl’, ‘all’, ‘,

			‘)	‘)/(‘obl’, ‘abl’, ‘, ‘)	‘, ‘)
<i>viima</i>	(‘nsubj’, ‘*’, ‘, ‘)	(‘obj’, ‘*’, ‘, ‘)	(‘obl’, ‘term’, ‘, ‘)		
<i>viima</i>	(‘nsubj’, ‘*’, ‘, ‘)	(‘obl’, ‘ill’, ‘, ‘)	(‘obl’, ‘abl’, ‘, ‘)		
<i>viima</i>	(‘nsubj’, ‘*’, ‘, ‘)	(‘obl’, ‘all’, ‘, ‘)	(‘obl’, ‘abl’, ‘, ‘)		
<i>viima</i>	(‘nsubj’, ‘*’, ‘, ‘)	(‘obj’, ‘*’, ‘, ‘)	(‘xcomp’, ‘ill’, ‘, ‘)	(‘obl’, ‘adit’, ‘, ‘)/(‘obl’, ‘all’, ‘, ‘)/(‘obl’, ‘ill’, ‘, ‘)	
<i>tooma</i>	(‘nsubj’, ‘*’, ‘, ‘)	(‘obj’, ‘*’, ‘, ‘)	(‘obl’, ‘adit’, ‘, ‘)	(‘obl’, ‘el’, ‘, ‘)/(‘obl’, ‘abl’, ‘, ‘)	(‘obl’, ‘all’, ‘, ‘)
<i>tooma</i>	(‘nsubj’, ‘*’, ‘, ‘)	(‘obj’, ‘*’, ‘, ‘)	(‘obl’, ‘ill’, ‘, ‘)	(‘obl’, ‘el’, ‘, ‘)/(‘obl’, ‘abl’, ‘, ‘)	(‘obl’, ‘all’, ‘, ‘)
<i>tooma</i>	(‘nsubj’, ‘*’, ‘, ‘)	(‘obj’, ‘*’, ‘, ‘)	(‘obl’, ‘all’, ‘, ‘)	(‘obl’, ‘el’, ‘, ‘)/(‘obl’, ‘abl’, ‘, ‘)	(‘obl’, ‘all’, ‘, ‘)
<i>tooma</i>	(‘nsubj’, ‘*’, ‘, ‘)	(‘obj’, ‘*’, ‘, ‘)	(‘obl’, ‘el’, ‘, ‘)		
<i>tooma</i>	(‘nsubj’, ‘*’, ‘, ‘)	(‘obj’, ‘*’, ‘, ‘)	(‘obl’, ‘abl’, ‘, ‘)		
<i>tooma</i>	(‘nsubj’, ‘*’, ‘, ‘)	(‘obj’, ‘*’, ‘, ‘)	(‘obl’, ‘es’, ‘, ‘)		
<i>tooma</i>	(‘nsubj’, ‘*’, ‘, ‘)	(‘obj’, ‘*’, ‘, ‘)	(‘xcomp’, ‘tr’, ‘, ‘)		

<i>panema</i>	(‘nsubj’, ‘*’, ‘’, ‘’)	(‘obj’, ‘*’, ‘’, ‘’)	(‘obl’, ‘adit’, ‘’, ‘’)		
<i>panema</i>	(‘nsubj’, ‘*’, ‘’, ‘’)	(‘obj’, ‘*’, ‘’, ‘’)	(‘obl’, ‘ill’, ‘’, ‘’)		
<i>panema</i>	(‘nsubj’, ‘*’, ‘’, ‘’)	(‘obj’, ‘*’, ‘’, ‘’)	(‘obl’, ‘all’, ‘’, ‘’)		
<i>panema</i>	(‘nsubj’, ‘*’, ‘’, ‘’)	(‘obj’, ‘*’, ‘’, ‘’)	(‘obl’, ‘adit’, ‘’, ‘’)	(‘obl’, ‘all’, ‘’, ‘’)	
<i>panema</i>	(‘nsubj’, ‘*’, ‘’, ‘’)	(‘obj’, ‘*’, ‘’, ‘’)	(‘obl’, ‘ill’, ‘’, ‘’)	(‘obl’, ‘all’, ‘’, ‘’)	
<i>panema</i>	(‘nsubj’, ‘*’, ‘’, ‘’)	(‘obj’, ‘*’, ‘’, ‘’)	(‘xcomp’, ‘ill’, ‘’, ‘ma’)		

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

Mina, Kertu Saul,

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) minu loodud teose “Eesti keele lausemallide automaatne tuvastamine liigutamisverbide näitel”, mille juhendajad on kaasprofessor Kadri Muischnek ja vanemteadur Jelena Kallas, reprodutseerimiseks eesmärgiga seda säilitada, sealhulgas lisada digitaalarhiivi DSpace kuni autoriõiguse kehtivuse lõppemiseni.
2. Annan Tartu Ülikoolile loa teha punktis 1 nimetatud teos üldsusele kättesaadavaks Tartu Ülikooli veebikeskkonna, sealhulgas digitaalarhiivi DSpace kaudu Creative Commons'i litsentsiga CC BY NC ND 4.0, mis lubab autorile viidates teost reprodutseerida, levitada ja üldsusele suunata ning keelab luua tuletatud teost ja kasutada teost ärieesmärgil, kuni autoriõiguse kehtivuse lõppemiseni.
3. Olen teadlik, et punktides 1 ja 2 nimetatud õigused jäävad alles ka autorile.
4. Kinnitan, et lihtlitsentsi andmisega ei riku ma teiste isikute intellektuaalomandi ega isikuandmete kaitse õigusaktidest tulenevaid õigusi.

*Kertu Saul*

**27.05.2024**