Humaniora

DISSERTATIONES
LINGUISTICAE
UNIVERSITATIS
TARTUENSIS
23

LIISI PIITS

Sagedamate inimest tähistavate sõnade kollokatsioonid eesti keeles

DISSERTATIONES LINGUISTICAE UNIVERSITATIS TARTUENSIS
23

DISSERTATIONES LINGUISTICAE UNIVERSITATIS TARTUENSIS
23

LIISI PIITS

Sagedamate inimest tähistavate sõnade kollokatsioonid eesti keeles

TARTU ÜLIKOOL
KIRJASTUS
1632

Tartu Ülikooli filosoofiateaduskond, eesti ja üldkeeleteaduse instituut

Väitekiri on lubatud kaitsmisele filosoofiadoktori kraadi taotlemiseks (üldkeeleteaduses) Tartu Ülikooli eesti ja üldkeeleteaduse instituudi nõukogu otsusega 13. jaanuaril 2015. a.

Juhendaja: prof Urmas Sutrop, filosoofiadoktor, Tartu Ülikool, Eesti Keele Instituut

Oponent: prof Pirkko Nuolijärvi, filosoofiadoktor, Kotimaisten kielten keskus

Kaitsmine toimub 21. aprillil 14.15 Tartu Ülikooli senati saalis.

Doktoritöö valmimist on toetanud Euroopa Liidu Sotsiaalfond.
Trükikulud on katnud Eesti Keele Instituut.

Euroopa Liit
Euroopa Sotsiaalfond

Eesti tuleviku heaks

ISSN 1406-5657
ISBN 978-9949-32-759-1 (trükis)
ISBN 978-9949-32-760-7 (pdf)

Autoriõigus Liisi Piits 2015

Tartu Ülikooli Kirjastus
www.tyk.ee

TÄNU

Kuigi õpingud doktorantuuris peaksid arendama eelkõige iseseisva uurimistöö tegemise oskust, on väitekirja valmimisele kaasa aidanud palju inimesi.

Tänan oma juhendajat prof Urmas Sutropit, kes kümne aasta jooksul ei kaotanud midagi oma optimistlikust hoiakust ega väsinud kinnitamast, et see väitekiri saab valmis. Tema initsiatiivi ja toetuseta poleks seda väitekirja sündinud.

Südamlik tänu kuulub Kristiina Rossile, kes on mind heade nõuannetega toetanud kogu ülikooliõpingute aja ja kelle sisulised märkused on mul aidanud oma kirjutatut paremini mõista ja mõtteid ning väljendust loogilisemaks muuta.

Hille Pajupuu ja Jelena Kallas on kolleegid Eesti Keele Instituudist, kes ei pidanud paljuks oma töökohustuste kõrvalt väitekirjaga tutvuda ning kelle tähelepanelikud kommentaarid aitasid väitekirja arusaadavamaks teha. Tänan Elgar Kudritskit, kes on mind aidanud mitmetes tehnilistes küsimustes ja Sirje Ainsaart, kes on inglise keelde tõlkinud nii minu väitekirjaga seotud artikleid kui ka töö kokkuvõtte. Tiia Margus ja Mare Valge Tartu Ülikoolist aitasid mul nende aastate jooksul ainepunktide süsteemis orienteeruda. Lisaks kuulub mu tänu Kadri Muischnekile, kes aastaid tagasi ohverdas pool ilusast suvepäevast, et minu tööks vajalikku korpust morfoloogilisest analüsaatorist läbi lasta. Aitäh Margus Luhale, kes selle töö kujundas ja Mari-Liis Kalvikule ning Kati Seinale, kes aitasid võidelda kirjavigadega.

Põhjalike retsensentide Katre Õimu, Jarmo Harri Jantuse ja Pirkko Nuolijärvi ettepanekute arvestamine andis tööle viimase lihvi. Tänan teid väitekirjale pühendatud tähelepanu eest!

Aitäh Meelis Mihklale, kes on mind kaasanud erinevatesse kõnetehnoloogia projektidesse, olenemata asjaolust, et märgatav osa tööajast kulus aina pikenevatele doktoriõpingutele. Lisaks mõtlen tänuga Helle Metslangi peale, kes mind 15 aastat tagasi nendele teemadele juhatas.

Tänan kõiki, kes minu töö valmimisele kaasa elasid!

Tallinnas 31. jaanuaril 2015. a
Liisi Piits

SISUKORD

1. Sissejuhatus 9
   1.1 Töö eesmärk ja ülesehitus 9
   1.2. Kollokatsioonide uurimise ajaloost ja mõiste definitsioonist 11
      1.2.1. Terminist *kollokatsioon* 11
      1.2.2. Kollokatsiooniseosest erinevatest eesmärkidest lähtuvalt 13
      1.2.3. Kollokatsioon kui tähendussuhe 14
      1.2.4. Kollokatsioonide liigitamisest 15
   1.3. Lühiülevaade töös kasutatavatest meetoditest 16

2. Inimest tähistavate sõnade tähendusest 21
   2.1. Inimest tähistavate sõnade tähenduslikust liigendamisest 22
      2.1.1. Inimest tähistavad sõnad sõnaraamatutes ja tesauruses 22
      2.1.2. Inimest tähistavate sõnade komponentanalüüs 31
   2.2. Inimest tähistavate sõnade paradigmaatilised ja süntagmaatilised leksikaalsed suhted 37
      2.2.1. Paradigmaatilised leksikaalsed suhted 38
      2.2.2. Süntagmaatilised leksikaalsed suhted 44

3. Sagedamate inimest tähistavate sõnade kollokatsioonikobarate võrdlus 45
   3.1. Hüpoteesid 45
   3.2. Uurimismeetod ja analüüsi aluseks olev materjal 47
      3.2.1. Materjal ja vahendid 47
      3.2.2. Kollokaatide leidmise kaugusest 50
   3.3. Vasakule hargnevad ühised kollokaadid 51
      3.3.1. Kollokatsioonikobarate suhestatus 51
      3.3.2. Vasakpoolsetest kollokaatidest lähemalt 56
      3.3.3. Vasakule hargnevad kollokaadid sõnaliigiti 59
   3.4. Paremale hargnevad ühised kollokaadid 61
      3.4.1. Kollokatsioonikobarate suhestatus 61
      3.4.2. Parempoolsetest kollokaatidest lähemalt 64
      3.4.3. Paremale hargnevad kollokaadid sõnaliigiti 65
   3.5. Vahekokkuvõte 68

4. Sagedamad inimest tähistavad sõnad kahes leksikogrammatilises konstruktsioonis 70
4.1. Hüpoteesid 70
4.2. Uurimismeetod ja analüüsi aluseks olev materjal 74
4.2.1. Leksikogrammatiliste konstruktsioonide lühitutvustus 74
4.2.2. Inimest tähistavate sõnade esindatus eesti keele koondkorpuses 75
4.2.3. Antonüümide esinemine süntagmaatilise tasandi konstruktsioonides 77
4.3. Adjektiivsed laiendid 79
4.3.1. Sõnade *poiss* ja *tüdruk* adjektiivsed laiendid 83
4.3.2. Sõnade *mees* ja *naine* adjektiivsed laiendid 87
4.3.3. Sõnade *poeg* ja *tütar* adjektiivsed laiendid 92
4.3.4. Sõnade *ema* ja *isa* adjektiivsed laiendid 95
4.3.5. Sõnade *inimene* ja *laps* adjektiivsed laiendid 98
4.4. Verb inimest tähistava sõna kollokaadina 102
4.4.1. Verbikonstruktsioonide tutvustus 102
4.4.2. Peamised tegevused, mille juures inimest tähistav sõna on subjektiks 106
4.4.3. Antonüümipaaride ja hüponüümipaari eriomased verbid võrdluses 112
4.5. Vahekokkuvõte 127
4.5.1. Soo avaldumine kollokatsioonisuhetes 127
4.5.2. Vanuse avaldumine kollokatsioonisuhetes 130
Kokkuvõte 132
Kirjandus 137
SUMMARY 149
ELULOOKIRJELDUS 158
CURRICULUM VITAE 160
LISAD (vt CD või http://hdl.handle.net/10062/45486)
Lisa 1. Vasakule hargnevad sagedamad kollokaadid
Lisa 2. Paremale hargnevad sagedamad kollokaadid
Lisa 3. Inimest tähistavate sõnade kõik vasakule hargnevad lihtkollokaadid
Lisa 4. Inimest tähistavate sõnade kõik paremale hargnevad lihtkollokaadid
Lisa 5. Inimest tähistavate sõnade adjektiivsed laiendid ja verbid, mille juures inimest väljendavad sõnad esinevad subjekti ja objekti funktsioonis

1. SISSEJUHATUS

1.1. Töö eesmärk ja ülesehitus

Siinses töös lähtutakse strukturalistlikust eeldusest, et keel kujutab endast kindlate vahekordadega seotud üksuste süsteemi, kus üksused (nt sõnad) ei oma kindlat väärtust mitte iseenesest, vaid ainult vastavate teiste üksustega võrdluses. Sõnade tähendusi ei saa määratleda teisiti, kui silmas pidades nende vahekordi teiste sõnadega keeles. Strukturalistlikule arusaamale, et tähendus on midagi lekseemile omast ja seda saab jagada komponentideks, on vastandatud kontekstuaalset arusaama, kus leitakse, et tähendus tekib vaid kindlas kontekstis. Selles töös tahetakse selgust saada, kuidas kaks lähenemist omavahel suhestuvad. Eesmärgiks on uurida inimest tähistavate sõnade kollokatiivseid suhteid teiste sõnadega ja nende sõnade omavaheliste paradigmaatiliste suhete väljendumist süntagmaatilisel tasandil. Tahetakse välja selgitada, kuidas sõna tähendus kajastub tema kollokatsioonides.

Keele kaudu me ainult ei nimeta asju, vaid anname koos nimetamisega edasi ka suhtumist ning kujundame hoiakuid: „Keelekasutus on sotsiaalselt ja kollektiivselt kujundatud, aga ta on ka sotsiaalselt kujundav" (Kasik 2002: 76). Seetõttu on oluline uurida, kuidas kollokatsioonimustrites paljastuvad soo, vanuse ja käitumisega seotud stereotüübid.

Korpuste ja korpuslingvistiliste vahendite kiire arengu tõttu on nii mujal maailmas kui Eestis hoogustunud sõnade süntagmaatiliste seoste uurimine. Kollokatsiooni kui süntagmaatilise tasandi nähtuse erinevaid aspekte on uuritud erinevatest eesmärkidest lähtuvalt, mistõttu erinevatel koolkondadel on välja kujunenud erinevad teoreetilised lähtekohad (vt 1.2.). Siinses töös lähtutakse briti neo-firthliku koolkonna (Sinclair 2004; Stubbs 2001) ja ameerika strukturalistide (Harris 1954, 1957) arusaamast, et sõna distributsioon peegeldab sõna tähendust.

Eestis on kollokatsioone uuritud nii keeleõppe seisukohast (Jaanits 2004, Heinsoo 2010), arvutilingvistika vaatepunktist (Kaalep, Muischnek 2009; Muischnek 2005), murdejoonte kaardistamise eesmärgil (Uiboaed 2010) kui ka leksikograafilisest aspektist (Kallas 2013). Käesolev töö püüab eesti keeleteaduses täita lünka, mis esineb sõna distributiivse käitumise ja semantika puutepunktide uurimises. Tahetakse välja selgitada, kuidas kollokatsioonisuhetes avalduvad otsisõnaga tähistatud inimese sugu ja vanus. Lähtutakse põhimõttest, et iga korduv koosesinemine kannab endas vihjeid kultuurikontekstile (Tognini-Bonelli 2001: 125) ning soo- ja/või vanusevastandusel põhinevate otsisõnade kollokaate uurides saab selgeks, milline on stereotüüpne nägemus mehest ja naisest, poisist ja tüdrukust.

Töös kontrollitakse järgnevaid hüpoteese (põhjalikumalt on hüpoteese kirjeldatud peatükkides 3.1 ja 4.1):

1. Mida rohkem sarnase väärtusega tähenduskomponente kahel sõnal esineb, seda suurem on kollokatsioonikobarate kattumine.

2. Paradigmaatilised suhted avalduvad kollokatsioonides.

3. Keele grammatikast sõltub piirkond, kust mingit tüüpi kollokatsiooniseoseid tasub otsida.

4. Inimest tähistavate sõnade kollokatsioonides avalduvad sugudevahelised erinevused.

5. Inimest tähistavate sõnade kollokatsioonid peegeldavad võimusuhteid ühiskonnas.

6. Inimest tähistavate sõnade kollokatsioonides avaldub otsisõnaga tähistatud tüüpilise isiku vanus.

Nende hüpoteeside kontrollimiseks antakse ülevaade inimest tähistavate sõnade paradigmaatilistest suhetest, analüüsitakse komponentanalüüsiga (Fodor, Katz 1963; Nida 1975) sõnade paiknemist inimese tähendusväljas ja uuritakse põhjalikult sõnade kollokatiivseid suhteid kahel teineteist täiendaval meetodil.

Väitekiri koosneb neljast osast, kokkuvõttest ja lisadest. Sissejuhatavas peatükis tutvustatakse kollokatsiooni mõistet ja definitsioone ning antakse väike ülevaade kollokatsioonide uurimise ajaloost. Samuti esitatakse lühiülevaade töös kasutatud meetoditest ja nende valiku seostest uurimiseesmärgiga.

Teise osa eesmärgiks on luua teoreetiline taust inimest tähistavate sõnade kollokatsioonide uurimiseks. Selles osas kirjeldatakse sõnade *ema*, *inimene*, *isa*, *laps*, *mees*, *naine*, *poeg*, *poiss*, *tüdruk* ja *tütar* tähenduste esitust erinevates sõnaraamatutes ja Eesti Wordnetis. Edasi vaadeldakse inimest tähistavaid sõnu tähendusvälja teooriast lähtudes ja uuritakse sõnade tähenduserinevusi komponentanalüüsile toetudes. Antakse ülevaade peamistest inimest tähistavate sõnade vahelistest paradigmaatilistest suhetest nagu hüponüümia, sünonüümia ja antonüümia erinevad liigid.

Kolmandas ja neljandas osas uuritakse korpusandmete (kasutatakse TÜ eesti keele koondkorpust) alusel inimest tähistavate sõnade distributsiooni ja otsitakse märke seostest keele süntagmaatilise ja paradigmaatilise tasandi vahel. Kolmandas peatükis analüüsitakse eraldi otsisõnast vasakule ja paremale jäävaid kollokaate. Otsisõnade kollokatsioonikobarate kattuvusi võrreldes vaadeldakse, millised paradigmaatilised suhted peegelduvad kollokatsiooniseostes. Tuuakse välja erinevused vasakule ja paremale hargnevate kollokaatide sõnaliigilises kuuluvuses. Neljandas peatükis käsitletakse inimest tähistavate sõnade kollokatsioone leksikogrammatilistest konstruktsioonidest lähtuvalt. Võrreldakse antonüümipaare, lähtudes eeldusest, et sagedamate kollokatsiooniseoste põhjal saab dekonstrueerida peidetud tähendusi ja viise, kuidas inimesi on tekstides kujutatud. Inimest tähistavate sõnade kollokaatide täielikud loendid ehk andmed, millele analüüs põhineb, on esitatud lisades.

Kokkuvõttes tuuakse välja töö peamised tulemused.

Tööga on kaasas 5 lisa, mis on kättesaadavad nii kaasasoleval CD-l kui veebiaadressil http://hdl.handle.net/10062/45486. Lisades on esitatud keelekorpuste põhjal koostatud inimest tähistavate sõnade kollokatsiooniseoste sagedusloendid.

1.2. Kollokatsioonide uurimise ajaloost ja mõiste definitsioonist

1.2.1. Terminist *kollokatsioon*

Termini *kollokatsioon* kasutuselevõttu moodsas keeleteaduses on seostatud kahe kuulsa briti lingvisti – John Rupert Firthi ja Harold E. Palmeri nimega. Neist viimane kasutas terminit *kollokatsioon* 1938. aastal ilmunud teoses „A Grammar of English Words" (1949 [1938]: x), kus kollokatsiooni nähti just keeleõppe seisukohast olulisena – „kollokatsioon on kahe või enama sõna järgnevus, mida tuleks õppida kui üht sõna". Teoreetilise tähtsuse ja olulise koha tähendusõpetuses andis kollokatsioonile esimesena siiski Palmeri kaasmaalane J. R. Firth, kes nimetas kollokatsiooni süntagmaatilise tasandi abstraktsiooniks. Kuulsaks on saanud tema lause: „You shall know a word by the company it keeps!" 'Tunne sõna tema kaaslaste järgi!'. Firth juhtis selle lausega tähelepanu tõsiasjale, et sõnad ei kombineeru juhuslikult ning ainult süntaktilisi reegleid järgides fraasidesse ja lausetesse, vaid et sõnade vahel valitsevad veel erilised seosed ja see sõnade koosesinemisviis annab nende kohta olulist informatsiooni. Just sellise koosesinemisviisi tähistamiseks rakendas Firth terminit *kollokatsioon*, mis tema kasutuses tähistas *harjumuslikku koosesinemist* (Firth 1957: 194). Kahjuks oli juba Firthi loodud määratlus ebaselge ega defineerinud mõistet piisavalt täpselt, mistõttu hiljem on segadus termini ümber ainult suurenenud. Laias laastus võib mõiste *kollokatsioon* määratlused jagada kaheks:

1. Kitsamalt määratletuna nähakse kollokatsiooni fraseoloogia osana. Mõistet üritatakse paigutada skaalale, kus ühes otsas on vabad sõnaühendid ja teises skaala otsas idioomid. Määratlemiseks kasutatakse mõisteid *transparentsus* ja *asendatavus*. Kus täpsemalt kollokatsioonid sel skaalal paiknevad ja kui transparentne või asendatav peab kollokatiivne ühend olema, selles ei olda aga sugugi ühel meelel.

Näiteks Palmeri definitsioonis ei saa kollokatsiooni tähendust tuletada seda moodustavate üksiksõnade tähendustest, vaid ta soovitab neid õpetada kui üksiksõnu ja peab oluliseks, et kollokatsioonid oleks esitatud sõnaraamatutes (Palmer 1949 [1938]: x). Sarnaselt määratleb kollokatsiooni tänapäeval ka Stefan Evert (2005: 17), kes defineerib kollokatsiooni kui sõnade kombinatsiooni, mille semantilised ja/või süntaktilised omadused ei ole täielikult ennustatavad üksikkomponentide põhjal ja mis seetõttu peaksid olema esindatud sõnaraamatutes. Nii Palmeri kui Everti definitsioonid peegeldavad hästi tõsiasja, et kollokatsiooni määratlus sõltub eesmärgist. Leksikograafilisest ja keeleõppe seisukohast määratletakse kollokatsiooni rohkem fraseoloogia osana. Pedagoogikas nähakse kollokatsioonis silda, mis aitaks ületada grammatika ja sõnavara vahelist lahusust (Krishnamurthy 2006: 596). Eestis on keeleõppes olulisi kollokatsioone uurinud oma magistritöös „Leksikaalsetest kollokatsioonidest soome ja eesti keeles" Kadri Jaanits, kes on aga transparentsuse seisukohalt Evertiga

vastupidisel arvamusel, defineerides kollokatsiooni kui väljendit „millest on võimalik aru saada sõnade üksiktähenduste kaudu; eeldades, et kombinatsiooni kõigil sõnadel on iseseisev tähendus" (Jaanits 2004: 7).

2. Laiem määratlus peab kollokatsiooniks kahe või enama sõna koosesinemist rohkem kui üks kord. Määratlemisel on oluline vaid sageduse arvestamine. Ühendi transparentsus või asendatavus ei mängi selles määratluses rolli. Kõige üldisemalt määratles kollokatsiooni briti lingvist John Sinclair (2004): „Collocation is the occurence of two or more words within a short space of each other in a text" 'kollokatsioon on kahe või enama sõna esinemine tekstis väikese vahega'. John Sinclair hakkas esimesena katsetama Firthi ideid korpusest saadud andmetel. Aastate jooksul koondas ta Birminghami ülikoolis enda ümber lingviste nagu Michael Hoey, Susan Hunston, Michael Stubbs, Wolfgang Teubert ja Elena Tognini-Bonelli, kellel oli sarnane nägemus kollokatiivsetest seostest ja keda peetakse neo-firthliku suuna esindajateks (McEnery, Hardie 2012: 122). Nt Michael Stubbs (2001: 24; 29) ütleb: „Kollokatsioon on leksikaalne suhe kahe sõna vahel, millel on tendents koos esineda". Stubbsi määratlus põhineb sõnade koosesinemissagedusel: kollokatsioon on sage koosesinemine.

Lisaks sagedusel põhineva koosesinemissageduse arvestamisele on kollokatsiooni mõistet seostatud ka mitmete assotsiatsioonimudelite kasutusega, mis arvestavad kollokatsioonide leidmisel koosesinemissageduse seost üksiksõnade sagedustega. Eesti autoritest määratleb Kadri Muischnek kollokatsiooni kui sõnaühendit, mis on defineeritud selle järgi, et teda moodustavad sõnad esinevad tekstides koos sagedamini, kui võiks eeldada nende eraldi esinemise sagedustest (Kaalep, Muischnek 2002b). Muischnek (2005) on edukalt kasutanud sõnadevaheliste seoste statistikuid määratlemaks verbikeskseid püsiühendeid. Põhjaliku ülevaate erinevatest assotsiatsioonimudelitest annavad Evert (2005) ja Bartsch (2004). Eesti keele murdekeelele sobivat sõnadevahelise seose statistikut on otsinud Kristel Uiboaed (2010: 307–326). Praktikas leiavad assotsiatsioonimudelid kasutust sõnatähenduste ühestamisel (vt nt Leacock jt 1996, Gale jt 1993).

Vahel on segaduste vältimiseks laiema kollokatsiooni mõiste asemel kasutatud terminit *koosesinemine* (*co-occurence*), vt näiteks Evert 2005. Ka 1998. aastal loodud Leipzigi ülikooli kollokatsioonide leidmise otsingumootor *Wortschatz*[1] on *kollokatsioonid* aastate jooksul ümber nimetanud *koosesinemisteks*. Samal ajal on briti traditsioonis mindud hoopis vastupidist teed, nimetades kõiki koosesinemisi *kollokatsioonideks*, aga selliseid koosesinemisi, mis esinevad sagedamini, kui võiks eeldada nende sõnade eraldi esinemise sagedusest, *tähenduslikeks kollokatsioonideks (significant collocations)*.

Siinses töös kasutatakse terminit *kollokatsioon* just laiemas tähenduses, tähistamaks sõnade leksikaalset koosesinemist neo-firthlikus mõttes. Väitekirjas lähtutakse Stubbsile (2001) ja Sinclairile (2004) toetudes järgnevast kollokatsiooni definitsioonist: kollokatsiooniks nimetatakse leksikaalset suhet kahe sõna vahel, mis esinevad sageli koos, samuti sellises suhtes sõnapaari.

---

1 *Wortschatz* http://corpora.informatik.uni-leipzig.de/ (9.09.14)

Töös vaadeldakse peamiselt kaheosalisi kollokatsioone, mille põhisõnaks ehk otsisõnaks on inimest tähistav sõna. Siiski tuleb mõnel juhul tõdeda, et kaheosalise kollokatsiooni puhul on tegelikult tegemist pooliku katkendiga suuremast üksusest. Näiteks kollokatsioonid *kõnelev mees*, *elav poeg*, *õppiv tütar* ja *käiv laps* on tegelikult katkendid pikematest kolme- või neljaosalistest ühenditest *vene keelt kõnelev mees*, *Pärnus elav poeg*, *esimeses klassis õppiv tütar* ja *koolis käiv laps*. Otsisõnaga kollokeeruvaid („Eesti keele käsiraamatus" (Erelt jt 2007) kasutatakse sama mõiste väljendamiseks terminit *kollodeeruvaid*) ehk sagedamini koos esinevaid sõnu nimetatakse töös *kollokaatideks*. Sõna koos oma sagedamate kollokaatidega on siinses töös nimetatud *kollokatsioonikobaraks*. Täpsemalt öeldes moodustavad siinse töö seisukohast kollokatsioonikobara otsisõna ja tema 30 sagedamat kollokaati.

1.2.2. Kollokatsiooniseosest erinevatest eesmärkidest lähtuvalt

Kollokatsiooniseoseid on uuritud erinevatel eesmärkidel. Ilmunud on mitmeid põhjalikke ülevaateid kollokatsioonide uurimisest erinevatest teooriatest lähtuvalt (nt Bartsch 2004; Granger, Paquot 2008; McEnery, Hardie 2012; Stefanowitsch, Gries 2009; Stubbs 2009), mistõttu siinses töös piirdutakse lühiülevaatega kahest olulisemast suunast.

Mitmesõnaliste üksuste uurimises eristuvad kaks peamist lähenemist: fraseoloogiline ja distributiivne uurimissuund. Fraseoloogiline uurimissuund sai alguse Ida-Euroopa traditsioonist, kus olulisel kohal oli fraseoloogiliste üksuste piiritlemine ja määratlemine. Algselt oli peamiseks eesmärgiks leida lingvistilisi kriteeriume, et tõelisi fraseoloogilisi üksusi eristada vabadest sõnaühenditest, mis uurimisainese hulka ei kuulunud (Granger, Paquot 2008: 28). Korpuslingvistika arenguga jõudis ka kollokatsiooni mõiste fraseoloogiasse, aga eelkõige huvituti (Cowie 1981; Mel'čuk 1998) sellest, kuhu paigutub kollokatsioon vabade sõnaühendite, piltlike väljendite ja idioomide suhtes.

Distributiivne uurimissuund ei ole tähelepanu pööranud fraseoloogiliste üksuste klassifitseerimisele, vaid lähtub Firthi nägemusest, et kollokatsiooniseosed avavad otsisõna tähendust. Selle uurimissuuna esindajad (Harris 1946, 1954, 1957; Sinclair 1991, 2003, 2004; Stubbs 1995, 2001, 2009; Evert 2005) lähtuvad mitmesõnaliste üksuste defineerimisel mitte lingvistilistest kriteeriumitest, vaid sagedusest.

Korpuslingvistika meetodite kaasamine fraseoloogiasse on kaht suunda lähendanud. Siiski on ka korpuspõhistes fraseoloogilistes uurimustes siiani tunda soovi fraseoloogilisi üksusi klassifitseerida (vt nt Heid 2008; Lindquist, Levin 2008). Kaht suunda hoiavad lahus nii erinevad eesmärgid kui terminoloogilised erinevused (vt eespool mõiste *kollokatsioon* erinev defineerimine). Fraseoloogiliste üksuste defineerimine ja leidmine on oluline nt keeleõppe seisukohast, kus kollokatsioonid pakuvad huvi peamiselt seetõttu, et emakeelekõnelejatele omaste kollokatsioonimustrite tundmine aitab keeleõppijal vältida ebaloomulikke

sõnaühendeid ning lisab kõnele sujuvust. Sellised uurimused tegelevad eelkõige keeleõppija jaoks oluliste kollokatsioonisuhete kaardistamisega (Jantunen 2004, 2009, Jantunen, Brunni 2012, Meunier, Granger 2008, Nesselhauf 2003, 2005). Eestis on keeleõppe seisukohast kollokatsioone uurinud nt Kadri Jaanits (2004), kes on koostanud keeleõppijale mõeldud soome-eesti kollokatsioonisõnastiku ja Heinike Heinsoo (2010), kes on uurinud keeleõppijale raskust valmistavaid adjektiive sisaldavaid kollokatsioone.

Distributiivse uurimissuuna meetodeid kasutatakse rohkem arvutilingvistikas, nt Eestis on Kadri Muischneki huvi kollokatsioonide vastu lähtunud eelkõige püsiühendite automaatse tuvastuse soovist (Kaalep, Muischnek 2009; Muischnek 2005. Kristel Uiboaed (2010, 2013) on statistilisi meetodeid kasutanud eelkõige murdes esinevate verbiühendite tuvastamiseks.

Kollokatsioonid pakuvad huvi ka leksikograafilisest aspektist (Siepmann 2005). Eestis on Jelena Kallas (2013) leksikograafidele mõeldes koostanud reeglid, mis tuvastavad kollokatsioone teatud leksikogrammatilistest konstruktsioonidest lähtuvalt ja pakuvad nii sõnaraamatute koostajatele sõna konteksti avamiseks sobivaid vahendeid.

Siinkirjutaja on kollokatsioonisuhetes näinud eelkõige otsisõna semantika avajat (nt Piits 2004, 2006, 2007, 2013). Käesolev töö ei otsi keeleõppe või leksikograafia seisukohast olulisi kollokatsioone, vaid eesmärgiks on sõna kollokaatide abil avada otsisõna tähendusaspekte.

1.2.3. Kollokatsioon kui tähendussuhe

Alates Firthist, kes leidis, et sõnade koosesinemisviis annab sõnade kohta olulist informatsiooni, on kollokatsioonisuhe olnud neo-firthide seas oluline tähendussuhe. Leiti, et leksikaalne tähendus ei esine lekseemis kui isoleeritud üksuses, vaid laiemas tähendusüksuses (Sinclair 1991). Stubbs (2001: 64, 65) toob Sinclairile toetudes välja sellise laiendatud tähendusüksuse neli koosesinemissuhet: kollokatsioon, kolligatsioon, semantiline preferents ja diskursuse prosoodia. Lühidalt võiks neid suhteid tutvustada nii:

- Kollokatsioon kui sõnade sage koosesinemine on puhtalt leksikaalne suhe. Nt *ema* kollokeerub sageli sõnadega *laps, kodu, isa*.
- Kolligatsioon kui suhe sõna ja grammatilise kategooria vahel on juba veidi abstraktsem suhe, mis näitab strukturaalseid eelistusi. Nt sõnale *ema* järgnevatest sõnadest moodustavad suurima grupi verbid.
- Semantiline preferents on suhe sõna ja mingi sõnagrupi vahel, kuhu kuuluvad sarnase tähendusaspektiga sõnad. Nt sõnaga *ema* võiks esineda koos verbid, mis tähistavad lapse eest hoolitsemist.
- Diskursuse prosoodia on suhe, kus sõna saab värvingu seetõttu, et tal on tendents esineda koos sõnadega, mis kannavad positiivset või negatiivset tähendust. Selliseid suhteid on hiljem hakatud nimetama semantiliseks prosoodiaks (Louw 1993; Stewart 2010).

Kuigi Stubbs on kontekstuaalsed suhted jaganud nelja gruppi, on selge, et tegelikus keelesituatsioonis on kõik esitatud suhted omavahel seotud. Näiteks kollokatsiooniseos kui kõige konkreetsem seos sõltub sõna sõnaliigilisest kuuluvusest ning kollokaate semantilistesse gruppidesse jagades võib teha järeldusi otsisõna semantilise preferentsi kohta. Ka semantiline prosoodia sõltub eelkõige kollokatsiooniseostest. Louw (1993: 157) semantilise prosoodia definitsioon ütleb, et see on sõnale iseloomulik tähendusaura, mis on kollokaatide tähendustest läbi imbunud.

Seega võib öelda, et otsisõna semantiline preferents võib tingida teatud tüüpi kollokaatide esinemise, samal ajal kui positiivset või negatiivset tähendust kandvad kollokaadid võivad muuta otsisõna aurat ehk värvingut ehk konnotatsiooni. Xiao ja McEnery on semantilist preferentsi pidanud kollokaadi omaduseks ja semantilist prosoodiat otsisõnale iseloomulikuks. Siiski möönavad nad, et kahe nähtuse vahel on vastastikune mõju (Xiao, McEnery 2006: 107).

Siinses töös uuritakse kollokatsiooniseosest abstraktsematest suhetest eelkõige semantilist preferentsi.

1.2.4. Kollokatsioonide liigitamisest

Kollokatsioone on jagatud ka grammatilisteks ja leksikaalseteks, pidades esimeste all silmas sõna kombinatsioone funktsioonisõnadega (nt pre- ja postpositsioonidega) ja teiste all leksikaalseid suhteid täistähenduslike sõnade vahel (Benson jt 1997). Grammatilisi kollokatsioone on käsitletud eesti keeleteaduses rektsiooniküsimusena (Kallas, Tuulik 2011: 60). Tänapäeval ei ole grammatiliste ja leksikaalsete kollokatsioonide eristamist peetud enam nii oluliseks. Sabine Bartsch nimetab kollokatsioonideks sageli korduvaid süntagmaatilisi kombinatsioone, mida reguleerivad süntaksi ja semantika kombinatoorsed reeglid (Bartsch 2004: 11). See tähendab, et leksikaalsete kollokatsioonide esinemist mõjutab grammatika: mingid sõnad esinevad teiste sõnade naabruses seetõttu, et süntaksireeglid lubavad neil selles kontekstis esineda. Näiteks esineb eesti keeles substantiivile eelnevate kollokaatide hulgas palju adjektiivatribuute, samas kui substantiivile järgnevate kollokaatide hulgas on neid märgatavalt vähem. Leksikaalseid ja grammatilisi kollokatsioone vaadeldakse tänapäeval tihti koos kui leksikogrammatilisi konstruktsioone või malle (nt Hunston, Francis 2000). Korpuspõhiselt koostatud süntagmaatilistes sõnastikes on leksikaalsete ja grammatiliste kollokatsioonide eristamisest loobutud, süntagmaatilisi suhteid kirjeldatakse leksikogrammatiliste konstruktsioonidena (Kallas 2013: 90). Konstruktsioonigrammatika ja kollokatsiooniteooria kokkupuutepunktis on tekkinud veel kollostruktsioonilise analüüsi meetod (Stefanowitsch, Gries 2003, 2009), mis uurib leksikaalsete üksuste esindatust erinevates grammatilistes konstruktsioonides.[2]

[2] Eestis on kollostruktsioonilisi meetodeid verbiühendite tuvastamisel kasutanud Uiboaed (2013).

Kollokatsioonid erinevad tähenduse transparentsuse ja ühendi püsivuse poolest. Neid saab liigitada vabadeks kollokatsioonideks (*free collocation*) ja piiratud või kinnistunud kollokatsioonideks (*restricted collocation, fixed collocation*).

Vabade kollokatsioonide all tuleks mõista sagedamini koosesinevaid sõnu, mis on kasutusel otseses tähenduses ja on ühendis vabalt asendatavad (Wang, Shaw 2008: 204), nt *väike laps, noor mees, ilus naine*. Mida pidada otseseks tähenduseks, pole küll alati selge: eesti keele üheköitelise sõnaraamatu tegijate sõnul võiks see olla tähendus, millele kujundlik laiendus toetub, mis tundub olevat tähendusmuutuse kõige tõenäolisem lähtekoht, mis osutab võimalikult konkreetsele objektile, tähistaja ja tähistatu kõige vahetumale seosele (Langemets jt 2010: 800).

Kinnistunud ehk piiratud kollokatsioonide alla kuuluvad sõnaühendid, milles üht või mõlemat sõna kasutatakse mitteotseses, tihti erilises või piltlikus tähenduses. Näiteks kollokatsioonides *kõva mees*, *puhas poiss* ja *taevane isa* on sõnad kasutusel piltlikus tähenduses. Tihti on sellistes ühendites raske sõnu asendada sama semantilise välja sõnadega, nt *taevane laps, taevane mees* või *taevane ema* kõlavad veidralt. David Alan Cruse ütleb piiratud kollokatsioonide kohta *bound collocations* 'seotud kollokatsioonid' ja defineerib neid sõnaühenditena, mille osi ei saa hästi lahutada (Cruse 2001: 41).

1.3. Lühiülevaade töös kasutatavatest meetoditest

Tänapäeval seostub mõiste *kollokatsioon* tugevalt korpuslingvistikaga, aga see pole alati nii olnud. Firthi välja pakutud kollokatsioonisuhted ei tuginenud mitte korpusstatistikal, vaid pigem tema enda keeletunnetusel. Üks tema lemmiknäiteid oli sõna *ass* 'tagumik, eesel' kollokeerumine väljendites *You silly ass* 'Sa rumal eesel' ja *He is an ass* 'Ta on vastik (siga)'. Tema eesmärk oli nende näidetega juhtida tähelepanu sellele, et sõna tähendus sõltub kontekstist (Esser 1999: 155).

Algselt oli *korpus* 1950-ndatel aastatel USA strukturaallingvistide kasutatav mõiste, mille all mõeldi vastava keele (allkeele, dialekti) kasutust esindavat tekstide kogu, millele kindlaid analüüsiprotseduure rakendades võis lingvist tuletada selle keele grammatika. Alles 1960. aastatel hakkasid korpused tähistama suuri tekstikogusid, mis koos arvutite arenguga hakkasid meenutama korpuseid tänapäevases mõttes (Leech 1991: 9).

Eesti keele arvutikorpuste loomisest hakati rohkem rääkima 1991. aasta kevadel (Õim 1991: 257). Samal aastal hakatigi Tartu Ülikooli eesti keele laboris korpust looma. Ka Eesti Keele Instituudis tekkis vajadus korpuse järele: 1994. aastal moodustas Indrek Hein digitaalsel kujul olemasolevatest tekstidest korpuse[3], milles tekstide valik oli juhuslik, korpus sisaldas ülekaalukalt ajakirjandustekste ja oli märgendamata, aga sellele vaatamata oli tegu pikka aega suurima eesti keele korpusega.

[3] Korpuspäringu lehekülg: http://www.eki.ee/corpus (9.09.2014)

Neist aastatest alates on koos erinevate korpuste loomise ja arendamisega jõudsalt kasvanud ka korpuslingvistika ja sellele seatud ootused. Käesolev töö näitab, kuidas on korpuslingvistika võimalused kümne viimase aasta jooksul paranenud: lisaks korpuste mahu suurenemisele, mis kajastub osaliselt ka sealt saadud andmete kvaliteedis, on tekkinud juurde tööriistu korpusest saadud andmete töötlemiseks ja erinevate süntagmaatiliste seoste leidmiseks poolautomaatselt süntaktilisi konstruktsioone arvestades. Kui esimesed uurimistulemused on saadud viie miljoni sõna suurust korpust ja kollokatsiooniseostena lihtsaid koosesinemissagedusi kasutades, siis hiljem võimaldas juba 203 miljoni sõna suurune korpus leida kollokaate, mis esinevad otsisõnaga juba mingis grammatilises seoses (nt adjektiivseid laiendeid või verbe, mille suhtes inimest tähistav sõna on subjektiks).

Korpusest kollokatsioonide leidmiseks kasutatakse mitmeid erinevaid meetodeid, mille valik sõltub eelkõige eesmärgist. Näiteks keeleõppes vajatakse loetelusid nii tavapärastest suure sagedusega ühenditest, mis aitaks keeleõppija keelekasutust sujuvamaks muuta, kui ka harvemini esinevatest idiomaatilistest sõnaühenditest, mille tähendus ei ole transparentne, mistõttu keeleõppija ei suudaks sõnaühendi tähendust üksiksõnade tähenduste põhjal tuletada. Arvutilingvistikas vajatakse kollokatsioone sõnatähenduste ühestamiseks ja mitmeosaliste püsiühendite leidmiseks. Leksikograafias ollakse huvitatud kollokatsioonidest kui sõna kasutuskonteksti näitajatest. Üldine tähendusõpetus peab aga kollokaate olulisteks sõnatähenduse täpsustajateks, mille esinemise järgi on võimalik saada infot sõna tähenduse ja semantiliste suhete kohta.

Eesmärgist sõltuvalt kasutatakse erinevaid meetodeid ja vahendeid. Kõige lihtsamaks meetodiks on kahe sõna koosesinemissageduse kokku lugemine. Sellise meetodiga leitakse kollokatsioonid, mis on sagedad. Meetod annab eelise suure esinemissagedusega otsisõnadele, mistõttu sobib see sõna kõige tavalisema konteksti leidmiseks ja kirjeldab otsisõna üldist distributsiooni ehk seda, milline on otsisõna tüüpiline esinemiskontekst. Sellise meetodiga võivad aga tähelepanu alt välja jääda harvaesinevad idiomaatilised väljendid. Viimaste leidmiseks on parem kasutada sõnadevahelise seose tugevuse mõõtmise statistikuid, mis arvestavad sõnade koosesinemissageduse suhet sõnade eraldi esinemise sagedusega. Selliste meetoditega saab leida eelkõige harvem esinevaid, sageli idiomaatilisi üksusi.

Lisaks meetoditele, mis arvestavad ainult sõnade esinemis- ja/või koosesinemissagedust, on võimalik leida kollokatsiooniseoseid kombineeritult süntaktiliste konstruktsioonidega. Näiteks on võimalik leida kõik otsisõnaga sagedamini koos esinevad adjektiivsed laiendid või leida kõik verbid, mille suhtes otsisõna on subjektiks. Selliste meetodite kasutamiseks on vaja korpust eelnevalt töödelda, morfoloogiliselt analüüsida, ühestada ja märgendada. Meetod annab paremini süstematiseeritud andmeid sõna käitumise kohta erinevates konstruktsioonides. Samas võivad tulemuste tõeväärtust mõjutada märgendusvead, mis automaatse märgenduse puhul on vältimatud. Sellist meetodit kasutava programmi Sketch

Engine'i evalveerijad on märkinud, et tulemuste kvaliteet sõltub korpuse suurusest ja kvaliteedist, lemmatiseerijast, sõnavisandite grammatikast ja sobiva statistiku valikust (Kilgarriff jt 2010: 4).

Spetsiaalselt kollokatsioonide leidmiseks loodud tehnilised abivahendid võib jagada kaheks: veebipõhisteks ja arvutisse allalaetavateks. Siinkirjutajale teadaolevalt on eestikeelsete kollokatsioonide leidmiseks kolm veebipõhist programmi. Neist vanim on 1998. Leipzigi ülikooli juures loodud rakendus Wortschatz, kuhu 2006. aastal lisati siinkirjutaja ettepanekul ka Tartu ülikooli koondkorpusest üks ajakirjanduskeelt sisaldav tolleaegne allkorpus. Sel ajal sisaldas Wortschatz 17 keelekorpust. 2014. aasta alguse seisuga võimaldab see otsimootor leida aga koosesinemissagedusi juba 230 keelekorpusest. Suure keelekorpuste hulga juures pakub veebipõhine rakendus tagasihoidlikke võimalusi otsingukriteeriumite muutmiseks – veebiversioonis pole võimalik muuta vahemikku, kust kollokatsioone otsitakse, valida statistikuid ega kasutada isiklikke korpusi.

Hoopis enam saab uurija määrata otsingutingimusi 2010. aastal Tartu Ülikoolis loodud kollokatsioonide tuvastamise programmis[4], kus lihtsa koosesinemissageduse kõrval saab otsinguteks kasutada ka kolme koosesinemistugevuse statistikut. Programm võimaldab otsida nii lemmade seoseid teiste lemmadega, lemmade seoseid teiste sõnavormidega kui ka sõnavormide seoseid teiste sõnavormidega. Nii sisestatava sõna kui ka otsitavate kollokaatide ringi saab piirata nende sõnaliigilise kuuluvusega. Siiski ei võimalda ka see programm isiklike korpuste kasutamist.

Kolmas veebipõhine programm on algselt eeskätt leksikograafidele mõeldud Sketch Engine (vt Kilgarriff jt 2004), mis võimaldab kollokatsioone otsida leksikogrammatilisi konstruktsioone arvestades. Programm võtab kollokatsioonide automaattuvastamisel aluseks morfoloogilise analüsaatori tulemused (sõnaliik ja grammatilised tunnused) ning kombineerib neid statistiliste meetoditega.

Spetsiaalselt lingvistidele loodud programmide kõrval on muidugi alati võimalik koosesinemissagedusi leida internetiotsinguks loodud tarkvara kasutades – näiteks sobib selleks Google'i otsimootor. Kuigi internetiotsing töötab suurima ja värskeima tekstimaterjali peal, on sellisel otsingul siiski puudusi, alates sellest, et otsingutulemuses ei kajastu otsisõna enda esinemisjuhud, vaid number näitab, mitmel leheküljel on sõna mainitud, kuni selleni, et n-ö korpus sisaldab hulgaliselt duplikaate ega ole morfoloogiliselt märgendatud. Guugeldamise puudustest on kirjutanud Adam Kilgarriff (2007) ja David Crystal (2011). Veebipõhiste programmide kõrval on olemas mitmeid spetsiaalselt kollokatsiooniseoste leidmiseks loodud arvutipõhiseid rakendusi. Näiteks Liverpooli ülikoolis Mike Scotti (2004) loodud rakendus WordSmith Tools (esimene versioon ilmus 1996. aastal) võimaldab otsida kollokatsioone oma valitud korpusest, määrata ise kollokatsioonide leidmise kaugust, valida kollokatsioonitugevuse leidmise statistikuid jms. Sarnaseid programme on loodud maailmas teisigi, nt MonoConc Pro, IMC Corpus Workbench, AntConc. Eestis on bakalaureusetööna valminud

[4] TÜ kollokatsioonide tuvastamise programm www.rabauti.ee/clc (09.09.2014)

Tallinna Ülikoolis tarkvara morfoloogiliste kollokatsioonide eraldamiseks (Ots 2012) ja Tartu Ülikoolis on magistritööna loodud kollokatsioonide tuvastamise programm, mis statistilisi meetodeid kasutades otsib koosesinevaid bigramme ühe lause piiridest tingimusel, et sõnade vahele ei jääks rohkem kui 4 sõna (Jets 2013: 8).

Käesolevas töös kasutatakse kaht teineteist täiendavat meetodit ja kaht korpusmaterjali analüüsiks sobivat rakendust. Kolmandas peatükis kasutatakse inimest tähistavate sõnade üldise distributsiooni kirjeldamiseks WordSmith Toolsi rakendust ja Tartu ülikooli eesti keele koondkorpuse tasakaalus korpuse ajakirjandustekste sisaldavat lemmatiseeritud allkorpust suurusega 5 miljonit sõna. Tänapäeva mõistes suhteliselt väikesest korpusest kollokatsioonide leidmisel ei kasutata etteantud grammatilis- süntaktilisi konstruktsioone ega kollokatsioonitugevuse arvestamise statistikuid, vaid leitakse eraldi vasakule hargnevad ja paremale hargnevad kollokaadid. Neist on võrdlusesse võetud 30 sagedamini esinevat. Kollokaadid on leitud kolme sõna suurusest vahemikust vasakult ja sama suurest vahemikust paremalt. Leidmisel on arvestatud absoluutset koosesinemissagedust, veidi suurem kaal on antud otsisõnale lähemal asetsevale kollokaadile, kuna eeldatakse, et mida lähemal otsisõnale, seda suurem on tõenäosus, et sõna kuulub ka süntaktiliselt samasse fraasi. Sel viisil saadud andmed on lihtsad (neid on vähe töödeldud) ning kollokaatide loend sisaldab rohkem asesõnu ja muid suurema esinemissagedusega sõnu, mis kollokatsioonitugevuse statistikuid kasutades oleks vaatluse alt välja jäänud. Lähemalt analüüsiti 10 inimest tähistava otsisõna 30 sagedamat kollokaati.

Neljandas peatükis on kollokaatide leidmiseks kasutatud Sketch Engine'i rakendust ja Jelena Kallase loodud fraasi moodustusstruktuurist lähtuvaid konstruktsioonimalle. Korpusena kasutatakse 203 miljoni sõna suurust Tartu Ülikooli eesti keele koondkorpust. Sel viisil saadud andmed sõltuvad küll konstruktsioonimallidest ega anna laiemat, ainult sagedusel põhinevat ülevaadet kõigist leksikaalsetest kollokaatidest, samas on andmed informatiivsemad ja paremini võrreldavad teiste keelte andmetega, sest grammatilisi konstruktsioone kasutades omame ülevaadet, millised süntaktilis-grammatilised suhted kollokatsiooniseose taga on.

Kasutatud korpused erinevad eelkõige oma suuruse poolest. 40 korda suurema korpuse kasutus neljandas peatükis oli tingitud peamiselt tehnilistest vajadustest: erinevate leksikogrammatiliste konstruktsioonide kohta piisava esinemisarvu tagamiseks ei oleks piisanud 5 miljoni sõna suurusest korpusest. Sisulise poole pealt on erinevused väiksemad. Kui väiksem korpus koosneb ainult ajakirjandustekstidest, siis suurem, eesti keele koondkorpus sisaldab väheses mahus ka ilukirjandust, teadustekste ja uue meedia tekste. Teiste tekstiliikide sisaldusele vaatamata on ka suuremas korpuses tugevalt ülekaalus ajakirjandustekstid. Väike erinevus on ka korpuste ajalises katvuses. Tasakaalus korpuse ajakirjandustekstide korpus sisaldab materjale aastatest 1990–2001, samal ajal kui suure koondkorpuse ajakirjanduse osa on veidi hilisemast perioodist: 1995–2008. Kuigi keel ja

keeles väljendatud teemad ei ole ajas püsivad ning suured ühiskondlikud muutused peegelduvad ka ühiskonnaliikmete keeles (Nuolijärvi 1986), võib eeldada, et suuremad erinevused esineks kaheksakümnendate ja üheksakümnendate aastate keelekasutuse vahel, kuna nende perioodide vahel toimusid Eesti ühiskonnas taasiseseisvumisega seotud murrangud.

Täpsemalt kirjeldatakse materjale ja kumbagi meetodit peatükkides 3.2 ja 4.2.

2. INIMEST TÄHISTAVATE SÕNADE TÄHENDUSEST

Teises peatükis kirjeldatakse kümne inimest tähistava sõna (*ema, inimene, isa, laps, mees, naine, poeg, poiss, tüdruk* ja *tütar*) tähenduste esitust sõnaraamatutes ja tesauruses ning nende tähenduste paradigmaatilisi suhteid ja paigutumist semantilistes väljades. Leitakse inimest tähistavate sõnade tähendusi eristavad komponendid ja tehakse valik peamistest põhitähendustest, mida kollokatsiooniseoste analüüsimisel võrdluseks võtta.

Ükskõik, kas tegu on keele käsiraamatu (nt Erelt jt 2007), kõrgkooli semantikaõpiku (nt Pajusalu 2009) või leksikograafiat käsitleva artikliga (nt Langemets 2003) – väga tihti armastatakse näidetena kasutada inimest tähistavaid sõnu, seda nii komponentanalüüsi võimaluste tutvustamiseks kui ka sõnatähenduste esituse näitlikustamiseks. Ka käesoleva töö autorit kannustas ühelt poolt huvi selle tähendusvälja sõnade vastu, teisalt tundus see olevat sobiv andmestik üldiste paradigmaatiliste ja süntagmaatilise tasandi seoste esile toomiseks.

Eesmärgiks oli inimest tähistavate sõnade kollokatsioonikobaraid võrreldes ja analüüsides uurida sõnadevahelisi leksikaalseid seoseid. Kuna sageduse aspekt on kollokatsiooniseoste leidmisel oluline, otsustas töö autor uuringuobjektiks valida 10 suurima esinemissagedusega inimest tähistavat sõna. Sõnad *mees, inimene, naine, laps, ema, poiss, isa, tüdruk* ja *poeg* (esinemissageduse kahanemise järjekorras) kuuluvad „Eesti kirjakeele sagedussõnastiku" andmetel 100 kõige sagedama substantiivi hulka (Kaalep, Muischnek 2002a). Kuna töös taheti uurida süntagmaatilise tasandi seoseid paradigmaatilise tasandiga, siis lisati otsisõnade sekka veel sõna *poeg* antonüüm *tütar*, mis on küll alles sagedussõnastiku 130. substantiiv.

Nimetatud sõnad kuuluvad inimest tähistavate põhisõnade hulka, kuna sõna esinemissagedust võib pidada monomorfeemsuse ja lühiduse kõrval peamiseks põhisõna tasandile omaseks tunnuseks (Taylor 1995: 49). Urmas Sutrop on defineerinud põhisõna nii: „põhisõna on psühholoogiliselt esiletulev, enamikul juhtudel morfoloogiliselt lihtne omasõna, mis kuulub samasse sõnaliiki ning millel on sama grammatiline potentsiaal nagu selle sõna semantilise välja prototüüpsetel esindajatel. Põhisõna osutab üldiselt mingile objektile, kvaliteedile või nähtusele põhitasemel ning on kasutatav kõigis olulistes kontekstides" (Sutrop 2011: 45). Sellest määratlusest paistab, et põhisõna ei tohi olla tuletatud. Sõna morfoloogilise lihtsuse üle tuleks aga otsustada sünkroonsest tasandist lähtuvalt (Sutrop 2011: 46). Seega, kuigi sõnad *naine, inimene* ja *poeg* on diakroonilisest aspektist tuletised (Eesti etümoloogiasõnaraamat 2012, Rätsep 2002: 102–107), ei takista see neid lugemast põhisõnavara hulka. Psühholoogiline esiletulek ilmneb ühelt poolt sõna kasutussageduses ja teisalt kiires meenumises nimetamiskatses (Sutrop 2000: 143). Eesti keele sugulussõnavara nimetamiskatsest on selgunud, et sugulaste või pereliikmetena nimetatakse alati ja kõige esimeste seas sõnu *ema* ja *isa*, pisut harvem sõnu *poeg, tütar* ja *laps*, ning sõnu *mees* ja *naine* on 28-st

katseisikust nimetanud vaid 3 (Klemm 2013: 41, 42). Viimaste väike esiletulek nimetamiskatses võib olla põhjustatud asjaolust, et neid ei tajuta sugulast või perekonnaliiget tähistava sõnana. Käesolevas töös ei ole nimetamiskatsetega testitud vaadeldavate sõnade kognitiivset esiletulekut, aga sõnade suur lingvistiline esiletulek (sõnade kuulumine 130 sagedama substantiivi hulka) tagab, et tegemist on psühholoogiliselt esiletulevate sõnadega.

Oma varasemates töödes on Sutrop (2000, 2002) eristanud põhisõnavara hulgas veel tuumsõnavara. Tuumsõnavara moodustub kognitiivsete (st kõige sagedamate või psühholoogiliselt esiletulevamate) ja lingvistiliste (st morfoloogiliselt lihtsate) põhisõnade ühisosast. Sutropi tuumsõnavara loendist (Sutrop 2000: 141), mis on tehtud eesti keele koondkorpuse 1980ndate korpuse põhjal, puudub küll sõna *naine*, aga see on seletatav sagedusloendi tegemisel esinenud lemmatiseerimisprobleemidega. Sagedussõnastiku (Kaalep, Muischnek 2002a) järgi kuulub 10 sagedaima nimisõna hulka ka *naine*, mistõttu võib tuumsõnavara definitsioonile vastavaks pidada sõnu *mees*, *naine* ja *inimene*.

2.1. Inimest tähistavate sõnade tähenduslikust liigendamisest

2.1.1. Inimest tähistavad sõnad sõnaraamatutes ja tesauruses

On hoiatatud, et sõnade ja sõnatähenduste kohta järelduste tegemine sõnaraamatute põhjal, kus tähendusi esitatakse isoleeritud üksustena, võib olla äärmiselt eksitav (Halliday jt 2004: 24–25). Siiski on sõnatähenduste empiirilist uurimist soovitatud alustada sõnaraamatust: leksikograaf Margit Langemets (2003: 98) möönab, et isegi kui tähenduste jagamist ja loetlemist on sõnaraamatutes peetud juhuslikuks, pakuvad need välja hulga sõnatähendusi, mida tuleks järgmise sammuna korpuse andmetega võrrelda. Samasuguseid või isegi kategoorilisemaid soovitusi jagab ka mitte-leksikograafist keeleteadlane Renate Pajusalu (2009: 34): „Iga eesti keele semantika uurija peaks alustama EKSSist[5] …"

Käesolevas töös antakse ülevaade inimest tähistavate sõnade tähenduste liigitamiserinevustest „Eesti keele seletavas sõnaraamatus", „Eesti keele põhisõnavara sõnastikus" ja Eesti Wordnetis, võtmata tähenduste sellist esitust ainuvõimaliku tõena, vaid pigem lähtepunktina, kust analüüsi alustada. Peatüki eesmärgiks on selgitada välja inimest tähistavate sõnade kõige olulisemad põhitähendused ja tähenduskomponendid, mis neid tähendusi eristavad, et luua teoreetine alus, millele kollokatsioone analüüsides toetuda.

[5] EKSS= „Eesti keele seletav sõnaraamat"

2.1.1.1. Inimest tähistavad sõnad „Eesti keele seletavas sõnaraamatus"

Sõnaraamatuanalüüsi alustatakse suurimast eesti keele sõnatähenduste kogust, „Eesti keele seletavast sõnaraamatust" (2009; edaspidi EKSS), mis sisaldab ligi 150 000 märksõna koos põhjalike seletustega. Sõnaraamatu esimene trükk ilmus 1988–2007 nime all „Eesti kirjakeele seletussõnaraamat", teises trükis on toimetamise ja täiendamise tulemusel muutunud nii märksõnade hulk (keelde lisandunud uute sõnade arvel) kui ka tähendusteks jaotumine. Näiteks sõnal *inimene* on esimeses trükis veel teise põhitähendusena välja toodud (vananenud) tähendus 'kelner', aga uues trükis on sellest loobutud ja sõnal *inimene* on seal ainult üks põhitähendus ning tähendus 'kelner' pole leidnud äramärkimist isegi tähendusvarjundina. Käesolevas peatükis kasutatakse peamiselt 2. trüki materjale. Vihikutena ilmunud esimesest trükist on võetud andmed sõnaartiklite esialgsete koostajate kohta.

Seletavas sõnaraamatus on tähenduste eristamiseks kasutatud mitmeid mooduseid: märksõna põhitähendused ehk sõna polüseemsed tähendused on eristatud araabia numbritega (1., 2. jne), alltähendused muude erimärkidega. Põhitähenduseks on nimetatud keeles kinnistunud sõnatähendust, mis jääb oma põhiosas muutumatuks kõigil sellega seostatud kasutusjuhtudel, k.a temaga seotud alltähendustes (Langemets 2009: 46). Kui sõnal on rohkem kui üks põhitähendus, siis on tegu polüseemse sõnaga. Leksikograafilise esitluse järgi saab monoseemseks pidada sõna, millel on üks põhitähendus. Lähemal vaatlusel küll selgub, et nn monoseemsel sõnal võib olla palju alltähendusi, mille kinnistumisaste on erinev ja tähendusvarjundi tekkemehhanism või tähendusülekande mall väga sarnane polüseemse sõna tähenduse tekkemehhanismi või malliga.

Analüüsitaval kümnel sõnal on EKSSi järgi kokku 37 põhitähendust, mis teeb keskmiselt 3,7 põhitähendust ühe lemma kohta. Võrdluseks olgu toodud, et Langemetsa (2009: 74) andmetel on eesti keeles nimisõnade polüseemia indeks 2,7, s.t keskmiselt on iga nimisõnalise tüvisõna kohta 2,7 põhitähendust. Seega võib öelda, et vaadeldavad nimisõnad on keskmisest polüseemsemad. Kuna tegu on sõnadega, mis kuuluvad 130 sagedama nimisõna hulka, siis kinnitab kõrgem polüseemiaindeks ka uuringutulemusi, mille järgi tekstides esinevad sagedamini need sõnad, millel on rohkem tähendusi (Langemets 2009: 71).

EKSSis esitatud inimest tähistavate sõnade põhitähenduste arv kõigub märkimisväärselt. Vaatluse all olevatest sõnadest saab leksikograafilise esitluse järgi monoseemseks nimetada ainult sõna *inimene*, millel on 1 põhitähendus. Seevastu näiteks sõnal *isa* on sõnaraamatus välja toodud 7 põhitähendust. Üleminekuid ühelt põhitähenduselt teisele, põhitähendustelt tähendusvarjunditele ja tarvitusjuhtumitele on peetud väga sujuvaks (Kont, Pikamäe 1962; Kilgarriff 1992), mis seab sõnatähenduste leksikograafilise liigenduse võimalikkuse üldse kahtluse alla ning tähendab, et tähenduspiiride tõmbamisel on subjektiivsetel teguritel oluline roll.

Lisaks põhitähendusele on EKSS (vt eessõna) kasutanud võimalust liigitada tähendusi ka erinevateks alltähendusteks: tähendusvarjunditeks (tähis ||), tarvitusjuhtumiteks (tähis |) ja kaasalluvussuheteks (a., b. jne). Tähendusvarjund on leksikaalselt kõige kinnistunum tähendusnihe. Juhuslikumat laadi on tarvitusjuhtum, mis sageli kirjeldab piltlikku kasutust. Vaadeldud sõnade tähenduste seas on ainult sõna *poeg* üks tähendus märgendatud tarvitusjuhtumiks ja saanud ka märgendi *piltlik*. Samas on sõnade *poiss* ja *ema* piltlikke tähendusi määratud ka põhitähendusteks. Kaasalluvussuhe (a., b. jne) kirjeldab tavaliselt olukorda, kus põhitähendus on jagunenud mitmeks võrdväärseks alltähenduseks, mis esinevad eri tarvituspiirkonnas. Sel moel on jagatud sõnade *poiss, tütar, poeg* ja *isa* põhitähendusi.

Inimest tähistavate sõnade puhul on EKSSis kasutatud erinevaid tähenduse liigitamise võimalusi. Näiteks sõnade *inimene* ja *laps* tähendused omavad palju alltähendusi. Sõnal *inimene* on ühe põhitähenduse alla liigitatud 7 alltähendust. Aga sõnade *ema* ja *isa* tähenduste jaotamisel on eelistatud tähendusi liigitada põhitähendusteks. Sõnal *ema* on 6 ja sõnal *isa* 7 erinevat põhitähendust, aga kummalgi on ainult üks alltähendus.

Kõige lihtsam seletus liigituserinevustele võiks olla viitamine erinevate märksõnaartikli koostajate erinevale nägemusele. Vaadeldud kümne sõna artiklid on kirjutanud neli erinevat autorit: Aino Kiindok (sõnade *ema, laps, mees* artiklid), Maimo Kristian (sõnade *inimene, isa* artiklid), Vilma Metstak (sõnade *tüdruk, tütar* artiklid) ja Feliks Vakk (sõnade *naine, poeg, poiss* artiklid) (EKSS 1988). Hiljem on muidugi mitmed leksikograafid materjali toimetanud. Võiks arvata, et kui tähenduspiirid on ähmased ja eristuspõhimõtted pole lühikese seletussõnaraamatute kirjutamise traditsiooniga välja kujunenud, siis leksikograafide arusaam tähenduste jagamisest põhi- ja alltähendusteks võib suuresti erineda. Siiski selgus, et toodud näidete juures ei saa seda põhjuseks tuua, kuna näiteks kõige rohkem erinevad, sõnade *inimene* ja *isa* märksõnaartiklid, on koostanud sama autor Maimo Kristian ning samuti erinevad, sõnade *ema* ja *laps* artiklid, on koostanud mõlemad Aino Kiindok.

Sarnasust on märgata antonüümsete sõnapaaride *mees–naine, poiss–tüdruk, isa–ema* tähenduste liigitamisel ja isegi sõnastamisel. Ilmselt on siin märksõnaartiklite koostajad võtnud eeskuju varem valminud antonüümsete sõnade artiklitest või on toimetamisprotsessi käigus antonüümipaaride sõnastust ja tähenduste liigitust ühtlustatud.

Eripärana võib välja tuua, et antonüümipaaride meessoost isikule viitavate lekseemide *mees, poiss, poeg* ja *isa* tähendused on eristatud detailsemalt kui naissoost isikutele viitavate lekseemide *naine, tüdruk, tütar* ja *ema* tähendused. St meessoost isikule viitavate sõnade puhul on rohkem välja toodud nii põhitähendusi kui ka alltähendusi. Mehi tähistavatel sõnadel on kokku eristatud 33 erinevat põhi- ja alltähendust, samal ajal naisi tähistavate sõnade tähendusi oli kokku 23. Kuna enamik artiklite autoreid on koostanud mõlemast soost isikuid tähistavate sõnade kohta artikleid, siis ei saa väita, et erinevused tähenduste eristamisel tuleneksid leksikograafide isiklikest eelistustest.

Osaliselt võib põhjuseks olla meessoost sõnade väiksem markeeritus, mis võimaldab sõnatähenduste laienemist – nt sõna *poeg* kasutatakse üldistavalt kõigi loomalaste kohta, aga sõna *tütar* selles tähenduses kasutada ei saa. Samuti on sõna *mees* kolmanda põhitähendusena esitatud võimalus kasutada sõna soolisust rõhutamata, inimese kohta (vt Piits 2007), aga *naine* jääb igas olukorras sooliselt markeerituks.

Kindlasti on suurem osa erinevustest tähenduste märkimisel tingitud sellest, et vaatluse all olevatel sõnadel ongi erinevad tähendused. Siiski tulevad võrdlusest välja ka aspektid, mis võiksid olla esitatud teisiti. Nt sõnal *laps* (vt näide 1) on eristatud kaks põhitähendust: 'inimene sündimisest sugulise küpsemise alguseni' ja 'kellegi järglane, poeg v. tütar (oma vanemate suhtes)'. Alltähendus 'meelitus- v kõnetlussõnana täiskasvanud inimese kohta' on seotud 1. põhitähendusega, kuigi see võiks pigem liigituda 2. põhitähenduse alla, sest näites kasutatakse omajatäiendit *mu*, mis viitab, et sõna, mille juurde täiend kuulub, peaks kandma endas sugulust või perekondlikku rolli märkivat tähendust (*laps* tähenduses 'kellegi järglane'). Sama tähendusülekanne on toimunud ka sõnade *poeg* ja *tütar* tähendustes ja seal on alltähendus 'hellitav kõnetlussõna' paigutatud just sugulussidemele viitava tähenduse alla.

(1) laps
1. inimene sündimisest sugulise küpsemise alguseni.
|| piltl (naiivse, lapse mõtte− ja hingelaadiga inimese kohta). *Eluvõõras inimene on suur laps.*
|| (meelitus− v. kõnetlussõnana täiskasvanud inimese kohta). *„*Kuhu siis .. mu kallis laps?" päris vana parun .. „Linna, linna!" nuttis proua ..*
2. kellegi järglane, poeg v. tütar (oma vanemate suhtes).
|| hrl. lastek (linnu- ja loomapoegade kohta).
|| piltl teat. ajajärgule tüüpilise vaimu− ja mõttelaadiga inimene; teat. ajajärgule tüüpiline nähtus. *Kõik me oleme nii või teisiti oma aja lapsed.*
|| piltl teat. rahvusest, paikkonnast v. keskkonnast pärit inimene. *Andide asukad on kõrgmägede lapsed. Karm põhjala pole oma lapsi hellitanud. Iisraeli lapsed.*
|| piltl hrv (millegi millestki tuleneva v. tekkiva kohta). **Need on sügislilled, udude ja raske kaste lapsed ..*
|| piltl (millegi omaks v. meelevalda sattumise kohta). **Kes meisse puutub, on surma laps.*

Kui antonüümipaaride juures on märgata, et tähenduste eristamisel on vastandsõnast malli võetud, siis rohkem erinevusi on hüponüümiasuhtes sõnade tähenduste eristuses. Näiteks võrreldes sõna *laps* ja *poeg* tähendusi, selgub et sarnaseid tähendusi on eristatud erinevalt. Märksõnaartiklis *laps* (1) on esitatud kaks põhitähendust, mille all on esitatud kokku 7 alltähendust (paljud neist märgendatud piltlikuks), aga märksõnaartiklis *poeg* (2) on välja toodud rohkem eraldiseisvaid põhitähendusi, mille all on esitatud üksikuid alltähendusi.

(2) poeg
1. meesisik oma vanema(te) suhtes, otsene meessoost järglane.
|| (hellitav kõnetlussõna). *„Kannata ära, poeg," rääkis äi väimehele.*
2. looma otsene järglane; noor loom, lind, kala vm.
| piltl *Pole penni poegagi* 'mitte mingit raha' *taskus*.
3. kõnek pistik, tütartaim vms. *Tahtsin su kaktuselt poega võtta ja kasvama panna.*
4. (poeetilisemas pruugis) mees oma maa, rahva, ajastu, teat. mõtteviisi vms. (väärika) esindajana. *Aafrika pojad. Mälestussammas isamaa, linna kuulsatele, ustavatele poegadele. Filosoofi vaimne poeg.*
5. hrl. kirikl **a.** Kristus. *Jeesus Jumala poeg. Isa, Poja ja Püha Vaimu nimel.*
b. (kirikuhierarhias alama kohta). *Kuningas on kiriku sõnakuulelik poeg.*

Mõnel juhul on sarnast tähendust liigitatud erinevalt, nt sõna *laps* ühe põhitähenduse alla on märgitud tähendused 'piltl teat. ajajärgule tüüpilise vaimu- ja mõttelaadiga inimene; teat. ajajärgule tüüpiline nähtus' ja 'piltl teat. rahvusest, paikkonnast v. keskkonnast pärit inimene'. Samas kui sõna *poeg* sarnane tähendus 'mees oma maa, rahva, ajastu, teat. mõtteviisi vms. (väärika) esindajana' on liigitatud eraldi üheks põhitähenduseks ja pole saanud ka märgendit *piltlik*. Mõnel juhul on erinevused osutunud ka põhjendatuks, nt sõna *laps* tähendus 'loomalaps' on esitatud tähenduse 'kellegi järglane' alltähendusena, aga sõna *poeg* tähendust 'looma järglane' pole võimalik liigitada tähenduse 'meesisik oma vanemate suhtes' alltähenduseks, kuna selles definitsioonis on sugu määratletud, aga sõna *poeg* saab looma kohta kasutada ka soost sõltumata.

Inimest tähistavate sõnade tähenduste liigitus on seletavas sõnaraamatus väga detailne ja jaotumine põhi- ning alltähendusteks kohati ebajärjekindel, mistõttu kõige olulisemate tähenduste väljaselgitamiseks võrreldi seletavas sõnaraamatus esitatud tähendusi „Eesti keele põhisõnavara sõnastikus" ja Eesti Wordnetis esitatud tähendustega.

2.1.1.2. Inimest tähistavate sõnade esituse võrdlus EKSSis, „Eesti keele põhisõnavara sõnastikus" ja Eesti Wordnetis

Järgnevalt tuuakse EKSSi analüüsi kõrvale võrdluseks näiteid, kuidas on tähendusi jagatud „Eesti keele põhisõnavara sõnastikus" (2014) ja mõistepõhises arvutitesauruses Eesti Wordnet[6].

„Eesti keele põhisõnavara sõnastik" on Eesti Keele Instituudis loodud ükskeelne õppesõnastik, mille peamiseks sihtrühmaks on keeleõppijad. Sihtrühmast lähtuvalt on võimalusel koondatud tähendusi ühe definitsiooni alla, sõnade alltähenduste esitamisel on piirdutud vaid sagedamatega ja põhitähendustega seotud termineid või metafoore lisatud vaid juhul, kui need on osutunud didaktiliselt otstarbekaks. Sagedusest lähtudes on jäetud kõrvale ka mitmed süstemaatiliselt polüseemsed tõlgendused (Kallas, Tuulik 2011: 67).

[6] Eesti Wordnet http://www.cl.ut.ee/ressursid/teksaurus/ (9.09.2014)

Eesti Wordnet on Tartu ülikoolis loodav leksikaal-semantiline andmebaas, mida koostatakse projekti EuroWordnet raames ja mille eeskujuks tuleb pidada Princetoni ülikooli psühholingvistide koostatud inglise Wordnetti[7]. Eesti Wordnetti on tehtud 1996. aastast alates (Orav jt 2011: 96). Aastate jooksul on tesaurust täiendatud nii käsitsi kui automaatselt, mis on tinginud ka vajaduse taksonoomiate kontrolliks. Inimesega seotud hierarhiasuhteid on analüüsinud ja parandanud Riin Kirt oma bakalaureusetöös (Kirt 2009). Eesti Wordnet on praegugi (2015. aastal) täiendatav andmestu. Siinses töös esitatud andmed on võetud andmebaasist 2012. aasta oktoobris.

Kui võrrelda inimest tähistavate sõnade erinevaid tähendusi põhisõnavara sõnastikus, EKSSis ja Eesti Wordnetis, jättes välja kõik loomale viitavad tähendused, siis võib täheldada erinevusi nii tähenduste arvus kui sisus (vt tabel 1). Tähenduste jagamise ja kokkuvõtmise alused on eri sõnastikes ja andmebaasides erinevad ning sõltuvad peale muu ka sõnaraamatu või andmekogu tüübist, püstitatud eesmärgist, kujuteldavast kasutajast, etteantud mahust, ajast jm, seega suures osas keelevälistest asjadest (Langemets 2003: 100).

Tabel 1. Inimest tähistavate sõnade polüseemsete põhitähenduste arvu võrdlus põhisõnavara sõnastikus, Eesti Wordnetis ja EKSSis

	inimene	mees	naine	poiss	tüdruk	poeg	tütar	laps	ema	isa	**kokku**
Põhisõnavara sõnastik	1	2	2	1	1	1	1	2	1	1	13
Wordnet	1	3	3	2	3	1	1	2	1	2	19
EKSS	1	3	2	3	3	3	2	2	4	6	29
Kokku erinevaid	1	4	3	4	4	3	2	2	4	6	33

Eesmärkidest ja sihtgrupist lähtuvalt on loomulik, et põhisõnavara sõnastikus on erinevaid tähendusi kõige vähem eristatud. Polüseemsed on vaid sõnad *mees*, *naine* ja *laps* (lisaks ka sõna *poeg*, mille polüseemsus siin tabelis ei kajastu, kuna teine tähendus on mitte inimese, vaid looma järglase kohta kasutatav).

Wordnetis paigutub tähenduste arvuga teiste vahele, seal on kümne sõna erinevaid tähendusi kokku 19.

Nagu võib eeldada, on kõige detailsemalt tähendusi eristanud EKSS, kus kümnel inimest tähistaval sõnal on kokku 29 inimese kohta kasutatavat põhitähendust (lisaks veel alltähendused). Kõigil sõnadel on seletavas sõnaraamatus sama palju või rohkem tähendusi eristatud kui põhisõnavara sõnastikus või Wordnetis. Ainult sõnal *naine* on tesauruses üks põhitähendus rohkem.

[7] Inglise Wordnet http://wordnet.princeton.edu/ (9.09.2014)

Erinevusi on ka Wordneti ja EKSSi tähenduste sisus, kuigi Wordneti leksikaalsed üksused pärinevad Keeleveebi sõnastike ja Tartu Ülikooli eesti keele korpuse kõrval peamiselt EKSSist (vt Eesti Wordnet[8]). Tähenduserinevusi väljendab tabeli viimane rida, kus on loetud kokku erinevad tähendused, ja kust on näha, et kolme andmekogu peale võib olla kokku rohkem erinevaid tähendusi. Näiteks sõnadel *mees* ja *naine* on Wordnet pidanud vajalikuks eraldada tähendused, mida saab kasutada ainult täiskasvanud isiku kohta, neist tähendustest, mida saab kasutada olenemata isiku east. EKSS on lisaks kokkulangevatele tähendustele pidanud oluliseks välja tuua sõna *mees* kasutuse, kus sõna saab kasutada soolisust rõhutamata ka üldiselt inimese kohta. Sõnadel *poiss* ja *tüdruk* puudub Wordnetis tähendus peigmehe või pruudi märkimiseks, aga EKSSis omakorda pole eraldi põhitähendusena välja toodud nende sõnade kasutust poja ja tütre märkimiseks, mis Wordnetis on olemas.

Selle alapeatüki eesmärk oli eelkõige näidata, et inimest tähistavate sõnade tähenduste liigitus erinevates sõnaraamatutes ja tesauruses on erinev ning sõltub suuresti neile esitatud ootustest. Kollokatsioonisuhete analüüsil ei saa arvatavasti tähendusi nii peenelt jagada, kui seda on teinud EKSS, siiski näitab järgnev analüüs (vt 3.3.2.), et mitme inimest tähistava sõna puhul satub 30 sagedama kollokaadi sekka sõnu, mille puhul saab kindlaks teha, millist polüseemsest tähendust otsisõna selles seoses kannab. On selge, et tähendused ei eksisteeri kindlapiiriliste muutumatute üksustena, mida oleks leksikograafil võimalik lõplikult rahuldaval ja ainuõigel viisil loetleda. Seega ei analüüsita siin enam, mille poolest sagedamate inimest tähistavate sõnade tähendusliigitused ja sisu erinevad, vaid üritatakse leida ühiseid jooni erinevate tähenduste vaheldus- ehk tekkemehhanismides, et välja selgitada iga polüseemse sõna kaks kõige olulisemat tähendust, mis kollokatiivsete suhete analüüsis oluliseks võiksid osutuda.

2.1.1.3. Inimest tähistavate sõnade kollokatiivselt relevantsed tähendused

Üldjuhul pole kollokaatide loendi põhjal ilma konkordantsiridu läbi vaatamata võimalik eristada kollokaate, mis kuuluvad otsisõna ühe või teise tähenduse juurde. Siiski võib lähemal konkordantsiridade uurimisel välja tulla, et mõned kollokaadid esinevad rohkem siis, kui sõna esineb ühes kindlas tähenduses. Kuna sagedamate kollokatsiooniseoste puhul on tegemist üldistusega, siis võib arvata, et sõna 30 sagedama kollokaadi analüüsis ei tõuse esile kollokaate, mis oleksid iseloomulikud otsisõna kasutusele mõnes harvemas tähenduses. Seega osutus otstarbekaks kõigi inimest tähistavate sõnade erinevate tähenduste hulgast välja tuua sõnade kõige olulisemad ehk kollokatiivselt relevantsed tähendused.

Et leida polüseemsete inimest tähistavate sõnade tähenduste hulgast kollokatiivselt relevantsed tähendused, toodi välja inimest tähistavate sõnade sarnased tähendusülekanded. Arvatakse, et ühe semantilise välja sõnad käituvad polüseemia

[8] http://www.cl.ut.ee/ressursid/teksaurus (10.10.2013)

seisukohalt sarnaselt. Alates 1970ndatest on räägitud regulaarsusest ja süsteemsusest polüseemias. Jurij Apresjan (1974), David Alan Cruse (2001 [1986]), James Pustejovsky (1995) ja Eestis Margit Langemets (2009) on leidnud, et tähendus võib muutuda või laieneda väga loogilisel ja süstemaatilisel viisil. Apresjan sõnastas esimesena regulaarse polüseemia definitsiooni: „Sõna A polüseemia on keeles regulaarne, kui selles keeles on sõna B, mille tähendusi b_i ja b_j eristatakse täpselt samal viisil nagu sõna A tähendusi a_i ja a_j ning kui a_i ja b_i ega a_j ja b_j pole sünonüümid" (Apresjan 1974: 16).

Langemetsa (2010: 295–304) „Süstemaatilise polüseemia kuldstandardi" põhjal konstrueeriti sagedamate inimest tähistavate sõnade jaoks neli tähendusülekande malli, mille alla sarnaseid tähendusvaheldusi koondada.

1. INIMENE – INIMENE_ROLL (lähisuhe). Sellise malliga võiks kirjeldada tähendusvahelduse tüüpi, kus ühest või teisest soost inimese kohta kasutatav sõna on võetud kasutusele samast soost abikaasa, elukaaslase või pruudi/peigmehe nimetamiseks. Nimetatud malli alla võiks liigitada nelja sõna tähendused:

mees$_1$ (meessoost isik) – *mees*$_2$ (abikaasa või elukaaslasena)
naine$_1$ (naissoost isik) – *naine*$_2$ (abikaasa või elukaaslasena)
poiss$_1$ (meessoost laps; noormees) – *poiss*$_2$ (peigmehe või elukaaslasena)
tüdruk$_1$ (naissoost laps; neiu) – *tüdruk*$_2$ (pruudi või elukaaslasena)

Sellist tähendusvahelduse järjekorda peab sõnade *mees* ja *naine* puhul usutavaks ka Eduard Vääri (2013: 35, 38), kes ütleb, et neid esialgu soolist erinevust tähistanud sõnu on hiljem hakatud kasutama paarisuhte osaliste kohta. Ilmselt on eraldi tähendus tekkinud elliptilisel moel: algselt omajatäiendiga kontekstuaalne tähendus on muutunud iseseisvamaks ega vaja enda kõrvale enam täiendeid *oma*, *minu, sinu* või *tema*. Sõnade *mees* ja *naine* puhul on tähendus kinnistunum ja iseseisvam kui kahe viimase sõna puhul, mis esinevad rohkem koos omajat näitava sõnaga. Kuigi ilmselt jääb omajatäiendi esinemine kõigi nende sõnade kollokaatide hulgas viitama just sõna kasutusele lähisuhtes oleva inimese tähenduses, aidates nii sõnatähenduste ühestamisel. Kui poleks sarnaselt käituvaid sõnu *mees* ja *naine*, siis võiks sõnade *tüdruk* ja *poiss* alternatiivse tähendustekke mehhanismina kujutleda, et polüseemne tähendus on tekkinud liitsõna põhiosa ellipsi teel sõnadest *tüdruksõber* ja *poiss-sõber*. Kuigi sõnu diakroonilisest vaatepunktist analüüsides jõuaksime ilmselt järeldusele, et viimaste puhul on tegemist hilisemate vastetega ingliskeelsetele sõnadele *girlfriend* ja *boyfriend*. Sõna *naine* erinevatele tähendustele on ligi sada aastat tagasi tähelepanu juhtinud Johannes Aavik, kes pidas sõna „tähenduslikult koormatuks" ja leidis, et tähenduse 'naisinimene' või 'abielunaine' tähistamiseks tuleks leida uus sõna (Aavik 1932: 19). Sama ajal sarnast tähendusvahelduse tüüpi kandev *mees* jääb tal tähelepanuta.

Eesti keele sugulaskeeled on sõna *naine* puhul käitunud eesti keelest erinevalt, püüdes mõisteid lahus hoida. Nt ungari keeles on tüvevokaali varieerides eristatud *naine* tähenduses 'abikaasa' ja *naine* tähenduses 'naissoo esindaja'. Mitmes läänemeresoome keeles on abielunaise tähistamiseks kasutusel hoopis eraldi sõna, nt soome, isuri ja karjala keeles *vaimo* (Rätsep 2009: 26).

2. INIMENE – INIMENE_ROLL (sugulus). Sellise malliga saab tähistada tähendusvaheldust, kus sugulusele mitteviitav inimest tähistav sõna on kasutatav ka suguluses oleva isiku tähistamiseks:

laps$_1$ (alaealine inimene) – *laps*$_2$ (järglane)

poiss$_1$ (meessoost laps) – *poiss*$_3$ (poja kohta)

tüdruk$_1$ (naissoost laps) – *tüdruk*$_3$ (tütre kohta)

Sõnade *poiss* ja *tüdruk* tähenduste 'alaealine' ja 'järglane' seotusest võib leida ka diakroonilist tõestust – sõnade *poiss* ja *poeg* ning *tüdruk* ja *tütar* tüved on keeleajalooliselt sama päritolu (Eesti etümoloogiasõnaraamat 2012). Kui tähendus *laps*$_2$ on kinnistunud ega vaja toetavat konteksti, siis sõnad *poiss* ja *tüdruk* järglase kohta kasutatuna vajavad omajatäiendit või muud suhtele viitavat sõna, nt ***Mul poiss läheb teise klassi.*** Tähendus *laps*$_2$ on EKSSis märgitud põhitähenduseks, aga *poiss*$_3$ ja *tüdruk*$_3$ tähendusvarjunditeks.

3. INIMENE_ROLL (sugulus) – IN_ROLL (rahvus) mall on eelnevatele selles mõttes vastupidine, et sugulusele viitavat sõna hakatakse kasutama laiemas tähenduses, otseses sugulussuhtes mitteoleva isiku kohta:

poeg$_1$ (järglane) – *poeg*$_2$ (mees mingi mõtteviisi või rahvuse vms esindajana)

tütar$_1$ (järglane) – *tütar*$_2$ (naine mingi mõtteviisi või rahvuse vms esindajana)

laps$_2$ (järglane) – *laps*$_3$ (inimene mingi mõtteviisi või rahvuse vms esindajana)

Tähendusülekanne on metafoorne, sõnu kasutatakse poeetiliselt teatud rahvusest, paikkonnast või keskkonnast pärit inimeste tähistamiseks. Erinevus seisneb selles, et sõnad *poeg* ja *tütar* jäävad markeerima ka sugu, aga sõna *laps* saab kasutada mõlemast soost rahvuse vms esindaja märkimiseks. EKSSis on *poeg*$_2$ ja *tütar*$_2$ märgitud iseseisvaks põhitähenduseks, aga *laps*$_3$ on märgitud tähendusvarjundiks.

4. INIMENE_ROLL (sugulus) – IN_ROLL (ideoloogia) mall on selline, kus sugulussuhtele viitav sõna on metafoorsena kasutusel ka vaimse juhi või eeskuju andva isiku tähistamiseks. Algselt lapsevanemale omistatud positiivsed tunnused, nt eeskujuks olemine, juhtimine, abistamine, on üle kantud isikule, kellega puudub sugulussuhe.

isa$_1$ (lapsevanem) – *isa*$_2$ (vaimne juht, õpetaja, algataja, looja)

ema$_1$ (lapsevanem) – *ema*$_2$ (eeskuju andev, juhatav v hooldav naine)

Sõnade *mees, naine, poiss, tüdruk, laps, poeg, tütar, ema* ja *isa* süsteemsemaid tähendusi mallide kaupa analüüsides tuli välja, et ühe malli piires võivad erinevate sõnade tähendused olla erineva tähtsusega: mõnel sõnal on sama malli tähendus liigitatud EKSSis põhitähenduseks, teisel sõnal aga vähem iseseisvaks tähendusvarjundiks. Liigitus võtab ilmselt arvesse sõna kasutussagedust selles tähenduses. Kindlasti ei kasutata sõnu kõikides tähendustes võrdse sagedusega. Kuna kollokatsiooniseostes peegelduvad eeldatavasti vaid otsisõna kõige sagedamad tähendusaspektid, siis valiti iga polüseemse sõna kaks kõige olulisemat tähendust, mida järgnevates peatükkides nii tähenduskomponentide määramisel kui kollokatsiooniseoste analüüsil aluseks võtta. Tähenduste definitsioonid on võetud EKSSist.

mees$_1$: meessoost inimene, meesisik
mees$_2$: abielumees
naine$_1$: naissoost inimene, naisisik
naine$_2$: abielunaine
laps$_1$: inimene sündimisest sugulise küpsemise alguseni
laps$_2$: kellegi järglane, poeg või tütar
poiss$_1$: meessoost laps
poiss$_2$: kavaler
tüdruk$_1$: naissoost laps
tüdruk$_2$: kellegi pruut, kallim, tüdruksõber
poeg$_1$: otsene meessoost järglane
poeg$_2$: mees oma maa, rahva, ajastu, teat. mõtteviisi vms. (väärika) esindajana
tütar$_1$: otsene naissoost järglane
tütar$_2$: naine oma maa, rahva, ajastu, teat. mõtteviisi vms. (väärika) esindajana
isa$_1$: meessoost vanem, mees oma lapse v laste suhtes
isa$_2$: vaimne juht, algataja, looja
ema$_1$: naissoost vanem, naine oma lapse v laste suhtes
ema$_2$: millegi naissoost algataja, looja

2.1.2. Inimest tähistavate sõnade komponentanalüüs

2.1.2. 1. Inimest tähistavad sõnad semantiliste väljade kontekstis

Inimest tähistavate sõnade tähenduste vahel esinev süstemaatiline polüseemia näitas, et ühe sõnavälja sõnadel on sarnased tähenduste tekkemehhanismid. Järgnevalt vaadeldakse, kuidas need sõnatähendused inimese tähendusväljas omavahel suhestuvad.

Semantiliste väljade teooria rajajateks võib pidada saksa ja šveitsi lingviste, kes 1920. ja 1930. aastatel lõid väljateooriaid, lähtudes saussurelikust põhimõttest, et keelemärgi tähendus selgub alles võrdluses teiste märkidega. Semantiliste väljade all on läbi aegade mõeldud nii sõnavälju, tähendusvälju kui ka mõistevälju. Just saksa lingvist Jost Trier (1931) muutis populaarseks sõna- ja mõisteväljade temaatika ja terminoloogia, kuigi kriitikute arvates ei suutnud ta ise hästi termineid *mõisteväli* (*Sinnfeld*) ja *sõnaväli* (*Wortfeld*) eristada ja defineerida (Lyons 1977: 251). Kui Trieri väljad põhinesid paradigmaatilistel suhetel, siis tema kaasmaalase W. Porzigi väli oli üles ehitatud süntagmaatilistele suhetele. Porzigi väljades olid olulised sõnadevahelised kollokatsiooniseosed, eriti pakkusid talle huvi kollokatsiooniseosed substantiivi ja verbi ning substantiivi ja adjektiivi vahel. Nt *koer: haugub, blondid: juuksed* (Lyons 1977: 251).

Hiljem on semantiliste väljade teooriat arendanud Ameerikas Adrienne Lehrer (1974), Floyd Lounsbury (1956) ja Eugene Nida (1975). Semantilise välja

teooriaga haakuvad ka Berlini ja Kay (1969) kategoriseerimise ning Roschi (1973) prototüübiteooriad.

Väljateooria arendamise kõrval on ilmunud suurel hulgal uurimusi, kus sõnu uuritakse väljade kaupa: nt värvinimede, liikumisverbide, sugulussuhete- ja emotsioonisõnavara kaupa. Tähendusväljade kaupa analüüsivad sõnade esinemust eri keeltes ka keeletüpoloogid: nii on tähendusväljade kaupa sõnu esitatud laensõnade andmebaasis WOLD[9].

Selles peatükis vaatlen kümne sagedama inimest tähistava sõna tähendusi tähendusväljas, kus esinevad paradigmaatilistes seostes sõnad. Jost Trieri paradigmaatilise väljateooria lähtekohti võib välja tuua kaks:

a. Sõnad on keele süsteemis omavahel seotud ja moodustavad tervikliku leksikaalse süsteemi, mis on pidevas muutumises.

b. Kuna sõnad on omavahel seotud, siis muutusi tähenduses ei peaks vaatlema üksiksõnade kaupa, vaid analüüsides ja võrreldes sõna semantilisi suhteid teiste sama välja sõnadega.

Paradigmaatilised seosed ei esine sõnade, vaid sõnatähenduste vahel: väljas on esitatud sõnade *inimene, mees, naine, poiss, tüdruk, laps, poeg, tütar, isa* ja *ema* tähtsamate põhitähenduste omavahelised suhted. Kuna esitatud on ainult sagedamate inimest tähistavate sõnade olulisemad tähendused (vt loetelu 2.1.1.3), siis nimetagem seda inimese semantilise välja tuumfragmendiks. Inimese tähendusväljas on esitatud kümne inimest tähistava sõna 19 kõige sagedamat inimest tähistavat tähendust. Polüseemsete sõnade erinevad tähendused on allindeksitega eristatud. Sõnade loomadele viitavad tähendused (nt *poeg* tähenduses 'loomalaps') jäävad looma tähendusvälja ja on eristatud alaindeksiga x.

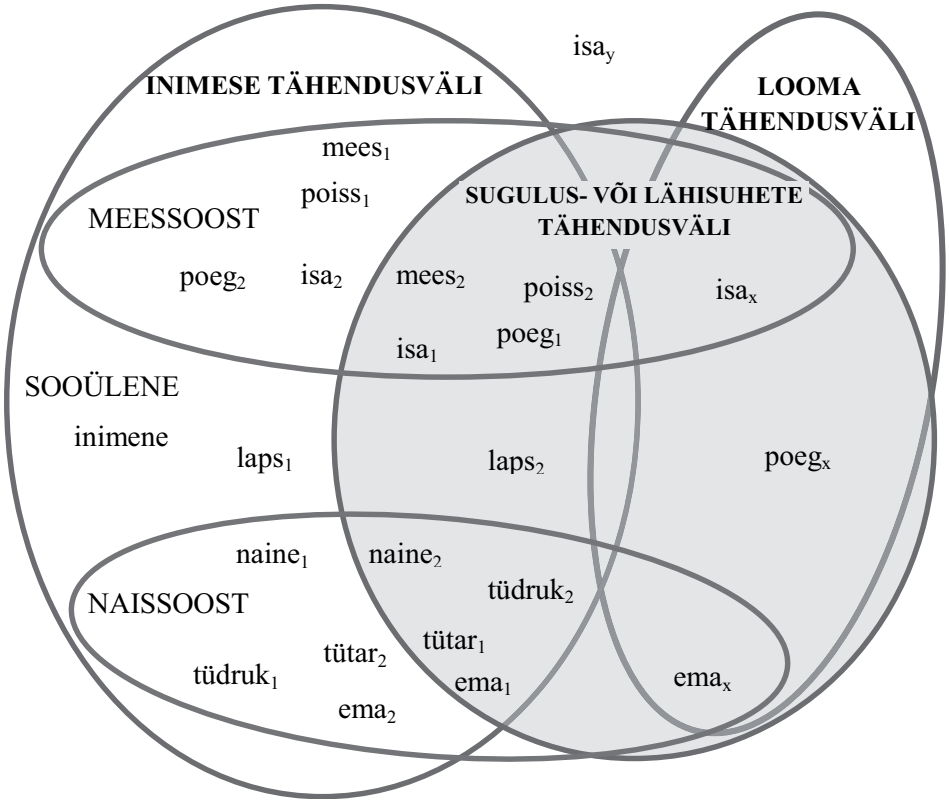
Nagu näha jooniselt 1, on peamiseks sõnatähenduste eristajaks sugulus- või lähisuhte olemasolu või puudumine. Kõiki vaatluse all olevaid sõnu v.a *inimene* on võimalik kasutada viitega sugulasele või elukaaslasele. Erinevus on vaid selles, kas esimene ehk nö algne tähendus viitab sugulus- või lähisuhte olemasolule (nt isa_1, $poeg_1$, ema_1, $tütar_1$) või on lähisuhte tähendus (nt $mees_2$, $poiss_2$) tekkinud põhitähenduse laienemisena, st algselt kasutati sõna koos omajatäiendiga, nt *minu mees, tema poiss, oma poiss* ja hiljem kujunes elliptiliselt sellest sõna eraldi tähendus.

Jooniselt on näha, kuidas sugulus- või lähisuhete tähendusväli ühendab inimese tähendusvälja looma tähendusväljaga. Nt sõnade *isa, poeg* ja *ema* loomale viitavad tähendused (märgitud alaindeksiga x) jäävad küll inimese tähendusväljast välja, aga on sugulussuhte tähendusvälja kaudu seotud inimest tähistavate sõnadega. Seega inimese tähendusväli haakub teiste väljadega lekseemide polüseemia tõttu.

Lisaks sõnatähenduste jagunemisele inimese, looma ja sugulus- või lähisuhete tähendusväljadesse, erinevad tähendused veel selle järgi, kas nad hõlmavad meessoost, naissoost või mõlemast soost isikuid. Ühe sõna neli tähendust võivad kuuluda kõik erinevatesse väiksematesse eri moel kattuvatesse väljadesse. Tähendus isa_1 kuulub nii sugulus- või lähisuhete tähendusvälja kui ka meessoost isikuid koondavasse ja inimese tähendusvälja. Samas kui isa_2 'millegi meessoost

[9] World Loanword Database http://wold.livingsources.org/semanticfield/2 (9.09.2014)

algataja, looja' ei oma enam ühisosa sugulussuhete väljaga. Tähendus *isa*$_x$ 'isane loom oma järglase suhtes' kuulub aga nii sugulussuhete- või lähisuhete kui looma tähendusvälja. Tähendus *isa*$_y$ 'taevane isa, jumal' on aga jäänud välja kõigist nimetatud väljadest.



Joonis 1. Inimese tähendusvälja tuumfragment ja selle seosed teiste väljadega

Selline joonis näitab, et väljad ei ole sugugi aredate piiridega, vaid tihti on erinevad väljad omavahel osaliselt kattuvad, mõnel juhul võib üks väli sisaldada ka teist. Jooniselt on näha, et sama sõna erinevad tähendused võivad kuuluda erinevatesse väljadesse. Semantiliste väljade teoorias on kritiseeritud mõistet *väli*, kuna see andvat vale ettekujutuse, justkui oleks tegu kahemõõtmeliste objektidega. Sellega tuleb nõustuda, sest joonisel 1 jääb osa tähenduserinevusi eristamata. Näiteks ei olnud enam võimalik eristada välja kuuluvaid tähendusi olenevalt sellest, kas tegu on täiskasvanu või alaealisega: ühte rühma sattusid *mees*$_1$ ja *poiss*$_1$ ning teises rühmas olid koos *naine*$_1$ ja *tüdruk*$_1$, samuti olid ühe välja sees *inimene* ja *laps*$_1$, mis kõik erinesid vanuselise aspekti poolest. Samuti ei olnud võimalik näidata peenemaid eristusi, nt *isa*$_2$ (piltlikus tähenduses millegi algataja) ja *poeg*$_2$ (kujundlikus tähenduses hoolealune) sattusid kokku tähendustega *mees*$_1$ ja *poiss*$_1$.

Kahemõõtmelise välja mudel ei võimalda analüüsida tähendusi nii, et kõik olulised tähendust eristavad aspektid saaksid tähelepanu. Sõnal on tähendus ainult tänu võrdlusmomendile ja vastandumisele teiste sõnadega, mis osaliselt jagavad temaga sarnaseid väärtusi, aga samas ka vastanduvad mingi komponendi poolest. Sellest ideest lähtub üks strukturaalsemantika peamisi meetodeid – komponentanalüüs. Inimese tähendusvälja tähendusi asutakse järgnevas peatükis analüüsima nii, et kõik sõnatähendust eristavad komponendid oleksid kindlaks määratud ja saaksid väärtuse.

2.1.1.2. Inimest tähistavate sõnade tähenduskomponendid

Semantilise komponentanalüüsi lähteks on ühelt poolt nimetatud 1930. aastatest fonoloogias rakendust leidnud distinktiivsete tunnuste teooriat ja teisalt 1950. aastatel Ameerika antropolingvistikas levinud sugulussuhete kirjeldamiseks loodud analüüsi. Hiljem kasutasid Ameerika teadlased Jerrold J. Katz ja Jerry A. Fodor komponentanalüüsi generatiivse grammatika semantikakomponendi osana. Tuntuks sai nende näide sõnaga *bachleor*, mille nelja erinevat tähendust 'noor rüütel', 'akadeemilise kraadi omanik', 'poissmees' ja 'noor hüljes paaritumisperioodil' püüti eristada semantiliste komponentide ehk markerite INIMENE, LOOM, MEESSOOST ja NOOR kaudu. Semantilised markerid olid nende arvates sellised süsteemsed teoreetilised moodustised, mis sobisid leksikaalsete üksuste tähenduste süsteemseks kirjeldamiseks. Sõnale pidi olema võimalik leida õige tähendus, liikudes hargmikku pidi ja valides, kas semantiline marker esineb antud tähenduses või mitte (Fodor, Katz 1963: 185–187).

Algselt oli komponentanalüüsi eesmärgiks ühe semantilise välja sõnu analüüsida nii, et tähenduskomponentidele antakse binaarne väärtus. Õige varsti hakati aga kasutama ka kolmest süsteemi, kus lisaks + ja - väärtusele kasutatakse ka +/- varianti, mis peaks näitama, et antud sõnatähenduse puhul on võimalikud komponendi mõlemad väärtused. Näiteks kuna sõnaga *inimene* võime viidata nii mees- kui naissoost isikutele, siis tähenduskomponendile MEESSOOST ei saa anda väärtuseks ei + ega -, vaid +/-. Sellist kolmest süsteemi kasutatakse ka selles töös (vt tabel 2).

Ameerika lingvist ja komponentanalüüsi propageerija Eugene Nida on oma raamatus „Componential analysis of meaning" eristanud kolme eri tüüpi tähenduskomponente (Nida 1975: 33–38):

a) ühised komponendid;
b) diagnostilised ehk eristavad komponendid;
c) täiendavad komponendid.

Ühiseks nimetatakse komponenti, mis esineb kõigil vaatluse all oleva välja sõnadel. Näiteks ainult sugulussuhete termineid uurides oleks ühiseks komponendiks SUGULUS, aga siinses töös on sugulussuhet väljendav komponent LÄHISUHTES OLEV eristavaks komponendiks, kuna vaatluse all on ka sõnad, mis sugulussuhet ei väljenda. Käesolevas töös on inimest tähistavatel sõnadel ühiseks komponendiks INIMENE, mis saab kõigi sõnade puhul väärtuseks +. Kuna see komponent saab

kõigi vaatluse all olevate sõnade puhul sama väärtuse ega erista tähendusi ühe tähendusvälja sees, siis pole seda tabelis enam esitatud.

Diagnostilised ehk eristavad komponendid on MEESSOOST, TÄISKASVANUD, LÄHISUHTES OLEV, EGOST KÕRGEMASSE GENERATSIOONI KUULUV, EGOST MADALAMASSE GENERATSIOONI KUULUV. Viimased kaks muutuvad oluliseks vaid siis, kui tegu on sugulust või muud lähisuhet märkiva sõnaga. Kui positiivse väärtuse saab komponent EGOST KÕRGEMASSE GENERATSIOONI KUULUV, siis on tegemist eellasega, kui plussi saab EGOST MADALAMASSE GENERATSIOONI KUULUV, siis järglasega. Kui komponent LÄHISUHTES OLEV saab positiivse väärtuse, aga kaks viimast komponenti saavad negatiivse väärtuse, siis on tegemist egoga sama tasandi generatsiooni ja partnerlussuhtega, nt *mees*$_2$, *naine*$_2$, *poiss*$_2$, *tüdruk*$_2$.

Tabel 2. Tähendusvälja inimene tuumfragmendi komponentanalüüs

	MEESSOOST	TÄISKASVANUD	LÄHISUHTES OLEV	EGOST KÕRGEMASSE GENERATSIOONI KUULUV	EGOST MADALAMASSE GENERATSIOONI KUULUV	polüseemse sõnatähenduse selgitus
inimene	+/-	+/-	+/-	+/-	+/-	
mees$_1$	+	+	+/-	+/-	+/-	meessoost inimene, meesisik
mees$_2$	+	+	+	-	-	abielumees
naine$_1$	-	+	+/-	+/-	+/-	naissoost inimene, naisisik
naine$_2$	-	+	+	-	-	abielunaine
laps$_1$	+/-	-	+/-	+/-	+/-	alaealine inimene
laps$_2$	+/-	+/-	+	-	+	järglane
poiss$_1$	+	-	+/-	+/-	+/-	meessoost laps
poiss$_2$	+	+/-	+	-	-	kavaler
tüdruk$_1$	-	-	+/-	+/-	+/-	naissoost laps
tüdruk$_2$	-	+/-	+	-	-	kellegi pruut, kallim
poeg$_1$	+	+/-	+	-	+	otsene meessoost järglane
poeg$_2$	+	+/-	+/-	+/-	+/-	mees oma maa, rahva esindajana
tütar$_1$	-	+/-	+	-	+	otsene naissoost järglane
tütar$_2$	-	+/-	+/-	+/-	+/-	naine maa, rahva esindajana
isa$_1$	+	+	+	+	-	meessoost lapsevanem
isa$_2$	+	+	+/-	+/-	+/-	vaimne juht, algataja, looja
ema$_1$	-	+	+	+	-	naissoost lapsevanem
ema$_2$	-	+	+/-	+/-	+/-	millegi naissoost algataja, looja

Tähenduskomponendi MEESSOOST määratlemine probleeme ei tekita. Sõnatähendustes *inimene, laps*$_1$ ja *laps*$_2$ on see komponent neutraliseerunud ja ülejäänud sõnatähenduste puhul ei ole samuti otsustamine raske. Probleeme tekitab komponendi TÄISKASVANUD väärtuse määramine. Esiteks pole kindel, mida selle komponendiga tähistada, kas vähemalt 18-aastast inimest või inimest, kes on saanud suguküpseks. Viimasel juhul määratakse *isa*$_1$ ja *ema*$_1$ komponendile TÄISKASVANUD positiivne väärtus, aga kui me peaksime täiskasvanuteks vaid 18-aastasi ja vanemaid isikuid, siis ei saaks nende tähenduste puhul sellele komponendile positiivset väärtust anda, kuna piisavalt on alla 18 aastaseid lapsevanemaid. Lisaks võib sõnu *mees* ja *naine* kasutada soomääratlusena olenemata inimese vanusest, mistõttu selles kasutuses on tähenduskomponent neutraliseerunud. Areng mittetäiskasvanust täiskasvanuks on graduaalne ja sõltub sellest, kas näeme täiskasvanuks saamist religioossest, sotsiokultuurilisest või juriidilisest aspektist (Panther, Thornburg 2012: 2). Seega lähtuti kõige tüüpilisemast kasutusest ja määrati sõnade *mees, naine, isa* ja *ema* tähenduskomponendile TÄISKASVANUD positiivne väärtus. Komponendile TÄISKASVANUD anti negatiivne väärtus sõnade *laps*, *poiss* ja *tüdruk* esimese tähenduse puhul, samas kui nende sõnade teistes tähendustes on see komponent neutraliseerunud.

Komponentidele antud väärtusi võrreldes hakkab silma, et sõna *inimene* kõik eristavad tähendused on neutraliseerunud (saanud väärtuseks +/-) , mis viitab, et sõnal on üldine tähendus, st sõna *inimene* saab kasutada nii laste kui täiskasvanute, nii meeste kui naiste kohta, nii suguluses olevate kui täiesti võõraste inimeste kohta.

Iga väljas esinev tähendus, mis pole sünonüümne mõne sama välja tähendusega, peaks mingi tähenduskomponendi väärtuse poolest teistest erinema. Kas selline komponentide arv on piisav, et eristada kõiki sõnatähendusi, ka piltlikke? Kindlasti mitte, kuna praegu langevad näiteks tähenduste *mees*$_1$ ja *isa*$_2$ ning *naine*$_1$ *ja ema*$_2$ tähenduskomponentidele antud väärtused kokku. Kuna selliseid juhte ei saa sünonüümseteks pidada ja tihti ei erine need omavahel ainult registrikuuluvuse poolest, siis peavad olemas olema veel võimalused, kuidas väljas esinevaid tähendusi eristada.

Nida väidab, et kuigi diagnostilised ehk eristavad komponendid sobivad tähistama *isa* kui bioloogilist loojat, pole need ainsad komponendid, mis tähendusega seostuvad. Näiteks võrdluses „Ta oli poisile nagu isa" muutuvad põhitähenduse diagnostilised komponendid sekundaarseks ja oluliseks muutuvad täiendavad komponendid TÄHELEPANELIK ja HOOLITSEV. Sarnane muutus toimub ka sõna *isa* kasutamisel jumala tähenduses. Oluliseks muutuvad täiendavad komponendid TÄHELEPANELIK, HOOLITSEV ja AUSTUSVÄÄRNE. Kusjuures nt diagnostilised komponendid nagu LÄHISUHTES OLEV, EGOST KÕRGEMASSE GENERATSIOONI KUULUV ja MEESSOOST muutuvad teisejärguliseks. Viimase komponendi teisejärguliseks muutumist näitab väga hästi tõsiasi, et mõnes matrilineaarses ühiskonnas on kristlased pidanud jumalast rääkima kui *meie emast*, mitte *isast* (Nida 1975: 36).

Nida arvates saab kujundlik tähendus tekkida kas diagnostiliste komponentide tähenduste muutumise ja/või täiendavate komponentide lisandumise toel. Seega sõnade *poeg*$_2$ 'mees oma maa, rahva, vms esindajana', *tütar*$_2$ 'naine oma maa, rahva,

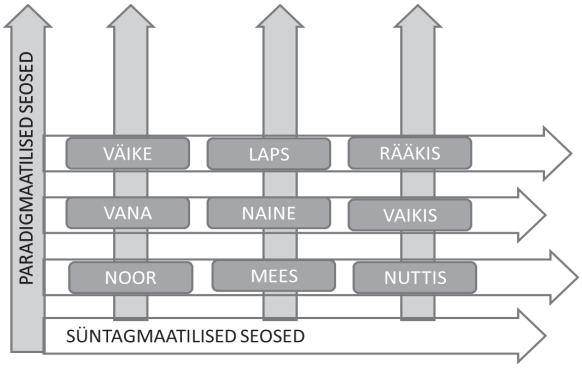
vms esindajana', *isa*$_2$ 'millegi algataja, looja, auväärt inimene' ja *ema*$_2$ ' millegi alus, tekitaja, esilekutsuja' kujundlikud tähendused on tekkinud täiendavate komponentide tähtsustumise teel. Põhitähenduse juures oluline diagnostiline tunnus LÄHISUHTES OLEV on taandunud, asemele on tulnud komponendid MINGIST KOHAST PÄRINEV algselt järglase kohta kasutatavate sõnade (nt *tütar* ja *poeg*) kujundliku kasutuse puhul või MILLEGI ALUS JA TEKITAJA algselt eellase kohta kastutava sõna (*ema* ja *isa*) kasutuse puhul.

Pöördudes tagasi küsimuse juurde, kas viiest tabelis esitatud diagnostilisest komponendist piisab 10 sagedama inimest tähistava sõna tähenduste eristamiseks, tuleb tõdeda, et need suudavad hästi eristada inimese tähendusvälja tuumfragmendi põhitähendusi, st tähendused nagu *inimene, mees*$_1$, *mees*$_2$, *naine*$_1$, *naine*$_2$, *laps*$_1$, *laps*$_2$, *poiss*$_1$, *tüdruk*$_1$, *poeg*$_1$, *tütar*$_1$, *isa*$_1$ ja *ema*$_1$ saavad teistest välja sõnadest komponentidele antud väärtusi varieerides eristatud. Need on samad sõnatähendused, mille väljatoomist on pidanud vajalikuks põhisõnavara sõnastik. Siiski piltlike sõnatähenduste *ema*$_2$ ja *isa*$_2$ eristamiseks tähendusest mees$_1$ ja *naine*$_1$ oleks vaja kasutada lisakomponente.

Samuti oleks vaja komponente lisada, kui soovitaks kümne sagedama inimest tähistava sõna tähenduste kõrval analüüsi lisada ka teisi inimest tähistavate sõnade tähendusi. Näiteks kui analüüsitavate sõnade hulgas oleks veel *onu*, siis tabelis esitatud sugulust kirjeldavate komponentide kõrval oleks vaja ka lisada eristav komponent SUGULUS OTSELIINIS, sest muidu saaksid *onu* ja *isa* kõigile komponentidele samad väärtused.

2.2. Inimest tähistavate sõnade paradigmaatilised ja süntagmaatilised leksikaalsed suhted

Sõnatähendused moodustavad semantilisi võrgustikke, kus tähenduste vahel valitsevad semantilised seosed. Saussure'ist alates on selliseid semantilisi seoseid jagatud kaheks: süntagmaatilisteks ja paradigmaatilisteks. Neist esimene suhe valitseb sõnade vahel, mis esinevad samas süntagmas üheaegselt. Näiteks väljendis *noor mees* on sõna *noor* süntagmaatiliselt seotud sõnaga *mees*. Paradigmaatilised seosed valitsevad aga sõnade vahel, mis ei esine korraga samas süntagmas, vaid on selles süntagmas asendatavad, nt väljendis *noor mees* on *noor* paradigmaatilises seoses sõnadega *vana, väike* jne (Lyons 1977: 240–241). Lyonsilt (2002: 220) pärineb lause: „Keel on kahetasandiline süntagmaatiliste ja paradigmaatiliste suhete süsteem". Selliste suhete lihtsustatud süsteemi on kujutatud joonisel 2.



Joonis 2. Paradigmaatiliste ja süntagmaatiliste seoste teljed

Saussure ise on küll paradigmaatilist suhet nimetanud *assotsiatiivseks suhteks,* öeldes, et seos valitseb sõnade vahel, mis assotsieeruvad mälus (Saussure 2000: 121). Seda võib väita nii 1916. aastal ilmunud C. Bally ja A. Sechehaye toimetatud „Cours de linguistique générale" (Saussure 2000) kui ka 1996. aastal leitud Saussure'i enda käsikirjaliste märkmete (Saussure 2006) põhjal. Termini *paradigmaatiline* tõi 1936. aastal keeleteadusesse Louis Hjelmslev (Matthews 2001: 50), kes eristas süntagmaatikat kui protsessi ehk teksti ja paradigmaatikat kui süsteemi ehk keelt (vt Hjelmslev 2012). Aga terminite *paradigmaatiline* ja *assotsiatiivne* (Saussure 2000; Hawkes 2003) kõrval on Zellig Harris alates 1946. aastast kasutanud sarnase suhte kohta terminit *substitutsionaalne* (Harris 1946: 1961). Ka Lyons on oma hilisemates töödes kasutanud sama väljendit, rõhutades, et see suhe kehtib elementide vahel, mis on kindlates süntagmades asendatavad (Lyons 2002: 96). Kuna tänapäeval on selliste suhete märkimiseks siiski kõige enam kasutusel sõna *paradigmaatiline* (vt nt Cruse 2001; Karlsson 2002; Pajusalu 2009), siis jääb ka siinse töö autor selle termini juurde.

Roman Jakobson on sarnasel dihhotoomial põhinevaid keelemärkide suhteid kirjeldanud terminitega *kombineerimine* ja *selekteerimine*. Jakobson ütleb, et iga märk koosneb märkidest ja on ühtlasi ka teiste märkidega kombinatoorses seoses. Süntagmaatilisel tasandil on märk suhtes kontekstiga. Paradigmaatilise tasandi mõistes on märk teiste märkidega seotud selektiivselt, st on seotud valikuvõimaluse ja asendatavuse kaudu (Jakobson 1956).

2.2.1. Paradigmaatilised leksikaalsed suhted

Paradigmaatilised suhted ei ole kõik ühte tüüpi, vaid neid on samuti liigitatud eri aspektidest. Näiteks hierarhilistest suhetest, kus üks tähendus sisaldub teises, võib nimetada hüponüümiat ja meronüümiat. Mittehierarhilistest suhetest on tuntumad proportsioonisuhted: paarisuhted nagu antonüümia ja kimpudena või klastritena esinevad suhted nagu sünonüümia. Järgnevates alapeatükkides vaadeldakse täpsemalt, kuidas sellised paradigmaatilised suhted nagu antonüümia, sünonüümia ja hüponüümia esinevad 10 sagedama inimest tähistava sõna tähenduste vahel.

2.2.1.1. Antonüümia

Antonüümiat võib pidada kõige esilduvamaks tähendussuhteks. Sõnaassotsiatsioonitestid on näidanud, et kõige sagedamini nimetavad katseisikud stiimulsõna antonüümi. Näiteks on tehtud katse, kus 1008 inimeselt küsiti assotsiatsioone teiste seas ka sõnadele *man* 'mees' ja boy 'poiss'. Sõnaga *man* seostus kõige rohkem tema antonüüm *woman* 'naine' (767 isikut nimetasid seda), järgnesid juba palju väiksema esinemissagedusega *boy* 'poiss' (65), *girl* 'tüdruk' (31) ja *dog* 'koer' (17). Sarnaselt assotsieerus ka sõnaga *boy* 'poiss' kõige rohkem

tema antonüüm *girl* 'tüdruk' (768 nimetamist), suure vahega järgnesid *man* (41), *scout* 'skaut' (37) ja *dog* 'koer'(10) (Panther, Thornburg 2012: 1). Sõnad *scout* ja *dog* assotsieeruvad sõnaga *boy* ilmselt süntagmaatiliste seoste tõttu.

Kui mõned autorid, nt J. Lyons (1977) ja D. A. Cruse (2000, 2001 [1986]), on terminiga *antonüümia* tähistanud kitsamalt ainult skalaarseid vastandusi, siis siinses käsitluses hõlmab termin erinevaid leksikaalseid vastandusi. Muus osas tuginetakse töös Lyonsi ja Cruse leksikaalsete suhete käsitlusele.

Antonüümia võib jagada neljaks suuremaks tüübiks:

a. Komplementaarsel vastandusel rajanev antonüümia. Opositsioonil puuduvad vaheastmed, on vaid polaarsed liikmed, nt *elus–surnud, täiskasvanu–alaealine, meessoost–naissoost*;

b. Skalaarsel vastandusel rajanev antonüümia, nt *külm–kuum, vähe–palju* on kõige tüüpilisem antonüümia liik, kus vastanduvad peamiselt adjektiivid või adverbid. Eesti temperatuurisõnade loetelukatsetega on kindlaks tehtud, et eesti keeles on *külma* kõige esilduvamaks antonüümiks *soe*, mitte *kuum* ja temperatuurisõnade kõige tugevam prototüüpne antonüümipaar on *soe–külm* (Sutrop 2005: 119–120). Cruse ja Lyons on ainult skalaarset vastandust nimetanud antonüümiaks;

c. Suunavastandusel põhinev antonüümia, nt *üles–alla, saabuma–lahkuma, enne–pärast.* Sellised opositsioonipaarilised väljendavad tegevuse, omaduse või tunnuse vastandsuunda;

d. Lisaks kolmele eelnevale on Lyons eraldi välja toonud ümberpööratavusel põhineva vastandussuhte *pöördvastandus* (*converseness*), mis esineb peamiselt vastandlike sotsiaalsete rollide *arst–patsient, õpetaja–õpilane* ja sugulussuhete *isa–poeg, isa–tütar, ema–laps* jne juures. Cruse liigitab sarnase suhte suunavastandusel põhineva vastanduse alla. Peale selle allrühma eristab ta skalaarsel ja suunavastandusel veel mitmeid allrühmi (vt Cruse 1986: 197–242), mida siinkohal tutvustama ei hakata, kuna analüüsitavas materjalis on eelkõige tegemist esimest ja viimast tüüpi vastandustega.

Lyons on märkinud, et alati ei pruugi olla ühtset vastust küsimusele, mis on mingi sõna antonüüm (Lyons 1977: 285). See ilmneb ka inimest tähistavate sõnade juures. Nt sõnale *ema* võib olla antonüümiks nii sooliselt vastanduv *isa* kui sugulussuhte põhjal vastanduv *laps*.

Sagedamate inimest tähistavate sõnade omavahelised antonüümiasuhted võib liigitada kas komplementaarsete vastanduste või pöördvastanduste alla. Mõnel juhul võivad sama sõna erinevad tähendused liigituda erinevalt. Komplementaarsel vastandusel põhinev antonüümia esineb tähenduste $mees_1$–$naine_1$, isa_1–ema_1, $poiss_1$–$tüdruk_1$, $poeg_1$–$tütar_1$ vahel. Nende puhul kehtib Lyonsi komplementaarse antonüümia kontroll-lause *Kui X on X, siis ta ei saa olla Y* ehk *Kui isa on isa, siis ta ei saa olla ema*.

Komplementaarsel opositsioonil peaksid puuduma vaheastmed, kuigi C. K. Ogden on juba 1932. aastal märkinud, et kunagi puhtalt kahepoolset määratlust meessoost–naissoost on võimalik näha „moodsate teadmiste ja vajaduste

valguses" juba skalaarsena ehk koos võimalike vaheastmetega (Ogden 1932: 86). Siiski on ka tänapäeval vastandusel meessoost–naissoost teravamad piirid, kui vastandusel täiskasvanud–mittetäiskasvanud. Komponentanalüüsis avaldus see samuti, kui selgus, et komponent TÄISKASVANUD osutus inimest tähistavate sõnade puhul kõige raskemalt määratletavaks (vt 2.1.2.2). Areng mittetäiskasvanust täiskasvanuks on Klaus-Uwe Pantheri ja Linda Thornburgi sõnul graduaalne ja sõltub ka sellest, kas näeme täiskasvanuks saamist religioossest, sotsiokultuurilisest või juriidilisest aspektist. Sellest tulenevalt on nad väitnud, et sõnale *mees* on „paremaks" antonüümiks soopõhiselt vastanduv *naine*, kui eapõhiselt vastanduv *poiss* (Panther, Thornburg 2012: 2).

Pöördvastanduse suhe esineb sõnapaaride vahel, mille liikmed tähistavad sama situatsiooni või sama situatsiooni osalisi eri perspektiivist. Sellise suhte alla lähevad vastandused *mees*$_2$*–naine*$_2$*, poiss*$_2$*–tüdruk*$_2$. Kui võime öelda lause *X on Y-i mees*, siis peame saama ütelda ka ümberpööratult *Y on X-i naine*. Samad laused kehtivad ka sõnade *poiss* ja *tüdruk* kohta, kui kasutame neid paarisuhte osaliste tähenduses. Ka sõnade vahel, mis tähistavad lapsevanemat ja järglast, on pöördvastandus olemas, nt kui *X on Y ema*, siis *Y on X poeg*. Sama vastandus on tähenduste *ema*$_1$*–laps*$_2$*, ema*$_1$*–poeg*$_1$*, ema*$_1$*–tütar*$_1$*, isa*$_1$*–laps*$_2$*, isa*$_1$*–poeg*$_1$*, isa*$_1$*–tütar*$_1$ vahel.

Antonüümisõnastikus (Õim 1995) on uurimisobjektiks olevast kümnest sõnast esitatud paarid *mees–naine, poiss–tüdruk, poeg–tütar.* Puudub komplementaarsel vastandusel põhinev antonüümipaar *isa–ema.* Peamiselt on vastandused soopõhised, ainult sõna *tüdruk* esineb nii soopõhiselt vastanduvas paaris *tüdruk–poiss* kui ka paaris *tüdruk–abielunaine*, kus vastanduse aluseks on abielus/mitteabielus. Siinses töös on vaatluse all eelkõige pöörd- või komplementaarsel soovastandusel põhinevad antonüümipaarid. 4. peatükis võrreldakse selliste paaride kollokatsiooniseoseid.

Komplementaarse vastanduse olulisus avaldus ka EKSSi inimest tähistavaid märksõnu võrreldes (vt 2.1.1.1), sest komplementaarsete antonüümipaaride juures oli märgata sarnasust nii märksõna põhitähenduste eristamise kui tähenduste sõnastamise puhul. Pöördvastanduses olevate sõnade artiklite ülesehituses sarnasust ei olnud.

Kui antonüümiat on traditsiooniliselt peetud paradigmaatiliseks suhteks, siis Lynne Murphy on väitnud, et antonüümia avaldub ka süntagmaatilisel tasandil. Ta toob kolm põhjendust – antonüümid esinevad koos samas lauses, nad moodustavad konstruktsioone ja antonüümia on erinevalt teistest paradigmaatilistest suhetest lisaks semantilisele nähtusele ka leksikaalne nähtus (Murphy 2006: 1). Mõned näited konstruktsioonidest, mida antonüümid moodustavad: *X ja Y, mõlemad X ja Y, X või Y, nii X kui ka Y* jne. Näiteks konstruktsiooni *nii X kui ka Y* 100 sagedamini korpuses esinenud juhtu sisaldavad kõik antonüümipaare, neist omakorda kõige sagedam on *men and women alike* 'nii mehed kui naised' (Panther, Thornburg 2012: 10–11). Peatükis 4.2.3 vaadeldakse, milliseid süntagmaatilise tasandi konstruktsioone moodustavad inimest tähistavad sõnad.

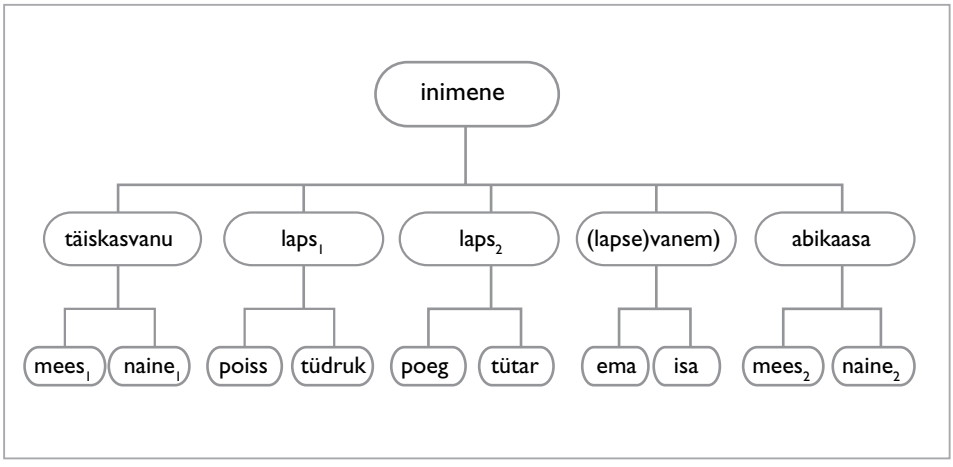
Kui Murphy otsis tõestust sellele, et antonüümia on nii paradigmaatilise kui süntagmaatilise tasandi nähtus, ainult konstruktsioonide analüüsist, siis siinse töö autor vaatleb 3. peatükis ka seda, kuidas antonüümiasuhted peegelduvad antonüümipaarikute kollokatsioonikobarate kattumises. Samuti võrreldakse soo- ja vanusepõhisel komplementaarsel vastandusel ja pöördvastandusel põhinevate sõnapaaride kollokatsioone 4. peatükis, et teada saada, kas otsisõnaga tähistatud inimese sugu ja vanus kajastuvad kollokatsiooniseostes.

2.2.1.2. Hüponüümia

Hüponüümia on hierarhiline paradigmaatiline suhe, mis esineb tähenduste vahel, kus üks tähendustest hõlmab teist, on teisest kõrgemal tasandil.

Tavaliselt puudub üks ja ainuõige võimalus, kuidas tähendusi hierarhiatesse liigitada.

Ka inimest tähistavate sõnade hierarhiapuid saab üles ehitada mitut moodi, see sõltub kindlasti sõnade hulgast, mis puusse tuleb paigutada. Selles peatükis esitatud hierarhiapuusse (joonis 3) on paigutatud ainult kõige olulisemad inimest tähistavate sõnade tähendused: need, mille esitamist on pidanud vajalikuks ka „Põhisõnavara sõnastik". Cruse (2001: 136–137) on tähelepanu juhtinud sellele, et hüponüümiasuhete liigitamisel ei pääse mööda ebajärjekindlusest: eri tasanditel võib liigitusalus erineda. Näiteks toob ta selle, et *looma* jagamine *lambaks* ja *hobuseks* erineb järgmise tasandi jagunemisest, kus *sheep* 'lammas' saab hüponüümideks *ram* 'oinas' ja *ewe* 'utt'. Ka joonisel 3 on näha, et esimese tasandi liigitus erineb järgmise tasandi liigitusest, mis on järjekindlalt soopõhine. Lisaks võib märgata, et esimese tasandi enda sees on liigitusalused erinevad. Nt *inimese* hüponüümid *täiskasvanu* ja *laps*$_1$ eristuvad vanuseliselt, aga sugulust näitavad hüponüümid *laps*$_2$, *lapsevanem* ja *abikaasa* hoopis teistelt alustelt.



Joonis 3. Sagedamate inimest tähistavate sõnade olulisemad tähendused hierarhiapuuna

Seega võib seada kahtluse alla, kas lõpuni loogiliselt ja samadelt alustelt liigituvat ühtset hierarhiapuud on üldse võimalik moodustada. Pigem tuleks hüponüümiasuhete vaatlemisel keskenduda väiksematele üksustele ehk mikroväljadele. Alates Eugenio Coseriust (1967) on eristatud makroväljade sees ka mikrovälju, mis võivad koosneda ülemmõistest ja tema hüponüümidest (Lipka 1980: 98). Näiteks Reinhold Peterwagner väidab, et semantiline väli võib koosneda ühest arhilekseemist, mis on jagatud hüponüümideks. Näiteks *täiskasvanu* hüponüümidega *mees* ja *naine* ning *laps* hüponüümidega *poiss* ja *tüdruk* on tema sõnul mikroväljad, mis sisalduvad omakorda suuremas makroväljas *inimene* (Peterwagner 2005: 165). Paistab tõesti, et mikroväljades on suhted pisut selgemad: sõnade *poiss* ja *tüdruk* hüperonüüm on kahtluseta $laps_1$, sõnade $mees_1$ ja $naine_1$ hüperonüüm aga *täiskasvanu*. Aga ka siin tekivad küsimused vaheastmete kohta: nt kas mõistete *lapsevanem* ja $laps_2$ hüperonüümiks on *inimene* või on seal veel vaheastmeks *sugulane*?

Kuna hüponüümia definitsiooni järgi saab iga entiteedi kohta loogilise vasturääkivuseta kasutada kõiki hüperonüüme, peaks iga sõna olema igas olukorras asendatav tema kõikide hüperonüümidega. Pajusalu (2009: 138) arvates päris nii see siiski ei ole. Ta leiab, et iga olukord määrab ära tähenduse spetsiifilisuse astme. Nt ei saa vanemaid nimetada sugulasteks: oleks veider ütelda *käisin sugulastel külas*, kui tegelikult käidi külas emal ja isal. Juhtumit, kus hüperonüümi üldse ei saa kasutada oma hüponüümi asemel, nimetatakse hüperonüümia blokeerimiseks. Muude sugulaste rühmas Pajusalu arvates blokeerimine ei toimi: võime öelda, et pulmas olid pruudi ja peigmehe sugulased, mõeldes rühma sisse ka mõlema poole vanemad. Selle näite põhjal tekib siiski küsimus, kas *sugulane* on üldse *ema* ja *isa* hüperonüüm või saab hüperonüümiks nimetada ainult otsest ülemmõistet. Sel juhul hüperonüümia blokeerimist ei toimu, kuna saab öelda *käisin oma vanematel külas*.

2.2.1.3. Sünonüümia

Sünonüümia on paradigmaatiline semantiline suhe, mis valitseb vormilt erinevate, aga sisult sarnaste leksikaalsete üksuste vahel. On selge, et sünonüümidel peab olema suur semantiline kattuvus, aga see ei saa olla ainus kriteerium. Cruse (2001: 266) teeb järgmisi paare esitades väga huvitava tähelepaneku:

olend	*filosoofia*
loom	*puu*
koer	*kass*
saksa lambakoer	*spanjel*

Ta leiab, et kuigi iga järgneva paari puhul on semantiline kattuvus üha suurem, ei tähenda see, et sõnapaaride sünonüümsusaste suureneks. Cruse järeldab, et sünonüümia ei avaldu ainult suure semantiline kattuvuse kaudu, vaid sünonüümidel peab olema ka madal sisemine kontrastiivsus.

Sünonüümiat võib pidada astmeliseks nähtuseks, eristatud on nii täis- kui osasünonüümiat. Täissünonüümideks saab pidada keelendeid, millel on ühesugused semantilised tunnused, hinnanguline karakteristika ja seostumine (Õim 2002: 309). Sünonüümia kahe lekseemi vahel on täielik, kui sünonüüme võib vastastikku teineteisega asendada igas kontekstis, ilma et muutuks denotatsioon ja väljendi tõeväärtus, stiil või afektiivne tähendus (Karlsson 2002: 250). Keele ökonoomsusprintsiibi tõttu on täissünonüümia haruldane. Aga isegi täissünonüümsetel sõnadel võib olla oma kasutusala ja kollokatiivsed piirangud. Heinike Heinsoo on toonud soome keele sünonüüme uurides välja, et *suuri* 'suur' ja *iso* 'suur' on küll absoluutsed sünonüümid, aga liitsõnades on kasutus kollokatiivne: *isovanhemmat* 'vanavanemad', *isosisko* 'vanem õde', aga *suurkaupunki* 'suurlinn' (Heinsoo 2010: 123).

Kui täissünonüüme võib keelest leida harva, ja ka need üksikud leiud erinevad kollokatiivsete suhete poolest, siis palju sagedamini saab kaht sõna pidada osasünonüümideks. Osasünonüümid on peaaegu sarnase tähendusega sõnad. Olemasolevale tähenduserinevusele viitab siiski see, et neid on võimalik ka teatud kontekstides vastandada, kasutades sidumiseks sõnu *pigem, täpsemalt öeldes, õigemini küll* jne.

Millised sünonüümiasuhteid saab välja tuua 10 sagedama inimest tähistava sõna vahel, sõltub sellest, kuidas määratleda nende 10 sõna tähendusi. Võttes arvesse ainult iga sõna esimese ehk kõige sagedama põhitähenduse, ei ole võimalik sünonüümiaseoseid välja tuua, aga vähem sagedamaid põhitähendusi võrreldes võib märgata ka sünonüümiasuhteid.

Osalist sünonüümiat võib leida sõnade *mees* ja *poiss* harvem esinevate tähenduste $mees_2$ ja $poiss_2$ vahel: mõlemad märgivad meessoost paarisuhte osalist. Seda ei saa küll pidada täissünonüümseks suhteks, kuna kõigis kontekstides ei ole sõnad vastastikku asendatavad: sõna *mees* kasutatakse enam pikemas suhtes või abielus oleva partneri kohta, aga sõna *poiss* pigem noorema ja mitteabielus partneri kohta. Seega võiks liigitada suhte osasünonüümiaks, kuna saab moodustada lause: *See on mu mees, õigemini mu poiss*. Sarnane sünonüümiasuhe esineb ka tähenduste $naine_2$ ja $tüdruk_2$ vahel.

Rohkem sagedamate inimest tähistavate sõnade peamiste põhitähenduste vahel valitsevaid sünonüümiasuhteid ei õnnestunud välja tuua, kuid nende sõnade kollokaate võrdlema asudes tuleb arvatavasti välja kollokaatidevahelisi sünonüümiasuhteid. Näiteks Heinsoo on leidnud, et soome keeles kasutatakse kena välimuse väljendamiseks naissoo esindaja puhul adjektiivina *upea* 'uhke, tore, suurepärane' ja meessoo esindaja puhul *komea* 'uhke, tore, hea välimusega, äge'. Siiski, president Halose kohta öeldakse *komea nainen*. Heinsoo väidab ka, et eesti keeles on sõna *kaunis* tarvitusel naissooesindaja iseloomustamiseks (Heinsoo 2010: 123). Etteruttavalt võib öelda, et seda väidet kontrollitakse peatükis 4.3. Sünonüümide uurijad on rõhutanud, et kollokatsiooniloendid annavad head materjali, mida teiste meetoditega veel edasi uurida (Vanhatalo 2005: 28).

2.2.2. Süntagmaatilised leksikaalsed suhted

Süntagmaatilised suhted kehtivad elementide vahel, mis võivad omavahel süntagmades kombineeruda (Lyons 2002: 96). Süntagmaatilised seosed on lineaarsed, neis põimub leksika ja süntaksitasandi reeglistik. See tähendab, et osaliselt on sõnadevahelised süntagmaatilise tasandi seosed tingitud süntaksist, aga lisaks on olemas ka semantikast tulenevaid seoseid, st samasse süntaktilisse konteksti sobib üks sõna paremini kui teine ja mõned sõnad esinevad süntaksist olenemata sagedamini koos kui teised.

Süntagmaatilise seose tugevuse järgi on eristatud kollokatsioone idioomidest: kui esimesed on sõnad, mis tüüpiliselt esinevad koos, siis viimased on juba kinnistunud väljendid. Siinse töö seisukohast selline range eristus ei osutu vajalikuks ja idioomid loetakse üldjuhul kollokatsioonide alla, vajadusel vaid täpsustades, et tegu on kinnistunud ehk piiratud kollokatsiooniga (vt 1.2.4). Samuti on süntagmaatilisi seoseid jagatud leksikaalseteks ja grammatilisteks. Nende eristamiseks on viimast tüüpi seoste kirjeldamiseks kasutatud ka sõna *kolligatsioon*. Nagu peatükis 1.2 kirjeldati, ei peeta tänapäeval enam oluliseks neid nähtusi rangelt eristada.

Süntagmaatiliste seoste olulisus võrreldes paradigmaatiliste suhetega avaldus afaasiahaigeid uurides: Roman Jakobson (1956) kirjeldab olukorda, kus teatud afaasiavormi põdevad haiged suutsid sõnu vaevata meenutada küll süntagmaatilistes konstruktsioonides (nt *poissmehe korter*), aga olid võimetud meenutama paradigmaatilistes seostes sõnu ega suutnud nimetada sõnade sünonüüme (nt vallalise mehe kohta ei suudetud öelda *poissmees*).

Halliday on süntagmaatilise tasandi suhted jaganud veel pindmise- ja süvatasandi suheteks, eristades lineaarseid sõnaliikide järgnevusi, nt *adjektiiv–substantiiv*, mittelineaarsetest funktsioonidest nagu *laiend–põhisõna* (Halliday 1966). Käesoleva töö 3. peatüki metoodika järgi otsides avalduvad eelkõige lineaarsed pindmise tasandi suhted – eraldi vaadeldakse otsisõnast vasakule jäävaid ja paremale jäävaid kollokaate, isegi kui mõlemad tegelikult on otsisõna suhtes sarnase laiendi funktsioonis. Samas lähtutakse 4. peatükis rohkem mittelineaarsetest süvatasandi suhetest, kuna olenemata positsioonist lauses on võimalik vaadelda verbe, mis kuuluvad otsisõna kui subjekti juurde, ja verbe, mis kuuluvad otsisõna kui objekti juurde. Seega 4. peatükis ei eristata enam vasak- ja parempoolseid kollokaate.

Kas kollokatsiooniseos esineb lekseemide või tähenduste vahel? Süntagmaatilised seosed on küll igal üksikul juhul vaadates samuti sõnade tähenduste vahel esinevad seosed, aga kollokatsioon on juba süntagmaatilise tasandi abstraktsioon, mis tähendab, et sõna erinevate tähenduste süntagmaatilised seosed on kollokatsioonides liidetud. Sellele vaatamata on võimalik leida kollokaate, mis esinevad ainult või peamiselt ühe tähendusega koos. Sellisest põhimõttest lähtuvalt ühestatakse arvutilingvistikas sõnatähendusi. Ka inimest tähistavate sõnade kollokaatide seas on neid, mis kuuluvad kokku vaid otsisõna ühe tähendusega. Nt sõnale *poeg* eriomased kollokaatverbid *kooruma* ja *lendama* esinevad vaid juhul, kui *poeg* kannab tähendust ‘looma järglane’.

3. SAGEDAMATE INIMEST TÄHISTAVATE SÕNADE KOLLOKATSIOONIKOBARATE VÕRDLUS

3.1. Hüpoteesid

Kolmandas peatükis kirjeldatakse, kuidas inimest tähistavate sõnade tähendus kajastub sõna üldises distributiivses käitumises. Otsitakse märke seostest keele süntagmaatilise ja paradigmaatilise tasandi vahel. Eraldi uuritakse otsisõnast vasakule ja paremale jäävaid kollokaate ja tuuakse välja erinevused vasakule ja paremale hargnevate kollokaatide sõnaliigilises kuuluvuses. Kontrollitakse kolme hüpoteesi.

Firth juhtis meie tähelepanu tõsiasjale, et sõnad ei kombineeru juhuslikult ning ainult süntaktilisi reegleid järgides fraasidesse ja lausetesse, vaid et sõnade vahel valitsevad veel erilised seosed ja see sõnade koosesinemisviis annab nende kohta olulist informatsiooni (vt 1.2). Samast ideest vaimustus ameerika viimaseks strukturalistiks tituleeritud Zellig Harris, kes lähtus sõnatähenduste liigitamisel distributiivsusest ja väitis, et samas kontekstis esinevatel sõnadel on sarnane tähendus: „Kui me hindame, et sõnade või morfeemide A ja B tähenduserinevus on suurem kui A-l ja C-l, siis me leiame tihti, et ka sõnade A ja B distributsioon erineb rohkem kui sõnade A ja C distributsioon." (Harris 1954: 156)

Üksuse distributsiooni nähakse kui kogu tema konteksti. Harrise enda töödes (Harris 1954, 1957) on kontekstina käsitletud vaheldumisi häälikuümbrust, sõnaümbrust ja ka kogu teksti. Siinses töös väljendub üksuse distributiivsus kollokatsioonides, mis on leitud vahemikust kolm sõna vasakule ja vahemikust kolm sõna paremale. Võrdlusesse lähevad kõigi kümne inimest tähistava sõna 30 sagedamat vasakule hargnevat ja 30 sagedamat paremale hargnevat kollokaati.

Harrise üldises määratluses jääb selgusetuks, mida ta peab silmas tähenduse all, mida üksuse distributsioon peaks kajastama. Siinses töös lähtutakse sellest, et sõna tähendus koosneb tähenduskomponentidest (vt komponentanalüüsist ptk 2.1.2.2). Inimest tähistavaid sõnu seob üheks tähendusväljaks komponent INIMENE. Distributiivsuse hüpoteesist lähtuvalt peaks seega kõigil inimest tähistavatel sõnadel juba olema suur osa ühist konteksti. Mida see ühine kontekst väljendab, seda kontrollitakse kolme hüpoteesiga:

1. hüpotees: **Mida rohkem sarnase väärtusega tähenduskomponente kahel sõnal esineb, seda suurem on kollokatsioonikobarate kattumine.** Kui leiab kinnitust, et suur osa inimest tähistavate sõnade kollokaatidest kattub, siis tuleks välja selgitada, mis rolli mängivad kattumise määras sõnade tähenduskomponendid. Inimest tähistavaid sõnu eristavad komponendid MEESSOOST, TÄISKASVANUD, LÄHISUHTES OLEV, EGOST KÕRGEMASSE GENERATSIOONI KUULUV ja EGOST MADALAMASSE GENERATSIOONI KUULUV. Tegelikult eristavad küll komponendid sõnade erinevaid tähendusi (vt 2.1.2.2), aga kuna kollokatsioonisuhe esineb sõnade vahel, olles

üldistus erinevatest tähendustevahelistest suhetest, siis arvestatakse sõna tähenduskomponentidest rääkides sõna esimese, kõige sagedama tähenduse komponentidele antud väärtusi. Uuritakse, kas neile komponentidele antud väärtused ehk sõnade tähendusaspektid kajastuvad ka kollokatsioonides.

2. hüpotees: **Paradigmaatilised suhted avalduvad kollokatsioonides.** Sahlgren on täpsustanud distributiivsuse hüpoteesi, öeldes et kui koosesinemissageduste põhjal tehtud distributiivne mudel väljendab sõnade süntagmaatilisi suhteid, siis ühiste kollokaatide võrdlusest saadud informatsioon peaks väljendama paradigmaatilisi suhteid (Sahlgren 2008: 7). Selles peatükis otsitakse vastust küsimusele, kuidas paradigmaatilised suhted süntagmaatilistes seostes ehk kollokatsioonides avalduvad. Paradigmaatilistest suhetest seovad vaatluse all olevaid sõnu peamiselt antonüümia ja hüponüümia. Lisaks võib märgata mõningast sünonüümiasuhet harvem esinevate tähenduste *mees*$_2$ ja *poiss*$_2$ ning *naine*$_2$ ja *tüdruk*$_2$ ning *inimene* ja *mees*$_3$ vahel. Nimetatud juhtudel ei esine aga sünonüümiasuhe sõna kõige sagedamate ja olulisemate tähenduste vahel. Kuna kollokatsioonid leitakse kogu sõnale, siis vaadeldi paradigmaatilisi suhteid eelkõige sõna peamisest ja kõige sagedamast tähendusest lähtuvalt, mistõttu sünonüümiasuhted jäävad paratamatult vaatluse alt välja[10].

3. hüpotees: **Keele grammatikast sõltub piirkond, kust mingit tüüpi kollokatsiooniseoseid tasub otsida.** Palju on vaieldud selle üle, kui kaugelt ja mis positsioonist tuleks kollokatsiooniseoseid otsida. Osaliselt sõltub see kindlasti keele grammatilisest struktuurist. Töös analüüsitakse inimest tähistavate otsisõnade paremale ja vasakule hargnevate sagedamate kollokaatide sõnaliigilisust, et kindlaks teha, kui palju sõltub kollokatsiooniseoste tuvastamine piirkonnast, kust neid otsida.

[10] Distributiivsuse hüpoteesi on küll sünonüümide peal testinud Herbert Rubenstein ja John B. Goodenough (1965) ning samu sünonüümipaare on hiljem uurinud samadest põhimõtetest lähtuvalt Georg A. Miller ja Charles Walter (1991). Mõlemad uurimused leidsid, et korrelatsioon semantilise läheduse ja konteksti sarnasusega oli olemas. Uurimustes kasutati sõnapaaride semantilise läheduse määramiseks küsitlusi ja läheduse indeks märgiti inimeste hinnangute põhjal. Neist lähtuvalt oleks antonüümipaarid ilmselt saanud väga madala väärtuse, kuna tegemist on paradoksaalse semantilise suhtega, kus tegelikult mitmest aspektist sarnased sõnad võivad esmapilgul tunduda väga erinevatena. Nimetatud uurimustes kinnitas hüpoteesi see, kui semantiliselt lähedasteks määratud sõnad olid katseisikute arvates teineteisega kontekstis asendatavad. Samas on just ideaalse antonüümia üheks definitsiooni osaks lisaks muudele aspektidele nimetatud ka asendatavust teineteisega samas konstruktsioonis: „Kaks leksikaalset üksust on antonüümid, kui nad on kontekstis teineteisega vormiliselt asendatavad, ilma et lause muutuks mittegrammatiliseks" (Panther, Thornburg 2012: 4). Tsitaadist lähtuvalt võiks eeldada, et antonüümsetel sõnadel on sarnane kontekst.

3.2. Uurimismeetod ja analüüsi aluseks olev materjal

3.2.1. Materjal ja vahendid

Kolmanda peatüki materjal on kogutud Tartu ülikooli eesti keele koondkorpuse tasakaalus korpuse ajakirjandustekstidest, milles oli korpuse allalaadimishetkel 5 miljonit tekstisõna aastatest 1990–2001. Ajakirjanduskorpus valiti seetõttu, et sama suures ilukirjanduskorpuses oleks liigselt rõhku andnud üksikud teosed ja autorid. Katse käigus tuli näiteks välja, et väiksema esinemissagedusega sõnade puhul jõudsid esikümnesse pärisnimedest kollokaadid. Enamasti olid sellised koosesinemised seotud kindla teosega. Seega jäädi ajakirjanduskorpuse kasutuse juurde. Kuigi ajakirjandustekstid kajastavad ehk kõige laiemalt keeles toimuvat, on selge, et nendes ei kajastu sugugi kogu keeles esinev kõnelejast ja olukorrast sõltuv varieerumine. Peamiselt sisaldab korpus kirjakeelt ja kirjutatud keelt, kuigi on märke, et kõnekeelsus on leidnud oma koha ka kirjalikus meedias, nt intervjuudes (Nuolijärvi 2005: 176).

Kümne inimest tähistava sõna esinemissagedused tasakaalus korpuse ajalehetekstides:

inimene	12277 korda
mees	7293 korda
laps	5583 korda
naine	4098 korda
ema	1504 korda
isa	1330 korda
poiss	1221 korda
poeg	914 korda
tüdruk	853 korda
tütar	612 korda

Sõnade esinemissagedusi võrreldes hakkab esimesena silma ligi 20-kordne erinevus sõnade *inimene* ja *tütar* esinemiskordades. Kollokaatide esinemist analüüsides erinevused nivelleeruvad, kuna kollokatsiooniseoste absoluutsageduste asemel arvestatakse sellega, mitmendal kohal mingi kollokaat sõna teiste kollokaatide hulgas esineb. Siiski võib otsisõna väga väike esinemissagedus tingida selle, et sagedamate kollokaatide hulka satub juhuslikke sõnu. Sellele võimalusele pööratakse edaspidi tähelepanu.

Sõnade esinemissagedusi võrreldes avaldus seaduspära, et soovastandusel põhinevatest antonüümidest esineb korpuses sagedamini meessoost isikut tähistav sõna. See avaldus nii sõnapaaride *mees–naine*, *poiss–tüdruk* kui *poeg–tütar* puhul. Ainult sõnade *ema* ja *isa* puhul see seaduspära ei kehti: naissoost lapsevanemat mainitakse tekstides veidi rohkem kui meessoost lapsevanemat.

Pärast korpuse valikut tuli otsustada, kas kasutada lemmatiseeritud või lemmatiseerimata korpust. On uurijaid, kes väidavad, et mõtet on vaadelda

vaid sõnavormide koosesinemist. Näiteks John Sinclair (2003, 2004) rõhutab, et erinevaid lemmavorme tuleks mõnel juhul käsitleda erinevate sõnadena, sest sageli võivad sama sõna erinevad vormid esineda hoopis erinevates kollokatsioonides. Stubbs on uurinud sõna *seek* 'otsima' erinevaid vorme ja leidnud korpuse andmetele toetudes, et sõnavormide *seek, seeking* ja *sought* 20 sagedamast kollokaadist on kattuvaid kollokaate küll 6, aga sõnavormil *seeks* on üks ühine kollokaat vaid sõnavormiga *seek*. Selleks kollokaadiks on *professional* 'erialane'. Teiste vormidega sõnavormil *seeks* ühised kollokaadid puuduvad. Stubbsil oli kasutada piisavalt mahukas Cobuildi allkorpus, mis sisaldas 200 miljonit sõna, aga sellele vaatamata kahtlustas ta ikka, et sõna kollokatsioonid on tihedalt seotud tekstitüübiga ning sõnavormi *seeks* eripärased kollokatsioonid on seotud tutvumiskuulutuste rubriigi olemasoluga korpuse ajalehetekstides (Stubbs 2001: 28).

Eesti keele uurimisel muudab aga keele muutevormirikas morfoloogiasüsteem sõnavormide koosesinemissageduse võrdluse keeruliseks. Seda esiteks seetõttu, et ühel sõnal võib olla väga palju muutevorme – substantiividel näiteks üle 28 erineva vormi, ja teiseks seetõttu, et esineb grammatilist homonüümiat. Isegi kui otsustame uurida vaid singulari nominatiivi vorme, siis soovides võrrelda sõnade *naine* ja *ema* kollokaate, peame arvestama, et neid sõnu otsisõnadena kasutades saame esimesel juhul tõesti kollokaadid, mis kuuluvad nominatiivi juurde, aga teisel juhul on grammatilise homonüümia tõttu võimatu eristada nominatiivse vormi juurde kuuluvaid kollokaate genitiivis ja partitiivis sõnavormide kollokaatidest. Samuti muutevormirikka soome keele uurimisel on leitud, et sünonüümsed sõnad võivad esineda eri vormides erineva sagedusega (Arppe 2001, 2002; Jantunen 2001) ning erinevused vormide esinemissageduses võivad olla tingitud ka sõna silbilisest koostisest (Arppe 2001: 13).

Inglise keeles on morfoloogiasüsteemist lähtuvalt kollokatsioone uurida lihtsam: võib vaadelda nii sõnavormide kollokeerumist eraldi kui ka liita erinevate sõnavormide analüüsimisel saadud andmed ja vaadelda lemmade kollokatsioone. Näiteks Sara Gesuato (2003) on uurinud sõnade *man* 'mees' ja *woman* 'naine' kollokatsioone inglise keeles, esitades kord näiteid sõnavormide *woman* ja *women* kollokatsioonide kohta, samas tehes ka üldistusi lemma *woman* kasutuskonteksti kohta.

Kindlasti on ka eesti keeles olemas seosed sõna vormi ja kollokaatide esinemise vahel. Näiteks on töö autor tähele pannud, et sõna *eesti* kollokeerub ajakirjanduskorpuse andmete põhjal enam mitmusevormiga *naiste* kui ainsusevormiga *naine*. Kahjuks ei oleks sagedamate inimest tähistavate sõnade kõigi üksikvormide uurimine nii tulemuslik. Mitmetel juhtudel ei lubaks sõnavormide esinemissagedus korralike tulemusteni jõuda, sest sõna kasutusjuhud ei jaotu mitte 4–5 erineva sõnavormi vahel nagu inglise keeles, vaid iga sõna juures tuleks arvestada vähemalt 28 võimaliku vormi kollokaatidega. Grammatilise homonüümia tõttu tekiksid raskused ka erinevate sõnade sõnavormide kollokeerumise võrdlemisel.

Seega otsustati koondada erinevad sõnavormid lemmade alla. Kaaluti erinevaid võimalusi sõnavormide lemmatiseerimiseks. Kasutati eesti keele morfoloogilist analüsaatorit Estmorf (Kaalep 1998), mis lemmatiseeris ja ühestas automaatselt kogu korpuse. Lemmatiseeritud korpuse kõrval kasutati mitmel juhul sama korpuse lemmatiseerimata varianti, et vajadusel saada andmeid ka lemma erinevate vormide sageduse kohta. Mõnel juhul on võrdlusmaterjaliks arvutatud välja ka sõnade koosesinemissagedusi internetis. Internetiotsinguks kasutati otsingumootorit Google[11] ja kõik võrreldavad otsingud tehti ühe päeva jooksul (3.11.2008), sest kollokatsioonide esinemissagedus sõltub otsingu tegemise hetkest. Näiteks on siinse töö autor leidnud, et kollokatsiooni *valged jõulud* esinemissagedus võis erineda ligi kümme korda olenevalt sellest, kas otsing tehti oktoobris või jaanuaris (Piits 2011: 116).

Kollokatsioonide leidmisel ja konkordantsiridade analüüsimisel kasutati Concordi programmi paketist WordSmith Tools 4,0.

Concordi programmiga tehtud lihtsates päringutes pole kasutatud etteantud leksikogrammatilisi konstruktsioone ega kollokatsioonitugevuse arvestamise statistikuid, vaid leitud on eraldi vasakule hargnevad ja paremale hargnevad sagedamad koosesinemised, millest võrdlusesse on võetud iga otsisõna 30 sagedamini esinevat kollokaati.

Algselt kaaluti ka kollokatsioonitugevuse arvestamist. WordSmith Toolsi Concordi programm oleks tol hetkel võimaldanud kasutada kolme erinevat kollokatsioonitugevuse mõõtmise statistikut: log-tõepära funktsiooni (*Log likelihood)*, vastastikuse informatsiooni mõõdikut *(Mutual Information)* ning *Z-skoori*. Statistikute puhul peab aga alati arvestama, millise aspekti nad võivad andmetes esile tuua ja mis jääb tähelepanuta. Nt klassikalist vastastikuse informatsiooni mõõdikut on kritiseeritud seetõttu, et see annab liiga suure kaalu harvaesinevatele nähtustele (Gómez 2013: 206; Xiao, McEnery 2006). Kollokatsioonitugevuse statistikute kriitikud (vt Stubbs 1995, 2001a; Kilgarriff 2005) on tõenäosustestidele ette heitnud, et need põhinevad juhuslikkuse eeldusel, aga nende sõnul pole keeleaineses midagi juhuslikku.

Siiski katsetati siinse töö andmete peal kõiki Concordi programmi välja pakutud statistikuid, aga tulemusi lihtsa koosesinemissagedusega võrreldes selgus, et kollokatsioonitugevuse statistikuid kasutades sattusid tabelites etteotsa haruldasemad koosesinemised, kusjuures tundus, et tulemust mõjutasid enam väikesed sageduserinevused. Võimalik, et kollokatsioonitugevuse statistikute kasutamiseks jäi korpuse maht liiga väikeseks. Kuna eesmärgiks ei olnud harvaesinevate ja idiomaatiliste kollokaatide leidmine, vaid sõnade tüüpilise kasutuskonteksti uurimine ja võrdlemine, siis jäädi lihtsa koosesinemissageduse võrdluse juurde. Lihtne koosesinemissageduse arvestamine on siiani populaarne neo-firthlikes uurimustes (vt nt Hoey jt 2007).

[11] Google otsingumootor www.google.ee (3.11.2008)

Otsisõnade endi esinemissageduste erinevused nivelleeruvad, kuna vaadeldakse kõigi sõnade 30 sagedamat kollokaati (vt tabeleid Lisades 1 ja 2). St suurima esinemissagedusega sõna *inimene* (esineb korpuses 12277 korda) 30. vasakule hargnev kollokaat *aasta* esineb koos sõnaga *inimene* 105 korda, mis moodustab 0,85% sõna *inimene* esinemissagedusest. Samal ajal kui väikseima esinemissagedusega sõna *tütar* (esineb korpuses 612 korda) 30. vasakule hargnev kollokaat *laps* esineb koos sõnaga *tütar* ainult 6 korda, moodustades aga siiski 0,98% protsenti sõna *tütar* esinemissagedusest.

3.2.2. Kollokaatide leidmise kaugusest

Üks raskusi kollokatiivsete suhete mõõtmisel avaldub probleemis, kui kaugel võib asetseda kollokaat otsisõnast, et seda saaks veel nimetada kollokatsiooniks. Sahlgren (2008: 11) nimetab seda „miljoni euro küsimuseks" ja toob hulgaliselt näiteid, kus erinevad autorid on kasutanud kollokatsioonide leidmisel väga erineva suurusega akent (*collocational span*). Kõige huvitavama näitena esitab ta Schütze (1992) välja pakutud variandi, kus akna laiust ei määrata sõnades, vaid tähemärkides, mis annab eelise pikkadele täistähenduslikele sõnadele. Schütze on sõnatähenduste automaatseks eristamiseks kasutanud 1000 tähemärgi laiuse akna kõrval ka nt 50 sõna laiust akent (Schütze 1998), samas kui näiteks Church ja Hanks (1989) on kasutanud ainult 5 sõna suurust akent. Sinclair (1991: 106) on leidnud, et kollokaatidel, mis on otsisõnast kaugemal kui neli sõna, on otsisõnaga vähe seost. Ilmselt määrab akna laiuse siiski eesmärk: Karlgren ja Sahlgren (2001) on leidnud, et paradigmaatiliste suhete tuvastamiseks on parem kitsam aken. Stubbsi sõnul on olemas konsensus, et olulised kollokatsioonid leitakse vahemikust 4:4, st neli sõna otsisõnast vasakule ja neli sõna paremale (Stubbs 2001: 29). Kindlasti oleneb aga vahemaa suurus palju keele tüübist: Stubbs on oma järeldusteni jõudnud rohkem analüütilist väljendusviisi kasutavat inglise keelt uurides ja näiteks Sinclair (2004) on tähele pannud, et inglise keeles täidab nimisõna kollokatsioonidest vasakult esimese positsiooni peaaegu alati artikkel *the*. Kuna eesti keeles artiklid puuduvad ja keel on sünteetilisem kui inglise keel, siis võiks eesti keeles optimaalne kollokaatide kaugus otsisõnast väiksem olla. Korrelatsioon on olemas ka korpuse suuruse ja optimaalse akna suuruse vahel. Sahlgren on leidnud, et suuremal materjalil võib akna suurus olla natuke laiem (pakub suuruseks 3:3), aga väiksema materjali peal annab paremaid tulemusi 2:2 aken (Sahlgren 2006: 115). Paul Baker (2006: 103) on võrrelnud sama materjali peal erinevaid aknaid ja leidnud, et aken 3:3 toob välja kõige paremini kollokaadid, mis kuuluvad otsisõnaga samasse nimisõnafraasi.

Lisaks on arutletud selle üle, kas otsinguaken peaks olema sümmeetriline või mitte (vt nt Jantunen 2004: 21). Kui üldiselt on levinud, et kollokaate vaadeldakse otsisõnast võrdselt kauguselt vasakult ja paremalt, siis leidub ka arvamusi (Mason 2000), et aken ei peaks olema sümmeetriline, kuna sõnad mõjutavad eelnevat

ja järgnevat konteksti erinevalt. Sümmeetrilisuse küsimus tuleb kõne alla aga peamiselt siis, kui mõlemale poole hargnevaid kollokaate vaadeldakse korraga, nt Jantunen (2004) on otsinud kollokaate vahemikust V3-P2. Siinses töös aga vaadeldakse paremale ja vasakule hargnevaid kollokaate eraldi, mistõttu on mõlemale poole hargnevaid kollokaate otsitud sama kaugelt.

Siinses peatükis on kollokaate otsitud vahemikust 3:3 ehk siis kolm sõna vasakule ja kolm sõna paremale. Lisaks on rakendatud kaale, mis annavad veidi suurema väärtuse kollokaadile, mis on sõnale lähemal. Lähtutud on põhimõttest, et mida kaugemal otsisõnast mingi sõna esineb, seda juhuslikumaks muutub seos otsisõnaga. 30 sagedama kollokaadi hulgast jäeti välja muutumatud sõnad ja korpuses esinenud allikaviited. Selleks kasutati 55-st sõnast või täheühendist koostatud *stoplist*'i, mis jättis kollokaatide hulgast välja konjunktsioone (nt *kui, et, aga, kuid, sest*), sagedamaid adverbe (nt *väga, juba, pärast*) ja adpositsioone (nt *üle, eest, enne, juurde, vastu, peale, koos*) ning viiteid, mis näitasid korpuses lause päritolu (nt *PM, EPL, AJAE, AJA, ML*). Muutumatud sõnad jäeti välja, kuna süntaktiline seotus otsisõnaga on väiksem ja nad kannavad ka vähem semantilist infot, mis võiks aidata kaasa inimest tähistavate sõnade tähenduse määratlemisel.

Peatükis vaadeldakse vasakule ja paremale hargnevaid kollokaate eraldi.

3.3. Vasakule hargnevad ühised kollokaadid

3.3.1. Kollokatsioonikobarate suhestatus

Sõna koos oma sagedamate kollokaatidega on nimetatud kollokatsioonikobaraks (*collocational cluster*). Lähtudes eelpool (vt 3.1) viidatud Harrise distributiivsuse hüpoteesist, siis mida rohkem on kokkulangevaid kollokaate kahel kollokatsioonikobaral, seda enam võib eeldada, et need sõnad on sarnase tähendusega (Harris 1954). Järgnevalt otsitakse seoseid otsisõnade kollokatsioonikobarate kattuvuse ja otsisõnade tähenduskomponentidele antud väärtuste vahel. Sõna tähenduskomponentidest rääkides võetakse aluseks sõna esimese, kõige sagedama põhitähenduse komponentidele antud väärtused (vt 2.1.2.2).

Arvestades et uuritakse kümne sõna 30 kõige sagedamat vasakule hargnevat kollokatsiooni, on kollokaatide koguarvuks 300. Siiski on sellest hulgast erinevaid kollokaate vähem kui kolmandik, st 96. Sellest hulgast on alla poole selliseid, mis esinevad vaid ühel otsisõnal, ülejäänud kollokaadid esinevad kahel või enamal inimest tähistaval otsisõnal.

Tabelis 3 esitatud arvud näitavad, kui palju mingi inimest tähistava sõna 30-st vasakule hargnevast kollokaadist kattub iga teise inimest tähistava sõna 30 sagedama kollokaadiga. Liikudes rida mööda paremale, on näha, milliste sõnadega on otsisõnal kõige rohkem (poolpaksus kirjas) ja millistega kõige vähem (kursiivis) ühiseid kollokaate.

Tabel 3. Otsisõnade ühised vasakule hargnevad kollokaadid. Suurimad kokkulangevused on esitatud poolpaksus kirjas, väikseimad kursiivis

	ema	inimene	isa	laps	mees	naine	poeg	poiss	tüdruk	tütar
ema		15	**20**	16	19	**20**	15	*14*	17	*14*
inimene	15		11	20	**21**	20	12	16	14	*11*
isa	**20**	*11*		12	14	16	14	12	14	13
laps	16	**20**	*12*		**20**	19	15	15	15	14
mees	19	21	14	20		**22**	14	16	17	*11*
naine	20	20	16	19	**22**		15	16	15	*12*
poeg	15	*12*	14	15	14	15		13	14	**18**
poiss	14	16	12	15	16	16	13		**19**	*11*
tüdruk	17	14	14	15	17	15	14	**19**		*12*
tütar	14	*11*	13	14	*11*	12	**18**	*11*	12	

Esimesest reast nähtub, et sõnal *ema* on kõige rohkem ühiseid kollokaate sõnadega *naine* ja *isa* ning kõige vähem ühiseid kollokaate sõnadega *poiss* ja *tütar*. Võrdluseks komponentanalüüsi tulemusi tuues (vt 2.1.2.2, tabel 2) selgub, et sõnade *ema* ja *naine* tähenduskomponentidest saavad samad väärtused MEESSOOST ja TÄISKASVANUD. Sõnadel *ema* ja *isa* on aga koguni neli sama väärtusega tähenduskomponenti ja vaid komponent MEESSOOST saab erinevad väärtused. Sõnaga *poiss* pole sõnal *ema* aga ühtki sama väärtusega tähenduskomponenti. Need tähendusaspektid kajastuvad kollokatsioonikobarate kattumises: sõnadel *poiss* ja *ema* on kõige vähem ühiseid kollokaate.

Teisest reast selgub, et sõna *inimene* kollokatsioonikobar sarnaneb kõige rohkem sõna *mees* kollokatsioonikobaraga. Neil on 30 kollokaadist 21 ühist kollokaati. Kollokaatide arvu sõna tähenduskomponentide väärtustega kokku viies selgub, et sõnal *inimene* on kõige rohkem ühiseid kollokaate sõnadega, mille tähenduses pole komponent LÄHISUHTES OLEV saanud positiivset väärtust. Kõige väiksem kollokatsioonikobarate kattumine (ainult 11 ühist kollokaati) on aga sõnade *poeg*, *tütar*, *isa* ja *ema* kollokatsioonikobaratega. Nende sõnade kõige olulisem tähendus väljendab lähisuhet. Neist viimastega pole sõnal *inimene* ühtki sama väärtusega tähenduskomponenti.

Kolmandast reast nähtub, et sõnal *isa* on suurim ühisosa sõnaga *ema* ja väikseim sõnaga *inimene*. Selline tulemus peegeldab sõnade tähendust: sõnade *isa* ja *ema* tähendused erinevad ainult komponendi MEESSOOST väärtuse poolest, samal ajal kui sõnadel *isa* ja *inimene* pole ühtki sama väärtusega tähenduskomponenti.

Sõnal *laps* on 20 ühist kollokaati sõnadega *inimene* ja *mees* ning ainult 11 ühist kollokaati sõnaga *isa*. Sõna *laps* peamine tähendus eristub sõna *inimene*

tähendusest vaid komponendi TÄISKASVANUD väärtuse poolest, samas kui *laps* erineb sõnast *isa* kõigi tähenduskomponentide väärtuste poolest.

Viiendalt ja kuuendalt realt nähtub, et sõnadel *mees* ja *naine* on omavahel kõige suurem kollokatsioonikobarate kattumine (22 ühist kollokaati) ja mõlemal sõnal on kõige vähem ühiseid kollokaate sõnaga *tütar* (sõnal *mees* on sõnaga *tütar* 11 ja sõnal *naine* 12 kokkulangevat kollokaati). Komponentanalüüs näitab, et sõna *mees* ja *naine* tähendusi eristab vaid komponendile MEESSOOST antud väärtus. Sõnal *mees* on sõnaga *tütar* kõik komponendid erinevad, sõnad *naine* ja *tütar* sarnanevad komponendi MEESSOOST negatiivse väärtuse poolest.

Sõna *poeg* kollokatsioonikobar kattub kõige rohkem sõna *tütar* kollokatsioonikobaraga. Kõige vähem on sõnal *poeg* ühiseid kollokaate sõnaga *inimene*. Sõnade *poeg* ja *tütar* tähendusi eristab vaid komponendile MEESSOOST antud väärtus, muud komponendid on sama väärtusega. Sõnadel *tütar* ja *inimene* on neli erinevat tähenduskomponenti ja vaid komponendile TÄISKASVANUD antud väärtus langeb kokku.

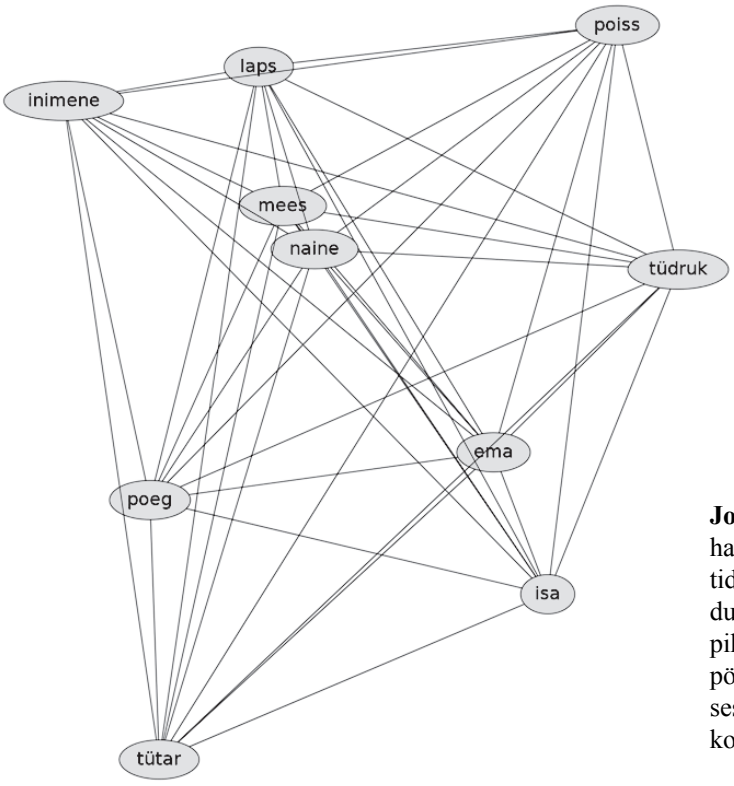
Kaheksandalt ja üheksandalt realt on näha, et sõnadel *poiss* ja *tüdruk* on omavahel kõige suurem kollokatsioonikobarate kattumine (19 ühist kollokaati). Mõlemal sõnal on kõige vähem ühiseid kollokaate sõnaga *tütar* (sõnal *poiss* on sõnaga *tütar* 11 ja sõnal *tüdruk* 12 kokkulangevat kollokaati). Sarnaselt teiste antonüümipaaridega erinevad sõnad *poiss* ja *tüdruk* vaid tähenduskomponendile MEESSOOST antud väärtuse poolest. Samas pole sõnal *poiss* sõnaga *tütar* ühtki sama väärtusega tähenduskomponenti ja sõnal *tüdruk* on sõnaga *tütar* üks sarnane tähenduskomponent.

Sõnal *tütar* on kõige rohkem ühiseid kollokaate sõnaga *poeg*, teiste sõnadega on ühisosa väga väike ja sõnadega *inimene*, *mees* ja *poiss* on kokkulangevusi kõige vähem. Kui sõnad *tütar* ja *poeg* erinevad vaid tähenduskomponendile MEESSOOST antud väärtuse poolest, siis sõnal *tütar* pole ühtki sarnast tähenduskomponenti sõnadega *mees* ja *poiss* ning sõnaga *inimene* on üks sama väärtusega tähenduskomponent.

Vasakule hargnevaid kollokatsioonikobaraid võrreldes selgus tendents, et otsisõnal on tavaliselt kõige rohkem ühiseid vasakule hargnevaid kollokaate antonüümse sõnaga: sõnal *isa* sõnaga *ema* ja vastupidi, sõnal *mees* sõnaga *naine* ja vastupidi, sõnal *tütar* sõnaga *poeg* ja vastupidi ning sõnal *poiss* sõnaga *tüdruk* ja vastupidi. See tendents on iseloomulik just vasakule hargnevatele kollokatsioonisuhetele: etteruttavalt olgu öeldud, et paremale hargnevate kollokatsioonide võrdlemisel (vt ptk 3.3) polnud sugugi kõigil sõnadel oma antonüümi kollokatsioonikobaraga suurimat kattumist.

Tabelis 3 esitatud informatsiooni visualiseerimiseks kasutati vabavaralist tarkvaraprogrammi Graphviz[12]. Joonisel 4 on esitatud inimest tähistavate sõnade omavahelised suhted, mis on välja arvutatud kollokatsioonikobarate kattumise ulatuse põhjal. Mida kaugem on distants sõnade vahel, seda erinevam on olnud ühiste kollokaatide osa. Sellelt jooniselt on hästi näha paradigmaatiliste ja süntagmaatiliste suhete seosed. Sõna kõige lähemaks naabriks on tema antonüüm, st antonüümipaarikutel on kõige suurem süntagmaatilise konteksti kattuvus.

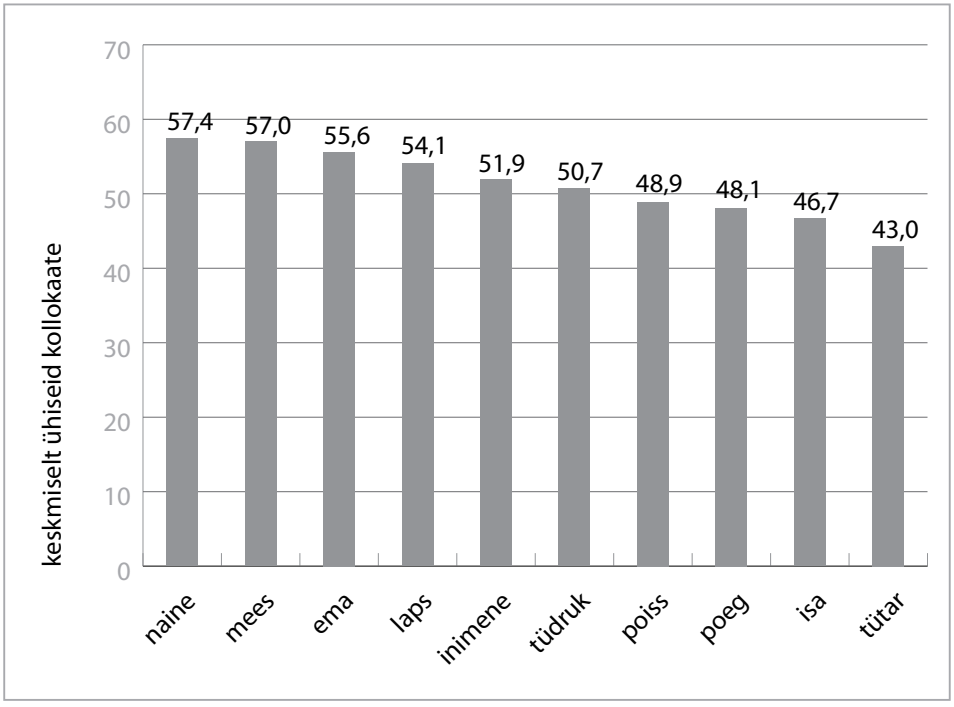
[12] Graphviz http://www.graphviz.org/ (9.09.2014)



Joonis 4. Vasakule hargnevate kollokaatide võrdluses peegelduvad seosed. Joonte pikkus on enamasti pöördvõrdelises seoses kokkulangevate kollokaatide arvuga

Kas antonüümsete sõnade sarnane kontekst peaks kinnitama või ümber lükkama distributiivsuse hüpoteesi, sõltub sellest, kuidas me antonüümiat määratleme. Antonüümiat võib pidada paradoksaalseks tähendussuhteks, sest see esineb tähenduste vahel, mis esmapilgul on tajutavad maksimaalselt erinevatena, aga lähemal vaatlusel selgub, et vastandlike tähenduste vahel on rohkem sarnast, kui erinevat. Vastandus esineb ühe tähendusvälja tähenduste vahel, mis erinevad ühe tähenduskomponendi poolest, samas ülejäänud komponentidele antud väärtused langevad kokku (vt 2.1.2.2). Seega on tegemist suurima võimaliku kokkulangemisega (jättes välja täissünonüümsed juhud), mistõttu võib öelda, et vasakule hargnevate kollokaatide analüüsil leidis kinnitust distributiivsuse hüpotees.

Järgmisena võrreldakse kõigi kümne vasakpoolse kollokatsioonikobara omavahelisi seoseid, et teada saada, millisel inimest tähistaval sõnal on kõigi teiste otsisõnadega kokku kõige rohkem ühiseid kollokaate (vt joonis 5). Keskmiselt kõige rohkem ühiseid kollokaate kõigi teiste kollokatsioonikobaratega on sõnadel *naine* ja *mees*. Keskmiselt 57% sõnade *naine* ja *mees* sagedamatest kollokaatidest langevad mõne teise inimest tähistava sõna kollokaadiga kokku. Seega võib öelda, et inimest tähistavate sõnade keskmes asuvad sõnad *naine* ja *mees*, mis süntagmaatilisel tasandil omavad teiste sõnadega kõige rohkem sarnast (sama väljendus ka joonisel 4).



Joonis 5. Otsisõnade ühiste vasakule hargnevate kollokaatide suhe protsentides

Järgnevad sõnad *ema* ja *laps*, mille kollokaatidest 55,6 ja 54,1 protsenti langevad teiste sõnade kollokaatidega kokku. Kui sõna *ema* on ühiste kollokaatide poolest kolmandal kohal, siis näiteks sõna *isa* on 14 ühise kollokaadiga alles eelviimane. Kõige vähem ühist teiste inimest tähistavate sõnade kollokaatidega on sõna *tütar* kollokaatidel, ainult 43 protsenti. Võimalik, et siin mängib rolli ka sõna *tütar* esinemissagedus korpuse materjalides. Kuna sõna esineb seal ainult 612 korda, siis võib juhtuda, et 30 sagedama kollokaadi hulka sattus ka juhuslikke sõnu, mis poleks suurema esinemissageduse korral sinna sattunud.

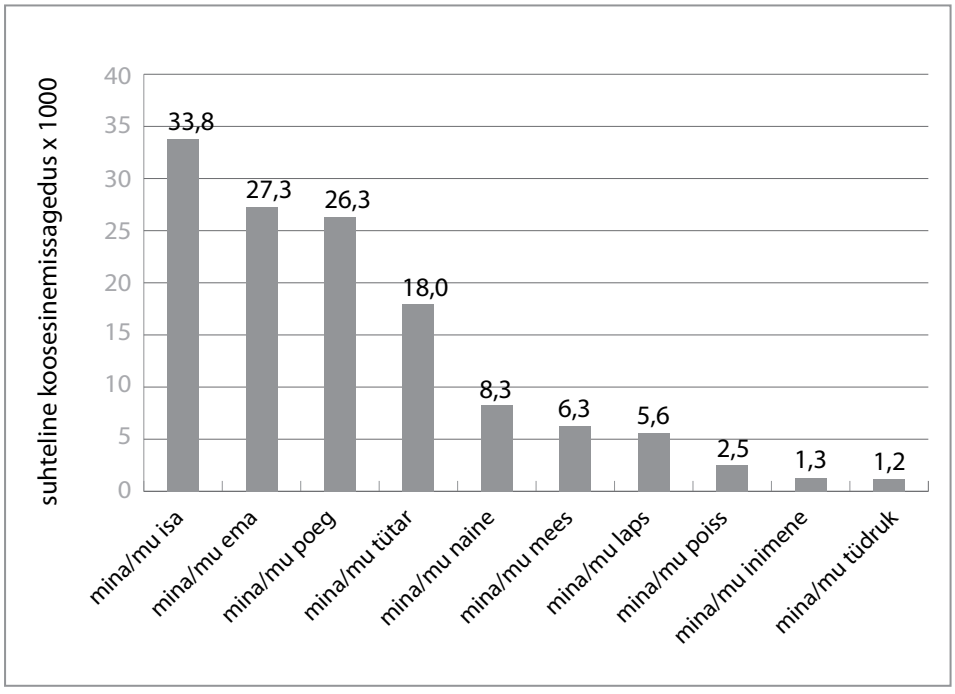
Inimest tähistavate sõnade kollokatsioonikobarate ühisosa on väga suur, st keskmiselt ligi pooled 30 sagedamast vasakule hargnevast kollokaadist kattuvad. Esmapilgul võib paista, et see tõestab distributiivsuse hüpoteesi kõige paremal moel. Kahtlused võivad aga tekkida, kui võtta võrdluseks valimiväliselt (*outgroup term*) suure esinemissagedusega sõna *auto*, mis enam ei oma teiste sõnadega ühist komponenti INIMENE. Kui võtta võrdlusesse *auto* 30 sagedamat kollokaati, siis selgub, et kokkulangevus inimest tähistavate sõnade kollokaatidega on samuti suur. Näiteks 8 kollokaati, mis esinevad kõigil inimest tähistavatel otsisõnadel, esinevad ka *auto* 30 sagedama kollokaadi seas. Võrdlus annab vihje, et need 8 kõige levinumat kollokaati pole ühised mitte ainult inimest tähistavatele sõnadele, vaid substantiividele üldiselt.

3.3.2. Vasakpoolsetest kollokaatidest lähemalt

Kõigil vaadeldud inimest tähistavatel sõnadel esinevad vasakpoolse kollokaadina sõnad *üks*, *kaks, kolm, mina, tema, see, olema, oma*. Kõigil otsisõnadel välja arvatud *tütar* on ühine kollokaat *saama*. Seitsmel otsisõnal on ühisteks kollokaatideks *eesti, noor, pidama* ja *teine*. Kuuel otsisõnal esinevad kollokaadina *laps, vanem, tegema, tulema, ütlema*. Viiel otsisõnal esineb kollokaadina *ise, aasta, naine, rääkima, sina, väike*. Kollokaadid *ema, iga, kõik, minema, uus* on ühised neljal sõnal. Kolmel inimest tähistaval sõnal on ühised kollokaadid *elav*, *isa, kodu, mõni, neli, poeg, poiss, tütar, vene, võima* ja *võtma.* Vaid kahele otsisõnale ühiste kollokaatide hulka kuuluvad sõnad *aastane, aeg, elama, hea, ilus, klass, noorem, pere, puue, selline, sündima, teismeline, tüdruk, vana, viimane*. Ülejäänud 41 sõna esinevad korraga vaid ühe otsisõna kollokaadina. Vt 30 sagedama kollokaadi esinemist lisast 1.

Milliseid järeldusi võiks teha 8 kollokaadi põhjal, mis esinevad kõigil kümnel otsisõnal? Ei saa väita, et tegu oleks kollokaatidega, mis oleksid iseloomulikud vaid inimest tähistavatele sõnadele. Numeraalid *üks, kaks* ja *kolm* viitavad keeles esinevale üldisele vajadusele substantiive loendada ja arvuliselt määratleda. Sõna *üks* esineb lisaks ka asesõnana positsioonides, kus artikleid kasutavad keeled kasutavad määramata artiklit. Määratud artikli ülesannetes esineb kollokaatide hulgas pronoomen *see*. Kuna arvestatud on puhast koosesinemissagedust, siis esineb kõigi sõnade ühise vasakule hargneva kollokaadi hulgas ka eesti keeles kõige sagedam lemma *olema*.

Lisaks esinevad kõigi kümne sõna kollokaatidena personaalpronoomenid *mina, tema* ja *oma*. Pärast korpuse lemmatiseerimist programmiga Estmorf tuleb arvestada, et personaalpronoomenite *mina* ja *tema* sagedusnäitajate hulgas on lisaks singulari vormidele esindatud ka pluurali *meie* ja *nemad* vormid. Jättes kõrvale mitmusevormid, võrreldi lemma *mina* erinevaid esinemisvorme. Selleks kasutati lemmatiseerimata korpust. Eesmärgiks oli teada saada, milliste inimest tähistavate sõnade puhul kasutatakse enam nominatiivivorme *mina* ja *ma* ning milliste sõnade kollokaatidena esinevad enam genitiivivormid *minu* ja *mu.* Kuna otsisõnade esinemissagedus on korpuses erinev ja hiljem sooviti korpusest saadud andmeid võrrelda ka internetist saadud andmetega, siis kasutati koosesinemissageduste võrdlemisel relatiivset ehk suhtelist sagedust, st absoluutne koosesinemissagedus jagati läbi inimest tähistava otsisõna sagedusega. Vormide *mina* ja *ma* ning *minu* ja *mu* suhtelist esinemissagedust võrreldes tuli välja, et kui nominatiivset vormi esines kõigi inimest tähistavate sõnade kollokaadina enam-vähem võrdselt, siis genitiivsete vormide esinemissagedus kollokaadina sõltus inimest tähistavast sõnast. Jooniselt 6 on näha, et genitiivses vormis personaalpronoomenit kasutatakse kõige rohkem sõnade puhul, mille tähenduskomponent LÄHISUHTES OLEV (vt 2.1.2.2 tabel 2) saab positiivse väärtuse. Eriti suur on *minu* ja *mu* koosesinemissagedus sugulussuhet tähistavate sõnadega *isa, ema, poeg, tütar*.



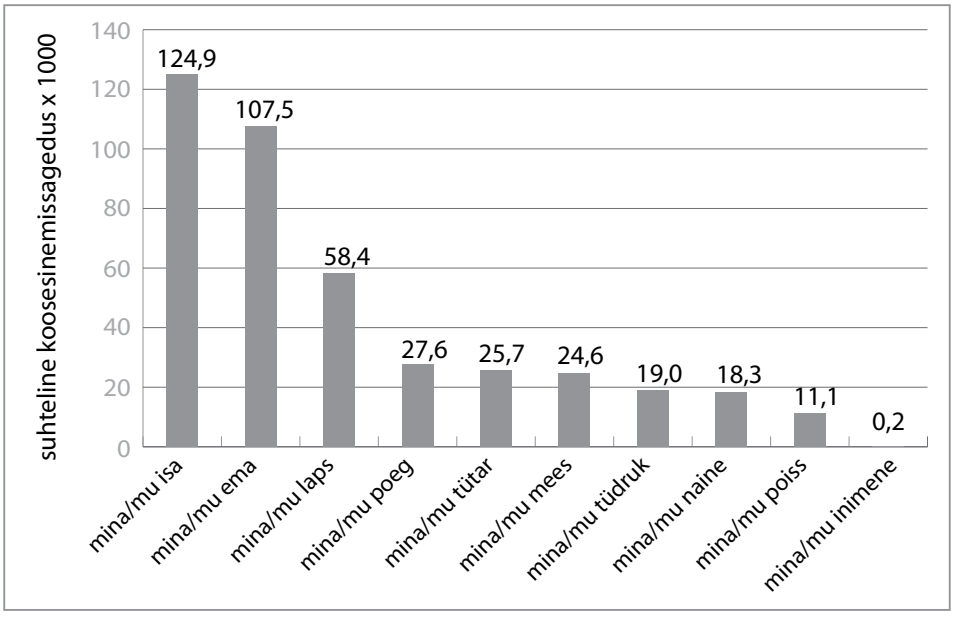
Joonis 6. Vormide *minu* ja *mu* suhteline koosesinemissagedus inimest tähistavate sõnadega. Kollokatsiooni absoluutne esinemissagedus on jagatud inimest tähistava sõna esinemissagedusega

Personaalpronoomeni 1. isiku genitiivivormi kasutatakse V1 (vasakult esimeses) positsioonis minakõneleja sugulussuhetele viitamiseks. Kõige suurema suhtelise esinemissagedusega on kollokatsioon *minu isa*. Kuigi sõna *ema* üldine kasutussagedus on korpuses suurem kui sõnal *isa*, siis kollokatsiooni *minu isa* absoluutsagedus on veidi suurem kui kollokatsioonil *minu ema* ja kollokatsiooni suhteline sagedus ehk kollokatsiooni suhe arvestades otsisõna sagedust on 20% suurem. Sellised andmed viitavad, et esimeses isikus kõnelejad märgivad personaalpronoomeniga oma suhet isasse sagedamini kui emasse.

Kollokatsioonide *minu ema* ja *minu poeg* esinemissagedus on peaaegu võrdne, veidi vähem räägitakse ajakirjandustekstides oma tütrest. Üheks põhjuseks, miks *mina* genitiivivorm esineb mainitud nelja sõna *isa, ema, poeg, tütar* kollokaadina, on nende inimest tähistavate sõnade tähendus – esimeseks ehk peamiseks tähenduseks on sugulussuhte, kas lapsevanemaks olemise või järglaseks olemise väljendamine. Sageduselt järgmise grupi moodustavad sõnad *naine, mees* ja *laps*. Nende sõnade üks tähendusi on samuti seotud mingi lähisuhte, kas abikaasaks või järglaseks olemise tähistamisega. Siiski pole sõna see tähendus, mille komponent lähisuhtes olev saab positiivse väärtuse, ajakirjandustekstides kõige olulisemaks ja seal kõige sagedamini esinevaks tähenduseks.

Kõige vähem esineb *mina* genitiivivorm koos sõnadega *inimene*, *poiss* ja *tüdruk*. Sõnal *inimene* puudub tähendus, mis võimaldaks kasutada genitiivis personaalpronoomenit atribuudina, ainult üks kollokaadi 16 esinemisjuhust esines V1 positsioonis, teised kasutusjuhud esinesid kõik V2 (vasakult teises) või V3 (vasakult kolmandas) positsioonis. Kuigi sõnu *tüdruk* ja *poiss* saab kasutada lähisuhte osaliste märkimiseks, siis ilmselt kasutatakse ajakirjandustekstides neid sõnu selles tähenduses harvem, mis kajastub ka personaalpronoomeni esinemuses.

Ajakirjanduskorpusest leitud materjali kõrvale otsiti internetist võrdlusmaterjali ja arvutati välja 10 sagedama inimest tähistava sõna koosesinemissagedused vormidega *minu* ja *mu* (vt joonis 7).



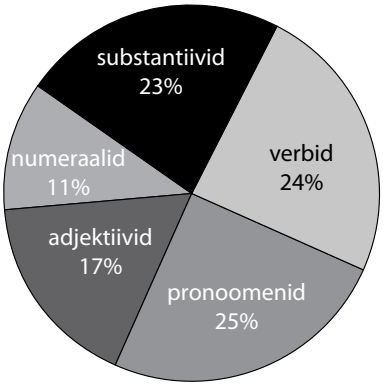
Joonis 7. Vormide *minu* ja *mu* suhteline koosesinemissagedus inimest tähistavate sõnadega. Kollokatsiooni absoluutne esinemissagedus om jagatud inimest tähistava sõna esinemissagedusega

Võrreldes jooniseid 6 ja 7 on näha, et internetis on *minu* ja *mu* vormi suhteline esinemissagedus kollokaadina palju kõrgem kui tasakaalus korpuse ajakirjanduskorpuses. Vahe on seda suurem, et korpusest leiti *mina* ja *mu* vormide esinemissagedused positsioonidest V1, V2 ja V3, aga internetist leiti kollokaadid ainult kaugusest V1. Soome keelt uurides on leitud, et personaalpronoomenite esinemine sõltub suurel määral teksti liigist. Näiteks laiemale lugejaskonnale mõeldud populaarkirjanduses esineb ligi 5 korda rohkem personaalpronoomeneid kui akadeemilistes tekstides (Jantunen 2012: 208). Üks suurema sageduse põhjuseid võib peituda selleski, et internetis on enam esimeses isikus suhtlust, kuna lisaks erinevatele ajakirjanduslikele veebiväljaannetele on seal palju foorumeid, kus kasutatakse isiklikumat suhtlemisstiili.

Üldised sagedusvahekorrad on aga kahel joonisel sarnased. Kõige sagedam on ka internetimaterjalides kollokatsioon *minu isa*, sellele järgneb *minu ema*. Võrreldes ajakirjanduskeele korpusega on sagedam kollokatsioon *minu laps*. Järgnevad kollokatsioonid enam-vähem samas järjestuses ja väikeste sageduserinevustega. Viimaseks jääb kollokatsioon *minu inimene*. Seega on andmete võrdlemisel märgata sarnaseid tendentse: sugulussuhtele viitavaid sõnu kasutatakse kõige enam koos genitiivatribuudiga *minu*, järgnevad sõnad, mida saab teises tähenduses kasutada muudele lähisuhetele viitamisel. Kõige harvem kasutatakse kollokaati koos sõnaga *inimene*, mille tähenduses on komponent LÄHISUHTES OLEV neutraliseerunud.

3.3.3. Vasakule hargnevad kollokaadid sõnaliigiti

Kollokaatide sõnaliigilise jaotuse analüüs ei kajasta adverbide, konjunktsioonide ja kaassõnade esinemist, sest muutumatud sõnad on andmete töötlemise esimeses etapis (vt 3.2.2) 30 sagedama kollokaadi hulgast välja jäetud. Kõigi kümne substantiivi vasakule hargneva 300 kollokaadi jaotust sõnaliigiti võib näha joonisel 8. Sõnadest neljandiku moodustavad pronoomenid, peaaegu sama palju on verbe (24%) ja substantiive (23%). Veidi väiksema osa kõigist 300 kollokaadist moodustavad adjektiivid (17%) ja numeraalid (11%).



Joonis 8. Vasakule hargnevad 300 kollokaati sõnaliigiti

Kollokaatide sõnaliikidesse jagunemist mõjutavad süntaksi sõnajärje reeglid ja otsisõna sõnaliik. 10 vaatluse all olevat inimest tähistavat sõna on substantiivid, mis võivad esineda lauses erinevates süntaktilistes funktsioonides. Süntaktiliste funktsioonide esitamisel on toetutud „Eesti keele grammatikale" (vt EKG I 1995, EKG II 1993). Näited funktsioonide juurde on leitud eesti keele koondkorpusest (203 miljonit sõna). Vaatluse all olevad inimest tähistavad sõnad esinevad lausetes subjektina (3), objektina (4), predikatiivina (5), atribuudina (6), adverbiaalina (7).

(3) Pärast tuli **mees** Lätti, ehitas elektrijaama. (Riigikogu stenogrammid 2000.05)

(4) Tõenäoliselt ootab **meest** mitme tuhande krooni suurune rahatrahv. (Eesti Päevaleht 1999.11.17)

(5) Ots on **mees**, kes kõnnib omapäi. (Lääne Elu 2000.06.13)

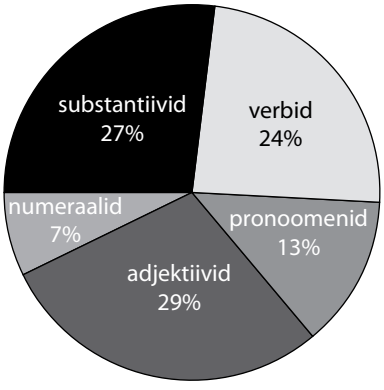
(6) **Mehe** sõnul ei ole vaja seadustesse teha muudatusi. (Eesti Päevaleht 2006.02.07)

(7) Politsei andis **mehele** kohapeal esmaabi. (Postimees 1996.12.20)

Lisaks moodustavad inimest tähistavad sõnad ka erinevaid väiksemaid üksusi, olles osalised substantiivifraasides nii põhisõna kui atribuudina ja kvantorfraasides atribuudina. Eelnevast lähtuvalt vaadeldakse, millised on otsisõnadest vasakule hargnevate kollokaatide võimalikud süntaktilised seosed otsisõnadega, seejuures peab mainima, et kõigil kollokaatidel ei pruugi üldse olla otsisõnaga süntaktilist seost, kollokaat võib kuuluda hoopis eelmisesse fraasi.

Kõige suurema grupi vasakpoolsetest kollokaatidest moodustavad pronoomenid. Need on sel juhul otsisõnaga seoses, kui esinevad adjektiivatribuudina (nt ***minu*** *mees,* ***tema*** *lapsed,* ***oma*** *poeg*). Koosesinemissageduselt järgmise grupi moodustavad verbid, mis otsisõnadest vasakul asetsevatena võivad näidata nii tegevusi, kus inimest tähistav sõna on objektiks või predikatiiviks, aga ebatüüpilise sõnajärje puhul ka neid, kus inimest tähistav sõna on subjektiks.

Vasakule hargnevad substantiivsed kollokaadid võivad esineda nii otsisõna laiendavate genitiivatribuutidena (***aasta*** *isa*) kui ka adverbiaalatribuutidena (***paanikas*** *inimesed*). Suure osa just vasakule hargnevatest kollokaatidest moodustavad adjektiivid. Peamiselt esinevad nad nimisõnafraasis adjektiivatribuutidena (***noor*** *naine,* ***aus*** *mees,* ***paha*** *poiss*).



Joonis 9. Vasakule hargnevad 96 erinevat kollokaati sõnaliigiti

Kui eemaldada 300 kollokaadi seast kõik kattuvad kollokaadid, siis järelejäänud 96 kollokaati analüüsides selgub, et vähenenud on pronoomenite ja numeraalide osakaal (vt joonis 9). See on ka ootuspärane, kuna otsisõnade kattuvate kollokaatide seas on kõige rohkem just nendest sõnaliikidest sõnu. Pronoomenitel

on võrreldes täistähenduslike sõnadega abstraktsem ja „tühjem” tähendus (EKG I 1995: 15), mis ilmselt võimaldabki samu pronoomeneid kasutada erinevate otsisõnadega seoses.

Jooniste 8 ja 9 võrdlemisel on näha, et verbide osakaal on kahel diagrammil sama, aga korduvate kollokaatide eemaldamisel on märgatavalt suurenenud substantiivide ja eriti adjektiivide hulk. See näitab, et varieeruvus on just nendest sõnaliikidest kollokaatide seas kõige suurem, ehk siis teisisõnu, just adjektiivsed ja substantiivsed kollokaadid erinevad otsisõnade kaupa kõige rohkem.

3.4. Paremale hargnevad ühised kollokaadid

Kui peatükis 3.3 võrreldi inimest tähistavate sõnade vasakule hargnevaid kollokaate, siis siin on vaatluse all paremale hargnevad kollokaadid. Kõigi sõnade peale kokku on erinevaid parempoolseid kollokaate 85. Neist 32 on sellised, mis esinevad vaid ühel otsisõnal, ülejäänud kollokaadid esinevad kahe või enama inimest tähistava sõna 30 sagedama kollokaadi seas. Vasakule hargnevaid erinevaid kollokaate oli 11 võrra rohkem, st 96, mis näitab, et sõnade paremale hargnevad kollokatsioonikobarad kattuvad enam.

3.4.1. Kollokatsioonikobarate suhestatus

Siinses peatükis võrreldakse 10 sagedama inimest tähistava sõna kollokatsioonikobaraid omavahel. Tabelis 4 on esitatud arvud, mis näitavad, kui suur osa paremale jäävast 30 kollokaadist erinevate inimest tähistavate sõnade puhul kattuvad.

Tabel 4. Otsisõnade ühised paremale hargnevad kollokaadid. Suurimad kokkulangevused on esitatud poolpaksus kirjas, väikseimad kursiivis

	ema	inimene	isa	laps	mees	naine	poeg	poiss	tüdruk	tütar
ema		19	**20**	*14*	16	17	18	*14*	17	*14*
inimene	19		15	17	**22**	21	17	17	18	*13*
isa	**20**	15		15	15	*14*	16	*14*	15	15
laps	*14*	17	15		17	**18**	17	16	15	15
mees	16	**22**	*15*	17		21	19	17	19	*15*
naine	17	**21**	*14*	18	**21**		18	18	17	16
poeg	18	17	16	17	**19**	18		**19**	17	*15*
poiss	*14*	17	*14*	16	17	18	**19**		18	17
tüdruk	17	18	*15*	*15*	**19**	17	17	18		*15*
tütar	14	*13*	15	15	15	16	15	**17**	15	

Rida mööda paremale liikudes on näha, milliste teiste sõnadega on otsisõnal kõige rohkem (poolpaksus kirjas) ja millistega kõige vähem (kursiivis) ühiseid kollokaate.

Esimeselt realt on näha, et sõnal *ema* on kõige rohkem ühiseid kollokaate antonüümiga *isa* (20 ühist) ja kõige vähem ühiseid kollokaate sõnadega *laps*, *poiss* ja *tütar*. Sarnased vahekorrad avaldusid ka vasakpoolseid kollokaate uurides (vt 3.3.1). Sõna *ema* suurim kollokatsioonikobarate kattumine oma antonüümi kollokatsioonikobaraga on mõistetav, kuna sõnu eristab vaid üks tähenduskomponent. Ootamatuna esineb aga vaid ühe võrra vähem ühiseid kollokaate ka sõnadel *ema* ja *inimene*. Need on sõnad, mille ükski tähenduskomponent pole saanud sama väärtust.

Teisest reast selgub, et sõnal *inimene* on kõige suurem kollokaatide ühisosa sõnaga *mees* ja kõige vähem ühiseid kollokaate sõnaga *tütar*. Sarnane tulemus saadi ka vasakpoolseid kollokaate võrreldes. Sõnade *inimene* ja *mees* 22 ühist kollokaati võivad näidata, et lisaks kolmele ühisele tähenduskomponendile võib inimese tähenduses komponent TÄISKASVANUD olla kaotanud oma neutraalsuse ja olla määratletud positiivse väärtusega, st tüüpiline inimene on pigem meessoost ja pigem täiskasvanud. Tulemusse võivad oma osa anda ka üksikud sõna *mees* geneerilise ehk sooülese kasutuse juhud. Sõnadel *inimene* ja *tütar* on aga vaid ühine üks komponent. Seega on mõistetav, et sõnadel on ainult 13 ühist kollokaati.

Sõna *isa* kollokaadid kattuvad kõige rohkem oma antonüümi *ema* kollokaatidega. Need sõnad erinevad tähenduskomponendile meessoost antud väärtuse poolest. Ülejäänud otsisõnade kollokaatidega on ühisosa ühtlaselt väike.

Sõnal *laps* on kõige rohkem ühiseid kollokaate sõnaga *naine* ja kõige vähem sõnaga *ema*. Samas pole erinevused kokkulangevate kollokaatide arvus kuigi suured – suurim kollokaatide kattuvus on 18 ühist ja väikseim on 14 ühist kollokaati. Sõnade *laps* ja *naine* tähendused erinevad kahe tähenduskomponendi poolest. Sõnadel *laps* ja *ema* pole aga ühtki ühist tähenduskomponenti.

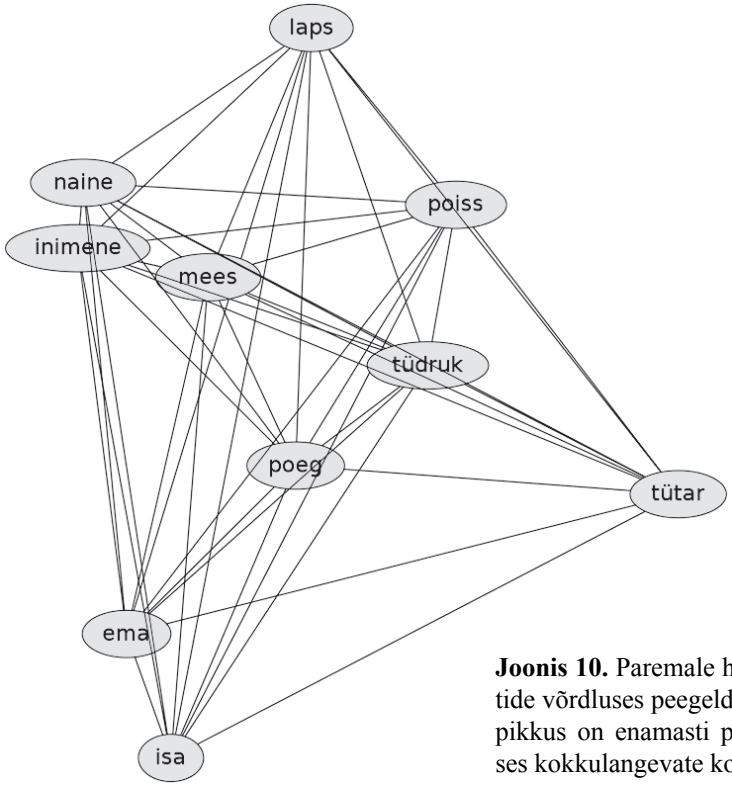
Sõnal *mees* on kõige rohkem ühiseid kollokaate sõnaga *inimene* (22 ühist), ühe võrra on ühiseid vähem antonüümse sõnaga *naine*. Kõige vähem on sõnal *mees* ühiseid kollokaate sõnadega *isa* ja *tütar* (15 ühist). Kui sõnal *mees* on sõnaga *isa* veel kaks sama väärtusega tähenduskomponenti, siis sõnadel *mees* ja *tütar* ühised komponendid puuduvad.

Sõnal *naine* on kõige rohkem ühiseid kollokaate sõnadega *inimene* ja *mees* (21 ühist) ning kõige vähem sõnaga *isa* (14 ühist). Sõnade *naine* ja *isa* tähendusi ühendab vaid tähenduskomponent TÄISKASVANUD.

Sõnal *poeg* on kõige rohkem ühiseid kollokaate sõnadega *mees* ja *poiss* ning kõige vähem ühiseid kollokaate sõnaga *tütar*. Selline tulemus on üllatav, kuna sõnal *poeg* on vaid üks ühine tähenduskomponent sõnadega *mees* ja *poiss* – see on komponent MEESSOOST. Ja vastupidi, sõnal *poeg* on sõnaga *tütar* kõik tähenduskomponendid samad, välja arvatud komponent MEESSOOST. Vasakpoolseid kollokaate võrreldes avaldus vastupidine tendents: teiste antonüümipaaride kõrval oli ka sõnadel *poeg* ja *tütar* kõige suurem kollokaatide kattuvus. Ka teiste sõnade puhul võib märgata, et paremale hargnevate kollokaatide puhul ei väljendunud nii selgelt antonüümide suurem kollokaatide kattumine. Ainult sõnad *ema*, *isa* ja

naine said maksimumarvu kokkulangevusi just oma antonüümiga. Sõnad *mees*, *poiss* ja *tüdruk* said maksimumilähedasi tulemusi, sõna *tütar* omas veel vähem ühiseid kollokaate sõnaga *poeg*. Sõna *poeg* kollokaatide kokkulangevus oli just kõige väiksem oma antonüümiga *tütar*.

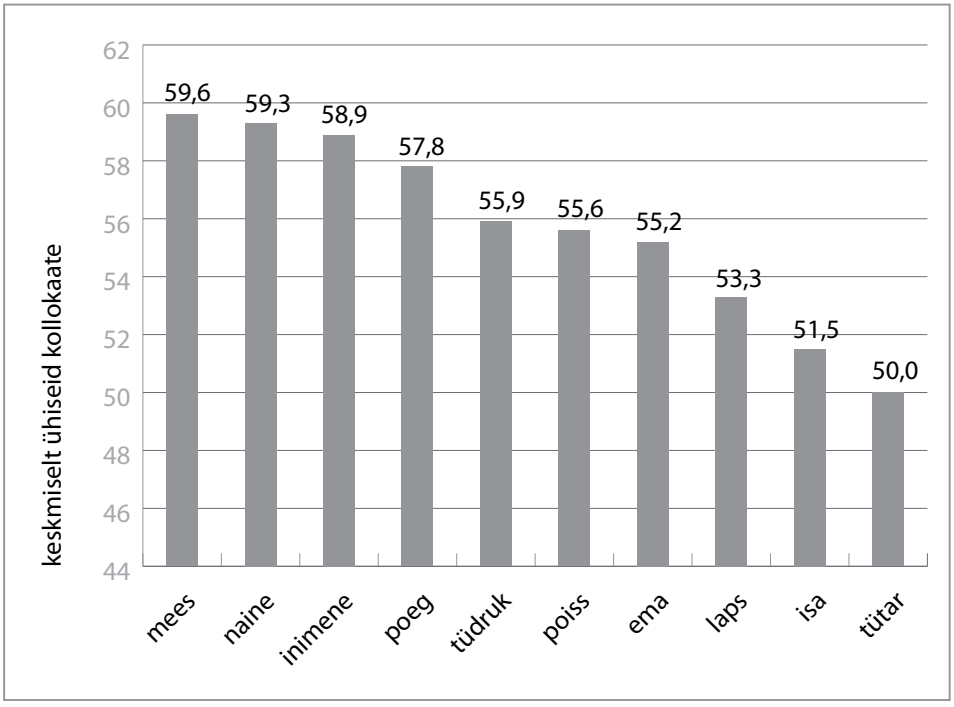
Joonisel 10 on näha paremale hargnevate kollokaatide võrdluses peegelduvad seosed. Antonüümiast tulenev lähedus on märgatav sõnapaaride *naine–mees, ema–isa* ja *tüdruk–poiss* vahel, kuigi kahel viimasel sõnal on veel lähedasemaid naabreid. Paremale hargnevate kollokaatide võrdluses ei peegeldu sõnade *tütar* ja *poeg* antonüümne seos.



Joonis 10. Paremale hargnevate kollokaatide võrdluses peegelduvad seosed. Joonte pikkus on enamasti pöördvõrdelises seoses kokkulangevate kollokaatide arvuga

Vasakule ja paremale hargnevate kollokaatide põhjal tehtud otsisõnade seoste graafe võrreldes võib märgata, et paremale hargnevate kollokaatide andmetele tuginev joonis on kokkusurutum, seosed sõnade vahel on tihedamad. Keskme moodustavad *mees*, *naine* ja *inimene*. Perifeeriasse jäävad sugulussuhetele viitavad sõnad nagu *ema, isa* ja *tütar*.

Sellist sõnade jagunemist tsentrisse ja perifeeriasse väljendab ka diagramm (vt joonis 11), kus kõigi kümne parempoolse kollokatsioonikobara omavahelised seosed on saanud arvulised väärtused. Number tulba kohal näitab, mitu protsenti otsisõna parempoolsetest kollokaatidest kattub keskmiselt teiste sõnade kollokaatidega.



Joonis 11. Otsisõnade ühiste paremale hargnevate kollokaatide suhe. Number näitab, mitu protsenti otsisõna kollokaatidest kattub keskmiselt teiste sõnade kollokaatidega

Kõige rohkem ühiseid kollokaate kõigi teiste kollokatsioonikobaratega on sõnal *mees*: keskmiselt 60% kollokaatidest langeb kokku mõne teise sõna kollokaatidega.

3.4.2. Parempoolsetest kollokaatidest lähemalt

Kõigi inimest tähistavate sõnade 30 sagedama paremale hargneva kollokaadi seas esinesid 7 sõna: *mina, minema, olema, oma, saama, see* ja *tema.* Võrreldes vasakule hargnevate kollokaatidega, on kõigi sõnade ühiste parempoolsete kollokaatide loendist puudu arvsõnad *üks, kaks* ja *kolm*, aga asemele on tulnud verbid *minema* ja *saama*. Ülejäänud pronoomenid *mina, tema, oma, see* ja eesti keele kõige sagedam sõna *olema* esinevad kõigi otsisõnade nii 30 sagedama vasakule hargneva kui ka 30 sagedama paremale hargneva kollokaadi hulgas. Verbide suurem esindatus ja arvsõnade puudumine parempoolsete kollokaatide seast on eeldatavasti eesti keele tüüpilisest sõnajärjest tingitud.

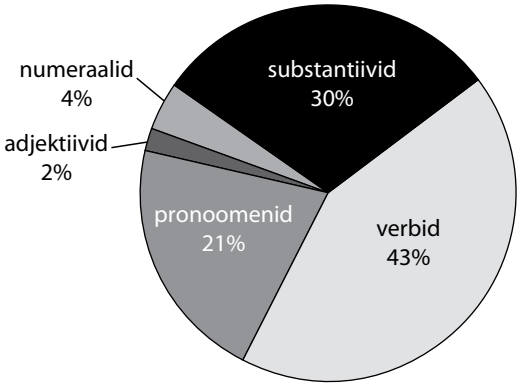
Peaaegu sama suure esinemissagedusega kollokaatidest tuleb nimetada peamiselt refleksiivpronoomeni funktsioonis esinevat sõna *ise* ja modaalverbi *pidama*, mis esinesid üheksa otsisõna kollokatsioonikobaras. Modaalverbidest olid sagedad veel *tulema* ja *võima*, mis esinesid kaheksa otsisõna sagedamate kollokaatide seas ja sama sage oli asesõna või numeraalina esinev *üks*.

Seitsmel otsisõnal on ühisteks kollokaatideks suhtlusverbid *ütlema* ja *rääkima*. Neid verbe kasutatakse kõige sagedamini otsekõne saateverbidena. Kumbki sõna ei esine sõna *laps* 30 sagedama paremale hargneva kollokaadi seas, mis viitab sellele, et ajakirjandustekstides edastatakse laste ütlusi harvem või kasutatakse refereerimisel juba sooliselt määratletud sõnu *poiss* ja *tüdruk*. Teiste inimest tähistavate sõnade kollokaatide seas on vähemalt üks verbaalset akti tähistavatest sõnadest esindatud, kusjuures meessoost isikute puhul (*isa, mees*) kasutatakse enam verbi *ütlema*, aga naissoost isikute puhul (*naine, tütar*) kasutatakse rohkem verbi *rääkima*. Samuti seitsme sõna sagedamate kollokaatide seas esineb sõna *kodu*. Ainult sõnade *inimene, mees* ja *tüdruk* 30 sagedama kollokaadi hulgast see sõna puudub. Nende sõnade kõige sagedamad tähendused ei saa komponendile LÄHISUHTES OLEV positiivset väärtust.

Kuuel otsisõnal esinevad kollokaadina *eesti, ema, hakkama, isa, jääma* ja *tegema*. Viiel otsisõnal esinevad *kaks, kõik, käima, laps, nimi* ja *tahtma*. Kollokaadid *elu* ja *võtma* on olemas neljal sõnal. Sõnad *aasta, aeg, auto, elama, naine, poeg, sina, suur, töö, tütar, õppima* esinevad korraga kolme otsisõna kollokaatide juures. Sõnad *andma, esimene, kool, leidma, mees, mängima, surema, surm, sõna, Tallinn, teadma, teine, tooma* ja *vanem* esinesid korraga kahe otsisõna 30 sagedama kollokaadi seas. Ülejäänud 32 sõna: *arv, arvama, eeskuju, jooksma, jutt, jätma, kaotama, kasvatamine, keha, käsi, mäss, noor, nägema, panema, pere, pilt, poiss, pärand, roll, sattuma, selline, sõber, sünd, sünnitama, tuba, töötama, tüdruk, uus, vaatama, vaim, vend* ja *õigus* esinevad vaid ühe otsisõna kollokaatide seas.

3.4.3. Paremale hargnevad kollokaadid sõnaliigiti

Vasakule ja paremale hargnevate kollokaatide sõnaliigilisust võrreldes on selgelt näha, et kollokaatide esinemist mõjutavad ka süntaktilised tegurid ja sõnajärg (vrd jooniseid 8 ja 12).



Joonis 12. Paremale hargnevad 300 kollokaati sõnaliigiti

Kui eeldame, et inimest tähistavad otsisõnad on kõige sagedamini subjekti funktsioonis, siis on loomulik, et verb on kõige sagedam sõnaliik kõigi otsisõnade paremale hargnevate kollokaatide seas, kuna eesti keeles eelneb subjekt verbile (seda nii juhul, kui tüüpiliseks sõnajärjeks pidada SVO-d (Erelt 2005: 8), kui ka juhul, kui tüüpiliseks pidada SOV-d (Ehala 2001: 43)). Lisaks on võimalikud ka muud sõnajärje tüübid ja samuti võib inimest tähistav sõna esineda substantiivi kõrval ka näiteks objekti, atribuudi või lisandina, aga paremale ja vasakule hargnevate kollokaatide võrdluses ilmnes, et tüüpilisem on verbi esinemine inimest tähistavatest sõnadest paremal. Verbid moodustavad 43% paremale hargnevatest kollokaatidest ja vaid 24% vasakule hargnevatest kollokaatidest.

Suure grupi paremale hargnevatest kollokaatidest moodustavad ka substantiivid. Need moodustavad kõigi sõnade sagedamatest paremale hargnevatest kollokaatidest 30%. Otsisõnast paremale jäävad ja otsisõnaga süntaktilises seoses olevad substantiivid võivad esineda:

a) nimisõnafraasi põhisõnana, kui inimest tähistav sõna on genitiivatribuut, nt *ema **raamat***.

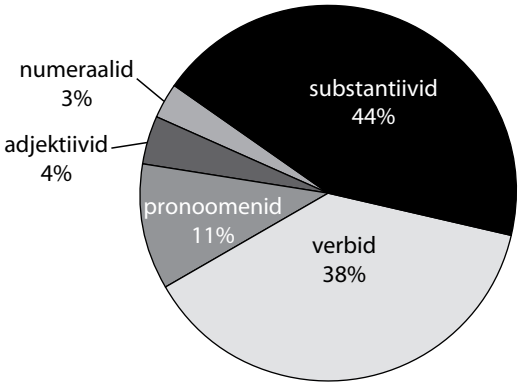
b) nimisõnafraasi põhisõnana, kui inimest tähistav sõna on adverbiaalatribuudi funktsioonis, nt *mehele **minek***.

c) nimisõnafraasis adverbiaalatribuudina, nt *mees **hobusega***.

d) apositsioonina, kui inimest tähistav sõna on eeslisandiks, nt *ema **Milvi***.

Adjektiivid moodustavad kõige väiksema osa paremale hargnevatest kollokaatidest ehk ainult 2%. Kuna paremale hargnevad kollokaadid on leitud vahemikust 3 sõna paremale, siis on nende hulgas kindlasti substantiive, mis ei esine otsisõnaga samas nimisõnafraasis, vaid on lauses hoopis iseseisvaks objektiks või määruseks. Sarnasel moel võib eeldada, et suur hulk parempoolseid adjektiivseid kollokaate ei esine otsisõnaga samas fraasis. Väike on tõenäosus, et parempoolne adjektiiv esineb otsisõna täiendina, sest küllalt harva ja sedagi peamiselt ilukirjandustekstides võib adjektiiv olla järeltäiendiks, nt *meri **ääretu*** (Tauli 1980: 159). Mõnel juhul võib siiski olla parempoolne adjektiiv koos otsisõnaga kaastäiendiks substantiivifraasis, nt *ema **uus** töökoht*. Siiski ei ole need sagedad juhud, mistõttu adjektiivid moodustavad paremale hargnevatest kollokaatidest ainult 2%. Võrdluseks: vasakule hargnevatest kollokaatidest on adjektiive 17%. See ongi suurim erinevus vasakule ja paremale hargnevate kollokaatide sõnaliigilises jaotuses.

Kui eemaldada 300 kollokaadi hulgast kokkulangevad ja analüüsida järele jäänud 85 kollokaadi sõnaliigilist kuuluvust, siis selgub, et substantiivide hulk koguarvust suureneb kõigi teiste sõnaliikide (v.a adjektiivide) vähenemise arvel (vt joonis 13). See näitab, et substantiivide varieeruvus on paremale hargnevate kollokaatide seas kõige suurem, samas kui numeraalidest, pronoomenitest ja isegi verbidest moodustavad kattuvad kollokaadid suurema osa.



Joonis 13. Paremale hargnevad 85 erinevat kollokaati sõnaliigiti

3.5. Vahekokkuvõte

Siinses vahekokkuvõttes antakse ülevaade, millised 3. peatüki sissejuhatuses püstitatud hüpoteesidest leidsid peatükis kinnitust. Lisaks põhjendatakse, miks lihtkollokatsioonide järgi on raske teha üldistavaid järeldusi selle kohta, milliseid konkreetseid tegevusi sooliselt ja vanuseliselt erinevatele inimestele ajakirjandustekstides omistatakse.

Küsimusele, kui suur osa inimese tähendusvälja sõnade vasakule ja paremale hargnevatest kollokaatidest kattusid, leiti vastus kõigi sõnade 30 sagedamat kollokaati võrreldes. Selgus, et kõigi uuritud inimest tähistavate sõnade 30 vasakpoolsest kollokaadist kattus keskmiselt 43–57% ja parempoolsetest koguni 50–60% teiste inimest tähistavate sõnade sagedamate kollokaatidega. Iga inimest tähistava sõna 30 sagedama vasakpoolse kollokaadi seas esinesid 8 sõna: *üks, kaks, kolm, mina, tema, see, olema* ja *oma*. Paremale hargnevatest kollokaatidest esinesid kõigi otsisõnade 30 sagedama kollokaadi seas 7 sõna: *mina, minema, olema, oma, saama, see* ja *tema.*

Kuigi inimest tähistavatel sõnadel esineb suur kollokatsioonikobarate kattumine, ei saa öelda, et kokkulangevad kollokaadid oleks iseloomulikud vaid inimese tähendusvälja sõnadele. Selleks oleks pidanud kaasama võrdlusesse piisaval hulgal inimese tähendusväljast välja jäävaid sõnu. Väike katse valimivälise substantiiviga *auto* näitas, et ka sõna *auto* 30 vasakpoolsest kollokaadist oli 23 selliseid, mis esinesid ühe või teise inimest tähistava sõna kollokaadina. Nii suur kokkulangevate kollokaatide hulk ei võimalda väita, et inimese tähendusvälja sõnu saaks ainult ühiste kollokaatide hulga järgi määratleda. Samas üksikute kollokaatide tasandile minnes on ka praeguses kollokaatide loendis sõnu (vt lisa 1 ja lisa 2), mis on omased vaid inimese tähendusvälja sõnadele, nt verbaalset akti tähistavad *ütlema* ja *rääkima*.

Kollokatsioonikobarate ja tähendusväljasiseste suhete võrdluses leiti, et iga tähenduskomponent eraldi ei omanud üldisele jaotusele mõju. Andmetest lähtuvalt ei saanud näiteks väita, et otsisõna tähenduskomponendile MEESSOOST antud sama väärtus oleks tinginud kollokatsioonikobarate suurema kattumise. Siiski avaldus tendents, et suurema kollokaatide kattuvusega otsisõnade tähenduses esines rohkem sama väärtusega tähenduskomponente. Peaaegu maksimaalselt avaldus see vasakule hargnevate kollokatsioonide puhul: kõigi suurima kollokatsioonikobarate kattumisega sõnade tähendused erinesid vaid ühele tähenduskomponendile antud väärtuse poolest ja kõigil väikseima kollokatsioonikobarate kattumisega otsisõnadel oli viiest komponendist vaid üks-kaks sama väärtusega komponenti või puudusid ühised tähenduskomponendid üldse. Leidis kinnitust hüpotees, et mida rohkem sarnase väärtusega tähenduskomponente kahel sõnal esineb, seda suurem on kollokatsioonikobarate kattumine.

Kui võrrelda inimese semantilise välja siseseid suhteid, siis tähendusvälja keskme moodustavad sõnad *mees* ja *naine*, mis omasid nii vasakule kui paremale hargnevate kollokaatide võrdluses kõige suuremat ühisosa teiste sõnade

kollokatsioonikobaratega. Seega võib öelda, et *mees* ja *naine* moodustavad inimese semantilise välja tuumiku. Võimalik, et see tuleb asjaolust, et nende sõnade tähendused kuuluvad nii üldisesse inimese tähendusvälja kui ka sugulus- või lähisuhete tähendusvälja, mistõttu ühisosa on suurem nii sugulust väljendavate sõnadega *ema* ja *isa* kui ka sugulust mitteväljendavate sõnadega *poiss* ja *tüdruk.*

Selgus, et lause süntaktiline struktuur mõjutab kollokaatide esinemist ja vasakpoolsete ning parempoolsete kollokaatide sõnaliigiline kuuluvus on erinev. See, et kollokatsioonisuhete esinemisvõimalus sõltub peale semantiliste tunnuste ka süntaksist, paljastus kollokaatide sõnaliigilisust uurides. Kõige suurem erinevus oli adjektiivide esinemissageduses – kui adjektiivid moodustasid vasakule hargnevatest kollokaatidest 17%, siis otsisõnast paremale jäi neid vaid 2%. Samuti suur erinevus oli verbide paiknemises: verbid moodustasid vasakpoolsetest kollokaatidest 24% ja paremale hargnevatest juba poole rohkem ehk 43%. Seega võib öelda, et kuigi kollokatsioonid on leksikaalsed süntagmaatilised suhted, sõltub nende esinemisvõimalus suurel määral ka keele grammatikast: süntaksist ja sõnajärje mallidest. Kinnitust leidis hüpotees, et piirkond, kust mingit tüüpi kollokatsiooniseoseid tasub otsida, sõltub keele grammatikast**.**

Paradigmaatilistest suhetest avaldusid kollokatsioonides peamiselt antonüümsed suhted. Kõik antonüümses suhtes sõnapaarid nagu *mees–naine, tütar–poeg, tüdruk–poiss, ema–isa*, mille osalised erinesid teineteisest komponendi MEESSOOST väärtuse poolest, omasid kõige rohkem ühiseid vasakpoolseid kollokaate. Sellega leidis kinnitust ka Sahlgreni väide, et sõnade süntagmaatilist konteksti, täpsemalt kollokatsioone võrreldes võib teha järeldusi paradigmaatiliste suhete (siinses töös peamiselt antonüümia) kohta. Kui antonüümiat võtta kui sarnasusseost – kahe sõna kõigist tähenduskomponentidest antakse ju vastandlik väärtus vaid ühele komponendile, siis võib nõustuda Harrisega, et sõnade sarnasus avaldub nende konteksti kokkulangevuses.

Selles peatükis võrreldi sagedamate inimest tähistavate sõnade kollokatsioonikobaraid eesmärgiga kirjeldada seoseid sõna distributsiooni ja tähenduse vahel. Algselt oli eesmärgiks uurida lisaks lähemalt üksikute kollokaatide esinemist erinevatel otsisõnadel, et kollokatsiooniseoste põhjal teha järeldusi selle kohta, kuidas erinevast soost ja erinevas vanuses inimesi meie kultuuris kujutatakse. Kollokaatidena esinevaid verbe lähemalt vaadeldes jõuti järeldusele, et süntaktilist struktuuri arvestamata on raske eristada verbe, mille suhtes on otsisõna subjektiks, neist verbidest, mille suhtes on otsisõna objektiks. Ilma lausekontekstita, ja eriti pärast kollokaatide lemmatiseerimist, ei saa kahjuks enam palju väita selle kohta, millised on kollokatsioonide süntaktilised seosed. Seetõttu peeti 4. peatükis vajalikuks kollokatsiooniseoseid vaadeldes võtta appi leksikogrammatilised mallid, mis annavad aimu, milline süntaktiline seos võiks otsisõna ja kollokaadi vahel esineda.

4. SAGEDAMAD INIMEST TÄHISTAVAD SÕNAD KAHES LEKSIKOGRAMMATILISES KONSTRUKTSIOONIS

4.1. Hüpoteesid

Selles peatükis uuritakse, kuidas kajastuvad kollokatsiooniseostes sõnadega *ema, inimene, isa, laps, mees, naine, poeg, poiss, tüdruk* ja *tütar* tähistatud isikute vanus ja sugu. On arvatud, et kollokatsioonimustrid paljastavad sotsiaalseid hoiakuid, eriti mis puudutab soo, vanuse ja käitumisega seotud stereotüüpe (Caldas-Coulthard, Moon 2010). Lähtutakse eeldusest, et sagedamate kollokatsiooniseoste põhjal saab dekonstrueerida peidetud tähendusi ja viise, kuidas inimesi on tekstides kujutatud.

Kui eelmises peatükis olid kollokatsioonid leitud vaid koosesinemissagedust arvestades, siis sel korral on koosesinemissagedus seotud kahe leksikogrammatilise konstruktsiooniga. Vaadeldakse täpsemalt inimest tähistavate sõnade adjektiivseid täiendeid ja verbe, mille juurde kuulub otsisõna subjektina. Tähelepanu all on kummaski konstruktsioonis esinevad 30 sagedamat kollokatsiooni, lisaks tuuakse välja vastanduvate paarikute 30 sagedamat eriomast verbi. Kollokatsiooniseoseid analüüsitakse peamiselt sõnapaaride kaupa kahe otsisõna kollokaate kõrvutades. Kui eelmises peatükis jõuti järeldusele, et antonüümipaaride kollokaadid kattuvad kõige enam, siis siinses peatükis vaadeldakse lähemalt, millised kollokaadid langevad neis paarikutes kokku ja millised on erinevad, ning mida see näitab selle kohta, kuidas erinevast soost ja vanusest inimesi kirjeldatakse. Seega tulevad vaatluse alla esimesena soovastandusel põhinevate antonüümipaaride *poiss–tüdruk, mees–naine*, *poeg–tütar, isa–ema* kollokaadid. Peamiselt on tegemist soovastandusel põhineva komplementaarse vastandusega, aga mõne sõnapaari tähenduste vahel esineb ka pöördvastandus, nt *mees–naine* ja *poiss–tüdruk* abikaasade või elukaaslaste tähenduses. Viimase paarina uuritakse sõnapaari *inimene–laps* kollokaate. Nende sõnade tähenduste vahel esineb nii vanuseline vastandus kui hüponüümiasuhe. Kontrollitakse kolme hüpoteesi.

1. hüpotees: **Inimest tähistavate sõnade kollokatsioonides avalduvad sugudevahelised erinevused.** Uurimisülesandeks on vaadelda, kas otsisõnaga tähistatud inimese sugu kajastub mingil moel kollokatsiooniseostes: kas on adjektiivseid täiendeid ja verbe, mida kasutatakse mingist soost inimeste kirjeldamisel sagedamini kui teisest soost inimeste kirjeldamisel.

Soost rääkides võetakse kasutusele mõisted bioloogiline ja sotsiaalne sugu, millest esimene tähistab anatoomilis-füsioloogilisi erinevusi meeste ja naiste vahel, aga teine kultuurilisi ehk nn õpitud erinevusi. Alates eelmise sajandi seitsmekümnendatest, kui ameerika antropoloog Gayle Rubin pidas vajalikuks eristada neid mõisteid eraldi terminitega *sex* ja *gender*, on sellist jaotamist nii pooldatud kui kritiseeritud.

Näiteks Judith Butler (1990) on seisukohal, et arusaam bioloogilisest soost on tihti kultuurist määratud. Mitmed uurijad on näidanud, et ka loodusteaduslikud väited põhinevad soolisustatud arusaamal, kuigi nägemus bioloogilise soo binaarsest jaotusest on ajaloo jooksul muutunud (Liljeström 2003: 118). Eeldatavasti peegeldub keeles eelkõige sotsiaalne sugu, sest keel on siinkirjutaja arvates üks sotsiaalse soo looja: sugu moodustavate soorollide ja soostereotüüpide kinnistaja ja taastootja. Töös tuuakse eraldi välja, kui sõnade kollokaadid on otseselt tingitud otsisõnaga tähistatud isiku bioloogilisest soost: näiteks reproduktiivseid funktsioone tähistavad verbid *sünnitama, viljastuma, viljastama, rasestama* või fertiilsusega seotud omadussõnad *rase, (last) ootav, imetav*.

Sotsiaalse soo olemust ongi määratletud stereotüüpide kogumina ehk uskumustena, mis inimestel grupi ja selle liikmete kohta on. Soostereotüübid avalduvad erinevates eluvaldkondades: iseloomujoontes, käitumises, ametivalikus, hobides, välimuses, perekonna funktsioneerimises ja kõikvõimalikes eelistustes (Huntoon 2009: 378).

Kuigi naised leiavad tänapäeval peaaegu võrdselt meestega rakendust väljaspool kodu, on visad muutuma Euroopas tööstusrevolutsiooni järgselt tekkinud stereotüübid, mis näevad meest perekonna rahateenijana ja naist koduste toimetuste eest vastutavana (Huntoon 2009: 379). Tegelikult on naiste palgatööle siirdumisega stereotüüpide kõrval vähe muutunud ka tegelik tööjaotus: Eesti ühiskonnas kehtib tööjaotus, kus naised teevad võrreldes meestega kaks korda rohkem kodutöid (Talves 2011: 103). 2009. aastal läbiviidud soolise võrdõiguslikkuse monitooringu andmetel on naised peamised vastutajad pesu ja riiete korrashoiu, söögitegemise, nõudepesu, tubade koristamise jms eest. Meeste õlul on aga erakorralisemad tööd nagu remonditööd ja auto sõidukorras hoidmine (Vainu jt 2010: 123). Naiste suuremat sidumist peresfääriga väljendab eesti keeles kõnetlussõnade temaatika: kui *härra* ei näita perekonnaseisu, siis kõnetlussõnade *preili* ja *proua* valiku võimalusega peetakse oluliseks määrata, kas kõnetatu on abielus või vallaline. Sellise ebavõrdse olukorra lahendamiseks võeti eelmise sajandi keskel ingliskeelsetes maades *Mrs* 'proua' ja *Miss* 'preili' kõrval kasutusele kõnetlussõna *Ms*, mis ei vaja abielulisuse määramist (Pauwels 2003).

Inimest tähistavate sõnade kollokaatidena esinevaid verbe analüüsitakse, et näha, kas naissoost isikuid tähistavad sõnad esinevad rohkem privaatsfääriga seotud tegevustes kui meessoost isikuid tähistavad sõnad. Sõnapaari *töötav ema* kasutus on George Lakoffi arvates seotud stereotüüpidega, mis meil on emast ja tööst. Ta väidab (Lakoff 1990: 81), et stereotüüpne ema on kodune, selle tõestuseks toob ta kaks lauset, mida laiendavad osalaused on seotud sidesõnaga *but* 'kuid', mis inglise keeles märgib tema sõnul normile vastandumist. Seega peaks need laused, kus vastandusena mõeldud lauseosa tegelikult sisaldub stereotüüpses ema mõistes, kõlama veidralt.

(8) *She is a mother, but she isn't a housewife.* 'Ta on ema, kuid ta pole koduperenaine'

(9) *She is a mother, but she is a housewife.* 'Ta on ema, kuid ta on koduperenaine'

Lakoffi arvates kõlab teine lause veidralt. Siinkirjutajal on raske hinnata ilma suuremate uuringuteta eestikeelsete lausete normaalust, aga ainult oma keeletunde põhjal öeldes kõlab teine lause ka eesti keeles veidramalt. Veidramalt kõlav vastandus peaks olema tõestuseks, et stereotüüpne ema on pigem kodune. Siiski peab ütlema, et Eesti kultuurikontekstis on kodused eelkõige värsked väikelaste emad. Seega oleks kontroll-lausetena kasutatavad laused:

(10) Ta sai emaks, kuid jäi koduseks.

(11) Ta sai emaks, kuid ei jäänud koduseks.

Neist lausetest on esimene veider, samas kui teine kõlab täiesti normaalsena. Sellest peaks järelduma, et stereotüüpne värske ema on kodune.

Kuna siinses töös kasutatavast korpusest moodustavad suurima osa ajakirjandustekstid, siis kajastavad tulemused peamiselt just seda, kuidas meedias inimest kujutatakse. Barbi Pilvre arvates on meedial oluline roll stereotüüpide tekkimises. Tema sõnul kiputakse naisi meedias kujutama privaatsfääriga seotud stereotüüpsetes rollides, samuti on olulisel kohal naise keha kirjeldamine nii naiste portreelugudes Eesti Ekspressi Persooni rubriigis kui mujal naistest kirjutavas meedias (Pilvre 2011a: 182). Meediale on üldse ette heidetud, et teemade käsitlus on muutunud isiklikumaks (Nuolijärvi, Tiittula 2000: 19).

Naisi on stereotüpiseeritud ka kui seksuaalobjekte, kes pööravad suurt tähelepanu oma välimusele (Huntoon 2009: 379). Inglise keele kohta on tehtud uuringuid, mille kohaselt sõnade võime omada erinevaid adjektiivseid kollokaate võib olla väga varieeruv. Näiteks on tähele pandud, et hulk adjektiive nagu *important* 'tähtis', *famous* 'kuulus' ja *rich* 'rikas' esinevad inglise keeles palju sagedamini seoses mehega, samas kui naist kirjeldatakse adjektiividega *busy* 'toimekas', *beautiful* 'ilus', *pretty* 'kena' (Mills 2002: 210). Selles peatükis otsitakse vastust küsimusele, kas eesti keeles kirjeldatakse naiste puhul samuti rohkem välimust ja meeste puhul sotsiaalset staatust ning kas on verbe, mis vihjavad grammatilise subjektiga väljendatud isikule kui seksuaalobjektile.

Janet Hyde'i (2005) sõnul on soostereotüüpse nägemuse järgi naised emotsionaalsemad ja verbaalselt võimekamad kui mehed, kuigi tema arvates puuduvad väidete kinnituseks valiidsed empiirilised tõendid. Hyde on oma uurimuses näidanud, et naiste ja meeste erinevused verbaalsetes võimetes ja käitumises on väikesed. Et näha, kuidas võimalikud erinevused peegelduvad kollokatsioonides, vaadeldakse lähemalt emotsioone ja verbaalset võimekust tähistavaid verbe.

2. hüpotees: **Inimest tähistavate sõnade kollokatsioonid peegeldavad võimusuhteid ühiskonnas.** Kui eeldada, et võim ühiskonnas ei jaotu võrdselt, nt Majandusliku Koostöö ja Arengu Organisatsiooni (OECD) aruande kohaselt kuulub maailmas peamine majanduslik ja poliitiline võim meestele (Jütting jt 2006), siis peaks see kuidagi kajastuma ka keeles. Eestis on samuti naiste töötasu madalam kui meestel ning palgavahe on Euroopa Liidu Statistikaameti andmetel[1]

[13] Eurostat ehk Euroopa Liidu Statistikaamet http://epp.eurostat.ec.europa.eu/portal/page/portal/statistics/themes

Euroopa Liidu suurim. Paremat juurdepääsu majanduslikele hüvedele peaksid väljendama rahalist seisu näitavad adjektiivid ja rahakasutust tähistavad verbid.

Meeste kesksele rollile ühiskonnas võib tõendeid leida ka lihtsalt meessoole viitavate sõnade suuremas esindatuses või asjaolus, et mehe kohta käivat sõna kasutatakse inimese tähistamiseks (nt mitmes indoeuroopa keeles on kasutusel sama sõna nii mehe kui inimese tähistamiseks). Piitsa (2004) ja Puna (2006) uurimused on näidanud, et eesti keeles on küll sooülese inimese tähistamiseks oma sõna, aga samas on sõna *mees* võimalik kasutada ka geneeriliselt mõlemast soost isiku kohta, eriti *mees*-lõpulistes liitsõnades. Sõna *mees* kasutus geneeriliselt viitab androtsentrismile, mis tähendab, et tüüpilise inimesena nähakse meest. Soome keeles on tehtud uurimus, kus katseisikutel paluti isikunimega laiendada fraase *tyypillinen suomalainen* 'tüüpiline soomlane', *tyypillinen peruskoululainen* 'tüüpiline põhikooli õpilane' *tyypillinen naapuri* 'tüüpiline naaber' jne. Tulemusi analüüsides vaadeldi, kas katseisikud olid rohkem kasutanud meessoole viitavaid isikunimesid või naissoole viitavaid. Naiskatseisikud eelistasid märkimisväärselt rohkem mehenimesid ainult kahel juhul 13-st (tüüpiline soomlane ja tüüpiline tv-vaataja olid mehed) ja naisenimesid eelistati märkimisväärselt rohkem samuti kahel juhul (tüüpiline kristlane ja tüüpiline patsient määrati naiseks). Meeskatseisikud valisid mehenimesid naisenimedest märkimisväärselt rohkem koguni üheksal juhul 13-st ja naisenimesid ei nimetatud märkimisväärselt rohkem ühelgi juhul (Engelberg 2002: 120–121). Seega nähakse just peamiselt soome meeste hulgas tüüpilise inimesena meest.

Hüpoteesist lähtuvalt vaadeldakse, kas meessoost isikule viitavaid otsisõnu kasutatakse korpuses rohkem kui nende soovastandusel põhinevaid antonüüme ja kas naistele viitavad sõnad esinevad lauses suurema tõenäosusega objektidena kui meestele viitavad sõnad.

Vägivaldsust ja kuritegelikku käitumist on samuti tõlgendatud kui võimu näitamist. Pearce'i inglise keele põhjal tehtud analüüs näitab, et British National Corpuse korpuses kollokeeruvad kriminalistika ja väärtegudega seostatavad sõnad tugevamalt sõnaga *man* 'mees' kui sõnaga *woman* 'naine' (Pearce 2008: 9). Selline jaotus on ka loogiline, kuna valdav enamik kuritegude toimepanijatest on mehed. Eestis moodustavad kuriteos kahtlustatavatest mehed 88% (Lindsalu 2012: 8) ja vangidest 95% (Ahven, Kruusmaa 2013: 103). Verbe analüüsides otsitakse vastust küsimusele, kas sõnad *mees, poiss, isa* ja *poeg* on sagedamini kui nende soovastandusel põhinevad antonüümid subjektiks tegevuste juures, mis tähistavad võimupositsioonilt sooritatud tegevusi (nt erinevad vägivallateod).

3. hüpotees: **Inimest tähistavate sõnade kollokatsioonides avaldub otsisõnaga tähistatud tüüpilise isiku vanus.**

Selle hüpoteesi kontrollimiseks võrreldakse peamiselt alaealisi tähistavate sõnade *laps*, *poiss* ja *tüdruk* adjektiivseid täiendeid ja verbe, mille suhtes need sõnad on subjektiks, nende sõnade kollokatsioonidega, mille tähenduses on tähenduskomponent TÄISKASVANUD saanud positiivse väärtuse.

4.2. Uurimismeetod ja analüüsi aluseks olev materjal

4.2.1. Leksikogrammatiliste konstruktsioonide lühitutvustus

Lisaks kollokatsioonide leidmise puht statistilistele meetoditele, mis arvestavad ainult koosesinemissagedust või koosesinemissageduse ja eraldiesinemissageduse suhteid, on loomulike keelte automaattöötluses kasutatud kombinatoorseid meetodeid, mis võtavad arvesse ka keele grammatikat. Niisugused lingvistilist teavet kasutavad meetodid vajavad morfoloogiliselt või süntaktiliselt märgendatud korpust. Näiteks Padó ja Lapata (2007) on kasutanud süntaktiliselt märgendatud korpust, et tuua välja sõnade vahel esinevaid sõltuvussuhteid. Leksikogrammatilisi malle või konstruktsioone on kasutatud ka hüponüümiasuhete automaatseks tuvastamiseks. See eeldas süntaktiliselt märgendatud korpust ja mallide koostamist, kus süntaksimärgendeid kombineeriti fraasidega *or other* 'või teised', *and ohter* 'ja teised', *including* 'kaasa arvatud' jpt, millest eeldati, et need võiks esineda koos hüponüümiasuhetes sõnadega (Hearst 1992: 541).

Keele automaattöötluse kõrval on hakatud kombinatoorseid meetodeid hindama ka leksikograafias. 2002. aastal esitleti (Kilgarriff, Rundell 2002) uut leksikograafidele mõeldud tarkvara Sketch Engine'it, kus kollokatsioonide automaattuvastamisel võeti aluseks morfoloogilise analüsaatori tulemused (sõnaliik ja grammatilised tunnused).

Siinses peatükis kasutataksegi Sketch Engine'i tarkvara ja Jelena Kallase loodud arvutigrammatikat (versioon 1.5, põhjalikku ülevaadet loodud reeglitest vt Kallas 2013), mille põhjal otsib programm grammatilistes suhetes olevaid sõnapaare ja reastab need koosesinemise sagedust (*raw frequency*) arvestades. Materjali on otsitud kolme leksikogrammatilise konstruktsiooni kohta:

1. *adjektiivne laiend* + *substantiiv*. See konstruktsioon liigitub Kallasel substantiivide leksikogrammatiliste konstruktsioonide hulka ja leitakse reegliga, mis Sketch Engine'i eesti moodulis kannab nime *Adj_modifier* (vt Kallas 2013: 36). Adjektiivsete laiendite kasutus peaks kõige otsesemalt näitama, kuidas inimest keeles kirjeldatakse.

2. *substantiiv subjektina* + *verb* liigitub verbi leksikogrammatiliste konstruktsioonide alla ja leitakse reegliga, mis kannab nime *subject_of* (vt Kallas 2013: 64). Sagedamad verbid, mille suhtes inimest tähistav sõna on subjektiks, näitavad seda, milliste tegevustega seoses inimest tähistavaid sõnu tüüpiliselt keeles kasutatakse.

3. *substantiiv objektina* + *verb* liigitub verbi leksikogrammatiliste konstruktsioonide alla ja leitakse reegliga, mis kannab nime *object_of* (vt Kallas 2013: 65). Sagedamad verbid, mille suhtes inimest tähistav sõna on objektiks, näitavad seda, millistes tegevustes inimest kujutatakse passiivse kogejana, objektina, kelle suhtes mingit tegevust sooritatakse.

4.2.2. Inimest tähistavate sõnade esindatus eesti keele koondkorpuses

Sketch Engine'i eesti mooduli sisendiks on u 203 miljoni sõna suurune Tartu ülikooli eesti keele koondkorpus, millest suurima osa moodustavad ajakirjandustekstid, aga kus on ka ilukirjandustekste, seadustekste, jututubade ja foorumite tekste ning teadusartikleid. Ilukirjandustekstid pärinevad alates 1990. aastatest ja ajakirjandustekstid perioodist 1995–2008. Teadusartiklid on pärit ajavahemikust 1997–2006. Uue meedia tekstid on internetist salvestatud aastatel 2003–2008. Üksikasjalikum korpuse kirjeldus asub eesti keele koondkorpuse kodulehel[2].

Tabelis 5 on esitatud erinevate inimest tähistavate sõnade esinemissagedused eesti keele koondkorpuses.

Tabel 5. Sagedamate inimest tähistavate sõnade esinemine eesti keele koondkorpuses.

otsisõna	sõna esinemissagedus	sõna esinemine arvus		
		ainsuses	mitmuses	koos numeraaliga
inimene	522436	30%	51%	19%
mees	307624	57%	34%	9%
laps	245399	34%	56%	10%
naine	179449	52%	41%	7%
ema	68944	92%	7%	1%
poiss	59315	50%	43%	7%
isa	57094	95%	4%	1%
tüdruk	44418	54%	38%	8%
poeg	40834	81%	13%	6%
tütar	28061	85%	9%	6%

Tabelist näeme, et otsisõnade esinemissagedus selles suures korpuses on võrreldav 3. peatükis analüüsitud väikese tasakaalus korpuse ajalehtede allkorpusest saadud andmetega (vt 3.2.1). Sarnaselt on kõige sagedamaks sõnaks *inimene*, samas järjekorras järgnevad *mees, laps, naine* ja *ema*. Ainult *poiss* ja *isa* ning *poeg* ja *tüdruk* on siinses tabelis kohad vahetanud. Samas jääb kehtima seaduspära, et meessoost isikule viitavad sõnad esinevad oma naissoost antonüümidega võrreldes sagedamini (v.a antonüümipaar *ema* ja *isa*). Sarnast tendentsi on märgatud ka muude keelte puhul, nt inglise keele kohta tehtud uuringutes on leitud, et lemma *man* 'mees' esineb 1,5 korda sagedamini kui lemma *woman* 'naine' (Pearce 2008: 2).

[14] Eesti keele koondkorpus kirjeldus http://www.cl.ut.ee/korpused/segakorpus (14.07.2014)

Eesti keeles on vahekord sarnane: koondkorpuses esineb lemma *mees* 1,7 korda sagedamini kui *naine*.

Lisaks absoluutse esinemissageduse erinevustele, võib oluliseks osutuda ka erinevas grammatilises arvus esinemise sagedus. Näiteks Pearce'i (2008) artiklist selgub, et kui sõna *mees* on rohkem kasutusel ainsuses, siis naistest räägitakse rohkem mitmuses: sõna *mees* ainsusevorm *man* on üle 1,5 korra sagedam kui mitmusevorm *men*, aga sõna *naine* mitmusevorm *women* on kaks korda sagedam kui ainsusevorm *woman.* Võimalusele, et *naine* esineb neutraalse tegevussubjektina pigem mitmuses kui ainsuses, on eesti keele puhul varem tähelepanu juhtinud Kristiina Ross, kes „Eesti keele grammatika" näitelausestikku analüüsides leidis, et kui neutraalse aktiivse tegutsejana nimetada naist, siis tundus mitmuse vorm kohasem, just nagu saaksid naised aktiivse tegevussubjektina esineda ainult koos tegutsedes ning üksik naine ei suudakski olla midagi muud kui meheliku teadvuse objekt (Ross 1996: 106). Et teada saada, kas nägemus naisest kui kollektiivsest tegutsejast tuleb esile ajakirjanduskeeles, selleks loeti kokku kõigi otsisõnade ainsus- ja mitmusvormid. Kokkulugemisel arvestati, et eesti keeles võib ainsusvorm koos arvsõnaga viidata ka mitmele tegutsejale, näiteks *tuhat inimest, kolme mehega* või *sajale mehele*. Inglise keeles kasutatakse ühest suuremate arvsõnade juures mitmusvormi. Seetõttu peeti vajalikuks eraldada kõik konkordantsiread, kus otsisõnale eelnes numeraal (regulaaravaldis sõna *inimene* ainsusvormide leidmiseks nägi välja näiteks selline [tag!="O"&tag!="N"] [lemma="inimene"& features="sg_.*"]). Seega on numeraale sisaldavad juhud määratud kui tuvastamata juhud. Need moodustasid sõnade *ema* ja *isa* esinemisjuhtudest 1%, aga sõna *inimene* esinemisjuhtudest 20% (vt tabel 5).

Eesti keele koondkorpuse andmetest selgus, et erinevalt inglise keelest kasutatakse eesti keeles lemmat *naine* sagedamini ainsuses (52%) kui mitmuses (41%). Lemma *mees* ainsusvorme kasutatakse küll veel sagedamini: üle 57% vormidest esines ainsuses ja vaid 34% mitmuses, aga suhe on inglise keelega võrreldes palju tasakaalustatum ning ei saa väita, nagu eesti ajakirjandustekstides kirjeldataks naisi kollektiivselt tegutsevatena, aga mehi individuaalsete isiksustena. Teisi soovastandusel põhinevaid paare võrreldes ei esinenud samuti reeglipära, et mehele viitavat sõna kasutataks rohkem ainsuses: nt sõnade *tüdruk* ja *tütar* ainsusvormide osakaal on suurem kui nende meessoole viitavatel antonüümidel. Soolise markeeritusega seoses ilmnes siiski üks eripära: mitmusvorme kasutatakse tunduvalt rohkem vaid sõnade *inimene* ja *laps* puhul, mis ei ole sooliselt markeeritud.

Samuti selgus, et väga harv on mitmusvormide kasutus sõnade puhul, mille kõige sagedam ja olulisem tähendus viitab sugulussidemele: nt sõnu *poeg, tütar, ema* ja *isa* kasutatakse 86–96 protsendil juhtudest ainsuses. Ainsusvormide ülekaal tuleneb ilmselt asjaolust, et inimestel on tavaliselt üks isa ja üks ema ning ka mitme samast soost järeltulija esinemistõenäosus on tänapäeva väikestes perekondades väiksem.

Erinevused inimest tähistavate sõnade esinemises grammatilises arvus võivad mõjutada teatud kollokaatide esinemist koos selle sõnaga (vt kollokaatide *ütlema*

ja *rääkima* esinemisest 4.4.1). Sellele on töös mõnel juhul tähelepanu pööratud, siiski pole kollokatsiooniseoseid otsitud ainsus- ja mitmusvormide kohta eraldi. Esitatud on iga otsisõna erinevate vormide kollokatsiooniseosed üheskoos.

4.2.3. Antonüümide esinemine süntagmaatilise tasandi konstruktsioonides

Enne kui asutakse võrdlema antonüümipaaride kollokaate, vaadeldakse, kas inimest tähistavad antonüümipaarid moodustavad ise süntagmaatilise tasandi konstruktsioone. Lähtutakse Lynne Murphy ideest (vt lähemalt 2.2.1.1), et antonüümia avaldub alati ka süntagmaatilisel tasandil ning antonüümipaarikud moodustavad konstruktsioone. Vaadeldakse täpsemalt antonüümipaarikute järjekorda konstruktsioonis *X ja Y*.

Sõnade *mees, naine, poiss, tüdruk, poeg, tütar, ema* ja *isa* 30 sagedama vasakule hargneva kollokaadi hulgas esines kõigil juhtudel sõna komplementaarsel vastandusel põhinev antonüüm (vt lisa 1). Samuti olid sõnade *ema* ja *isa* sagedamate kollokaatide seas esindatud pöördvastanduses olevad sõnad *laps*, *poeg*, *tütar* ning sõnade *poeg* ja *tütar* kollokaatide seas omakorda *ema* ja *isa*. Ainult sõna *laps* kollokaatide seas oli pöördvastanditest esindatud vaid *ema* ja puudus *isa*.

Täpsemalt on siinses töös andmeid kogutud konstruktsiooni *X ja Y* kohta. Asta Õimu (1995) sõnul seatakse sootunnusel rajanevates antonüümipaarides tavaliselt esikohale isast märkiv liige: *isa–ema; poeg–tütar; mees–naine; kukk–kana*, v.a opositsiooni *matriarhaat–patriarhaat* järjestus, mis on Õimu arvates tingitud ajalistest teguritest. Lisaks võib tema arvates järjestuse määrata ka harjumuspärane kasutus, nt sõnade *õde–vend* puhul, ja oma rolli mängib antonüümipaariliste järjestamisel ka subjektiivsuse moment. Õimu väite uurimiseks leiti 203 miljoni sõna suurusest eesti keele koondkorpusest inimest tähistavate sõnade komplementaarsel vastandusel põhinevate paaride *X ja Y* konstruktsioonid, et selgitada kumb antonüümidest on esikohal. Tulemused kajastavad lemmade esinemist neis konstruktsioonides, st esindatud pole mitte ainult nominatiivsed sõnavormid *mees ja naine*, vaid ka *mehed ja naised*, *meestele ja naistele* jne.

Tabel 6. Soolisel vastandusel põhinevate inimest tähistavate sõnade *X* ja *Y* konstruktsioonide sagedusnäitajad

Meessugu märkiv sõna esimesena	Sagedus	Naissugu märkiv sõna esimesena	Sagedus	Suhe meessugu esimesena/ naissugu esimesena
mees ja naine	4425	naine ja mees	1521	2,9/1
poeg ja tütar	505	tütar ja poeg	215	2,3/1
poiss ja tüdruk	897	tüdruk ja poiss	270	3,3/1
isa ja ema	725	ema ja isa	1430	1/1,9
vend ja õde	329	õde ja vend	490	1/1,4

Võrreldes konstruktsioonis *X ja Y* esinevate antonüümide järjestust, selguski (vt tabel 6), et meessugu märkiv sõna esines rohkem esikohal: nt *mees ja naine* esines 4425 korda, aga *naine ja mees* ainult 1521 korda. Sama seaduspära avaldus ka konstruktsioonide *poeg ja tütar* (505) ja *tütar ja poeg* (215 korda) ning *poiss ja tüdruk* (897) ja *tüdruk ja poiss* (270) suhetes. Teistest eristus ainult antonüümipaar *ema–isa*, kus poole sagedamini (1430 vs 725) esines konstruktsioonis esikohal naissugu märkiv sõna. Lisaks kontrolliti väidet, et harjumuspärane järjestus opositsioonipaaris *õde–vend* seab esikohale naissoost liikme. Konstruktsioonis *X ja Y* osutus see väide tõeseks, kuna *õde* esines esikohal 1,4 korda sagedamini kui vend.

Kuidas seletada konstruktsioonides *ema ja isa* ning *õde ja vend* sellist sõnade järjestust? Üldistatult võib kõik järjestusparameetrid (ka vastanduse meessoost–naissoost) viia tasandile, kus vastanduvad markeerimata ja markeeritud liige. Antonüümipaari esimeseks sõnaks valitakse tavaliselt semantiliselt ja morfoloogiliselt markeerimata positiivne liige. Semantiline markeeritus tuleneb pragmaatilistest asjaoludest, selles kajastub meie teadmine maailma tegelikest suhetest (Õim 1995). Tavaliselt on markeerimata liikmeks meessugu märkiv sõna. Lisaks on William Cooper ja John Ross esitanud erinevaid semantilisi ja fonoloogilisi hierarhiaid, millest peaks sõltuma paarissõnade järjekord. Eelpoolmainitud järjestuse meessoost–naissoost, positiivne–negatiivne kõrval ka täiskasvanu–alaealine, elus–elutu, ainsus–mitmus, ees–taga, agent–patsient jne (Cooper, Ross 1975: 65–66). Siiski ei suuda ka need järgnevushierarhiad põhjendada, miks *ema* on sagedamini esikohal kui *isa*. Cooper ja Ross esitavad lisaks ka mitmeid fonoloogilisi tunnustehierarhiaid, nt teisele kohale satub sõna, millel on rohkem silpe, silbil pikem heliline tuum, rohkem alguskonsonante, vähem lõpukonsonante. Sõnad *ema* ja *isa* ei erine silpide arvu ja konsonantide esindatuse poolest, mistõttu fonoloogiliste tunnuste järgi ei peaks sõnapaaril *ema–isa* sellist järjestust tekkima. Järjestuse fonoloogilisi põhjuseid pole mõtet ilmselt otsida ka seetõttu, et Cooperi ja Rossi andmetel on teistsuguse fonoloogilise struktuuriga ingliskeelne järjestus samuti *mother and father*, mitte vastupidi. Nad ei osanud sellist järjestust muul moel seletada, kui väites: ”We believe that mothers are special” ‘Me usume, et emad on erilised’ (Cooper, Ross 1975: 105). Jääb selgusetuks, mida selle lausega on tahetud öelda. Ilmselt tuleb järjestuse põhjust otsida siiski semantiliste faktorite seast, sest on leitud, et need kaaluvad üles fonoloogilised põhjused (Cooper, Ross 1975: 103). Arvatavasti nähakse ema rolli laste kasvatamisel olulisemana. Seega on tegemist järjestusega, kus tähtsam eelneb vähemtähtsamale. Sõnaraamatumärksõnu analüüsides on Kerli Puna (2006: 34) jõudnud samuti järeldusele, et ema rolli laste kasvatamisel on peetud ühiskonnas tähtsamaks kui isa rolli. Põhjendusena toob ta välja, et emadust märkivaid sõnu (*ema, eit* jne) on kõikides sõnaraamatutes rohkem kui isadust (*isa, taat* jne) märkivaid sõnu.

Cooperi ja Rossi (1975) üks järjestushierarhiatest ütleb, et täiskasvanu tuleb sõnapaarides enne alaealist. Selline järjestuspõhimõte kehtis kõigi pöördvastanduses olevate inimest tähistavate sõnapaaride puhul (vt tabel 7).

Tabel 7. Pöördvastandusel põhinevate inimest tähistavate sõnade X ja Y konstruktsioonide sagedusnäitajad

Täiskasvanu esikomponendina	Sagedus	Alaealine esikomponendina	Sagedus	Suhe täiskasvanu esimesena/ alaealine esimesena
ema ja laps	438	laps ja ema	103	4,3/1
ema ja poeg	203	poeg ja ema	25	8,1/1
ema ja tütar	317	tütar ja ema	21	15/1
isa ja laps	75	laps ja isa	14	5,3/1
isa ja poeg	734	poeg ja isa	17	43/1
isa ja tütar	129	tütar ja isa	20	6,5/1

Sõna *ema* ja sõna *isa* olid konstruktsioonides *X ja Y* mitmeid kordi sagedamini esikohal kui sõnad *laps, poeg* ja *tütar*. Tabeli viimases veerus on esitatud konstruktsioonide *X ja Y* ning *Y ja X* ligikaudne suhe, mis näitab mitu konstruktsiooni, kus täiskasvanu esikohal, tuleb iga konstruktsiooni kohta, kus alaealine esikohal. Eriti hakkas silma konstruktsiooni *isa ja poeg* sage kasutus, mis pole ainult 43 korda sagedam kui *poeg ja isa*, vaid edestab mitmekordselt kõiki *ema ja Y* konstruktsioone. Teiste pöördvastanduses sõnadega esineb *isa* koos palju harvem. Tekkis kahtlus, et koosesinemissageduse suur hulk võib olla seotud levinud ärinimetustega *isad ja pojad*, aga pooltes läbivaadatud lausetes oli ärile või aktsiaseltsile viidatud vaid paaril korral. Hoopis suurema hulga (ligi 16%) moodustasid erinevad kunstiteoste nimed, nt Ivan Turgenevi raamat „Isad ja Pojad" või Jaan Toomiku videoinstallatsioon „Isa ja Poeg". Sellele vaatamata jääb konstruktsiooni *isa ja poeg* esinemissagedus sellisele tasemele, mis lubab aimata isade ja poegade vahelise suhte olulisust. Ja muidugi on see ka parimaks tõestuseks Cooperi ja Rossi järgnevushierarhiale, mis ütleb, et täiskasvanum on esikohal.

4.3. Adjektiivsed laiendid

Siinses peatükis on vaatluse all inimest tähistavate sõnade adjektiivsed laiendid. Kui adjektiivatribuudiks on eesti keele akadeemilises grammatikas (EKG II 1993: 115) määratletud omadussõna kõrval ka järgarvsõnu, põhiarvsõnade mitmusevorme, kaassõnafraase, määrsõnu, omadus- ja arvsõnalisi asesõnu jms, siis Kallase loodud sõnavisandite grammatikas on adjektiivseteks laienditeks märgendatud vaid need nimisõna juurde kuuluvad sõnad, mis morfoloogilise analüsaatori Estmorfiga on saanud adjektiivi märgendi. Näiteks sõna *poiss* adjektiivseks laiendiks loetakse ainult sõna *oma* need juhud, kui see on adjektiiviks märgendatud (12), aga välja on jäänud *oma* asesõnaline kasutus (13).

(12) Me oleme siin kõik **omad** /A/pl_n poisid. (Kustumatu valguse maailm)
(13) Nad ei tea praeguseni, et mul on **oma** /P/sg_g poiss. (Postimees Extra 1996.06.22)

Samuti on erinevalt liigitatud *nud*-partitsiipe ja *v*-partitsiipe. Kui peatükis 3.3 leiti, et *kadunud* kuulus sõna *poeg* 10 sagedama vasakule hargneva kollokaadi hulka (vt lisa 1, tabel L. 7), siis adjektiivsete laiendite loendis sõna ei esinenud, sest *kadunud* on märgendatud verbiks. Samas *v*-partitsiibid (näiteks *õppiv, töötav, käiv*) on esitatud adjektiivsete laienditena, kuna Estmorf on need märgendanud omadussõnaks.

Adjektiivsete laiendite loend ei sisalda keskvõrdes, ülivõrdes ega käändumatuid adjektiive, kuna Estmorfi märgendusest tingituna on Kallase loodud sõnavisandite grammatikas nende tuvastamiseks eraldi reeglid (Kallas 2013: 37).

Seega peab andmete analüüsil arvestama Estmorfi märgendussüsteemi ja sõnavisandite grammatika reeglite omapäraga, sest sarnases süntaktilises funktsioonis sõnad võivad paigutuda erinevate leksikogrammatiliste konstruktsioonide alla. Tulemusi võivad mõjutada reeglites kirjas olevad kontekstipiirangud. Näiteks on asesõnaliste laiendite reeglis kirjas keeld otsida personaalpronoomeneid, mistõttu asesõnaliste laiendite seas ei esine vorme *minu*, *tema* jne. Need on aga pronoomenid, mis sõnade *mees, naine, poiss* ja *tüdruk* kollokaatidena võiksid olla abiks sõnatähenduste ühestamisel, sest nagu selgus peatükist 3.3.2, kasutatakse genitiivis personaalpronoomenit peamiselt vaid neil juhtudel, kui sõna märgib lähisuhte osalist. Niisiis peab arvestama, et see, mida leksikograafilises kirjelduses on peetud ebaoluliseks ja mis on sõnavisandite grammatika reeglitega kõrvaldatud, võib laiemast semantilisest vaatepunktist ja sõnatähenduste ühestamisel osutuda oluliseks.

Siinses peatükis on vaatluse all algvõrdes ja käänduvad adjektiivid eestäiendina: kõik adjektiivid on leitud otsisõnast kuni kahe sõna kauguselt vasakult. Kallase moodustatud reegel tugineb sõnaliigimärgenditele ja määrab adjektiivseks laiendiks kokkuvõtlikult öeldes kõigis käänetes algvõrdes adjektiivid, mis ühilduvad järgneva substantiiviga käändes ja arvus. Lisaks keelatakse grammatikas Estmorfiga adjektiiviks märgendatud sõnavormi *olnud* lugemine adjektiivsete laiendite hulka.

Tabel 8. Sagedamate inimest tähistavate sõnade adjektiivsete laiendite esinemine eesti keele koondkorpuses

otsisõna	otsisõna esinemissagedus	adjektiivsete laiendite absoluutarv ja protsent otsisõna esinemissagedusest
inimene	522436	73346 (14%)
mees	307624	43922 (14%)
laps	245399	22965 (9%)
naine	179449	23997 (13%)
ema	68944	4636 (7%)
poiss	59315	10943 (18%)
isa	57094	2914 (5%)
tüdruk	44418	8541 (19%)
poeg	40834	4509 (11%)
tütar	28061	3784 (13%)

Tabelist on näha, kui paljudel juhtudel esineb sõna koos adjektiivse laiendiga. Selgub, et kõige harvem laiendatakse adjektiiviga sõnu *ema* ja *isa* (7–8%). Kolm kuni neli korda sagedamini aga kasutatakse adjektiivseid laiendeid sõnade *poiss* ja *tüdruk* kirjeldamisel.

Järgnevalt analüüsitakse adjektiivseid laiendeid otsisõnapaaride kaupa, tuues välja nii paarile iseloomulikke kui paarikutele eripäraseid adjektiive. Võrreldakse sõnapaaride *poiss–tüdruk, mees–naine, poeg–tütar, ema–isa* ja *inimene–laps* 30 sagedamat adjektiivset laiendit. Antonüümipaaride kollokaate on ka varasemalt võrreldud, tuues välja nii sõnaga tugevalt kollokeeruvaid ehk tõmbuvaid sõnu ja selliseid, mis tõukuvad ehk mille koosesinemissageduse tõenäosus on kordi suurem antonüümse otsisõnaga (Renouf, Banerjee 2007).

Et uurida inimest tähistavate sõnade semantilist preferentsi (vt mõiste kohta 1.2.3), loodi adjektiivsete laiendite analüüsiks Robert Dixoni (1982) adjektiivide semantiliste tüüpide VANUS, MÕÕDE, HINNANG, FÜÜSILISED OMADUSED, KIIRUS ja INIMESEGA SEOTUD OMADUSED põhjal kuus selle töö materjali analüüsiks sobivat adjektiivide semantilist tüüpi:

1. VANUS koondab adjektiive, mis tähistavad nii inimese suhtelist (nt *noor, keskealine, vana, alaealine*) kui ka täpset vanust (nt *50aastane, 3aastane*).
2. FÜÜSIS koondab adjektiive, mis tähistavad isiku mõõtmeid (nt *suur, väike, pikk*), kehalisi võimeid (nt *tugev, kiire*) või atraktiivsust (nt *ilus*, *kena*, *kaunis*).
3. PSÜÜHIKA koondab adjektiive, mis kirjeldavad inimese emotsioone (*vihane, armastav*), iseloomu või vaimseid võimeid (nt *hea, halb, aus, probleemne, andekas*).
4. HINNANG koondab adjektiive, mis annavad inimesele üldisema hinnangu, mida ei saa täpsemalt füüsise või psüühiliste omadustega seostada (nt *tavaline, süütu, haige, terve*).
5. SOTSIAAL koondab adjektiive, mis tähistavad sotsiaalset staatust (*rikas, vaene, kuulus*), aga näitavad ka, mis moel ollakse ühiskonnas aktiivsed (nt *õppiv*, *töötav*, *(koolis) käiv*).
6. PEREKOND koondab omadussõnu, mis näitavad perekonnaseisu (nt *vallaline, üksik*), seost teiste pereliikmetega (nt *keskmine, ainus, lähedane*) või isiku fertiilsust (nt *rase, imetav*).

Lisaks esineb inimest tähistavate sõnade adjektiivsete laienditena sõnu, mis ei liigitu loodud tüüpidesse, aga mille jaoks ei olnud ka eraldi tüübi loomine mõttekas, nt sõnad *elav, kõnelev, olev*. Näiteks *elav* kuulub sõnade *inimene, mees, naine, ema, isa, tütar* ja *poeg* 30 sagedama kollokaadi hulka. Seejuures pole *elav* kasutusel tähenduses ‘elavaloomuline’, vaid see moodustab tegelikult osa kolme- või enamaliikmelisest ühendist, kus sõnale *elav* eelneb veel kohta või ehitist näitav kohamäärus või määrusfraas, nt *Harjumaal elavat meest, Kastani tänava puumajas elav mees*. Sõnad, mis ei liigitu kuude peamisse tüüpi, tuuakse vajadusel lihtsalt eraldi välja.

Samuti tuleb rõhutada, et mõni adjektiiv võib kuuluda mitmesse tüüpi, kusjuures tüüpidesse kuulumine võib oleneda ka otsisõnast. Nt adjektiivid *suur* ja *väike* sõnade *poiss* ja *tüdruk* kollokaatidena võivad viidata nii laste kasvule (tüüp FÜÜSIS) kui vanusele (tüüp VANUS), aga *suur* sõna *mees* kollokaadina võib tähistada nii kasvu (tüüp FÜÜSIS) kui tähtsust (tüüp SOTSIAAL).

Tabelites 9–13 on esitatud adjektiivsete kollokaatide absoluutsagedused. Selleks, et erineva esinemissagedusega otsisõnade kollokatsioonide sagedused oleksid võrreldavad, arvutati välja normaliseeritud koosesinemissagedus $NS(os_1,k_1)$ valemiga:

$$NS(os_1, k_1 = \frac{S(os_1,k_1)}{S(kn)} * 1000$$

$S(os_1,k_1)$ = otsisõna (os_1) ja tema kollokaadi (k_1) absoluutne koosesinemissagedus;
$S(k_n)$ = kõikide otsisõnaga kollokeeruvate sõnade absoluutsagedus ehk otsisõna adjektiivsete laiendite üldarv.

Normaliseeritud koosesinemissageduse kasutamine nivelleerib erinevused, mis võiksid tuleneda üksiksõnade erinevast esinemissagedusest (vt Biber jt 1998: 263, 264).

Tabeliveerus *erinevus paarikust* (vt tabelid 9–13 ja tabelid 18–22) esitatud number näitab, mitu korda on adjektiivi normaliseeritud koosesinemissagedus suurem või väiksem kui tema paariku normaliseeritud koosesinemissagedus sama adjektiiviga. Selle leidmiseks jagatakse omavahel mõlema otsisõna normaliseeritud koosesinemissagedused antud adjektiiviga:

$$\text{erinevus paarikust} = \frac{NS(os_1,k_1)}{NS(os_2,k_1)}$$

Vastuseks saadud number 1 näitab, et erinevust ei ole, sellest suurem number näitab, mitu korda sagedamini esineb adjektiivne kollokaat (võrdluses oma paarikuga) just selle otsisõnaga koos. Numbri asemel olev tärn tähistab seda, et adjektiiv ei esine koos teise otsisõnaga ühtegi korda. Kui adjektiiv esineb vähemalt kaks korda sagedamini ühe paarikuga, siis on see tabelites esile toodud paksema kirjaga. Kui sõna esineb väga sageli ühe paariku kollokaadina, siis on tegu sõnale eriomase kollokatsiooniseosega.

4.3.1. Sõnade *poiss* ja *tüdruk* adjektiivsed laiendid

Sõnade *poiss* ja *tüdruk* 30 sagedamast adjektiivist on ühiseid kollokaate 23 ehk ligi 77% (vt tabel 9).

Tabel 9. Sõnade *poiss* ja *tüdruk* 30 sagedamat adjektiivset laiendit esinemissageduse järjekorras. Tärniga tähistatakse adjektiive, mis vastandsõna kollokaatide hulgast puuduvad

	POISS			TÜDRUK		
	Adjektiivne laiend	Erinevus paarikust	Kollokatsiooni absoluutsagedus	Adjektiivne laiend	Erinevus paarikust	Kollokatsiooni absoluutsagedus
1.	väike	0,9	1209	väike	1,1	1046
2.	noor	0,8	582	noor	1,3	590
3.	paha	5,4	391	ilus	2,9	425
4.	14aastane	1,1	278	teismeline	1,9	255
5.	13aastane	1,3	263	14aastane	0,9	195
6.	12aastane	1,3	258	15aastane	0,9	166
7.	15aastane	1,1	244	alaealine	1,7	164
8.	suur	2,3	231	12aastane	0,8	157
9.	tubli	1,7	210	13aastane	0,8	154
10.	puhas	26,1	201	16aastane	0,9	129
11.	hea	1,9	192	kena	2,4	116
12.	ilus	0,3	185	11aastane	0,9	112
13.	16aastane	1,1	180	pisike	0,9	101
14.	tore	1,7	178	tubli	0,6	96
15.	10aastane	1,5	174	10aastane	0,7	88
16.	teismeline	0,5	173	tore	0,6	84
17.	11aastane	1,2	169	hea	0,5	80
18.	pisike	1,1	141	kuueaastane	0,9	79
19.	alaealine	0,6	126	tavaline	1,7	79
20.	kümneaastane	1,7	122	suur	0,4	77
21.	kuueaastane	1,1	116	viieaastane	1,2	75
22.	oma	*	111	17aastane	1,2	71
23.	seitsmeaastane	1,5	95	8aastane	1,2	68
24.	aastane	1,7	94	kaheksa-aastane	1,0	63
25.	pikk	3,3	89	uus	1,1	59
26.	kaheksa-aastane	1,0	83	paha	0,2	56
27.	kohalik	0,9	82	kümneaastane	0,6	55
28.	viieaastane	1,5	82	nelja-aastane	0,9	53
29.	9aastane	1,3	80	õppiv	1,8	53
30.	nelja-aastane	1,2	79	üheksa-aastane	0,9	49

Semantiline tüüp VANUS. Nii sõna *poiss* kui *tüdruk* adjektiivsetest täienditest moodustavad kaks kolmandikku vanust näitavad omadussõnad – peamiselt erinevad *aastane*-lõpulised sõnad *nelja-aastane, viieaastane, kuueaastane, seitsmeaastane, kaheksa-aastane, 8aastane, 9aastane, üheksa-aastane, kümneaastane, 10aastane, 11aastane, 12aastane, 13aastane, 14aastane, 15aastane, 16aastane, 17aastane, aastane,* aga neile lisaks ka üldisemalt vanusele viitavad *noor, teismeline* ja *alaealine*. Nimetatutele lisaks esinesid veel kollokaatidena *väike, pisike* ja *suur*, mis kasvu tähistamise kõrval võivad mitmetel juhtudel tähistada ka vanust (vt näited (14) (15) ja (16)).

(14) Kui mina **väike** poiss olin, räägiti meil kodus ikka Eesti ajast. (Riigikogu stenogrammid 05.1997)

(15) Ise oled juba **suur** poiss, aga ei tea selliseid lihtsaid asju! (Postimees Extra 03.09.1996)

(16) Aga **pisikese** poisina tahtis Indrek Tarand saada aadlikuks. (Eesti Päevaleht 12.04.2003)

Vanust tähistavate adjektiivide täpsem analüüs näitab, et tüüpiline *poiss* või *tüdruk* on teismeline: kollokaatide sagedustabeli eesotsas on just teismeeale viitavad *aastane*-lõpulised täiendid (nt kõige sagedam on 14-*aastane*). Harvem mainitakse alla kümne aastaste laste vanust ja hoopiski pole kummagi sõna 30 sagedama kollokaadi sekka sattunud kolme esimest eluaastat tähistavaid täiendeid. Kollokaatide hulgas on küll sõna *aastane*, aga kasutusnäidete järgi otsustades ei ole sellega tähistatud üheaastasi, vaid peaaegu kõigil ülevaadatud juhtudel oli sõnal veel ees arvuline täiendosa, mis ortograafiavea tõttu või muul põhjusel asus liitsõna järelosisest eraldi.

Semantiline tüüp FÜÜSIS. Vanuse kõrval on määratletud poiste ja tüdrukute välimust, kasutades adjektiive *väike, suur, pikk, ilus, kena* ja *pisike*. Nii suurusele kui vanusele viitav *väike* on mõlema otsisõna kõige sagedam kollokaat, mis esineb kummagi otsisõnaga koos üle 1000 korra. Nii *väike* kui *pisike* pole kummalegi otsisõnale eriomased, vaid esinevad võrdselt mõlema otsisõna kollokaadina. Atraktiivsust tähistavad *ilus* ja *kena* esinevad üle kahe korra sagedamini *tüdruku* kollokaadina, aga mõõtmeid tähistavad *suur* ja *pikk* üle kahe korra sagedamini sõna *poiss* kollokaadina.

Semantiline tüüp PSÜÜHIKA. Välimuse kõrval on määratletud ka iseloomu või hinnangut inimese iseloomule. Nii poissi kui tüdrukut iseloomustatakse samade sõnadega: *hea, paha, tore, tubli.* Samas on suured erinevused kasutussageduses – kollokatsioon *paha poiss* esineb 391 korda, aga *paha tüdruk* ainult 56 korda. Üldse võib öelda, et kõik iseloomule viitavad sõnad esinevad sõna *poiss* kollokaadina 1,7–5,4 korda sagedamini kui sõna *tüdruk* kollokaadina.

Semantiline tüüp SOTSIAAL. Sõnade *poiss* ja *tüdruk* sagedamatest kollokaatidest kuulub siia tüüpi vaid *õppiv*. Kollokaati laiendab enamikel juhtudel sõna, mis näitab, kus klassis või koolis poiss või tüdruk õpib. Seega väljendatakse kollokaadiga pigem õppimiskohta või vanuseastet, mitte õppimisprotsessi ennast.

Semantilisse tüüpi HINNANG kuuluvad adjektiivid *oma, kohalik, uus, tavaline, puhas.* Need kannavad üldisemat hinnangut, ega ole seostatavad kitsamalt isiku iseloomu või välimusega. Esmapilgul välimust tähistava adjektiiviga *puhas* ei kirjeldata isiku välimust, vaid sõna kasutatakse enamikel juhtudel piltlikus tähenduses inimese kirjeldamiseks, kes on süütu või keda ei õnnestu süüdi mõista. Adjektiiv *puhas* esineb koguni 26 korda sagedamini koos sõnaga *poiss* kui sõnaga *tüdruk*. Ka teiste inimest tähistavate sõnadega kombineeritult esineb *puhas* vähemalt kümme korda harvem. Seega vastab *puhas poiss* kinnistunud ehk piiratud kollokatsiooni definitsioonile, mille järgi kinnistunud kollokatsioonis esinevaid sõnu on raske asendada sama tähendusvälja sõnaga (vt 1.2.4). Ka sõnal *poiss* on selles kollokatsiooniseoses laiem tähendus: sõnaga pole tähistatud mitte ainult alaealist meessoost inimest, vaid ühendiga *puhas poiss* on viidatud nii täiskasvanud isikule (17), grupi liikmetele, kelle seas on ka naissoost isikuid (18), ning isegi linnale (19).

(17) Oma esinemises Riigikogus rääkis Tiit Vähi, kui **puhas poiss** ta on. (Postimees 1997.02.13)

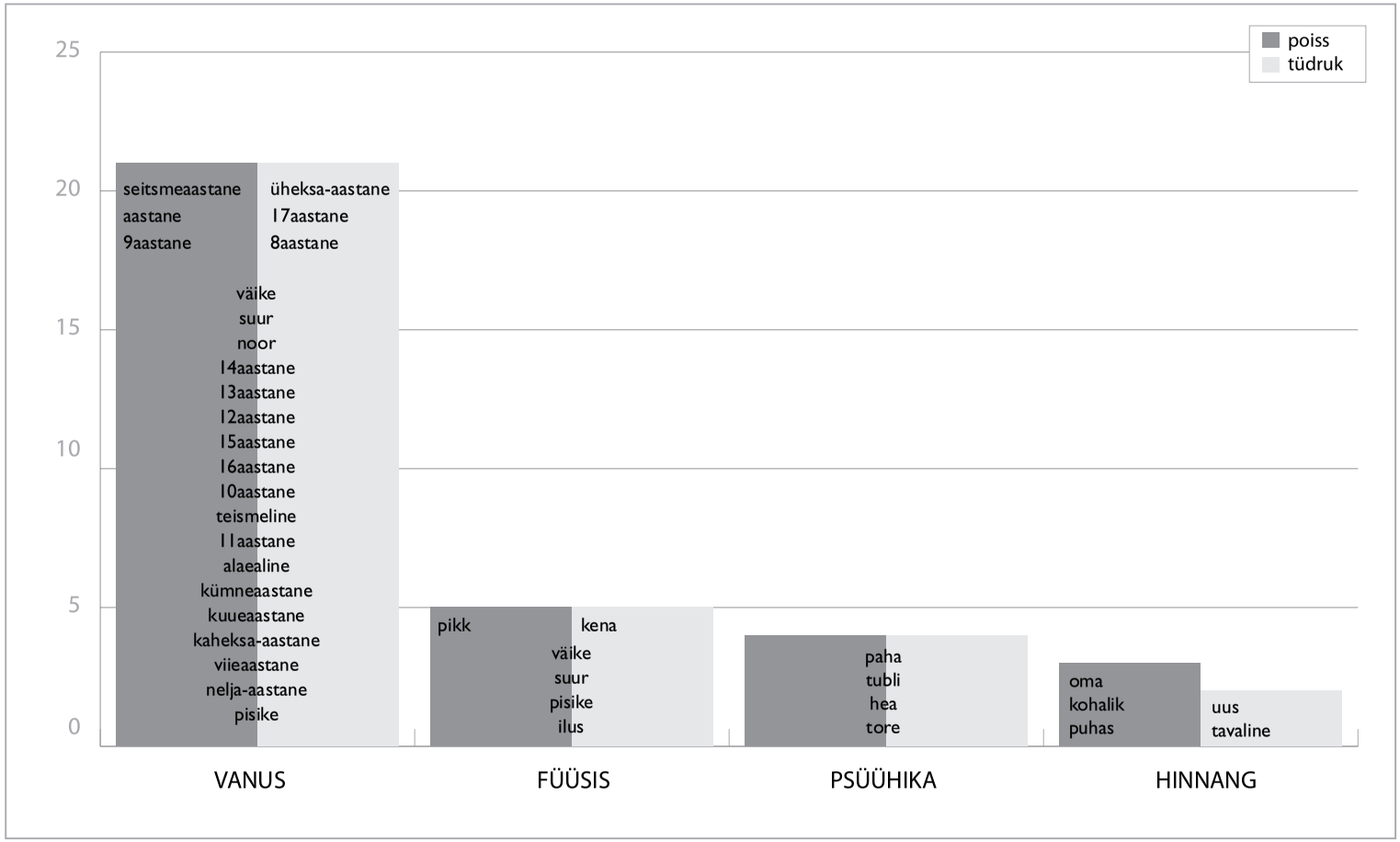
(18) Üldse, kõikide meie olümpiakoondise liikmete dopinguanalüüsid andsid negatiivse märgi, st positiivse tulemuse – kõik korras. Millest ei pea automaatselt tuletama, et eranditult kõik meie tugevamad sportlased n-ö **puhtad poisid** on. (Postimees 1996.10.29)

(19) Kui tõestatakse, et tegu oli kuritegeliku sissetungiga Leedu – NATO õhuruumi, jääb Vilnius **puhtaks poisiks** ning Leedu – Baltikum – NATO saavad endale mõjusa trumpkaardi Venemaa-suunalistes poliitilistes mängudes. (Eesti Päevaleht 28.09.2005)

Samal moel on kollokatsiooni *oma poiss* võimalik kasutada ka täiskasvanud mehe tähistamiseks. Sõna *oma* esineb tähenduses ‘usaldusväärne ja omasugune’ ainult *poisi* kollokaadina. Possessiivse pronoomenina esineks *oma* ka sõna *tüdruk* kollokaadina, aga seda adjektiivsete laiendite loend ei kajasta.

Sõnale *tüdruk* kõige iseloomulikumad adjektiivatribuudid on *ilus* ja *kena*, mis esinevad sõna *tüdruk* kollokaadina 2,9 ja 2,4 korda sagedamini kui koos sõnaga *poiss*. Siiski ei ole need numbrid võrreldavad sõna *puhas* 26-kordse esinemissageduse erinevusega sõna *poeg* ja *tütar* kollokaadina. Seega ei saa väita, et kollokatsioonide *ilus tüdruk* ja *kena tüdruk* puhul oleks tegu kinnistunud kollokatsiooniga. Väikesele esinemissageduse erinevusele vaatamata pole neis ühendites kasutatud sõnu piltlikus tähenduses ning korpuses kollokeeruvad *ilus* ja *kena* ka teiste sama tähendusvälja sõnadega.

Joonisel 14 on näha, et ülekaalukalt suurima grupi sõnade *poiss* ja *tüdruk* adjektiivsetest kollokaatidest moodustavad vanust tähistavad sõnad. Teistesse gruppidesse kuuluvaid kollokaate on palju vähem. Suuri erinevusi sõnade *poiss* ja *tüdruk* kollokaatide gruppidesse jagunemises ei eksisteeri.



Joonis 14. Sõnade *poiss* ja *tüdruk* sagedamate adjektiivsete kollokaatide jagunemine peamistesse semantilistesse gruppidesse

4.3.2. Sõnade *mees* ja *naine* adjektiivsed laiendid

Sõnade *mees* ja *naine* 30 sagedamast adjektiivist on kokkulangevaid kollokaate 18 ehk 60% (vt tabel 10).

Tabel 10. Sõnade *mees* ja *naine* 30 sagedamat adjektiivset laiendit esinemissageduse järjekorras. Tärniga tähistatakse adjektiive, mis vastandsõna kollokaatide hulgast puuduvad

	MEES			NAINE		
	Adjektiivne laiend	Erinevus paarikust	Kollokatsiooni absoluutsagedus	Adjektiivne laiend	Erinevus paarikust	Kollokatsiooni absoluutsagedus
1.	noor	0,7	4436	noor	1,5	3649
2.	vana	1,0	1441	vana	1,0	781
3.	uus	2,0	995	ilus	5,3	630
4.	õige	7,3	906	elav	1,9	491
5.	tundmatu	4,1	905	kaunis	39,0	426
6.	keskealine	1,0	714	töötav	3,6	416
7.	kõva	88,4	647	keskealine	1,0	381
8.	suur	5,7	630	rase	*	348
9.	hea	2,6	544	uus	0,5	274
10.	võõras	1,2	509	endine	3,4	247
11.	vaba	6,1	504	tark	0,9	227
12.	valge	2,5	497	võõras	0,8	224
13.	elav	0,5	462	üksik	3,3	221
14.	kohalik	2,0	462	ettevõtlik	5,5	214
15.	tark	1,1	455	ainus	2,2	184
16.	vaene	3,8	418	kena	2,7	181
17.	rikas	3,2	379	olev	0,9	181
18.	must	5,1	365	paljas	4,3	154
19.	olev	1,1	357	30aastane	1,0	135
20.	tubli	1,8	349	eakas	1,5	129
21.	tugev	1,6	332	kohalik	0,5	129
22.	pikk	2,0	312	tundmatu	0,2	122
23.	30aastane	1,0	258	35aastane	1,2	119
24.	vihane	3,7	234	tugev	0,6	117
25.	ilus	0,2	217	40aastane	1,0	116
26.	kõnelev	3,1	217	hea	0,4	115
27.	töötav	0,3	211	vallaline	3,9	113
28.	tähtis	8,1	208	50aastane	1,2	110
29.	40aastane	1,0	207	valge	0,4	109
30.	terve	1,8	194	tubli	0,5	104

Sõnade *mees* ja *naine* sagedamatest kollokaatidest kuuluvad semantilisse tüüpi VANUS adjektiivid *noor, vana, keskealine, eakas, 30aastane, 35aastane, 40aastane* ja *50aastane.* Kui *aastane*-lõpulised sõnad määratlevad vanust täpsemalt, siis sõnade *noor*, *vana, keskealine* ja *eakas* puhul pole kindel, mis vanusegrupile täpselt iga sõnaga viidatakse. Need kategooriad sisaldavad subjektiivseid hinnanguid ja olenevad tihti ka kõneleja vanusest: eeldatavalt võivad nooremad inimesed viidata kollokatsiooniga *vana mees* ka keskealisele mehele, kusjuures vanemad inimesed võivad sama isiku kohta kasutada sõnaühendit *noor mees*. Mõlema otsisõna kaks kõige sagedamat adjektiivset laiendit on *noor* ja *vana*. Neist *noor* on omakorda ülekaalukalt sagedaim, esinedes kolm korda sagedamini kui *vana* sõna *mees* atribuudina ja ligi viis korda sagedamini sõna *naine* atribuudina.

Semantiline tüüp FÜÜSIS. Siia kuuluvad adjektiivid *tugev, valge, must, suur, pikk, ilus, kaunis, kena* ja *paljas*. Kui vanust määratlevatest adjektiividest ei saanud ühele või teisele otsisõnale eriomaseid kollokaate välja tuua, siis välimust kirjeldavad sõnad jagunevad meestele ja naistele eriomasteks. Ainult *tugev* esines enam-vähem võrdselt mõlema otsisõna kollokaadina. Kuigi lausekonteksti lähemalt uurides selgus, et sõna kasutati erinevate otsisõnadega erinevas tähenduses. Kui *tugev* tähistab sõna *mees* kollokaadina enamikel juhtudel füüsilist tugevust (20), siis sõna *naine* kollokaadina tähistab see hoopis meelekindlust ja vaimujõudu (21).

(20) Lembit on 110 kilo kaaluv musklites **tugev mees** .. (SL Õhtuleht 24.05.2000)

(21) Eesti ühiskond toetub **tugevatele naistele**, kes kõigega hakkama saavad. (Eesti Päevaleht 14.06.2000)

Michael Pearce on sarnase tähelepaneku teinud ka inglise keelt uurides. Ta leidis, et sõnad nagu *broad-shouldered* 'laiaõlgne', *big* 'suur' ja *strong* 'tugev', mis mehe adjektiivatribuutidena esinesid füüsise kohta kasutatuna, said sõnaga *naine* koos esinedes piltlikuma ja üldisema tähenduse (Pearce 2008: 8). Niisiis ei pruugi sõna esinemine mõlema otsisõna kollokaadina veel näidata, et erinevused puuduvad, sest lausekonteksti uurides võib selguda, et kollokaati kasutatakse erinevas tähenduses.

Sõnale *mees* eriomased kollokaadid on *valge, must, suur* ja *pikk.* Europiidset rassi ja nahavärvi tähistav *valge* esineb 2,5 korda sagedamini koos sõnaga *mees* ja mustanahalistele viitav *must* 5,1 korda sagedamini sõna *mees* kollokaadina. See atribuut ei viita ainult neegrirassile, vaid mõnel juhul ka teistele tumedama nahavärviga inimestele, vt allolevaid näiteid.

(22) See on **must mees**, pakistanlane, ta ei tohi minna, ära teda sisse lase! (Eesti Päevaleht 03.12.2005)

(23) Belgradi jaama ümbruses tiirleb **musti mehi** nagu kärbseid. (Areen 1998)

Kuigi sõna *must* saab kasutada ka tähenduses määrdunud, ebapuhas (vt EKSS), siis kollokatsiooniseoses *must mees* ei õnnestunud selliseid kasutusjuhte korpuses esinenud lausete põhjal tuvastada. Iseasi, kas saame väita, et näites (23) on sõnaühendit kasutatud neutraalsel moel või kannab võrdlus *musti mehi nagu kärbseid* siiski teataval määral negatiivset konnotatsiooni.

Sõnale *mees* eriomased füüsist kirjeldavad kollokaadid *pikk* ja *suur* on ka sõnale *poiss* eriomased (vt 4.3.1).

Sõnale *naine* eriomased füüsist kirjeldavad kollokaadid on atraktiivsust tähistavad *ilus, kaunis, kena* ja alastust tähistav *paljas*. Sõna *kaunis* esineb koos sõnaga *naine* 39 korda sagedamini kui sõnaga *mees*. Paika peab Heinike Heinsoo väide: „Eesti keeles on sõna *kaunis* tarvitusel naissooesindaja iseloomustamiseks" (Heinsoo 2010: 123). Selle adjektiivi kasutus näitab, et esmapilgul täissünonüümsetel sõnadel võivad olla kollokatiivsed piirangud, mis lubavad adjektiivi kasutada vaid ühest soost isiku määratlemiseks. Sõnu *ilus* ja *kena* kasutatakse samuti sagedamini naise iseloomustamiseks, aga vahe enam nii suur ei ole: *ilus* esineb 5,4 korda sagedamini koos sõnaga *naine* kui sõnaga *mees* ja *kena* esineb 2,7 korda sagedamini koos sõnaga *naine*. Kolm atraktiivsust märkivat täiendit *ilus, kaunis* ja *kena* moodustavad sõna *naine* kõigist adjektiivsetest laienditest üle 5 protsendi, samad sõnad moodustavad mehe adjektiivatribuutidest alla 1 protsendi.

Sõna *paljas* esineb 4,3 korda sagedamini koos sõnaga *naine*. Alastusest räägitakse üldse seoses naistega mitu korda sagedamini – ka analüsaatoriga Estmorf peamiselt määrsõnaks märgendatud *alasti* esineb 107 korda koos sõnaga *mees* ja 181 korda koos sõnaga *naine*.

Semantiline tüüp PSÜÜHIKA. Siia tüüpi kuuluvad *tark, tubli, ettevõtlik, hea, kõva* ja *vihane*. Sõnade *mees* ja *naine* 30 sagedama adjektiivatribuudi sekka on sattunud vaid positiivseid iseloomuomadusi tähistavad kollokaadid. Neist tark ja tubli esinevad mõlema otsisõnaga koos enam-vähem võrdselt. Sõnale *naine* eriomaseks kollokaadiks on *ettevõtlik*, mis esineb 5,5 korda sagedamini sõnaga *naine* kui sõnaga *mees*.

Kollokatsiooni *kõva mees* võib liigitada kinnistunud kollokatsioonide hulka: sõnad ei esine ühendis oma kõige otsesemas tähenduses. Sõna *kõva* on kasutusel piltlikus tähenduses 'tubli, heatasemeline, silmapaistev' ja *mees* tähendus ähmastub niivõrd, et saame kollokatsiooni *kõva mees* kasutada mööndustega ka naise kohta. Lisaks on kinnistunud kollokatsiooni tunnuseks see, et sõna pole kollokatsioonis võimalik asendada teiste sama tähendusvälja sõnadega. Sõna *mees* ei ole selles kollokatsioonis asendatav sõnaga *naine*. Samas on siiski võimalik asendus sõnaga *poiss*: 54 korral on korpuses kasutatud kollokatsiooni *kõva poiss*.

Semantiline tüüp PEREKOND. Perekondlikku seisu või fertiilsust märkivad sõnad *vallaline, üksik, endine* ja *rase* esinevad ainult sõna *naine* 30 sagedama kollokaadi seas. Adjektiivne laiend *rase* kuulub nende sõnade hulka, mille esinemine sõna *mees* atribuudina võiks olla välistatud. Siiski esineb korpuses ka kolm korda *rase mees* – kahel juhul meeste rasedusest kui võimalikust tulevikuperspektiivist rääkivas lauses ja ühel juhul neist tulevikuperspektiividest rääkiva filmi pealkirjas. Kollokatsioon *rase naine* esineb üle 100 korra sagedamini.

Lisaks kuulub sellesse semantilisse tüüpi adjektiiv *uus* sõna *naine* kollokaadina, samal ajal kui sõna *mees* kollokaadina liigitub uus teise semantilisse tüüpi. Kui kollokatsiooniga *uus naine* tähistatakse üheksal juhul kümnest uut

elukaaslast (24), siis kollokatsiooni *uus mees* kasutatakse üheksal juhul kümnest seoses mõne ametikoha täitmisega (25).

(24) „Siis sa pead **uue naise** võtma," naljatas Tiina toona. (Eesti Päevaleht 24.10.2000)

(25) Ta värbas sel aastal **viis uut** meest. (SL Õhtuleht 22.09.2007)

Semantilisse tüüpi SOTSIAAL kuuluvad adjektiivid *vaene, rikas, tähtis, töötav* ja *elav*. Kui perekondlikku seisu märkivad sõnad ei mahu *mehe* sagedamate kollokaatide sekka, siis sotsiaalset staatust märkivad sõnad nagu *tähtis, rikas, vaene* esinevad ainult sõna *mees* sagedamate kollokaatide hulgas. Ainult *töötav* esineb sõnaga *naine* 3,6 korda sagedamini kui sõnaga *mees*.

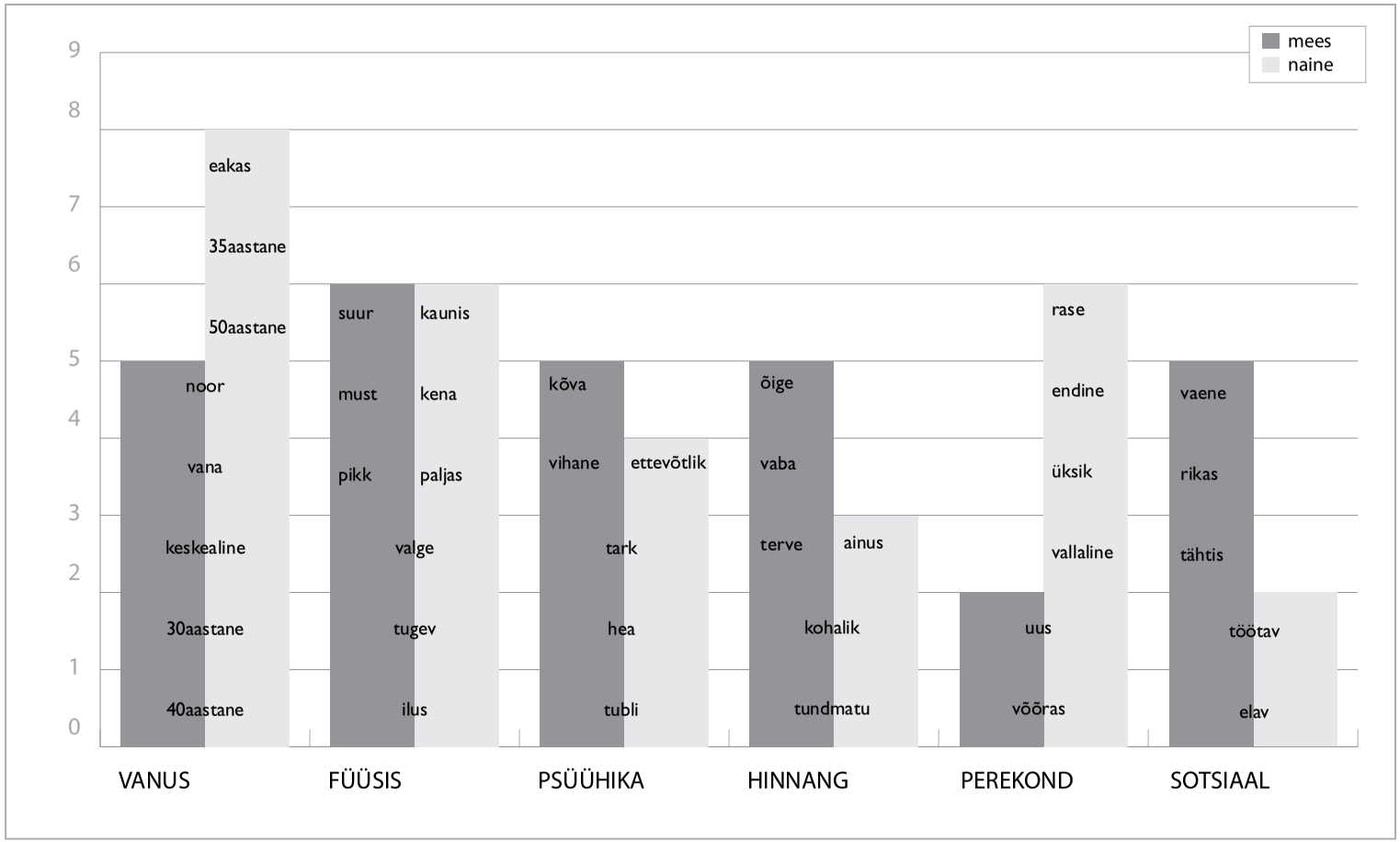
Lisaks peetakse mehe puhul palju olulisemaks näidata, mis ametikohal mees töötab. Kui võrrelda kollokaadi *töötav* kasutust, siis sõna *mees* puhul on üheksal juhul kümnest tegu kolmeliikmelise kollokatsiooniseosega, kus fraasi *töötav mees* laiendab veel töökohta näitav sõna (26). Naise puhul on sõna *töötav* üle kolme korra sagedam, aga pooltel juhtudel ei esine sõna koos ametikohta näitava sõnaga, vaid kollokatsioon *töötav naine* viitab vastandina kodusele naisele lihtsalt töötavale naisele (27).

(26) Klienditeenindajana **töötav mees** ei näe asja tumedates toonides, .. (Eesti Päevaleht 10.11.2004)

(27) Uurimus näitas, et **töötavatel naistel** oli koduprouadest madalam kolesteroolitase, vererõhk ja kehakaal. (Eesti Päevaleht 25.01.1999)

Semantilisse tüüpi HINNANG kuuluvad adjektiivid *terve, võõras, ainus, õige, vaba, kohalik, kõnelev* ja *tundmatu*. Kriminaalse temaatikaga võib seostada sõna *tundmatu* kasutust nii *mehe* kui *naise* adjektiivatribuudina. Peamiselt on selle sõna esinemine ajakirjanduses seotud politseiuudiste rubriigiga. Sõna on üle nelja korra sagedam *mehe* kollokaadina ja esineb seal rohkem kuriteo toimepanija kirjeldustes (üle poolte juhtudest). Samas kui *tundmatu naine* esineb 75 protsendil juhtudest kuriteo ohvri kohta kasutatuna. Peamiselt kuritegevusega seoses kasutatakse mehe kirjeldamisel kollokaati *kõnelev*, mis omab peaaegu alati laiendeid näitamaks, mis keelt kõneletakse. Kaheksal juhul kümnest viidatakse vene keele kõnelejatele. Ühel juhul kümnest eelneb kollokatsioonile *kõnelev mees* laiend *eesti keelt* ja sama palju esineb mõne muu keele kõnelejatele viitamist. Miks mainitakse enam vene keele kõnelejaid? Seda ei saa seletada ainult asjaoluga, et kurjategijatest on suurem osa vene keele kõnelejaid (vangidest moodustavad eesti emakeelega vangid 40% ja vene emakeelega vangid 57% (Ahven, Kruusmaa 2013: 104)). Pigem tuleks põhjust otsida markeeritusest. Eestis on riigikeel eesti keel ilmselt normiks, markeerimata nähtuseks, millest kõrvalekallet on mõttekas markeerida.

Kokkuvõtteks on joonisel 15 näha, et kollokaadid jagunevad mitme grupi vahel ja ükski grupp ei tõuse teistest rohkem esile. Vanust väljendavaid adjektiive on veidi enam ja eriti just sõna *naine* kollokaatide hulgas. Füüsist kirjeldavaid adjektiive on mõlema sõna puhul sama tihti kasutatud, aga erinevus on sõnavalikus. Teiste gruppide puhul paistab, et sõna *naine* kollokaatide seas on enam perekondlikku seisu või fertiilsust kirjeldavaid adjektiive ja sõna *mees* kollokaadid kirjeldavad rohkem sotsiaalset staatust.



Joonis 15. Sõnade *mees* ja *naine* sagedamate adjektiivsete kollokaatide jagunemine peamistesse semantilistesse gruppidesse

4.3.3. Sõnade *poeg* ja *tütar* adjektiivsed laiendid

Sõnade *poeg* ja *tütar* 30 sagedamast adjektiivist on ühiseid kollokaate 25 ehk 83% (vt tabel 11). Lisaks suurimale kollokaatide ühisosale paistab teiste tabelitega võrdlusest silma, et paarikutele eriomaseid kollokaate on vähe. Üle kahe korra sagedamini esinevad vaid kaks kollokaati: *20aastane* sõna *poeg* kollokaadina ja *elav* sõna *tütar* kollokaadina.

Tabel 11. Sõnade *poeg* ja *tütar* 30 sagedamat adjektiivset laiendit esinemissageduse järjekorras

	POEG			TÜTAR		
	Adjektiivne laiend	Erinevus paarikust	Kollokatsiooni absoluutsagedus	Adjektiivne laiend	Erinevus paarikust	Kollokatsiooni absoluutsagedus
1.	väike	0,9	246	väike	1,1	229
2.	nelja-aastane	1,0	127	kolmeaastane	1,6	142
3.	kaheaastane	0,9	123	teismeline	1,6	133
4.	aastane	0,9	113	viieaastane	1,5	130
5.	11aastane	1,0	106	kaheaastane	1,1	114
6.	viieaastane	0,7	104	aastane	1,2	110
7.	kolmeaastane	0,6	103	nelja-aastane	1,0	106
8.	kuueaastane	1,1	100	elav	2,1	91
9.	12aastane	1,1	99	11aastane	1,0	85
10.	teismeline	0,6	98	13aastane	1,5	81
11.	ainus	0,9	86	ainus	1,1	78
12.	kaheksa-aastane	1,3	77	12aastane	0,9	74
13.	14aastane	1,0	72	kuueaastane	0,9	73
14.	17aastane	0,9	72	10aastane	1,2	71
15.	10aastane	0,8	68	õppiv	1,3	69
16.	13aastane	0,7	65	17aastane	1,1	65
17.	15aastane	1,2	65	seitsmeaastane	1,3	65
18.	õppiv	0,8	65	14aastane	1,0	61
19.	seitsmeaastane	0,8	61	16aastane	1,2	58
20.	16aastane	0,8	57	9aastane	1,9	52
21.	8aastane	1,0	57	18aastane	1,7	48
22.	kümneaastane	1,1	57	kaheksa-aastane	0,7	48
23.	alaealine	1,6	55	üheksa-aastane	1,9	48
24.	ühine	1,6	53	8aastane	1,0	47
25.	elav	0,5	52	15aastane	0,8	44
26.	20aastane	2,7	48	kümneaastane	0,9	44
27.	keskmine	1,3	46	7aastane	1,1	38
28.	5aastane	1,3	45	4aastane	1,0	36
29.	7aastane	0,9	42	6aastane	1,0	35
30.	4aastane	1,0	41	käiv	1,3	31

Semantiline tüüp VANUS. Peamise osa sõnade *poeg* ja *tütar* sagedamatest adjektiivsetest laienditest moodustavad vanusele viitavad sõnad. Täpsemalt määratlevad vanust erinevad *aastane*-lõpulised sõnad *kaheaastane, kolmeaastane, nelja-aastane, 4aastane, viieaastane, 5aastane, kuueaastane, 6aastane, seitsmeaastane, 7aastane, 8aastane, kaheksa-aastane, üheksa-aastane, 9aastane, kümneaastane, 10aastane, 11aastane, 12aastane, 13aastane, 14aastane, 15aastane, 16aastane, 17aastane, 18aastane, 20aastane* ja *aastane*. Lisaks määratlevad vanust üldisemalt kollokaadid *alaealine, väike* ja *teismeline*.

Võrreldes neid sõnu antonüümipaari *poiss* ja *tüdruk* kollokaatidega, hakkab silma, et *aastane*-lõpulistest kollokaatidest kasutatakse sõnade *poeg* ja *tütar* atribuutidena sagedamini sõnu, mis tähistavad väiksemaid, alla kümne aastasi lapsi, samal ajal kui tüüpilised poiss ja tüdruk on teismelised. Siiski on sõnade *poiss, tüdruk, poeg* ja *tütar* kollokatsiooniseosed väga sarnased ja peamiselt määratletaksegi nende puhul isiku vanust.

Semantiline tüüp PEREKOND. Siia tüüpi kuuluvad adjektiivid *keskmine, ainus* ja *ühine.* Seostele teiste õdede-vendadega viitavad täiendid *keskmine* ja *ainus*. Seost vanematega väljendab täiend *ühine*, mis viitab, et tegemist on kahe lapsevanema ühise lapsega.

Semantiline tüüp SOTSIAAL. Lisaks vanuse ja perekondlike suhete määratlemisele on veel üksikud adjektiivid, mis viitavad mõnele muule aspektile, nt elukohale. Täiend *elav* ei olnud kasutuses tähenduses elavaloomuline, vaid märgib ligi 90 protsendil juhtudest koos kohanimelise laiendiga elukohta vt (28). Kui üldiselt määratlevad kollokaadid poega ja tütart alaealisena, siis adjektiivatribuuti *elav* kasutatakse peamiselt täiskasvanud lapse elukoha märkijana.

(28) Raamatupidamistööd teeb **Tartus elav tütar**, ehituste eesotsas on olnud aga ehitusinsenerist poeg. (Maaleht 20.09.2001)

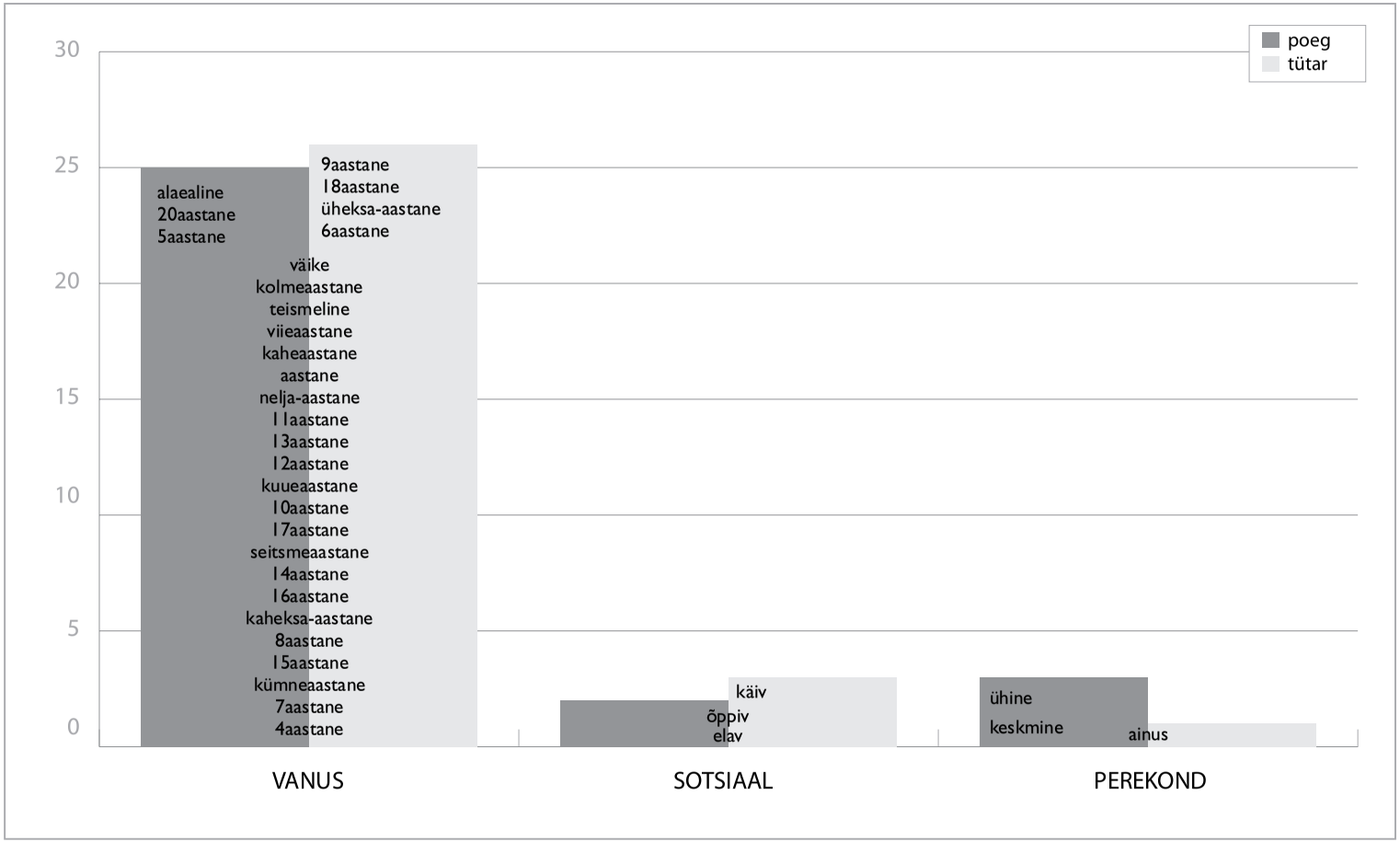
Lisaks määratletakse õppimiskohta või taset: v-kesksõnad *õppiv* ja *käiv* esinevad koos laienditega, mis näitavad kooli või klassi.

(29) Olen kahe **kutsekoolis õppiva poja** ema. (Postimees 1997.01.11)

(30) **Neljandas klassis käiva tütre** hariduse pärast tal põhjust muretseda polnud. (Eesti Päevaleht 02.09.2004)

Sõnade *poeg* ja *tütar* 30 sagedamat adjektiivset laiendit on silmatorkavalt sarnased, erinevusi võib märgata alles keskmise esinemissagedusega kollokatsiooniseostes, mis 30 sagedama hulka ei mahu. Kui võrrelda kõiki sõnu *tütar* ja *poeg* laiendavaid adjektiive (vt kogu sagedusloetelu Lisas 5) ja võtta välja sõnad, mis ühe sõna kollokaadina esinevad vähemalt 6 korda sagedamini kui teise sõna kollokaadina, siis ilmnevad ka mõned erinevused. Sõnale *tütar* iseloomulikud adjektiivsed laiendid, mis esinevad vähemalt 6 korda sagedamini sõna *tütar* kollokaadina, on peamiselt välimust kirjeldavad: *kena, kaunis, imekaunis, ilus, imeilus* ja *tilluke.* Samal ajal ainult sõnaga *poeg* esinevad koos adjektiivid *suur, hea, viimane* ja *ustav* ning vähemalt 10 korda sagedamini esinevad sõnaga *poeg* koos kollokaadid *kange* ja *vapper*.

Joonis 16 rõhutab ilmekalt vanust tähistavate adjektiivide tähtsust nii sõnade *poeg* kui *tütar* kollokaatide seas.



Joonis 16. Sõnade *poeg* ja *tütar* sagedamate adjektiivsete kollokaatide jagunemine peamistesse semantilistesse gruppidesse

4.3.4. Sõnade *ema* ja *isa* adjektiivsed laiendid

Antonüümipaari *ema* ja *isa* 30 sagedamast adjektiivatribuudist langeb kokku 21 kollokaati ehk 66 protsenti (vt tabel 12).

Tabel 12. Sõnade *ema* ja *isa* 30 sagedamat adjektiivset laiendit esinemissageduse järjekorras. Tärniga tähistatakse adjektiive, mis vastandsõna kollokaatide hulgast puuduvad

	EMA			ISA		
	Adjektiivne laiend	Erinevus paarikust	Kollokatsiooni absoluutsagedus	Adjektiivne laiend	Erinevus paarikust	Kollokatsiooni absoluutsagedus
1.	noor	4,3	959	vaimne	17,2	248
2.	tulevane	2,6	239	hea	2,2	192
3.	hea	0,5	139	püha	58,9	185
4.	elav	1,1	134	noor	0,2	139
5.	töötav	1,5	103	bioloogiline	2,3	81
6.	olev	3,5	101	elav	0,9	78
7.	haige	2,0	86	kuulus	19,4	61
8.	kasvatav	6,7	85	tulevane	0,4	58
9.	vana	2,0	75	õnnelik	1,4	57
10.	imetav	*	67	värske	2,7	48
11.	õnnelik	0,7	67	töötav	0,7	43
12.	ootav	8,8	56	taevane	31,0	39
13.	bioloogiline	0,4	56	suur	2,5	34
14.	lihane	1,1	49	uus	1,6	33
15.	hoolitsev	1,2	44	rikas	23,1	29
16.	kallis	1,0	43	lihane	0,9	28
17.	kodune	26,4	42	haige	0,5	27
18.	murelik	3,0	38	kallis	1,0	27
19.	paljulapseline	4,3	34	armastav	1,3	26
20.	uus	0,6	33	tegelik	6,6	25
21.	toitev	*	32	hoolitsev	0,8	23
22.	armastav	0,8	32	tõeline	1,7	23
23.	vastne	1,0	31	vana	0,5	23
24.	eakas	1,5	28	halb	1,5	21
25.	värske	0,4	28	ideaalne	6,0	19
26.	võimukas	4,8	23	tubli	1,3	19
27.	tubli	0,8	23	vastne	1,0	19
28.	vaimne	0,1	23	olev	0,3	18
29.	õige	0,9	23	vägivaldne	9,5	18
30.	õnnetu	1,5	22	õige	1,2	17

Semantilisse tüüpi VANUS kuulub kolm kollokaati: *noor, vana* ja *eakas*. Sõnade *ema* ja *isa* 30 sagedama adjektiivatribuudi seas polnud ühtki vanust täpsemalt määratlevat *aastane*-lõpulist liitsõna. Siiski oli *noor* kõige sagedam adjektiivatribuut sõna *ema* kollokaatide seas, moodustades viiendiku kõigist sõna *ema* adjektiivatribuutidest ning esinedes ka sõna *isa* kollokaatide tabelis kõrgel kohal. Tundub, et selle atribuudiga ei tähistata aga enamikel juhtudel lapsevanema vanust, vaid seda, kui värske lapsevanemaga on tegu. Seega kuulub sõna samal ajal ka tüüpi PEREKOND.

Semantilisse tüüpi PSÜÜHIKA kuuluvad *hea, tubli, armastav, hoolitsev, halb, vägivaldne, võimukas, õnnelik, murelik* ja *õnnetu.* Nende seas annavad inimese käitumisele või iseloomule positiivse hinnangu adjektiivid *tubli, armastav, hoolitsev* ja *hea.* Neist viimane esineb kaks korda sagedamini sõna *isa* kollokaadina. Negatiivset hinnangut kandvad adjektiivid jagunevad kaheks: sõnale *isa* on eriomased *halb* ja *vägivaldne* ning sõnale *ema* on eriomane *võimukas*. Emotsioone kirjeldavatest adjektiividest kasutatakse mõlema sõna puhul võrdselt sõnu *õnnelik* ja *õnnetu*, samas kui *murelik* esineb kolm korda sagedamini *ema* kollokaadina.

Semantilisse tüüpi PEREKOND kuuluvad sõnad *noor, värske, vastne, tulevane, ootav, imetav, lihane, kodune, toitev*, *kasvatav* ja *paljulapseline.* Võrdselt sageli viidatakse nii isa kui ema uuele rollile lapsevanemana sõnadega *värske* ja *vastne*. Tulevasele lapsevanema rollile viitav *tulevane* on kaks korda sagedam sõna *ema* kollokaadina ja samas tähenduses kasutatav *(last) ootav* on üle 8 korra sagedam sõna *ema* kollokaadina. Lisaks on sõnale *ema* eriomased veel *imetav, kodune, toitev*, *kasvatav* ja *paljulapseline*. Selles tüübis sõnale *isa* eriomaseid adjektiivseid laiendeid ei saa välja tuua.

Semantilisse tüüpi SOTSIAAL kuuluvad *töötav, kuulus, suur* ja *rikas.* Nii *ema* kui *isa* kollokaatide hulgas esineb võrdse sagedusega *töötav*. Selle sõna kasutus kollokaadina on siiski mõlema sõna puhul erinev: kui *ema* kollokaadina märgib sõna *töötav* vastandina sõnale *kodune* (mis on sõnale *ema* eriomane, esinedes 26 korda sagedamini koos sõnaga *ema*) peamiselt lihtsalt tööl käivat ema (31), siis *isa* kollokaadina esinev *töötav* omab suuremal osal juhtudest laiendit, mis näitab töökohta või ametit (32).

(31) .. hoopis **töötavate** emade lapsed on ärksamad. (Eesti Päevaleht 27.11.1999)

(32) **Energeetikuna töötaval** isal on dramaatilise tenori hääl. (SL Õhtuleht 1999.12.10)

Sotsiaalset staatust näitavad kollokaadid *kuulus, rikas* ja *suur* on sõnale *isa* eriomased. Kollokatsiooniga *suur isa* tähistatakse pigem sotsiaalses staatuses kõrgel kohal olevat meest– väejuhti või riigitegelast (33).

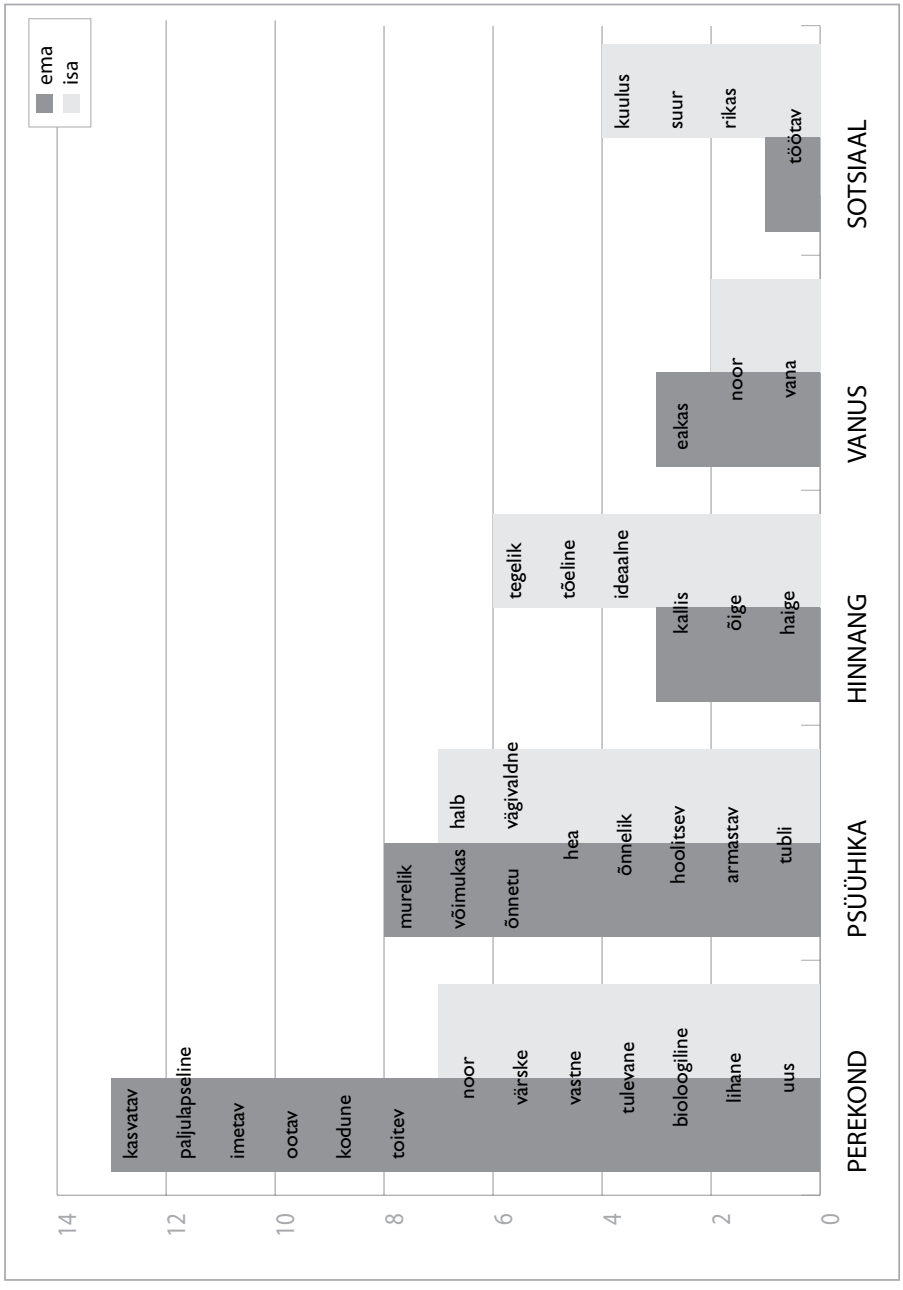
(33) .. kus on nõukogude rahvaste **suurest** isast, lapsi paitavast Leninist päris mitu ilusat pilti .. (Eesti Päevaleht 10.06.2006)

Semantilisse tüüpi HINNANG kuuluvad *õige, bioloogiline, tõeline, ideaalne, tegelik, uus, kallis* ja *haige.* Mitmed kollokaadid on kasutusel isaduse tähistamiseks, nt bioloogilist isadust tähistavad täiendid *bioloogiline, tegelik, tõeline* ja *õige*. Neist sõna *ema* kollokaadina esineb sama sageli vaid sõna *õige*. Ülejäänud adjektiivid koos sõnaga *ideaalne* on vaid *isale* eriomased. Sõnale *ema* on selles grupis eriomane vaid *haige.*

Muudest, esitatud semantilistesse tüüpidesse liigitumatutest kollokaatidest, tulevad esile sõnale *isa* eriomased kollokaadid *püha, taevane* ja *vaimne*. Need sõnale *isa* eriomased kollokaadid viitavad *isa* peamisele kasutusele kujundlikus mõttes vaimse juhi, vaimuliku või jumala kohta. Sõna *vaimne* on kasutusel ka *ema* kollokaadina, seda küll 17 korda harvem. See on ka ainuke sagedam täiend, mis

viitab, et sõna *ema* saab kasutada otsese eellase tähistamise kõrval vaimse eeskuju kohta. Seega on märgata, et metafoorses tähenduses vanema rollis (jumala, vaimulike juhi, eeskuju andjana) esineb peamiselt meessoost lapsevanemat märkiv *isa*.

Joonisel 17 on näha, et sõnaga *ema* kollokeeruvatest adjektiividest kuulub suurim osa tüüpi PEREKOND, samal ajal kui tüüpidesse SOTSIAAL ja HINNANG kuulub rohkem sõna *isa* kollokaate.



Joonis 17. Sõnade *ema* ja *isa* sagedamate adjektiivsete kollokaatide jagunemine peamistesse semantilistesse gruppidesse

4.3.5. Sõnade *inimene* ja *laps* adjektiivsed laiendid

Sõnapaari *inimene* ja *laps* 30 sagedamast adjektiivatribuudist langeb kokku 8 kollokaati ehk ligi 27% (vt tabel 13).

Tabel 13. Sõnade *inimene* ja *laps* 30 sagedamat adjektiivset laiendit esinemissageduse järjekorras. Tärniga tähistatakse adjektiive, mis teise otsisõna kollokaatide hulgast puuduvad

	INIMENE			LAPS		
	Adjektiivne laiend	Erinevus paarikust	Kollokatsiooni absoluutsagedus	Adjektiivne laiend	Erinevus paarikust	Kollokatsiooni absoluutsagedus
1.	noor	28,5	4545	väike	14,5	2864
2.	elav	2,1	2329	alaealine	*	610
3.	tavaline	6,7	1944	hea	1,2	561
4.	uus	2,2	1467	aastane	33,5	462
5.	töötav	22,8	1456	haige	2,3	410
6.	hea	0,8	1446	kolmeaastane	*	399
7.	olev	1,4	1344	elav	0,5	343
8.	kohalik	4,4	1104	terve	1,5	340
9.	normaalne	4,5	1083	olev	0,7	306
10.	vana	7,1	981	14aastane	*	298
11.	lihtne	*	887	sündiv	150,6	283
12.	erinev	11,9	877	käiv	7,4	280
13.	võõras	1,4	815	õppiv	23,7	267
14.	eakas	*	810	15aastane	35,7	235
15.	lähedane	*	763	12aastane	*	234
16.	terve	0,7	738	ühine	*	227
17.	tundev	32,3	722	vanune	10,4	225
18.	korralik	10,3	692	andekas	1,7	224
19.	aus	71,6	686	tulevane	69,6	218
20.	konkreetne	9,9	666	uus	0,5	208
21.	tore	2,4	661	kooliealine	*	193
22.	õige	14,2	635	kaheaastane	*	190
23.	väike	0,1	632	eelkooliealine	*	186
24.	tark	6,8	605	võõras	0,7	184
25.	mõtlev	8,5	599	3aastane	*	182
26.	üksik	2,0	568	16aastane	83,0	182
27.	haige	0,4	567	18aastane	10,3	181
28.	süütu	3,5	521	järgmine	11,1	174
29.	rikas	14,7	469	probleemne	36,6	172
30.	suur	0,9	463	10aastane	*	171

Kui eelnevalt analüüsitud sõnapaaride vahel valitses soovastandusel põhinev antonüümia, siis järele jäänud paari *laps* ja *inimene* vahel võib näha vanuselist vastandust ja hüponüümiat (kui võtta sõna *laps* tähenduses 'alaealine inimene'). Sõnadel *laps* ja *inimene* on palju vähem ühiseid adjektiivatribuute kui antonüümipaaridel. Sõnade *poeg* ja *tütar* kollokaatide 83% kattumise kõrval on sõnade *laps* ja *inimene* kollokaatide 27% kattumine väike. Kõige suurema mittekattuvate kollokaatide grupi moodustavad kõikvõimalikud vanusele viitavad sõnad sõna *laps* sagedamate kollokaatide seas.

Semantilisse tüüpi VANUS kuulub suur hulk sõnale *laps* eriomaseid adjektiive. Eelkõige on nendeks erinevad *aastane*-lõpulised sõnad *kaheaastane, kolmeaastane, 3aastane, 10aastane, 12aastane, 14.aastane, 15aastane, 16aastane 18aastane* ja *aastane.* Veel sündimata lapse tähistamiseks kasutatakse sageli täiendeid *sündiv* ja *tulevane.* Üldist vanust märgitakse sõnadega *vanune* (koos vanust täpsustava laiendiga)*, kooliealine, eelkooliealine, alaealine*, *väike* ja *suur*. Neist viimased kaks võivad kirjeldada sõna *laps* kollokaadina nii vanust kui füüsist. Vanust tähistavate adjektiivide suur osakaal sõna *laps* kollokaatide seas on sarnane osatähtsusega, mida mängivad vanusele viitavad sõnad sõnade *poeg, tütar, poiss* ja *tüdruk* sagedamate kollokaatide hulgas. Nende ko-hüponüümidega on sõnal *laps* suurem kollokatsioonikobarate kattuvus kui hüperonüümiga *inimene*. Nt sõnadel *laps* ja *poeg* on koguni 14 ühist adjektiivset kollokaati (vt tabeleid 11 ja 13). Kui sõna *laps* sagedamatest kollokaatidest moodustavad erinevad vanusele viitavad sõnad ligi poole, siis sõna *inimene* sagedamate kollokaatide seas on vanusele viitavaid sõnu vaid kolm: *noor, vana* ja *eakas*. Need kõik on sõnale *inimene* eriomased, st sõna *noor* esineb 26 korda sagedamini koos sõnaga *inimene* ja *eakas* puudub sõna *laps* kollokaatide hulgast sootuks.

Semantilisse tüüpi PSÜÜHIKA kuuluvad *hea, andekas, tark, korralik, aus, konkreetne, tore, probleemne.* Neist kaks esimest esinevad peaaegu võrdselt nii sõna *inimene* kui *laps* kollokaadina. Lisaks on sõnale *inimene* eriomased positiivseid hinnanguid kandvad kollokaadid *tark, korralik, aus, konkreetne* ja *tore*. Sõnale *laps* on eriomane negatiivset hinnangut kandev *probleemne*.

Semantilisse tüüpi PEREKOND kuuluvad *lähedane, üksik, ühine* ja *järgmine.* Neist kaks esimest esinevad sõna *inimene* kollokaatidena ja kaks viimast sõna *laps* kollokaatidena.

Semantilisse tüüpi SOTSIAAL kuuluvad *suur, väike, rikas, töötav, õppiv, käiv.* Sõnad *suur* ja *väike* kuuluvad siia tüüpi just sõna *inimene* kollokaatidena, tähistades pigem tähtsust ja olulisust (34), mitte füüsilisi mõõtmeid või vanust. Sõna *laps* kollokaatidena näitavad need pigem vanust (35) ja on liigitatud semantilisse tüüpi VANUS.

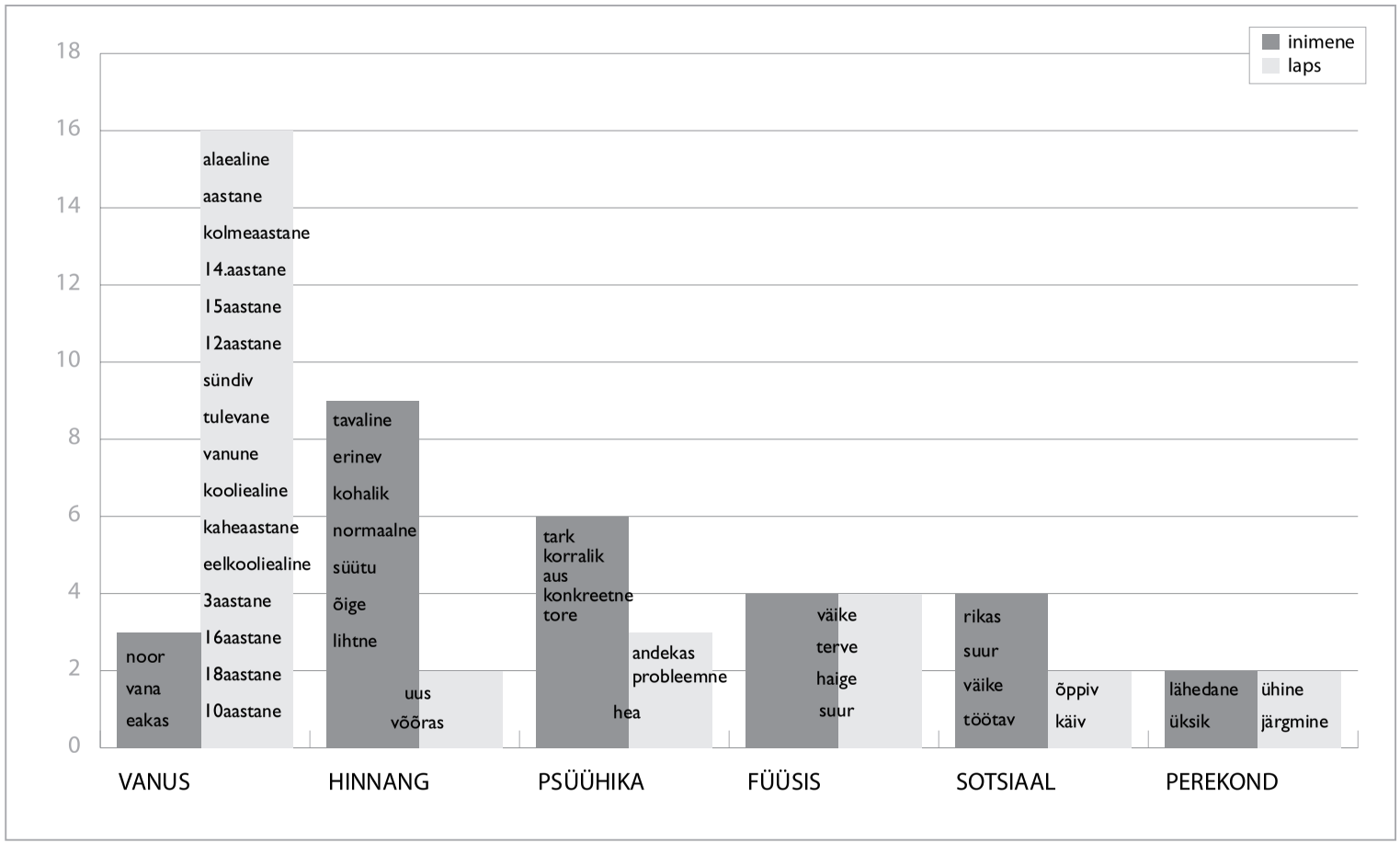
(34) Inimkonna ajalugu on tegelikult **suurte inimeste** suurte tegude lugu. (Riigikogu stenogrammid: märts 2001)

(35) Väga **suuri lapsi** ekraanil ei näe, sest 16–18aastasi pidasime juba täiskasvanuks ja neil oli õigus saates osalemisest ka keelduda. (SL Õhtuleht 2004.10.19)

Sõnale *inimene* on veel eriomased *rikas* ja *töötav*. Sõnale *laps* on eriomased *õppiv* ja *käiv*.

Semantilisse tüüpi HINNANG kuuluvad *võõras, uus, terve, tavaline, erinev, kohalik, normaalne, süütu, õige, lihtne, haige.* Adjektiivid *võõras, uus* ja *terve* esinevad võrdselt nii sõna *inimene* kui *laps* kollokaadina. Sõnale *laps* on eriomane *haige*. Sõnale *inimene* on eriomased *tavaline, erinev, kohalik, normaalne, süütu, õige* ja *lihtne.*

Jooniselt 18 on näha suur vanust tähistavate adjektiivide ülekaal sõna *laps* kollokaatide hulgas. Sõna *inimene* kollokaatidest kuulub aga mitu korda rohkem adjektiive semantilistesse gruppidesse HINNANG, PSÜÜHIKA ja SOTSIAAL.



Joonis 18. Sõnade *inimene* ja *laps* sagedamate adjektiivsete kollokaatide jagunemine peamistesse semantilistesse gruppidesse.

4.4. Verb inimest tähistava sõna kollokaadina

4.4.1. Verbikonstruktsioonide tutvustus

Siinses peatükis lähtutakse eeldusest, et sagedamini inimest tähistava sõna kollokaadina esinevad verbid peaksid näitama, millistes tegevustes inimesi kõige tüüpilisemalt nähakse või millistes tegevustes on inimene passiivseks kogejaks, kelle suhtes mingit tegevust sooritatakse. Selliste verbide leidmiseks kasutatakse kaht erinevat leksikogrammatilist konstruktsiooni: verbid, mille suhtes otsisõna on subjektiks, leitakse grammatikas reeglitega, mis koonduvad nimetuse alla *subject_of,* ja verbid, mille suhtes otsisõna on objektiks, leitakse reeglitega *object_of* (vt Kallas 2013). Kui *subject_of* reeglid toovad välja verbid, mis peaksid näitama, millistes tegevustes inimesi nähakse tegutsejatena, siis *object_of* reeglitega saab välja tuua verbid, mis näitavad, kelle suhtes mingit tegevust teostatakse.

Tabel 14. Erinevate inimest tähistavate sõnade juurde kuuluvate verbide hulk, mis tuvastatud *subject_of* ja *object_of* reeglitega

otsisõna	otsisõna esinemissagedus	verbikonstruktsioonid, kus otsisõna subjektiks	verbikonstruktsioonid, kus otsisõna objektiks
inimene	522436	61801 (11,8%)	10798 (2,1%)
mees	307624	52914 (17,2%)	5823 (1,9%)
laps	245399	24536 (10,0%)	5594 (2,3%)
naine	179449	26160 (14,6%)	3055 (1,7%)
ema	68944	13096 (19,0%)	253 (0,4%)
poiss	59315	10705 (18,0%)	998 (1,7%)
isa	57094	9219 (16,1%)	259 (0,5%)
tüdruk	44418	7879 (17,7%)	804 (1,8%)
poeg	40834	3801 (9,3%)	526 (1,3%)
tütar	28061	2779 (9,9%)	282 (1,0%)

Tabelis on esitatud verbikonstruktsioonide absoluutsed esinemissagedused erinevate otsisõnadega. Protsentides näidatakse, kui suure osa mingi verbikonstruktsioon moodustab vastava inimest tähistava otsisõna kõigist esinemisjuhtudest. Näiteks sõna *inimene* esineb korpuses kokku 522436 korral, millest subjektina 12% juhtudest ehk 61801 korral ja objektina 2,1% juhtudest ehk 10798 korral.

Kui võrrelda esitatud arve, siis hakkab kõigi sõnade puhul silma, et kuigi otsisõna esinemissagedus on suur, esineb sõna subjekti ja objektina kokku ainult 10–20 protsendil juhtudest. See võib tuleneda asjaolust, et paljud inimest tähistavad sõnad võivad esineda hoopis adverbiaalina kogejalausetes (*Naist huvitab muusika. Poisil hakkas halb*), omajalausetes (*Pojal on maja*) või tulemuslausetes (*Tütrest sai ehitaja*), samuti subjekti laiendava genitiivatribuudina lausetes nagu

Ema raamat kukkus põrandale või *Inimese surm vapustas teda*. Seega peab arvestama, et valides teatud grammatilised konstruktsioonid, kitsendame märgatavalt kollokatsioonide leidmise konteksti. Samas on meil siiski pisut selgem, mis seos on sõnal tema kollokaadiga.

Osaliselt tuleneb inimest tähistavate sõnade subjekti ja objektina esinemise väike protsent ka asjaolust, et kasutatud leksikogrammatilisi struktuure tuvastavate reeglitega pole võimalik kõiki esinemisjuhte tuvastada. Lisaks võimalikele lünkadele sõnavisandite grammatikas, mida kasutaja ei pruugi tähele panna, võivad tulemust mõjutada ka lemmatiseerimisest ning märgendamisest tulenevad vead, mis automaatse märgenduse juures on vältimatud.

4.4.1.1. Verbikonstruktsioon *subject_of*

Sõnavisandite grammatika tuvastab vaid täissubjekte. Kallase sõnul loetakse verbi subjektiks nominatiivis substantiive, mis asuvad lauses vahetult enne või pärast verbi. Kontekstipiiranguteks on, et tegemist peab olema finiitverbiga ja verb ei tohi olla umbisikulise tegumoe vormis (Kallas 2013: 66).

Verbid, mille juures inimest tähistav sõna on subjektiks, leitakse kahe reegli abil:

1. reegliga leitakse isikulises tegumoes pöördelised verbivormid (v.a lemma *olema* ja *ei*), millele ei järgne verbi ega afiksaaladverbi ning millele eelneb nimetavas käändes otsisõna ja võib eelneda adverb või eitussõna.

2. reegliga leitakse isikulises tegumoes verbivormid (v.a lemma *olema* ja *ei*), millele järgneb nimetavas käändes otsisõna ega järgne verbi ega afiksaaladverbi.

Nagu reeglitest nähtub, ei ole alati tegu vahetu järgnevusega: esimene reegel lubab substantiivi juurde kuuluvaks verbiks märkida verbi ka siis, kui nende vahel on adverb. Teisalt nähtub reeglitest, et muid kitsendusi on rohkem, näiteks reeglites oleva afiksaaladverbe keelava loendiga *alla, alt, edasi, eemale, esile, ette, juurde, järele, kaasa, kinni, kokku, kõrvale, külge, lahku, lahti, ligi, läbi, maha, mööda, otsa, peale, pealt, ringi, sisse, taga, tagant, tagasi, täis, vahele, vastu, välja, ära, üle, üles, üleval, ümber* üritatakse kõrvaldada tulemustest ühendverbid. Samuti on tahetud tulemustest välja jätta ahelverbid.

Ühest küljest on üksik- ja perifrastiliste verbide (perifrastiline verb tähenduses ühend-, ahel-, või väljendverb, vt EKG 1993: 18) eristamise soov mõistetav, kuna finiitne verbivorm kannab perifrastiliste verbide puhul ainult osa verbi tähendusest. Samas on reeglid poolikud ja eemaldavad ainult need juhud, kus perifrastilise verbi teine komponent järgneb vahetult finiitsele verbivormile. Kui vahetut järgnevust ei ole, siis jäävad ahel- ja ühendverbid alles nagu allolevates näidetes:

(36) Ema **pidi** mind ja õde Sirjet üksinda **kasvatama** ning minema palgatööle. (Eesti Päevaleht 06.11.1999)

(37) Muututi kidakeelsemaks ja peljati, et naine **kannab** kaasale **ette. (**SL Õhtuleht 1999.12.16).

Kui ülalolevates näidetes oli probleemiks, et reegel on liiga kitsas eemaldamaks kõiki sarnaseid juhte, siis afiksaaladverbide loendid, mis on esitatud tavaliste sõnaloenditena, tekitavad vastupidiseid probleeme: need reeglid jätavad välja ka juhud, kus verbile järgneb hoopis prepositsioon nagu näites (38)

(38) Šerif kinnitab, et ükski politseinik ei tulistanud, rehv aga lõhkes siis, kui naine **sõitis üle** muru. (SL Õhtuleht 2003.11.03)

Seega tuvastab *subject_of* reegel ainult osa kõigist verbidest, mille juures otsisõna on subjektiks. Reeglite sarnane rakendumine kõigi otsisõnade puhul on aga omakorda seotud morfoloogilise märgendamise täpsusega. Andmeid analüüsides tuleb arvestada vigadega Estmorfi morfoloogiaanalüsaatori väljundis, kuna analüsaator ei suuda vormihomonüümia tõttu eristada kõikidel juhtudel sõnade *isa* ja *ema* vorme. Estmorf on näites (39) sõna *ema* omastava käände vormi määranud nimetavaks, mistõttu on WordSketchi reeglid selle sõna ekslikult määranud subjektiks. Samas ei ole subjekti tuvastatud lauses (40), sest seal on ekslikult sõnavorm määratud osastavaks.

(39) .. tuli Hargo Johanson kohale (kuigi elas siis Tõrvas) ja võttis **ema** (/S/ema/sg_n) ka kaasa. (Valgamaalane 10.02.2007)

(40) Väikest Margust iseloomustab **ema** (/S/ema/sg_p) kui tõsist last, kes õppis vara lugema. (Eesti Päevaleht 11.09.1999).

Kuigi *subject_of* reegel tuvastab ainult osa kõigist verbidest, mille juures otsisõna on subjektiks, võib öelda, et kui kõigi sõnade jaoks on reeglid samad, siis võib eeldada, et tulemused peaksid olema võrreldavad.

4.4.1.2. Verbikonstruktsioon *object_of*

Verbid, mille suhtes inimest tähistav sõna on objektiks, leitakse viie reegli järgi:

1. reegliga leitakse umbisikulises tegumoes verbid, millele eelneb nimetavas käändes otsisõna.

2. reegliga leitakse käskivas kõneviisis verbid (v.a ainsuse 2.pöörde vormid ja lemmad *olema* ja *andma*), millele järgneb otsisõna nimetavas käändes ning millele ei järgne afiksaaladverbi.

3. reegliga leitakse jaatavas kõnes verbid (v.a lemma *olema*), mis pole käskivas kõneviisis, infinitiivivormides ega umbisikulises tegumoes ja millele võib järgneda omadussõna või nimisõna omastavas käändes ning millele järgneb otsisõna omastavas käändes koos interpunktsioonimärgiga.

4. reegliga leitakse jaatavas kõnes verbid (v.a kesksõna vormid ja lemma *olema*), millele järgneb otsisõna osastavas käändes ja millele ei tohi järgneda verbi, adverbi, kaassõna ega sõnu, mis antud afiksaaladverbide loendina.

5. reegliga leitakse verbid, mis on kindla kõneviisi lihtminevikus ja oleviku umbisikulises tegumoes ja millele eelneb otsisõna osastavas käändes.

Objekti juurde kuuluvate verbide leidmiseks on reegleid rohkem, aga sõnavisandite grammatika tuvastab neid suhteid palju vähem (vt tabel 14). Võiks arvata, et inimest tähistav sõna esinebki lauses sagedamini subjektina. Kui subjektiks

määratakse keskmiselt ligi 15 protsenti otsisõnadest, siis objektina kümme korda vähem, st keskmiselt 1,5 protsenti. Tabelist 15 on näha, et kõige rohkem on objektiks määratud sõna *laps* (2,3% sõna esinemisjuhtudest) ja kõige vähem sõna *ema* (0,5% sõna esinemisjuhtudest). Kui usaldada morfoloogilist analüsaatorit, siis võiks öelda, et sõna *ema* ja *isa* esinevad nende reeglitega tuvastatavatel juhtudel ligi neli korda harvem objektina kui sõnad *laps* ja *inimene*. Siiski annavad sellesse tulemusse suure panuse märgendusprobleemid: vormihomonüümia tõttu eksitakse sõnade *ema* ja *isa* käände määramisel. Seda võib öelda sõnavormide sageduse võrdluse põhjal. Kui sõnade *mees* ja *naine* iga osastava käände vormi kohta esineb korpuses 3,5 nimetava käände vormi ja sõna *laps* puhul isegi ainult 1,5 nimetava käände vormi, siis sõnade *ema* ja *isa* iga osastavaks käändeks määratud vormi kohta tuleb 30 nimetavaks märgitud vormi. Tabel 15 näitab hästi, kuidas just osastavat käänet nõudvate reeglite (4. ja 5. reegel) järgi tuvastatud suhteid on sõna *ema* puhul üle 50 korra vähem tuvastatud.

Tabel 15. Erinevate *object_of* reeglitega tuvastatud verbide hulk, kui sõnad *ema* ja *laps* on objektiks

	ema	laps
1. reegel: otsisõna nimetav + umbisikuline tm	152	828
2. reegel: käskiv kv + otsisõna nimetav	7	65
3. reegel: kindel kv + otsisõna omastav + interpunkt.	21	236
4. reegel: kindel kv + otsisõna osastav	69	3961
5. reegel: otsisõna osastav + umbisikuline tm	9	569
KOKKU	258	5659

Kui üks reeglitest hakkab prevaleerima, siis on tulemus näha väljundis. Esimese reegli suurem osakaal sõnade *ema* ja *isa* verbikonstruktsioonide tuvastamisel tekitas olukorra, kus *object_of* reeglitega leiti enam neid verbe, mis esinesid lausetes umbisikulises tegumoes. Selliste verbivormide esiletõus tingis omakorda selle, et sõnade *ema* ja *isa* kollokaatidena esines palju vägivaldseid tegevusi tähistavaid verbe (vt lisa 5, lehed ema (2) ja isa (2)), nt *tapma, karistama, küüditama, peksma, süüdistama, vahistama, sundima, diskvalifitseerima, hukkama, arreteerima, kahtlustama, represseerima* ja *mõrvama.* Samas kui sõnade *laps* ja *inimene* kollokaatideks oli sellistest verbidest vaid *lööma, sundima, tapma.* Kuna paistis, et nii sõnavormide homonüümiast kui muudest põhjustest tulenev objektisuhete väike tuvastushulk mõjutab tulemust, siis jäeti siinsest tööst *object_of* reegliga saadud verbide täpsem analüüs välja. Kõigi otsisõnade *object_of* reeglitega saadud kollokaadid on esitatud lisas 5.

4.4.2. Peamised tegevused, mille juures inimest tähistav sõna on subjektiks

Selles peatükis võrreldakse kümne inimest tähistava sõna *subject_of* verbikonstruktsioonis esinevat 30 sagedamat verbi omavahel. Seega on vaatluse all kõige sagedamad verbid: iga inimest tähistava sõna 30 sagedama verbi esinemissagedus kokku moodustab keskmiselt 50% selle sõna *subject_of* reegliga leitud kõikide verbide esinemissagedusest. Väiksema esinemissagedusega verbe vaadeldakse peatükis 4.4.3, kus on tähelepanu all antonüümipaarikutele eriomased verbid.

Kokku 300 verbist on erinevaid 62 (vt tabel 16), kuna palju verbe esineb mitme otsisõna kollokaadina. Tabelis esitatud verbide järjekord lähtub sellest, mitme otsisõna kollokaadina verb esineb. 15 esimest verbi esinevad kõigi otsisõnade 30 sagedama kollokaadi seas ja 18 viimast esinevad ainult ühe otsisõna 30 sagedama kollokaadi seas.

Tabel 16. *Subject_of reeglite* järgi tuvastatud sagedamad verbid

	EMA	ISA	LAPS	INIMENE	MEES	NAINE	POEG	TÜTAR	POISS	TÜDRUK	mitmel sõnal esineb	suhteline sagedus kokku
saama	370	312	1312	879	417	511	626	601	620	536	10	6185
ütlema	751	684	64	84	257	341	271	291	297	313	10	3355
pidama	264	236	433	603	291	348	242	266	259	230	10	3172
tulema	185	246	279	257	238	171	300	248	252	241	10	2418
rääkima	339	238	75	105	176	247	87	227	191	228	10	1913
tegema	173	158	159	212	222	183	174	158	269	201	10	1909
käima	132	142	304	132	94	98	331	317	182	176	10	1908
minema	160	162	207	146	186	152	197	245	222	174	10	1851
jääma	98	114	264	145	155	153	237	162	188	162	10	1677
tahtma	169	142	181	254	129	174	132	140	173	152	10	1646
elama	145	167	156	251	81	95	208	288	73	95	10	1558
võima	110	99	270	329	140	176	129	72	99	119	10	1542
võtma	147	137	89	105	160	128	121	101	138	103	10	1228
teadma	137	101	145	221	94	105	95	90	125	95	10	1207
hakkama	83	76	157	184	98	97	103	115	146	142	10	1201
leidma	104	90	80	88	79	86	105		86	88	9	806
küsima	118	103			77	136	113	101	82	90	8	821
tundma	78		130	192	67	71		101	64	75	8	777
surema	166	219	69		112	101	105	112			7	883
vastama	78	88			91	94	76		93	98	7	618
nägema		79	117	102		69	74		76	84	7	600
töötama	121	177		71		87	89	90			6	635

	EMA	ISA	LAPS	INIMENE	MEES	NAINE	POEG	TÜTAR	POISS	TÜDRUK	mitmel sõnal esineb	suhteline sagedus kokku
sõitma		71			87	83	82	65	79		6	466
sündima			342				592	662	85	91	5	1773
õppima			279				295	313	93	96	5	1077
mängima		79	127		75				178	66	5	526
arvama	128	115		97				72		91	5	502
tunnistama	78				117	92			77	90	5	453
panema	86	118			79				97	72	5	452
istuma					92	62	97		72	81	5	404
vaatama			74	79					78	114	4	344
suutma			63	123	77	77					4	340
kasvama			139				105	97			3	342
andma	95	145			73						3	313
meenutama	118					73		65			3	255
tooma		72					76	79			3	227
kinnitama					72	66				79	3	216
pääsema			69					65	66		3	200
jõudma			64		68					65	3	196
lubama	78	68									2	146
viima	92	98									2	190
oskama			87	85							2	173
vajama			186	140							2	327
lõpetama							92	108			2	200
lisama	76										1	76
kasvatama	81										1	81
õpetama		73									1	73
sööma			77								1	77
soovima				63							1	63
kartma				70							1	70
kasutama				75							1	75
muutuma				81							1	81
mõtlema				97							1	97
ostma				105							1	105
väitma					70						1	70
kandma						72					1	72
sünnitama						78					1	78
hukkuma							74				1	74
teatama							76				1	76
abielluma								76			1	76
helistama								101			1	101
jooksma									85		1	85

Sagedamate verbide analüüsiks rühmitati verbid, võttes aluseks Trageli ja Pajusalu tuumverbide liigituse (vt Pajusalu jt 2004) ja lisades sinna siinse töö vajadusest tingitud rühmasid (6. ja 7. rühm). Sel kujul rühmitamine on mõnevõrra tinglik ja lähtub eelkõige sagedamate verbide liigitamise vajadustest. Eriomaste verbide semantika nõuab täpsemat liigitust, seda on kirjeldatud ptk 4.4.3.

Sagedamate verbide analüüsis eristatakse järgnevaid verbigruppe:

1. Modaalset tähendust kandvad verbid: *pidama, võima, saama* jne. Siia gruppi on koondatud modaalverbid ja verbid, mis sagedamini kannavad muu tähenduse kõrval ka modaalset tähendust.

2. Suhtlusverbid: *ütlema, rääkima, küsima* jne. Siia gruppi koondati verbid, mis esinevad kas ainult suhtlusverbidena või esinevad muu tähenduse kõrval sagedamini otsekõne saateverbidena (nt *kinnitama, lisama, meenutama, tunnistama, leidma*).

3. Kognitiivverbid tähistavad erinevaid kognitiivseid protsesse, nt *nägema, tahtma, mõtlema*.

4. Liikumis- ja asendiverbid *jooksma, tulema, istuma* jne.

5. Omandamisverbid *võtma, andma, panema, viima, tooma* jne.

6. Inimese eluetappe tähistavad verbid *sündima, elama, surema* jne.

7. Aktiivseid tegevusi tähistavad verbid *mängima, töötama, tegema* jne.

Ülejäänud verbe, nt *suutma, vajama, oskama, muutuma, lubama* ja *pääsema* ei olnud otstarbekas erinevatesse gruppidesse jaotada.

Modaalset tähendust kandvad verbid. Kuigi *subject_of* reeglitega üritati ahelverbe moodustavaid finiitseid verbivorme loendist välja jätta, pole see siiski eespool kirjeldatud põhjustel õnnestunud ja suure osa ühistest verbidest moodustavad erinevad modaalsust väljendavad verbid: *saama, pidama, võima, jääma, hakkama, tahtma*. Need sõnad esinevad peamiselt ahelverbides, milles finiitverb väljendab infiniidiga tähistatud situatsiooni modaalsust, aspekti, kausatiivsust või viisi, aga mille sisuliseks tuumaks on hoopis infiniit (EKG II: 19). Millist modaaltähendust need tuumverbid võivad kanda, sellest on pikemalt kirjutanud Ilona Tragel ja Renate Pajusalu (Pajusalu jt 2004). Ilma lausekontekstita on aga võimatu määrata, millist tähendust need kannavad, kuna kollokaatide loendis esinevad modaalverbid ilma sisulist tähendust kandva infiniidita. Seetõttu ei hakata pikemalt arutlema, millist rolli need sõnad inimest tähistavate sõnade kollokaatidena mängivad.

Suhtlusverbid. Modaalsust väljendavate verbide kõrval on inimest tähistavate sõnade juures olulisel kohal suhtlusverbid. 30 sagedama kollokaadi hulka kuuluvad verbid *ütlema, rääkima, küsima, vastama, tunnistama, teatama, kinnitama, meenutama, väitma, arvama* ja *lisama.* Kui modaalsust väljendavad verbid esinevad sama suure esinemissagedusega ka näiteks sõna *auto* kollokaatidena, siis suhtlusverbid enam mitte: kollokatsiooniseostes väljendub, et verbaalne väljendusoskus on vaid inimesele omane.

Suhtlusverbidest kõige sagedamad on *ütlema* ja *rääkima*, mis esinevad iga inimest tähistava sõna 30 sagedama verbi hulgas. Enamikel juhtudel esinevad need

sõnad kollokaatide sagedusloendi tipus, ainult sõnade *laps* ja *inimene* juures on selgelt näha, et sõnade normaliseeritud koosesinemissagedused sõnadega *ütlema* ja *rääkima* jäävad mitu korda väiksemaks. Ilmselt on põhjus sõnade *inimene* ja *laps* üldisemas tähenduses, mis ei sobi kokku vajadusega ütlejat täpsemalt (nt sooliselt) määratleda. Otsisõnade mitmus- ja ainsusvormide esinemissageduse võrdluses selgub, et sõnu *laps* ja *inimene* kasutatakse enam siis, kui ütlejaid või rääkijaid on mitu (vt tabel 17).

Tabel 17. Verbide *ütlema* ja *rääkima* juurde kuuluvate nimetavas käändes otsisõnade mitmus- ja ainsusvormide jagunemine

otsisõna	verb	otsisõna ainsuse nimetavas	otsisõna mitmuse nimetavas	mitmusvormide osakaal
inimene	ütlema	169	286	63%
	rääkima	135	389	74%
laps	ütlema	56	63	53%
	rääkima	45	89	66%
mees	ütlema	557	105	16%
	rääkima	356	147	29%
naine	ütlema	362	43	11%
	rääkima	210	58	22%

Tabelist on näha, et sõna *naine* kõigist nimetava käände vormidest, mis koos verbiga *ütlema* esinevad, on vaid 11% mitmuses, samas kui sõna *inimene* vormidest, mis verbiga *ütlema* koos esinevad, on mitmusvorme 63%. Selline jaotus vastab keelefilosoof Paul Grice'i koostööprintsiibi kvantiteedimaksiimile, mis ütleb, et suhtlus peab olema nii informatiivne kui võimalik, aga mitte informatiivsem kui vajalik (Grice 1991). Sugu on ilmselt kategooria, mis kannab endas kõnelejate jaoks olulist informatsiooni ja mida tuleks võimalusel määratleda. Seega, kui tegevuses osalejaid on üks, siis üritatakse teda sooliselt määratleda, aga kui osalejaid on rohkem kui üks, muutub väiksemaks tõenäosus kõnelejaid mingi sooga siduda ja kasutada tuleb sooüleseid (nt *laps*) või isegi soo- ja vanuseüleseid (nt *inimene*) sõnu.

Ilmneb veel üks erinevus: kui kõigi teiste sõnade puhul on sõna *ütlema* sagedam kui *rääkima*, siis sõnade *laps* ja *inimene* puhul on vastupidi. Samuti on kõigi sõnade juures sõna *rääkima* seotud rohkem mitmusevormidega kui *ütlema*. Arvatavasti tuleb see verbide erinevast tähendusest: *ütlema* kasutatakse rohkem konkreetsema ja piiritletuma teatamisakti kohta, samal ajal kui rääkima on verb, mida saab kasutada retsiprookse teatamise ja laiema suhtlusakti kohta. Seega ühekordse teatamise kohta on sobivam kasutada ainsust ja vastastikuse teatamise kohta mitmust. Kui sõna *laps* esinebki sagedamini vaid verbide *ütlema* ja *rääkima* subjektina ja sõna *inimene* juurde kuulub veel verb *arvama*, siis teiste inimest

tähistavate sõnade 30 sagedama kollokaadi seas esineb 5–7 suhtlusverbi. Lisaks sõnadele *ütlema* ja *rääkima* on sagedased veel *küsima*, *vastama, tunnistama, arvama, kinnitama*. Sõna *mees* kollokaadina esineb 30 sagedama seas veel lisaks *väitma*, sõna *ema* kollokaadina *lisama* ja sõna *poeg* kollokaadina *teatama*.

Eluetappe tähistavad verbid. Lisaks suhtlusverbidele on inimest tähistavate sõnade kollokaatide seas olulisel kohal inimese eluetappe kirjeldavad verbid *sündima, elama, kasvama, abielluma, surema* ja *hukkuma*. Kui *elama* esineb kõigi vaatluse all olevate inimest tähistavate sõnade juures, siis *surema* esineb rohkem nende sõnade juures, mis väljendavad sugulus- või lähisuhet, puududes sõnade *inimene*, *poiss* ja *tüdruk* 30 sagedama kollokaadi seast. Suurima normaliseeritud sagedusega esineb *surema* egost kõrgema generatsiooni sugulasi tähistavate sõnade *ema* ja *isa* kollokaadina. Verb *sündima* esineb aga just ainult egost madalamat generatsiooni tähistavate sõnade *laps*, *poeg* ja *tütar* ning samuti mittetäiskasvanuid tähistavate sõnade *poiss* ja *tüdruk* 30 sagedama kollokaadi seas. Sõna *kasvama* esineb kõige sagedamini samuti järglasi tähistavate *laps*, *poeg* ja *tütar* kollokaadina.

Kognitiivverbidest esinevad *nägema, tahtma, mõtlema, teadma, vaatama, tundma, meenutama, õppima, kartma, soovima.* Kõige sagedamini esinevad nägemismeelega seotud *nägema* ja *vaatama*. Mõlemad kollokaadid esinevad sõnade *laps, inimene, poiss, tüdruk* sagedamate kollokaatide seas. Ainult *nägema* esineb sõnade *isa, naine* ja *poeg* kollokaadina. Mõlemad verbid puuduvad aga sõnade *ema, mees* ja *tütar* 30 sagedama kollokaadi seast. Teiste meelte kaudu saadavate aistingutega seostatavaid verbe ei esine 30 sagedama verbi seas üldse. Kaheksa sõna kollokaadina esineb verb *tundma*, mida võib kasutada nii erinevate aistingute kui emotsioonide kirjeldamiseks. Mentaalset tegevust väljendav verb *mõtlema* esineb vaid sõna *inimene* sagedama kollokaadina, samal ajal kui meenutamine on tegevus, mis on omane rohkem naissoole: verb *meenutama* esineb ainult sõnade *ema, naine* ja *tütar* 30 sagedama kollokaadi seas. Tundub, et õppimine on tegevus, millega tegeletakse vaid noorena: *õppima* esineb vaid sõnade *laps, poeg, tütar, poiss* ja *tüdruk* kollokaadina. Need on sõnad, mida kasutatakse tüüpiliselt alaealiste inimeste kohta.

Liikumis- ja asendiverbidest esinevad *jooksma, tulema, minema, käima, sõitma, jõudma, istuma.* Verbid *tulema, minema* ja *käima* esinevad kõigi otsisõnade 30 sagedama kollokaadi seas. Need on muidugi verbid, mis ei tähista ainult asukoha muutust, vaid võivad ahelverbi koostises väljendada tegevuse algust, lõppu või muud tähendust (41).

(41) Näiteks üks naine **läks** mehele **maksma** kümme merekarpidest keed (kokku 12 000 Eesti krooni), kümme prisket siga, 3000 Eesti krooni ja pidustuse külalistele. (Eesti Päevaleht 13.11.1999)

Lausete ülevaatamisel jääb mulje, et peamiselt kasutatakse neid verbe siiski asukoha muutuse tähistamiseks. Kui *tulema* on verb, mis tähistab liikumist kõneleja poole ja *minema* liikumist vastassuunas, siis *käima* peaks hõlmama nii minemist kui tulemist.

Sõnade *isa, mees, naine, poiss, poeg* ja *tütar* 30 sagedama verbi seas esineb ka kollokaat *sõitma*. Seda verbi saab kasutada erinevate sõiduvahenditega liiklemise kohta. Lisaks esineb sõna ainult sõna *poiss* sagedamate verbide juures liikumisverb *jooksma*, mis esinenud liikumisverbidest määratleb kõige täpsemalt liikumise tüüpi. Asendiverbidest esineb sõnade *mees, naine, poiss, tüdruk* ja *poeg* sagedamate kollokaatide seas *istuma*.

Omandamisverbidest esinevad *võtma, andma, panema, viima, tooma*. Sõna *võtma* esineb kõigi sõnade sagedamate verbide seas, esinedes seal sageli erinevates perifrastilistes verbikombinatsioonides (nt *ema võttis ühendust, võttis kohtutee ette, võtab hinge*), mistõttu ei tasu ilma lähemalt lausekonteksti analüüsimata omandamisverbide semantika kohta kaugeleulatuvaid järeldusi teha. Verbide analüüs näitab, et inimestele on võtmine omasem kui andmine (*andmine* esineb vaid sõnade *ema, isa* ja *mees* juures), siiski ei saa teha järeldust skaalal negatiivne-positiivne, kuna andmine ei tähista alati heategu (42).

(42) Võõrad mehed **andsid** kannatanule ühel korral ka **peksa**. (Eesti Päevaleht 18.06.1999)

Omandamisverbidest esineb *tooma* sõnade *isa, poeg* ja *tütar* juures. *Viima* aga ainult *ema* ja *isa* juures. Omandamisverbe esinebki kõige rohkem koos sõnadega *isa* ja *ema*, samal ajal kui sõnade *laps, inimene* ja *naine* juures esineb omandamisverbidest vaid *võtma*.

Aktiivseid tegevusi tähistavad verbid *mängima, töötama, sööma, ostma, kasutama, kandma, tegema, sünnitama, kasvatama, õpetama.* Sõna *mängima* esineb kõige rohkem nende sõnade kollokaadina, mis tähistavad alaealisi: *laps, poiss, tüdruk*. Neist kõige rohkem esineb verb sõna *poiss* kollokaadina (s.o ligi kolm korda sagedamini kui *tüdruku* kollokaadina). Täiskasvanutest mängivad eelkõige meessoost isikud: verb on sage sõnade *isa* ja *mees* kollokaadina, puududes sõnade *ema* ja *naine* 30 sagedama kollokaadi seast. Uuriti lähemalt ka verbi *mängima* esinemist lausekontekstis. Selleks võeti iga otsisõna kohta 20 juhuslikku lauset, kus kollokatsioon esines. Selgus, et kõige rohkem mängitakse sportmänge ja erinevaid pille, lisaks mängitakse näidendites, vähem nimetatakse hasartmängude mängimist, nukkudega mängimist jms. Erinevatel inimest tähistavatel sõnadel esines see kollokatsioon erinevas lauseümbruses, näiteks koos sõnaga *mees* kasutati seda kollokatsiooni sportmängudega seoses kolmveerandil uuritud juhtudest, aga sõna *isa* puhul mainiti seda kollokatsiooni ligi pooltel juhtudel pillimänguga ja veidi vähem näitemänguga seoses. Sõnade *poiss* ja *tüdruk* puhul mainiti ka ligi pooltel juhtudel sportmängude mängimist, vähem kasutati kollokatsiooni pillimängu, hasartmängu ja näitemänguga seoses ning sõna *tüdruk* juures mainiti mitmel korral nukkudega mängimist. Teistest erines kollokatsiooni *laps mängib* kasutus, sest üldjuhul ei kasutata seal sõna mingite sport- või pillimängude kohta, vaid lihtsalt mängimise kui mängulise tegevuse kirjeldamiseks.

Sõna *töötama* esineb just nende inimest tähistavate sõnadega, millega koos *mängima* ei esine (erandiks on sõna *isa, millega koos* mainitakse sageli nii töötamist kui mängimist). Seega need, kes ei mängi, töötavad. Täiskasvanuid

tähistavatest sõnadest puudub *töötama* ainult sõna *mees* 30 sagedama kollokaadi hulgast. Toitumisega seotud verb *sööma* esineb vaid sõna *laps* verbide hulgas. Joomist ei ole ühegi inimest tähistava sõna 30 sagedama verbi seas mainitud. Sõna *ostma* esines vaid sõna *inimene* verbide hulgas.

Sõna *kandma* esineb vaid sõna *naine* sagedamate kollokaatide seas. Lausete lähemal vaatlusel selgub, et naised ei kanna koormaid või kandameid, vaid kolmveerandil juhtudest kantakse riideid ja ehteid. Ainult ühel juhul kahekümnest mainitakse fertiilsusega seotud kollokatsiooni *naine kannab last*.

Laste kasvatamise ja õpetamisega seotud verbidest esineb *kasvatama* ainult sõna *ema* sagedama kollokaadina, samal ajal kui *õpetama* esineb vaid *isa* kollokaadina. Verbi *sündima* kausatiivne variant *sünnitama* esineb vaid sõna *naine* 30 sagedama kollokaadi seas, puududes teiste naissoost isikuid tähistavate sõnade sagedamate kollokaatide hulgast. See on ainus loendis esinev sõna, mille esinemise determineerib bioloogiline sugu: sõna *sünnitama* saab otseses tähenduses esineda ainult naisele viitava sõna kollokaadina.

4.4.3. Antonüümipaaride ja hüponüümipaari eriomased verbid võrdluses

Antonüümi- ja hüponüümipaaride paarikuid võrreldi omavahel, et välja tuua kummalegi paarikule eriomased verbid. Võrdluses osalesid kõigi sõnapaaride 999 sagedamat verbi (vt lisa 5), kuna suuremat hulka ei võimaldanud Sketch Engine'i sõnavisandite programm leida. Kui subjekti *tütar, poeg, tüdruk, ema* ja *isa* juurde kuuluvad verbid mahtusid kõik 999 sagedama verbi hulka, siis sõnade *mees, inimene, laps, naine* ja *poiss* puhul võis osa verbe välja jääda. Seega, kui väidan, et mingit verbi otsisõna juures ei esine, siis tuleb viimaste sõnade puhul arvestada, et mingi hulk harvaesinevatest (peamiselt 1–2 korda esinevatest) kollokaatidest jäi tehnilistel põhjustel loendist välja.

Vaatluse all on iga inimest tähistava sõna 30 sagedamat eriomast verbi, mille normaliseeritud koosesinemissagedus ühe otsisõnaga on neli või enam korda suurem kui tema normaliseeritud koosesinemissagedus (enamasti antonüümse) paarilisega. Normaliseeritud koosesinemissagedus väljendab otsisõna ja verbi koosesinemissageduse suhet verbide üldarvu, mille suhtes otsisõna on subjektiks. See peaks nivelleerima otsisõnade erinevast esinemissagedusest tulenevaid erinevusi. Normaliseeritud koosesinemissagedus $NS(os_1,k_1)$ arvutati valemiga:

$$NS(os_1, k_1 = \frac{S(os_1,k_1)}{S(kn)} * 1000$$

$S(os_1,k_1)$ = otsisõna (os_1) ja tema kollokaatverbi (k_1) absoluutne koosesinemissagedus

$S(k_n)$ = kõikide otsisõna kui subjektiga kollokeeruvate verbide absoluutsagedus

Tabelites 18–22 näitab tabeliveerus *erinevus paarikust* esitatud number, mitu korda on verbi normaliseeritud koosesinemissagedus suurem või väiksem kui tema paariku normaliseeritud koosesinemissagedus sama verbiga. Selle leidmiseks jagatakse omavahel mõlema otsisõna normaliseeritud koosesinemissagedused antud verbiga:

$$\text{erinevus paarikust} = \frac{NS(os_1,k_1)}{NS(os_2,k_1)}$$

Vastuseks saadud number 1 näitab, et erinevust ei ole, sellest suurem number näitab, mitu korda sagedamini esineb verb (võrdluses oma paarikuga) just selle otsisõnaga koos. Numbri asemel olev tärn näitab, et verb ei esine selles korpuses koos teise otsisõnaga ühtegi korda.

Tabelites on esitatud ainult eriomased kollokaadid ehk need, mille normaliseeritud koosesinemissagedus selle otsisõnaga on vähemalt neli korda suurem kui normaliseeritud koosesinemissagedus tema paarikuga. Kui selles peatükis räägitakse mingile sõnale eriomastest verbidest, siis peetakse silmas eriomasust tema antonüümse või hüponüümse paarilise suhtes. Nt verb *võitlema* on eriomane sõnale *poiss*, kuna esineb selle sõnaga koos üle nelja korra sagedamini kui sõnaga *tüdruk*. Samas ei oleks see sõnale *poiss* iseloomulik, kui võrdluses oleksid *poiss* ja *mees*, kuna verbi *võitlema* normaliseeritud koosesinemissagedus mõlema sõnaga on peaaegu võrdne.

Eriomaste verbide grupeerimisel tuli võtta kasutusele kitsamad verbide semantilised grupid kui 30 kõige sagedama verbi analüüsil. Gruppide moodustamisel lähtuti konkreetse materjali analüüsimise vajadusest: püüti luua semantilised grupid, mis koondaksid enda alla suurema osa vaadeldud paarikute 30 sagedamast eriomasest verbist.

1. agressiivseid tegevusi tähistavad verbid, nt *tapma, röövima, lööma*;
2. suhtlusverbid või verbaalset väljendust tähistavad verbid;
3. emotsionaalset käitumist tähistavad verbid, nt *muretsema, kiljatama;*
4. tööde tegemist tähistavad verbid, nt *saagima, kuduma;*
5. mängulisi ja loovaid tegevusi tähistavad verbid;
6. aktiivseid füüsilisi tegevusi tähistavad verbid;
7. vaimseid tegevusi tähistavad verbid;
8. võimu omamisele viitavad verbid, raha kulutamisega seotud verbid;
9. laste kasvatamisega seotud verbid;
10. fertiilsusega seostatavad verbid, nt *sünnitama, viljastama, viljastuma.*

4.4.3.1. Sõnade *poiss* ja *tüdruk* juurde kuuluvad verbid võrdluses

Sõnale *poiss* eriomastest kollokaatidest moodustavad kõige suurema grupi agressiivseid tegevusi tähistavad verbid: *võitlema, tapma, lõhkuma, kaklema, taguma, kiusama, röövima, tabama* ja *sundima*. Lisaks mahuvad 30 *poisile* eriomase verbi

hulka lihtsalt aktiivseid füüsilisi tegevusi tähistavad verbid *tegutsema, loopima, tassima, rabama, vehkima, juhtima, kaitsma* ja *peitma* ning erinevate tööde tegemist tähistavad verbid *ehitama, saagima* ja *lõikama.*

Suhtlusverbidest on poisile eriomased *kõnelema, hõikama* ja *käskima*. Neist viimane tähistab võimupositsioonilt lähtuvat suhtlusakti. Sõnale *poiss* eriomaseid emotsionaalsust väljendavaid verbe ega vaimseid tegevusi tähistavaid verbe pole võimalik välja tuua.

Tabel 18. Antonüümidele *poiss* ja *tüdruk* eriomased kollokaatverbid. Tärniga tähistatakse verbe, mis vastandsõna kollokaatide hulgast puuduvad

	SÕNALE POISS ERIOMASED			SÕNALE TÜDRUK ERIOMASED		
	Verb	Erinevus paarikust	Kollokatsiooni absoluutsagedus	Verb	Erinevus paarikust	Kollokatsiooni absoluutsagedus
1.	võitlema	4,3	29	naeratama	7,2	32
2.	tegutsema	5	27	itsitama	4,5	20
3.	tapma	4	22	kiljuma	23,1	17
4.	lõhkuma	14,7	20	raputama	4,6	17
5.	ehitama	4,4	18	maksma	5,4	16
6.	kaklema	*	12	nentima	5,9	13
7.	põdema	4,4	12	näima	5	11
8.	loopima	8,1	11	sünnitama	*	11
9.	tassima	6,6	9	tajuma	7,5	11
10.	suunduma	6,6	9	kihistama	6,1	9
11.	rabama	*	8	muretsema	4,1	9
12.	taguma	5,9	8	tarvitama	4,1	9
13.	peitma	5,9	8	tohtima	4,1	9
14.	vehkima	5,2	7	tutvustama	4,1	9
15.	kiusama	*	7	avaldama	5,4	8
16.	üllatama	*	7	katma	*	8
17.	kõnelema	5,2	7	loetlema	*	8
18.	kaitsma	5,2	7	mõtisklema	*	8
19.	juhtima	*	7	osutama	5,4	8
20.	puuduma	*	7	sobima	5,4	8
21.	lõdisema	*	6	lehvitama	4,1	6
22.	hõikama	4,4	6	rippuma	4,1	6
23.	rahunema	4,4	6	iseloomustama	*	5
24.	käskima	*	6	jumaldama	6,8	5
25.	röövima	*	6	määratlema	*	5
26.	mahtuma	*	6	rokkima	*	5
27.	lõikama	4,4	6	sädistama	6,8	5
28.	sundima	*	6	värvima	6,8	5
29.	tabama	*	6	anuma	5,4	4
30.	saagima	*	5	ehtima	*	4

Sõnale *tüdruk* eriomastest verbidest moodustavad kõige suurema grupi emotsionaalsusega seotud tegevusi tähistavad verbid *itsitama, kiljuma, kihistama, muretsema, sädistama, jumaldama* ja *rokkima*. Lisaks on *tüdrukule* eriomane verb *naeratama*, mis aga erinevalt eelnevatest ei väljenda otseselt subjekti emotsiooni, vaid annab subjektist pildi kui avatud suhtlejast. Samas on naeratamises kui mitteverbaalses suhtlusaktis peidus soov meeldida. Sellist kuvandit tüdrukust kui teadlikust ihalusobjektist toetavad veel verbid *ehtima* ja *värvima*, mis tähistavad tegevusi, millega tüdrukud püüavad end kaunimaks muuta. Tüdrukule kui seksuaalobjektile viitab aga kõige selgemalt verb *maksma*, mis esineb vaid üksikutel kordadel tähenduses 'millegi eest tasuma', ning omab enamikes lausetes tähendust 'mingit hinda väärt olema', viidates sel viisil tüdrukule mitte kui müügisubjektile, vaid kui prostitutsiooniäri müügiobjektile:

(43) Pääskülas Laulu tänava bordellis **maksab tüdruk** tavaliselt 1000 krooni (Õhtuleht 1999).

Suhtlusverbidest esinevad sõna *tüdruk* eriomaste kollokaatidena *nentima, tutvustama, avaldama, loetlema, iseloomustama* ja *anuma*. Neist viimane tähistab koos verbiga *tohtima* selliseid tegevusi, mille subjektil puudub võim.

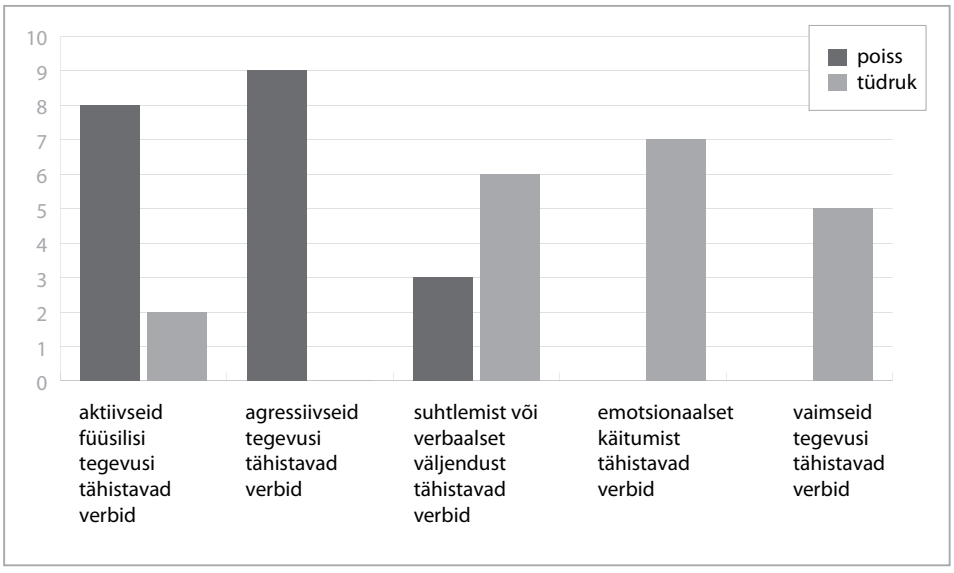
Tüdrukutele eriomaseid agressiivseid tegevusi ei saa välja tuua. Agressiivsemale käitumisele tundub esmapilgul viitavat *raputama*, mis esineb vaid sõna *tüdruk* kollokaatide hulgas. Siiski selgub kollokatsiooni sisaldavaid lauseid üle vaadates, et verbi kasutatakse hoopis passiivse suhtlussituatsiooni kirjeldamiseks: viieteistkümnel esinemisjuhul seitsmeteistkümnest raputavad tüdrukud keeldumise märgiks pead:

(44) **Tüdruk raputas** pead ning pühkis põllenurgaga silmi (Kasemaa, ilukirjandus).

Tüdrukule eriomaseid aktiivseid füüsilisi tegevusi on vähe, välja võib tuua vaid *lehvitama* ja *rippuma*. Samas on *tüdrukule* enam iseloomulikud erinevad vaimseid tegevusi tähistavad verbid: *tajuma, loetlema, mõtisklema, osutama* ja *määratlema.*

Fertiilsusega seotud verbidest saab välja tuua vaid *sünnitama*, mis esineb 30 sagedama *tüdrukule* iseloomuliku verbi seas. See on verb, mis on otseselt bioloogilisest soost tingitud ega saaks seetõttu esineda sõna *poiss* kollokaadina. Sõna *poiss* eriomaste verbide hulgas pole aga esindatud meeste fertiilsusega seotud verbe.

Eriomaste verbide jagunemise semantilistesse gruppidesse võtab kokku joonis 19, millelt on näha, et sõnaga *poiss* kollokeeruvad verbid tähistavad peamiselt agressiivseid või aktiivseid tegevusi, samal ajal kui sõnaga *tüdruk* esinevad rohkem koos suhtlemist, emotsionaalset käitumist ja vaimseid tegevusi tähistavad verbid.



Joonis 19. Antonüümidele *poiss* ja *tüdruk* eriomaste sagedamate kollokaatverbide jagunemine peamistesse semantilistesse gruppidesse

4.4.3.2. Sõnade *mees* ja *naine* juurde kuuluvad verbid võrdluses

Antonüümipaari *mees–naine* verbide võrdlusest hakkab silma, et sõnale *mees* on väga iseloomulikud agressiivseid tegevusi tähistavad verbid (vt tabel 19). Sõnad *tulistama, võitlema, ähvardama, röövima, vägistama, pooma, tabama, vahistama, kaaperdama, sõdima, haavama* ja *purustama* moodustasid 30 sagedamast sõnale *mees* eriomasest verbist 40%, esinedes vähemalt neli korda sagedamini sõna *mees* kui sõna *naine* kollokaadina. Agressiivseid tegevusi tähistavate verbide kõrval esines ka verbe, mis tähistasid aktiivseid füüsilisi tegevusi, nt *treenima, hiilima, rüüpama* ja *vehkima*.

Sõnale *mees* eriomastest suhtlusverbidest esinevad *pomisema, kohmama* ja *käsutama*. Kui verb *käsutama* tähistab verbaalset akti, mille subjektil on ka võim, siis verbid *pomisema* ja *kohmama* tähistavad vastupidiselt verbaalset akti, milles puudub soov aktiivselt teisi mõjutada, veel vähem võimuga sundida.

Tööde tegemist tähistavad kollokaadid. Sõnale *mees* eriomaste verbide hulgas esineb *ehitama*, mille normaliseeritud koosesinemissagedus sõnaga *mees on* üle 16 korra suurem kui sõnaga *naine*. Napilt jääb 30 sagedama mehele eriomase tegevuse hulgast välja saagimine.

Suurima sõnale *naine* eriomase grupi moodustavad verbaalset väljendust või üldisemalt häälitsemist kirjeldavad verbid nagu *kiljuma, kriiskama, sädistama, nuuksuma, ahastama, nähvama, karjatama, manitsema, tänitama*. Neist esimesed kolm kirjeldavad peamiselt hääletooni kõrgust, samas kui neile järgnevad verbid *nuuksuma, ahastama, nähvama* ja *karjatama* on pigem seotud emotsioonide väljendamisega. Loetelus viimaseid verbe *manitsema* ja *tänitama* kasutatakse

situatsioonide kirjeldamiseks, kus kedagi korrale kutsutakse, midagi õpetatakse või seletatakse. Sõna *tänitama* kannab endas lisaks halvustavat hinnangut, verbi kasutades antakse teada, et seletused on tüütud, etteheitvad ja parastavad. Kui sõnaga *mees* sageli koos esinev *käsutama* tähistab tegevust, mille subjektil on võimu, siis verbe *manitsema* ja *tänitama* kasutatakse subjektide juures, kellel on soov sekkuda, aga puudub võim.

Tabel 19. Antonüümidele *mees* ja *naine* eriomased kollokaatverbid. Tärniga tähistatakse verbe, mis vastandsõna kollokaatide hulgast puuduvad

SÕNALE MEES ERIOMASED				SÕNALE NAINE ERIOMASED		
	Verb	Erinevus paarikust	Kollokatsiooni absoluutsagedus	Verb	Erinevus paarikust	Kollokatsiooni absoluutsagedus
1.	tulistama	5,4	141	sünnitama	102,7	203
2.	võitlema	4	138	moodustama	6,2	105
3.	ähvardama	6,1	124	rasestuma	*	37
4.	röövima	4,7	96	kiljuma	*	27
5.	ehitama	16,3	66	kogema	6,1	21
6.	tabama	4,6	56	roolima	6,4	19
7.	vägistama	*	46	ahastama	5,7	17
8.	muhelema	7,1	43	karjatama	4,9	17
9.	vehkima	5,6	34	söandama	4,3	15
10.	kaitsma	4,1	33	kuduma	*	14
11.	pomisema	4,9	30	nuuksuma	*	11
12.	pooma	4,6	28	värvima	4,4	11
13.	vahistama	*	27	õmblema	4,4	11
14.	treenima	5,9	24	libistama	*	10
15.	õnnestuma	*	22	manitsema	*	10
16.	kaaperdama	*	21	toitma	5,1	10
17.	leiduma	*	21	puhkema	*	9
18.	käsutama	*	20	tähtsustama	*	9
19.	valdama	4,9	20	viljastuma	*	9
20.	organiseerima	4,4	18	leinama	5,4	8
21.	rüüpama	4,2	17	tänitama	*	8
22.	liituma	4,2	17	väärtustama	4	8
23.	sõdima	*	15	kriiskama	*	7
24.	haavama	*	15	fantaseerima	*	6
25.	hiilima	*	14	müksama	*	6
26.	purustama	*	14	nähvama	4	6
27.	ennustama	*	14	poseerima	*	6
28.	tikkuma	*	12	puhastama	*	6
29.	kohmama	*	11	sädistama	*	6
30.	komistama	*	11	teadvustama	*	6

Vaimseid tegevusi tähistavatest verbidest esinevad naisele eriomastena *tähtsustama, fantaseerima, väärtustama, teadvustama* ja *kogema*.

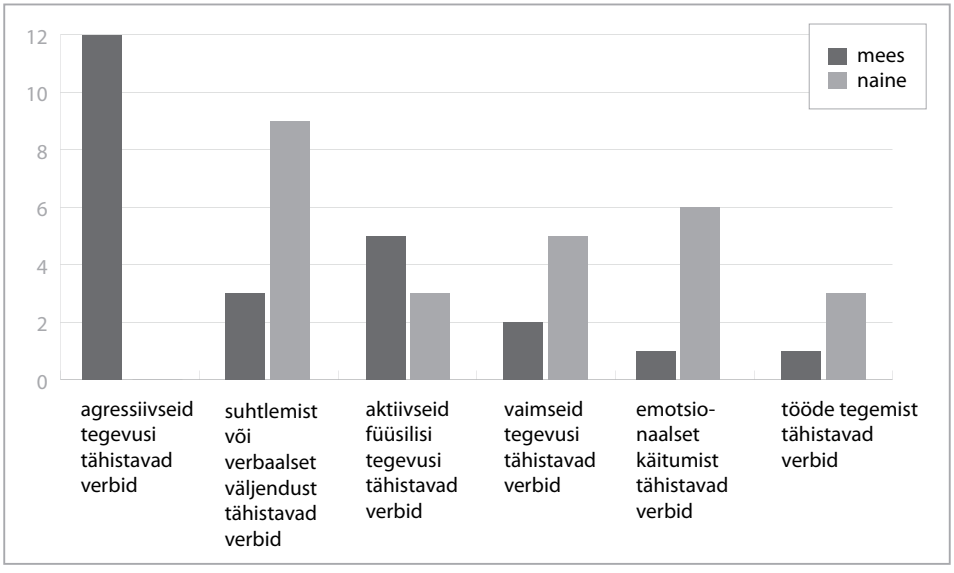
Erinevaid töid tähistavatest verbidest on sõnale *naine* eriomased verbid *kudu-ma, õmblema* ja *puhastama*.

Mõne verbi esinemine *naise* kui subjekti kollokaadina on bioloogilisest soost tingitud. Need on naise fertiilsusega seotud verbid *sünnitama, rasestuma* ja *vil-jastuma*, mis puuduvad sõna *mees* kollokaatide seast üldse või esinevad seal üle 100 korra harvem. Kahe refleksiivse verbi *rasestuma* ja *viljastuma* kausatiivsed variandid *rasestama* ja *viljastama* võiksid teoreetiliselt esineda sõna *mees* kollo-kaatide seas, aga peamiselt ajakirjandustekste sisaldavas korpuses neid siiski ei esine. Lisaks fertiilsusega seostatavale sõnavarale on sõnale *naine* iseloomulike kollokaatide seas järglaste hooldamisele viitav *toitma*.

Naisele kui teadlikule ihalusobjektile viitavad verbid *värvima* ja *poseerima*.

Ühtegi naisele eriomast vägivallategu 30 naisele iseloomulikuma verbi hulgas ei esinenud. Muid aktiivseid tegevusi tähistavad kollokaadid *roolima, müksama, moodustama*.

Kokkuvõtteks on jooniselt 20 näha, et sarnaselt sõnaga *poiss* moodustavad sõna *mees* eriomastest kollokaatverbidest samuti suurima grupi erinevaid agressiivseid tegevusi tähistavad verbid. Väiksem on sõnale *mees* eriomaste verbide seas muid aktiivseid tegevusi tähistavate kollokaatide osakaal, siiski on neid rohkem kui na*i*-sele eriomaste verbide hulgas. Sõnale *naine* eriomased verbid tähistavad eelkõige verbaalset väljendust, vaimseid tegevusi või emotsionaalset käitumist. Jooniste 19 ja 20 võrdlusest on näha, et sõnadele *mees* ja *naine* eriomaste verbide jaotumine semantilistesse gruppidesse on sarnane sõnade *poiss* ja *tüdruk* eriomaste verbide jaotusega, viimaste puhul on erinevused aga veelgi suuremad.



Joonis 20. Antonüümidele *mees* ja *naine* eriomaste sagedamate kollokaatverbide jagune-mine peamistesse semantilistesse gruppidesse

4.4.3.3. Sõnade *poeg* ja *tütar* juurde kuuluvad verbid võrdluses

Sõnadele *poeg* ja *tütar* eriomaste verbide tabel (tabel 20) eristub teistest selle poolest, et eriomaste verbide koosesinemissagedused subjekti funktsioonis otsisõnaga on väga väikesed.

Tabel 20. Antonüümidele *poeg* ja *tütar* eriomased kollokaatverbid. Tärniga tähistatakse verbe, mis vastandsõna kollokaatide hulgast puuduvad

SÕNALE POEG ERIOMASED				SÕNALE TÜTAR ERIOMASED		
	Verb	Erinevus paarikust	Kollokatsiooni absoluutsagedus	Verb	Erinevus paarikust	Kollokatsiooni absoluutsagedus
1.	tapma	*	22	lausuma	8,2	12
2.	kooruma	*	16	kavatsema	4,6	10
3.	seisma	4,4	12	kuulama	4,1	9
4.	võitma	4,4	12	meeldima	*	8
5.	langema	*	10	esinema	*	7
6.	kaduma	5,8	8	julgema	9,6	7
7.	väljuma	*	6	osalema	8,2	6
8.	lendama	4,4	6	muretsema	6,8	5
9.	peksma	*	5	nentima	6,8	5
10.	põgenema	*	5	rahustama	6,8	5
11.	paranema	*	5	väärima	*	5
12.	kannatama	*	5	kõndima	5,5	4
13.	lükkama	*	5	reisima	5,5	4
14.	muutuma	*	5	taipama	*	4
15.	käskima	*	4	eelistama	*	3
16.	puhuma	*	4	igatsema	*	3
17.	tõmbama	*	4	ilmutama	*	3
18.	ehitama	*	4	järgnema	*	3
19.	korraldama	*	4	kahetsema	4,1	3
20.	maksma	*	4	kahtlustama	*	3
21.	mässama	*	3	keelama	4,1	3
22.	virutama	*	3	kiljuma	*	3
23.	sammuma	*	3	loobuma	4,1	3
24.	ihkama	*	3	läbima	4,1	3
25.	hiilima	*	3	lööma	4,1	3
26.	lavastama	*	3	märkima	*	3
27.	meenuma	*	3	rivistama	*	3
28.	hoolitsema	*	3	teadustama	*	3
29.	äratama	*	3	tutvustama	4,1	3
30.	tulistama	*	3	tõstma	*	3

Kõige sagedam *tütrele* iseloomulik verb (st verb, mille subjektina esineb *tütar* vähemalt neli korda sagedamini kui *poeg*) *lausuma* esineb korpuses koos *tütrega* vaid 12 korda ja eriomaste verbide sagedustabeli 15. verb *eelistama* vaid ainult 3 korda. Ka sõnale *poeg* iseloomulike verbide absoluutsagedused on väikesed. Väikesed esinemissagedused ei tule sellest, et sõnadele *poeg* ja *tütar* oleks iseloomulikud väga harvaesinevad sõnad, ega ka sellest, et nende sõnade puhul oleks *subject_of* reeglit raskem rakendada – pigem peitub põhjus selles, et need sõnad esinevad koondkorpuses kümnest vaatluse all olevast inimest tähistavast sõnast kõige harvem. Seetõttu võib tekkida küsimus, kas tabeli lõpupoole esinevad kollokaadid kannavad olulist informatsiooni poistele ja tüdrukutele iseloomulike käitumisviiside kohta või on sinna sattunud juhuslikke koosesinemisi. Siiski võib sõna *poeg* kollokaadiloendis märgata sarnasusi teiste, suurema esinemissagedusega inimest tähistavate sõnade kollokatsiooniloenditega. Sõna *poeg* esineb sarnaselt sõnadega *mees*, *poiss* ja *isa* subjektina agressiivseid tegusid tähistavate verbide juures. Sõna *poeg* kollokaadina esinevad *tapma, peksma, mässama, virutama* ja *tulistama*. Samuti on *pojale* eriomaste verbide loendis meestetööde hulka kuuluv *ehitama*, mis on iseloomulik ka teistele meessoost inimesi tähistavatele sõnadele.

Kõige suurema grupi sõna *poeg* eriomastest verbidest moodustavad erinevad aktiivsetele tegevustele viitavad sõnad *lükkama, tõmbama, väljuma, võitma, korraldama, sammuma, hiilima, puhuma, lavastama, äratama* ja *hoolitsema.* Ühtegi sõnale *poeg* eriomast vaimset tegevust tähistavat verbi 30 sagedama eriomase kollokaadi seas ei esinenud, suhtlusverbidest esineb vaid *käskima*.

Sõna *poeg* kollokaatidest märgivad võimusuhetele allajäämist verbid *langema, kaduma, põgenema* ja *kannatama.* Finantsvõimu omamisele viitab aga verb *maksma*.

Suure koosesinemissagedusega verbid *kooruma* ja *lendama* esinevad vaid juhul, kui *poeg* kannab tähendust 'looma järglane'. Seega ilmnevad iseloomulikes kollokatsioonides ka sõnatähenduste erinevused: kui sõna *tütar* looma järglase kohta ei kasutata, ei saa ka need verbid *tütre* kollokaadina esineda.

Kõige suurema grupi sõnale *tütar* eriomastest kollokaatidest moodustavad suhtlusverbid *lausuma*, *märkima, teadustama, tutvustama, keelama, rahustama, nentima* ja veidi emotsionaalsemat väljendust tähistav *kiljuma*. Samal ajal esineb sõna *poeg* vaid koos suhtlusverbiga *käskima*, mis näitab ühtlasi, et tegevuse sooritajal on võimu. End tahavad kehtestada ka tütred, kes esinevad subjektina verbi juures *keelama*. Erinevuseks on vaid see, et kui pojad sunnivad käskimisega aktiivsele tegevusele (45), siis tütred sunnivad keelates passiivsusele (46).

(45) .. **poeg** käskis tuletõrjujatel kiiresti ukse maha lüüa. (Postimees 1998.04.18)

(46) **Tütar keelas** emal maksta. (Eesti Ekspress 1998)

Passiivsusele viitab veel *tütre* kollokaadina esinev *loobuma*.

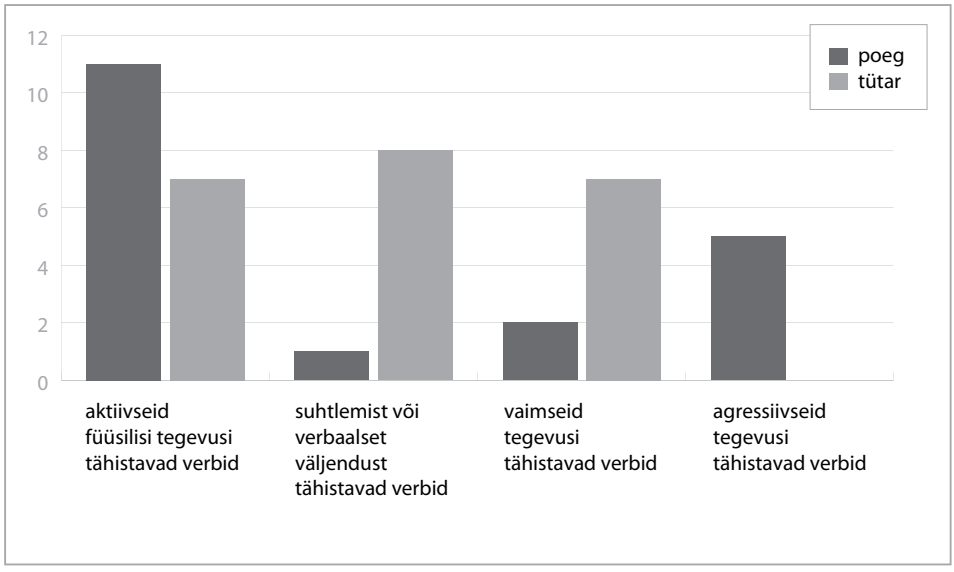
Sõnale *tütar* eriomaseid agressiivseid tegevusi ei saa välja tuua: esmapilgul agressiivset tegevust tähistav *lööma* pole kasutusel otseses löömise tähenduses, vaid piltlikus tähenduses: *tütar lööb üle mehe* või *lööb segi elukorralduse*.

Sõna *tütar* on subjektiks aktiivseid füüsilisi tegevusi tähistavate verbide juures *osalema, kõndima, reisima, järgnema, läbima, rivistama* ja *tõstma*.

Sama suure osa sõnale tütar eriomastest verbidest moodustavad vaimseid tegevusi tähistavad verbid *eelistama, kahetsema, kahtlustama, kuulama, taipama* ja *kavatsema, igatsema.*

Tütrele kui ihalusobjektile viitavad verbid *meeldima* ja *esinema*.

Kokkuvõtteks võib jooniselt 21 näha, et nii sõnale *poiss* kui *tüdruk* eriomastest kollokaatverbidest kuulub suur osa aktiivseid tegevusi tähistavate verbide hulka. Ainult agressiivsed tegevused on iseloomulikud vaid poegadele. Samal ajal nimetatakse tütreid ülekaalukalt rohkem verbaalset väljendust või vaimseid tegevusi tähistavate verbide subjektina.



Joonis 21. Antonüümidele *poeg* ja *tütar* eriomaste sagedamate kollokaatverbide jagunemine peamistesse semantilistesse gruppidesse

4.4.3.4. Sõnade *ema* ja *isa* juurde kuuluvad verbid võrdluses

Bioloogilisest soost tingitud verbidest esinevad ainult sõna *ema* kollokaatidena *sünnitama* ja *imetama* (vt tabel 21). Need on verbid, mille puhul *isa* ei saakski puht bioloogilistel põhjustel subjektina esineda. Verbi *imetama* kõrval moodustavad sõnale *ema* eriomastest kollokaatidest suure osa verbid, mis on seotud laste kasvatamisega: *toitma, jootma, keelitama, tõrelema, tirima* ja *silitama.* Teistest sõnale *ema* iseloomulikest verbidest hakkavad silma kollokaadid, mis tähistavad kodutöid: lisaks laste eest hoolitsemisele üldisi koduseid toimetusi tähistavad verbid *küpsetama, keetma, pakkima, pesema, kuduma* ja *askeldama*.

Tabel 21. Antonüümidele *ema* ja *isa* eriomased kollokaatverbid. Tärniga tähistatakse verbe, mis vastandsõna kollokaatide hulgast puuduvad

SÕNALE EMA ERIOMASED				SÕNALE ISA ERIOMASED		
	Verb	Erinevus paarikust	Kollokatsiooni absoluutsagedus	Verb	Erinevus paarikust	Kollokatsiooni absoluutsagedus
1.	kurtma	5,4	46	teenima	4,0	48
2.	nutma	5,1	36	ehitama	16,3	23
3.	sünnitama	*	35	jalutama	6,6	14
4.	toitma	15,5	22	kostma	5,7	12
5.	pesema	12,7	18	veetma	8,5	12
6.	kahtlustama	6,0	17	pärandama	12,8	9
7.	küpsetama	11,3	16	tulistama	6,4	9
8.	keetma	4,2	12	puhkama	5,7	8
9.	kuduma	*	11	kahtlema	5,7	8
10.	keelitama	7,0	10	tervitama	5,0	7
11.	tirima	7,0	10	sõdima	4,3	6
12.	imetama	*	9	nurisema	*	6
13.	söötma	6,3	9	juhendama	4,3	6
14.	nähvama	5,6	8	ronima	8,5	6
15.	pihtima	*	8	päästma	4,3	6
16.	pakkima	*	8	vägistama	*	5
17.	taipama	5,6	8	edestama	*	5
18.	hüüatama	4,9	7	maalima	*	5
19.	aimama	4,9	7	ujuma	*	5
20.	taotlema	*	7	kiigutama	*	4
21.	tõrelema	4,2	6	patsutama	5,7	4
22.	jootma	*	6	meisterdama	5,7	4
23.	silitama	4,2	6	valdama	5,7	4
24.	solvuma	4,2	6	röövima	*	4
25.	raputama	4,2	6	nautima	5,7	4
26.	laenama	4,2	6	omama	5,7	4
27.	katma	4,2	6	seadma	*	4
28.	askeldama	*	5	tekkima	*	4
29.	piiluma	*	5	hooplema	4,3	3
30.	nuuksuma	*	4	pussitama	*	3

Kolmanda suure grupi moodustavad erinevaid emotsionaalseid tegevusi väljendavad verbid. Eriti sagedad on kurbust väljendavad verbid *kurtma, nutma, nuuksuma*, aga ka muude emotsioonidega seostatavad sõnad *solvuma* ja *nähvama*. Neist mitmed on sageli kasutusel ka saatelausetes suhtlusverbidena, mistõttu *kurtma* ja *nähvama* on koos sõnadega *keelitama, pihtima, hüüatama* ja *tõrelema* liigitatud ka suhtlusverbide alla.

Ainult isadele omaseid tundepuhanguid tähistavaid verbe ei saa välja tuua. Võib-olla väljenduvad vihaemotsioonid vägivallategudes: verbid *tulistama, sõdima, vägistama, röövima* ja *pussitama* esinevad sõna *isa* kollokaadina vähemalt 4 korda sagedamini kui sõna *ema* kollokaadina. Iseloomulike vägivallategude esindatuse poolest ei erine *isa* teistest meessoole viitavatest sõnadest.

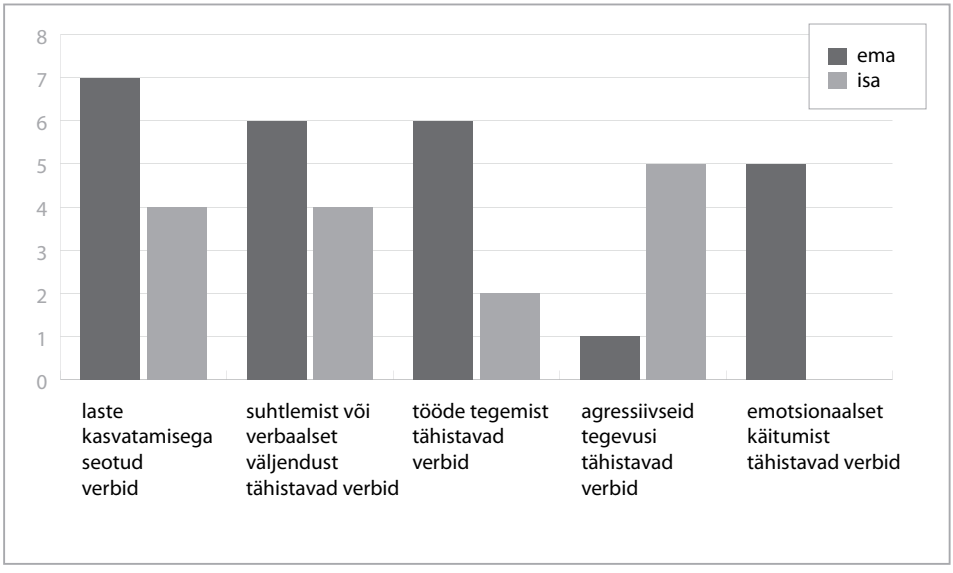
Isale eriomased tööd on erakorralisemad, väga sage on *ehitama*, mis esineb sõnaga *isa* 16 korda sagedamini kui sõnaga *ema*, samuti *meisterdama*, mille esinemissagedus on aga juba väiksem.

Kõige sagedam sõnale *isa* eriomane verb on *teenima*, mis esineb korpuses koos sõnaga *isa* 48 korda. Teenimise kõrval peetakse isade puhul oluliseks ka pärandamist – sellele viitab *pärandama*, mis esineb sõnaga *isa* 13 korda sagedamini kui sõnaga *ema*. Nii sõnad *teenima*, *pärandama* kui *omama* kuuluvad verbide hulka, mis tähistavad võimu või raha omamist.

Suhtlusverbidest kuuluvad isale eriomaste kollokaatide hulka sõnad *kostma, hooplema, nurisema, tervitama.*

Võrreldes laste kasvatamisega seotud verbe, hakkab silma, et isa tegevused lapsega tunduvad verbide põhjal mängulisemad: sõnaga isa kollokeeruvad *kiigutama, patsutama* ja *meisterdama*). Lisaks aga ka *juhendama.*

Antonüümidele *ema* ja *isa* eriomaste verbide hulgas tõusis võrreldes teiste antonüümipaaridega esile laste kasvatamisega seotud eriomaste verbide suur hulk (vt joonis 22) Sõnale *ema* on ülekaalukalt eriomased erinevaid koduseid töid ja emotsionaalset käitumist tähistavad verbid, samal ajal kui sõnale *isa* eriomastest kollokaatidest moodustavad suurima grupi agressiivseid tegevusi tähistavad verbid.



Joonis 22. Antonüümidele *ema* ja *isa* eriomaste sagedamate kollokaatverbide jagunemine peamistesse semantilistesse gruppidesse

4.4.3.5. Sõnade *inimene* ja *laps* juurde kuuluvad verbid võrdluses

Jõudes soovastandusel põhinevate antonüümipaaride analüüsi juurest peamiselt hüponüümiasuhtes sõnapaari *inimene–laps* juurde, on märgata, et iseloomulike verbide absoluutsagedused on palju suuremad kui teistel paaridel (vt tabel 22).

Tabel 22. Sõnadele *inimene* ja *laps* eriomased kollokaatverbid. Tärniga tähistatakse verbe, mis vastandsõna kollokaatide hulgast puuduvad

SÕNALE INIMENE ERIOMASED				SÕNALE LAPS ERIOMASED		
	Verb	erinevus paarikust	Kollokatsiooni absoluutsagedus	Verb	erinevus paarikust	Kollokatsiooni absoluutsagedus
1.	ostma	9,6	650	sündima	19,2	840
2.	mõtlema	5,1	601	õppima	5,2	685
3.	töötama	4,4	439	kasvama	14,6	342
4.	uskuma	4,0	377	mängima	9,6	312
5.	maksma	5,3	332	omandama	4,2	103
6.	helistama	5,6	254	laulma	6,4	89
7.	pöörduma	5,2	224	nutma	5,0	84
8.	lootma	4,9	123	joonistama	*	73
9.	sobima	4,8	122	tutvuma	6,2	32
10.	juhtima	*	90	hüüdma	4,7	26
11.	hääletama	17,5	88	võistlema	7,3	26
12.	teadvustama	5,1	77	kilkama	*	24
13.	jagunema	4,3	76	meisterdama	*	21
14.	jaksama	4,8	73	tassima	5,7	18
15.	vastutama	4,9	62	lustima	*	17
16.	väärtustama	5,1	51	jäljendama	*	15
17.	leppima	4,6	46	põlvnema	4,4	14
18.	arvestama	8,9	45	kisama	17,6	14
19.	tellima	4,4	44	rippuma	7,1	14
20.	lähtuma	7,3	37	lõhkuma	5,0	14
21.	taotlema	4,8	36	sundima	4,4	14
22.	sõltuma	7,1	36	pugema	8,2	13
23.	ümbritsema	*	34	hullama	*	12
24.	seostama	6,7	34	koristama	6,9	11
25.	kandideerima	*	34	jonnima	*	10
26.	kahtlema	4,5	34	matkima	*	10
27.	leiduma	*	32	hulkuma	*	10
28.	vabanema	*	30	võimlema	*	9
29.	annetama	*	29	röökima	*	9
30.	pahandama	5,4	27	kihutama	7,6	9

Peamiselt tuleneb see asjaolust, et nende otsisõnade esinemissagedus on korpuses suurem ja *subject_of reegliga* on saadud rohkem tulemusi (vt 4.4.1). Samas võib tuua võrdluseks, et kui ülejäänud otsisõnadest moodustavad 30 eriomast

verbi *subject_of reegliga* leitud verbidest keskmiselt 3,3%, siis sõna *inimene* 30 eriomast verbi moodustavad juba 6,7% ja sõna *laps* 30 eriomast verbi koguni 11,7%. Seega võib öelda, et hüponüümses suhtes sõnade võrdluses leitud eriomased verbid on suurema esinemissagedusega kui antonüümipaarikutele eriomased verbid. Nii kuuluvad ka tabelis 22 esitatud sõna *inimene* iseloomulikud verbid *ostma, mõtlema* ja *töötama* ning sõna *laps* iseloomulikud verbid *sündima*, *õppima* ja *kasvama* mõlema otsisõna 30 sagedama *subject_of* reegliga tuvastatud verbi hulka (vt ptk 4.4.1).

Sõnale *inimene* on iseloomulikud täiskasvanutele omased tegevused, mis seotud raha teenimise ja kulutamisega, nt verbid *töötama, ostma, maksma, tellima* ja *annetama.* Sõnaga *inimene* esinevad sagedamini koos verbid *juhtima, hääletama, kandideerima, vastutama, taotlema* ja *pahandama,* mis samuti on eelkõige võimu omamise ja teostamisega seostatavad.

Lisaks on sõnale *inimene* iseloomulikud arutlust ja analüüsi tähistavad verbid *mõtlema, väärtustama, teadvustama, arvestama, lähtuma, seostama, kahtlema, uskuma, leppima* ja *lootma.*

Lapse eluetappe kirjeldavad verbid *sündima, kasvama* ja *põlvnema*. Lisaks on sõnale *laps* eriomased õppimise ja arenemisega seostatavad verbid: *õppima, omandama, jäljendama, tutvuma* ja *matkima*. Samuti mänguliste ja loovate tegevuste kohta kasutatavad verbid: *mängima, laulma, joonistama, meisterdama, tassima, lustima* ja *hullama.*

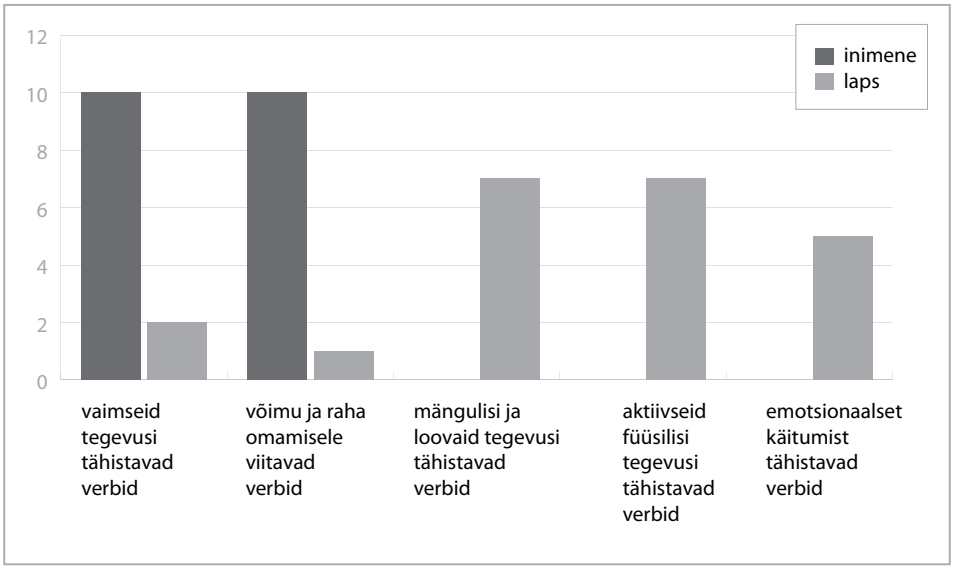
Tekib kontrast inimesele iseloomulike analüüsivate tegevuste ja lastele iseloomulike aktiivsete tegevuste vahel. Sõnale *laps* on eriomased aktiivset liikumist tähistavad kollokaadid *võistlema, rippuma, pugema, koristama, hulkuma, kihutama* ja *võimlema*.

Sõna *laps* juures esinevad vähemalt 4 korda sagedamini emotsioone tähistavad verbid *nutma, kilkama, kisama, jonnima* ja *röökima.*

Kahe *lapsele* iseloomuliku verbi *lõhkuma* ja *sundima* puhul väljendub veel agressiivsus, viimase puhul võiks eeldada võimu taotlemist.

Sõnale *laps* ja *inimene* eriomaste verbide võrdlusest hakkas silma, et kui sõnale *laps* iseloomulikud verbid tähistasid tegevusi, mida võib pidada eelkõige alaealistele omasteks, siis inimese kollokaatidena esinesid verbid, mis tähistasid eelkõige täiskasvanulikke tegevusi nagu töötamine, analüüsimine, ostmine ja juhtimine. Seega, kuigi sõna *inimene* tähenduses on komponent TÄISKASVANU saanud väärtuseks +/- ja ainult sõna *laps* esimene tähendus on määratletud mittetäiskasvanuks, siis tuleb võrdlusest välja, et tüüpiliselt peetakse last alaealiseks ja sõnaga *inimene* tähistatakse eelkõige täiskasvanuid.

Jooniselt 23 on näha, et võrreldes antonüümipaarikutele eriomaste kollokaatverbide jaotumisega, on siin selgemalt eristunud sõnale *inimene* ja sõnale *laps* eriomaste kollokaatide semantilised grupid. Sõna *inimene* on subjektiks peamiselt vaimseid tegevusi tähistavate ning võimu ja raha omamisele viitavate verbide juures, samal ajal kui *laps* on subjektiks mängulisi, aktiivseid tegevusi ning emotsionaalset käitumist tähistavate verbide juures.



Joonis 23. Hüponüümidele *inimene* ja *laps* eriomaste sagedamate kollokaatverbide jagunemine peamistesse semantilistesse gruppidesse.

4.5. Vahekokkuvõte

4.5.1. Soo avaldumine kollokatsioonisuhetes

Sagedamate inimest tähistavate sõnade analüüs kahes leksikogrammatilises konstruktsioonis kinnitas hüpoteesi, et sugudevahelised erinevused avalduvad inimest tähistavate sõnade kollokatsioonides. Inimest tähistavate sõnade kollokaatidena esinevad adjektiivid ja verbid tõid esile eelkõige stereotüüpse nägemuse naissoost ja meessoost inimestest.

4.5.1.1. Aktiivsus privaatsfääris ja ühiskondlikus elus

Kollokatsioonisuhete analüüs kinnitas, et naisi nähakse enam privaatsfääris ja mehi avalikus sfääris aktiivsetena. Suur osa 30st sõnale *ema* eriomasest verbist tähistab rutiinseid kodutöid, nt *küpsetama, keetma, pakkima, pesema* ja *askeldama*. Samal ajal kui isade kanda on erakorralisemad kodutööd, millele viitavad verbid *ehitama* ja *meisterdama*. Selline verbide jaotus peegeldab Eesti ühiskonnas kehtivat tööjaotust, kus naised teevad võrreldes meestega kaks korda rohkem kodutöid. Kuigi tänapäeval teevad naised meestega peaaegu võrdselt palgatööd, on rollijaotus ja stereotüübid säilinud.

Lisaks üldistele kodutöödele on lastega peres peamiselt naiste ülesandeks ka lasteaia ja kooliga suhtlemine, lastega mängimine, nende abistamine ja laste transportimine (Vainu jt 2010: 123). Sõnale *ema* eriomastest kollokaatverbidest moodustavadki suure osa lastega tegelemise ja nende eest hoolitsemisega seotud verbid: *sünnitama, imetama, toitma, jootma,* aga ka *keelitama, tõrelema, tirima, raputama* ja *silitama.* Samas kuuluvad peamiselt isa ülesannete hulka sellised lapsega seostatavad tegevused nagu *jalutama*, *juhendama*, *kiigutama*, *patsutama* ja *meisterdama*. Kollokatsioonid näitavad, et lapse esmavajaduste rahuldamine on meie kultuuris ema kanda ja isa nähakse pigem lapse lõbustaja rollis. Kuigi juhendamine ja õpetamine tunduvad sõnavara järgi olevat samuti pigem isa ülesanded, võib märgata, et isa distsiplineeriv roll on asendumas mängukaaslase rolliga. See vastab üldistele tendentsidele, kuidas isaroll lääne ühiskonnas on muutunud (Burgess 2000).

Seda, et isadelt oodatakse peamiselt rahalist panust ja palgatöö tegemist, näitab verbi *teenima* kasutus, mis on isale eriomastest verbidest kõige suurema esinemissagedusega. Raha teenimine on omakorda eelduseks, et hiljem oleks, mida pärandada: sõna *pärandama* esineb *isa* kollokaadina 12 korda sagedamini kui *ema* kollokaadina. Sõna *töötav* kasutus näitab, et meeste puhul peetakse olulisemaks nende tööalasid. Kuigi nii sõnade *ema* kui *naine* kollokaadina esines *töötav* sagedamini kui meessugu tähistavate sõnade kollokaadina, ilmnesid erinevused, kuidas sõna kasutati. Kui naiste puhul kasutati sageli fraasi *töötav naine/ema* vastandina fraasile *kodune naine/ema*, siis meeste puhul eelnes sõnale

töötav peaaegu alati laiend, mis näitas töökohta või ametit. Selles avaldub, et naiste puhul arvestatakse enam võimalusega, et naine ei tööta, aga meeste puhul vastandus *töötav–kodune* ei ole mõeldav ning oluliseks saab hoopis täpsem määratlus, mis ametikohal või alal mees töötab. Stereotüüpne nägemus tööst endast sisaldab Ameerika kultuuris kodust väljaspool töötamist, mille alla kodutööd ning laste kasvatamine ei kuulu (Lakoff 1987: 80). Sarnast arusaama tööst kinnitavad ka eesti keele kollokatsiooniseosed: korpuses olid vastandatud *kodune ema* ja *töötav ema*. Viimane tähistas väljaspool kodu töötavat lapsevanemat, aga verbide analüüs näitas, et mõlema emarolli juurde kuulub suur hulk kodutöid nagu küpsetamine, keetmine, pakkimine, pesemine jne.

Naiste suuremat seostamist privaatsfääriga näitab asjaolu, et naise puhul peetakse palju olulisemaks märkida, milline on tema perekondlik seis. Sõna *naine* 30 sagedama adjektiivse laiendi hulgas esinevad *vallaline, üksik, endine* ja *uus*. Mehe perekondlikku seisu ei peeta vajalikuks rõhutada: sõna *mees* sagedamate kollokaatide hulgas perekondlikku seisu või abielustaatust näitavaid kollokaate ei esine. Selle asemel esinevad seal sotsiaalset staatust märkivad sõnad *tähtis, rikas* ja *vaene*. Ka meessoost lapsevanema sotsiaalset staatust peetakse oluliseks, seda näitavad sõna *isa* kollokaadid *kuulus, rikas* ja *suur*. Neist esimesed esinevad sõna *isa* kollokaadina isegi 20 ja 30 korda sagedamini kui *ema* kollokaadina. Baker on inglise keelt uurides leidnud, et kui meeste kirjeldamiseks on läbi aegade kasutatud hulgaliselt võimule ja edukusele viitavaid sõnu nagu *rich* 'rikas', *wealthy* 'jõukas', *grand* 'väljapaistev', *famous* 'kuulus', *distinguished* 'silmapaistev', siis naise kirjeldamiseks on neid kasutatud väga harva (Baker 2010: 138). Ilmselt peegeldavad sagedamad adjektiivatribuudid seda, mida kummagi soo esindaja puhul oluliseks peetakse. Naiste puhul peetakse oluliseks mainida, milline on tema suhe mehega – kas tegemist on endise või uue naisega või hoopis vallalise naisega. Mehe puhul peetakse oluliseks määrata, milline on tema sotsiaalne staatus: kas mees on rikas või vaene, ja kas tal on ühiskondlikus elus oluline roll täita.

4.5.1.2. Atraktiivsus ja selle rõhutamine

Adjektiivsete laiendite analüüs näitas, et naiste puhul peetakse olulisemaks välimuse kirjeldamist, samal ajal kui meeste puhul on olulisemad iseloomujooned. Sagedamad atraktiivsust märkivad sõnad *ilus, kaunis, kena* moodustavad sõna *naine* kõigist adjektiivsetest laienditest üle 5%, samad sõnad moodustavad mehe adjektiivsetest laienditest alla 1%. Seega võib öelda, et naise välimuse kirjeldamisel rõhutatakse ilu mitu korda enam. Siiski ei saa öelda, et meeste puhul ilust üldse ei räägitaks, nt sõnade *poiss* ja *mees* 30 sagedama adjektiivse kollokaadi sekka kuulub sõna *ilus*. Paul Baker, kes on sõna *man* 'mees' kollokatsioone uurinud diakroonilisest aspektist, on küll leidnud, et inglise keeles on võrreldes eelmise sajandi tekstidega hakatud üha rohkem meest kirjeldama sõnadega, mis viitavad atraktiivsusele, nt *pretty* 'ilus', *dapper* 'elegantne', *handsome* 'kena'

(Baker 2010: 138). Kas sarnane tendents esineb ka eesti keeles, selle kohta praegu diakroonilised uuringud puuduvad.

Alastiolekut mainitakse samuti naise puhul mitmeid kordi sagedamini, nt *paljas* esineb üle 4 korra sagedamini koos sõnaga *naine* kui koos sõnaga *mees*.

Simone de Beauvoiri (1997: 185) palju tsiteeritud lause „naiseks ei sünnita: naiseks saadakse" saab kinnitust sõnade *tütar* ja *poeg* kollokaatide analüüsil. Tütreid kirjeldatakse eelkõige väliste tunnuste põhjal kui pisikesi ja kauneid, samal ajal kui poegade puhul kirjeldatakse nii välimust (kasutades selleks omadussõna *suur*) kui ka iseloomu. Viimast kirjeldatakse samuti stereotüüpsel viisil: pojad on erinevalt tüdrukutest kanged, vaprad ja ustavad. Sellised andmed toovad esile erinevad ootused, mis järglastele esitatakse. Ilu märgitakse tüdruku puhul neli korda sagedamini kui poisi puhul.

Eriomaste verbide analüüs näitas, et kõigi naissugu tähistavate sõnade v.a *ema* kollokaatide seas esines mõni verb, mis kinnitab stereotüüpset nägemust naisest kui (seksuaal)objektist. Näiteks sõna *tüdruk* sagedamate eriomaste kollokaatide seas esinesid verbid *maksma* 'mingit hinda väärt olema', *värvima*, *ehtima.* Sõna *naine* kollokaatide seas esinesid *värvima* ja *poseerima*. Sõna *tütar* kollokaadina esinesid *meeldima, esinema*.

Lisaks esinevad ainult naisi tähistavate sõnade kollokaatidena reproduktiivseid funktsioone tähistavad verbid *sünnitama, viljastuma, rasestuma* ja fertiilsusega seotud adjektiivid *rase, (last) ootav, imetav*. Meeste fertiilsusega seotud verbe ja adjektiive sagedamate kollokaatide seas ei esinenud.

4.5.1.3. Verbaalne väljendus ja emotsionaalsus

Suhtlusverbe analüüsides selgub, et naissoost isikute verbaalse väljenduse ja häälitsuste kohta on kasutusel rohkem erinevaid verbe. Näiteks kui sõnale *tütar* 30 eriomase kollokaadi seas on suhtlusverbe ja emotsionaalsust väljendavaid verbe 13, siis sõna *poeg* esineb vaid kolme sellist tähendust kandva eriomase verbi subjektina. Nii sõnade *tüdruk, tütar, naine* kui ka *ema* eriomastest kollokaatidest moodustavad suure grupi suhtlusverbid, mis väljendavad lisaks ka seda, kuidas midagi öeldakse. Sageli kannavad need ka halvustavat hinnangut, nt *tänitama, näägutama, kriiskama, kiljuma, nähvama*. Samas on kõigile meessoost subjektidele eriomased verbid *käsutama* ja *käskima*, mis näitavad võimu. Ainult vanemlikul võimupositsioonil olevaid isikuid tähistavad *ema* ja *isa* esinevad võrdselt sageli subjektina verbi *käskima* juures. Seevastu sõna *tüdruk* eriomased kollokaadid *tohtima* ja *anuma* näitavad just võimu puudumist.

4.5.1.4. Ühiskonnas kehtivate võimusuhete avaldumine

Kinnitust leidis hüpotees, et kollokatsioonid peegeldavad ühiskonnas kehtivaid võimusuhteid. Paremat juurdepääsu majanduslikele hüvedele väljendavad

rahalist seisu näitavad adjektiivid ja rahakasutust väljendavad verbid. Adjektiivide analüüsist selgus, et kollokaat *rikas* on eriomane eelkõige meessoost isikuid tähistavatele sõnadele *mees, isa* ja *poeg,* nt *rikas isa* esines 23 korda sagedamini kui *rikas ema*. Ainukese erandina saab välja tuua kollokatsiooni *rikas tüdruk*, mis esineb pisut sagedamini kui *rikas poiss*.

Võim perekonnas sõltub samuti paljuski sellest, kes otsustab rahakasutuse üle. Ressursiteooria kohaselt on suurem võim sellel osapoolel, kes panustab perekonda suurema sissetulekuga (Talves 2011: 111). Võimusuhted ei lähe aga ainult erinevate sugude vahelt. Sõnale *inimene* väga iseloomulikud verbid võrreldes sõnaga *laps* on *ostma* ja *maksma* mis näitavad, et majanduslik võim kuulub täiskasvanutele. *Tüdruku* kollokaadina esinev *maksma* tundub esmapilgul viitavat samuti võimu omamisele, aga lausekonteksti tasandil selgub, et enamikel juhtudel ei viidata tüdrukule kui müügisubjektile, vaid kui prostitutsiooniäri müügiobjektile, mistõttu näitab selle sõna esinemine just täielikku võimu puudumist.

Vägivaldsust ja kuritegelikku käitumist on samuti tõlgendatud kui võimu näitamist (Pearce 2008: 8). Nii sõnade *mees*, *isa*, *poiss* kui *poeg* 30 sagedama eriomase kollokaatverbi seas esineb palju agressiivseid ja vägivaldseid tegevusi tähistavaid verbe. Nt sõna *mees* 30 sagedamast eriomasest kollokaatverbist tähistasid 40% erinevaid agressiivseid tegevusi: *tulistama, võitlema, ähvardama, röövima, vägistama, pooma, tabama, vahistama, kaaperdama, sõdima, haavama* ja *purustama.* Ka sõnale *poiss* eriomastest kollokaatidest moodustavad kõige suurema grupi agressiivseid tegevusi tähistavad verbid *võitlema, tapma, lõhkuma, kaklema, taguma, kiusama, röövima, tabama* ja *sundima*. Vägivallategusid tähistavad verbid ei puudu ka sugulusseosele viitavate sõnade *isa* ja *poeg* eriomaste kollokaatide hulgast.

Suure esinemissagedusega adjektiivsed laiendid *vaimne* ja *püha* sõna *isa* kollokaatidena näitavad, et *isa* esineb sageli metafoorses tähenduses vanema rollis (vaimuliku juhi, eeskuju andjana), samal ajal kui *ema* esineb peaaegu kõigil juhtudel vaid lapsevanemana. Selline erinevus sõnade *ema* ja *isa* kollokatsiooniseostes peegeldab patriarhaalseid ühiskonnasuhteid, kus vaimsed liidrid on enamasti meessoost ning isegi jumalat personifitseeritakse meessoost isikuna.

4.5.2. Vanuse avaldumine kollokatsioonisuhetes

Kui sõnade *inimene, isa, ema, mees* ja *naine* kollokaatide seas on vanusele viitavate adjektiivsete täiendite hulk tühine, siis sõnade *poeg, tütar, poiss, tüdruk* ja *laps* sagedamatest adjektiivatribuutidest moodustavad 80–90% vanusele viitavad täiendid. Loetelust kolm viimast on sõnad, mille komponent TÄISKASVANU on saanud negatiivse väärtuse. Sõnad *mees, naine, isa* ja *ema* on määratletud täiskasvanuks. Sõnade *inimene*, *poeg* ja *tütar* tähenduskomponent TÄISKASVANU saab väärtuseks +/-, mis tähendab, et neil juhtudel ei ole alaealisus või täiskasvanuks olemine sõnatähenduses määratletud.

Sagedamad vanusele viitavad adjektiivid näitavad, milline on otsisõnaga tähistatud isiku tüüpiline vanus. Nt vanusele viitavad *aastane*-lõpulised sõnad ei jagune ühtlaselt kõigi inimest tähistavate sõnade vahel, vaid teatud vanusegruppe tähistavad adjektiivid kuuluvad kindlate otsisõnade kollokaatide hulka. Poiste ja tüdrukute puhul nimetatakse enim teismeliste eagruppi, tüüpiliste poegade ja tütardena nähakse aga rohkem alla kümne aasta vanuseid lapsi, samas kui mehe ja naise kirjeldamiseks kasutatakse peamiselt keskeale viitavaid sõnu. Kuigi sõnade *poeg* ja *tütar* tähendused ei sisalda vanuselisi piiranguid (tähenduskomponent TÄISKASVANU on määratletud +/-) ja mõlema sõna kollokaatide hulgas on ka üks *aastane*-lõpuline sõna, mis viitab täiskasvanueale, siis kollokatsiooniseostele tuginedes võib väita, et tüüpilised *poeg* ja *tütar* on alaealised, kusjuures vanuse märkimine osutus nende sõnade puhul üliolulisek s.

Sõnade *inimene*, *ema* ja *isa* 30 sagedama adjektiivatribuudi sekka polnud jõudnud ühtki vanust täpsemalt määratlevat *aastane*-lõpulist liitsõna.

Nii adjektiivide kui verbide põhjal võib öelda, et õppimine on vaid alaealistele omane tegevus. Sõna *õppiv* kuulub ainult sõnade *poiss, tüdruk, poeg, tütar* ja *laps* kollokaatide hulka. Seda, et õppimine on omane vaid mittetäiskasvanutele, näitab ka verbide analüüs. Kollokaat *õppima* esineb vaid sõnade *laps, poiss, tüdruk, poeg* ja *tütar* 30 sagedama *subject_of* reegliga tuvastatud verbi seas.

KOKKUVÕTE

Väitekirjas analüüsiti sõnade *ema, inimene, isa, laps, mees, naine, poeg, poiss, tüdruk* ja *tütar* kollokatsioone, et teada saada, kas ja kuidas sõna tähendusaspektid kollokatsioonides avalduvad. Väitekirja kahele teoreetilisele peatükile järgneva analüütilise, korpusuuringul põhineva osa saab jagada kaheks:

1. Kolmandas peatükis tegeleti eelkõige inimest tähistavate sõnade üldise distributiivse käitumise uurimisega. Erinevate sõnade 30 sagedamat vasakule ja 30 sagedamat paremale hargnevat lihtkollokatsiooni võrreldes ja analüüsides otsiti seoseid sõnade tähenduskomponentide ja kollokatsioonikobarate kattumise vahel.

2. Neljandas peatükis uuriti, kuidas sõnadega *ema, inimene, isa, laps, mees, naine, poeg, poiss, tüdruk* ja *tütar* tähistatud isikute vanus ja sugu kajastuvad kollokatsiooniseostes. Lähtuti sellest, et kollokatsioonimustrid peaksid paljastama sotsiaalseid hoiakuid, eriti soo ja vanusega seotud stereotüüpe. Vaadeldi täpsemalt inimest tähistavate sõnade adjektiivseid laiendeid ja verbe, mille suhtes inimest tähistavad otsisõnad esinesid subjektina. Tähelepanu all oli kummaski konstruktsioonis esinevad 30 sagedamat kollokaati, lisaks toodi välja vastanduvate paarikute 30 sagedamat eriomast verbi. Kollokatsiooniseoseid analüüsiti peamiselt sõnapaaride kaupa kahe otsisõna kollokaate kõrvutades.

Kolmandas peatükis selgus, et suur osa inimese semantilisse välja kuuluvate sõnade vasakule ja paremale hargnevatest kollokaatidest kattusid: kõigi uuritud inimest tähistavate sõnade 30 vasakpoolsest kollokaadist kattus keskmiselt 43–57% ja parempoolsetest koguni 50–60% teiste inimest tähistavate sõnade sagedamate kollokaatidega. Süntagmaatiliste seoste analüüs näitas, et inimese sõnaväljas moodustavad keskme sõnad *mees* ja *naine*, millel on kõige suurem kollokatsioonikobarate kattumine teiste inimest tähistavate sõnadega.

Kattuvas osas esines kollokaate, mis olid ühised kõigile kümnele otsisõnale. Kõigi inimest tähistavatel sõnade 30 sagedama vasakpoolse kollokaadi seas esinesid 8 sõna: *üks, kaks, kolm, mina, tema, see, olema* ja *oma*. Paremale hargnevatest kollokaatidest esinesid kõigi otsisõnade 30 sagedama kollokaadi seas 7 sõna: *mina, minema, olema, oma, saama, see* ja *tema.* Seega suur osa inimest tähistavate sõnade kollokaatidest kattusid.

Kollokatsioonikobarate ja tähendusväljasiseste suhete võrdluses leiti, et iga tähenduskomponent eraldi ei omanud üldisele jaotusele mõju. Andmetest lähtuvalt ei saanud näiteks väita, et otsisõna tähenduskomponendile MEESSOOST antud sama väärtus oleks tinginud kollokatsioonikobarate suurema kattumise. Siiski avaldus tendents, et suurema kollokaatide kattuvusega otsisõnade tähenduses esines rohkem sama väärtusega tähenduskomponente. Peaaegu maksimaalselt avaldus see tendents vasakule hargnevate kollokatsioonide puhul. Leidis kinnitust hüpotees, et mida rohkem sarnase väärtusega tähenduskomponente kahel sõnal esineb, seda suurem on kollokatsioonikobarate kattumine.

Kolmandast peatükist selgus veel, et lause süntaktiline struktuur mõjutab kollokaatide esinemist: vasakpoolsete ning parempoolsete kollokaatide sõnaliigiline kuuluvus on erinev. Kõige suurem erinevus oli adjektiivide esinemissageduses – kui adjektiivid moodustasid vasakule hargnevatest kollokaatidest 17%, siis otsisõnast paremale jäi neid vaid 2%. Samuti suur erinevus oli verbide paiknemises: verbid moodustasid vasakpoolsetest kollokaatidest 24% ja paremale hargnevatest juba poole rohkem ehk 43%. Seega võib öelda, et kuigi kollokatsioonid on leksikaalsed süntagmaatilised suhted, sõltub nende esinemisvõimalus suurel määral ka keele grammatikast: süntaksist ja sõnajärje mallidest. Kinnitust leidis hüpotees, et keele grammatikast sõltub piirkond, kust mingit tüüpi kollokatsiooniseoseid tasub otsida**.** See tähelepanek osutub oluliseks, kui soovitakse inimest tähistavate sõnade kollokatsioonisuhteid eri keeltes võrrelda.

Paradigmaatilistest suhetest avaldusid kollokatsioonides peamiselt antonüümsed suhted. Kõik antonüümses suhtes sõnapaarid nagu *mees–naine, tütar–poeg, tüdruk–poiss, ema–isa*, mille osalised erinesid teineteisest komponendi MEESSOOST väärtuse poolest, omasid kõige rohkem ühiseid vasakpoolseid kollokaate. Sellega leidis kinnitust ka Sahlgreni väide, et sõnade süntagmaatilist konteksti, täpsemalt kollokatsioone võrreldes võib teha järeldusi paradigmaatiliste suhete (siinses töös peamiselt antonüümia) kohta. Kui antonüümiat võtta kui sarnasusseost – kahe sõna kõigist tähenduskomponentidest antakse ju vastandlik väärtus vaid ühele komponendile – siis võib nõustuda Harrisega, et sõnade sarnasus avaldub nende konteksti kokkulangevuses.

Kolmandas peatükis jõuti järeldusele, et süntaktilist struktuuri arvestamata on raske eristada verbe, mille suhtes on otsisõna subjektiks, neist verbidest, mille suhtes otsisõna on objektiks. Seetõttu peeti 4. peatükis vajalikuks kollokatsiooniseoseid vaadelda leksikogrammatilisi malle appi võttes, mis annavad aimu, milline süntaktiline seos võiks otsisõna ja kollokaadi vahel esineda.

Neljandat peatükki alustati inimest tähistavate sõnade adjektiivsete laiendite analüüsiga (vt 4.3). Uurimus näitas, et väga olulise osa moodustavad inimest tähistavate sõnade adjektiivsetest laienditest vanusele viitavad sõnad. Eriti suure osa moodustavad need sõnade *poeg, tütar, poiss, tüdruk* ja *laps* kollokaatidest. Kui sõna *laps* sagedamatest adjektiivsetest laienditest tähistab vanust 57%, siis sõnade *poeg* ja *tütar* adjektiivsetest laienditest tähistab vanust koguni 86%. Kuigi sõnade *poeg* ja *tütar* tähendused ei saanud komponentanalüüsis komponendile TÄISKASVANUD negatiivset väärtust nagu sõnad *poiss, tüdruk* ja *laps*, võib kollokaatide põhjal öelda, et tüüpiline poeg ja tütar on alaealised. Kuigi sõnade *poeg* ja *tütar* ning *poiss* ja *tüdruk* kollokatsioonikobarate kattumine oli väga suur, leidus siiski ka juba nende puhul kollokaate, mis tõid esile soostereotüüpse nägemuse: poissi kirjeldati sõnadega *pikk, paha* ja *suur* ning tüdrukut adjektiividega *ilus* ja *kena*. Tüdrukute kõrval peetakse ka naiste puhul olulisemaks välimuse kirjeldamist, samal ajal kui meeste puhul on olulisemad iseloomujooned. Sagedamad atraktiivsust märkivad sõnad *ilus, kaunis, kena* moodustavad sõna *naine* kõigist adjektiivsetest laienditest üle 5 protsendi, samad sõnad moodustavad mehe

adjektiivsetest laienditest alla 1 protsendi. Seega võib öelda, et naise välimuse kirjeldamisel rõhutatakse ilu mitu korda enam.

Lisaks atraktiivsuse rõhutamisele peetakse naise puhul palju olulisemaks märkida, milline on tema perekondlik seis. Sõna *naine* 30 sagedama adjektiivse laiendi hulgas esinevad *vallaline, üksik, endine* ja *uus*. Mehe perekondlikku seisu ei peeta vajalikuks rõhutada: sõna *mees* sagedamate kollokaatide hulgas perekondlikku seisu või abielustaatust näitavaid kollokaate ei esine. Selle asemel esinevad seal sotsiaalset staatust märkivad sõnad *tähtis, rikas* ja *vaene*. Ka isa sotsiaalset staatust peetakse oluliseks, seda näitavad kollokaadid *kuulus, rikas* ja *suur.* Neist esimesed esinevad sõna *isa* kollokaadina isegi 20 ja 30 korda sagedamini kui *ema* kollokaadina. Naiste puhul peetakse oluliseks mainida, milline on tema suhe mehega – kas tegemist on endise või uue naisega või hoopis vallalise naisega. Mehe puhul peetakse oluliseks määrata, milline on koht sotsiaalsel redelil: kas mees on rikas või vaene, ja kas tal on ühiskondlikus elus oluline roll täita.

Inimest tähistavate sõnade 30 sagedama verbi semantilisest analüüsist selgus, et modaalset tähendust kandvate verbide kõrval moodustavad olulise grupi suhtlusverbid, kognitiivverbid, liikumis- ja asendiverbid, aga ka inimese eluetappe tähistavad verbid jpm. Kui modaalsust väljendavad verbid esinevad sama suure esinemissagedusega ka näiteks sõna *auto* kollokaatidena, siis suhtlusverbid enam mitte: kollokatsiooniseostes väljendub, et verbaalne väljendusoskus on vaid inimesele omane. Lisaks suhtlusverbidele on inimest tähistavate sõnade kollokaatide seas olulisel kohal inimese eluetappe kirjeldavad verbid *sündima, elama, kasvama, abielluma, surema* ja *hukkuma*. Kui *elama* esineb kõigi vaatluse all olevate inimest tähistavate sõnade kollokaadina, siis *surema* kollokeerub rohkem nende sõnadega, mis väljendavad sugulus- või lähisuhet, puududes sõnade *inimene*, *poiss* ja *tüdruk* 30 sagedama kollokaadi seast. Kollokatsiooniseoste analüüsist selgus, et õppimine on tegevus, millega täiskasvanud ei tegele: *õppima* esineb vaid sõnade *laps, poeg, tütar, poiss* ja *tüdruk* 30 sagedama kollokaadi seas. Seega juba kõige sagedamate verbide analüüsist selgus, et kuigi suur osa sagedamatest kollokaatidest kattusid, esines ka erinevusi. Järgnevalt uuriti aga eriomaseid verbe, st verbe, mis esinesid ühe inimest tähistava sõna kollokaadina vähemalt neli korda sagedamini kui tema paariku kollokaadina.

Eriomaste kollokaatverbide analüüs kinnitas, et naisi nähakse enam privaatsfääris ja mehi avalikus sfääris aktiivsetena. Suur osa 30st sõnale *ema* eriomasest verbist tähistab rutiinseid kodutöid, nt *küpsetama, keetma, pakkima, pesema* ja *askeldama*. Samal ajal on isade kanda erakorralisemad kodutööd, millele viitavad verbid *ehitama* ja *meisterdama*. Selline verbide jaotus peegeldab Eesti ühiskonnas kehtivat tööjaotust, kus naised teevad võrreldes meestega kaks korda rohkem kodutöid. Kuigi tänapäeval teevad naised meestega peaaegu võrdselt palgatööd, on rollijaotus ja stereotüübid säilinud.

Sõnale *ema* eriomastest kollokaatverbidest moodustavad suure osa lastega tegelemise ja nende eest hoolitsemisega seotud verbid: *sünnitama imetama, toitma, jootma,* aga ka *keelitama, tõrelema, tirima, raputama* ja *silitama.* Samas

peamiselt isa ülesannete hulka kuuluvad lapsega seostatavad tegevused *jalutama*, *juhendama*, *kiigutama*, *patsutama* ja *meisterdama*. Kollokatsioonid näitavad, et lapse esmavajaduste rahuldamine on meie kultuuris ema kanda ja isa nähakse pigem lapse juhendaja ja lõbustaja rollis.

Seda, et isadelt oodatakse peamiselt rahalist panust ja palgatöö tegemist, näitavad sõnale *isa* eriomased verbid *teenima* ja *pärandama*. Sõna *töötav* kasutus näitab, et meeste puhul peetakse olulisemaks nende tööalasid. Kuigi nii sõnade *ema* kui *naine* kollokaadina esines *töötav* sagedamini kui meessugu tähistavate sõnade kollokaadina, ilmnesid erinevused, kuidas sõna kasutati. Kui naiste puhul kasutati sageli fraasi *töötav naine* vastandina fraasile *kodune naine*, siis meeste puhul eelnes sõnale *töötav* peaaegu alati laiend, mis näitas töökohta või ametit. Selles avaldub, et naiste puhul eeldatakse enam võimalust, et naine ei tööta, aga meeste puhul vastandus *töötav–kodune* ei ole mõeldav ning oluliseks saab hoopis täpsem määratlus, mis ametikohal või alal mees töötab.

Eriomaste verbide analüüs näitas, et kõigi naissugu tähistavate sõnade (v.a *ema*) kollokaatide seas esines mõni verb, mis kinnitab stereotüüpset nägemust naisest kui (seksuaal)objektist. Näiteks sõna *tüdruk* sagedamate eriomaste kollokaatide seas esinesid verbid *maksma* 'mingit hinda väärt olema', *värvima*, *ehtima.* Sõna *naine* kollokaatide seas esinesid *värvima* ja *poseerima*. Sõna *tütar* kollokaadina esinesid *meeldima, esinema*.

Lisaks esinevad ainult naisi tähistavate sõnade kollokaatidena reproduktiivseid funktsioone tähistavad verbid *sünnitama, viljastuma, rasestuma* ja fertiilsusega seotud omadussõnad *rase, (last) ootav, imetav*. Meeste fertiilsusega seotud verbe ja adjektiive sagedamate kollokaatide seas ei esinenud.

Suhtlusverbide analüüs kinnitas stereotüüpset nägemust, et naisi nähakse emotsionaalsemate ja verbaalselt võimekamatena kui mehi. Naissoost isikute verbaalse väljenduse ja häälitsuste kohta on kasutusel rohkem erinevaid verbe. Näiteks kui sõna *tütar* 30 eriomase kollokaadi seas on suhtlusverbe ja emotsionaalsust väljendavaid verbe 13, siis sõna *poeg* esineb vaid kolme sellist tähendust kandva verbi subjektina. Nii sõnade *tüdruk, tütar, naine* kui *ema* eriomastest kollokaatidest moodustavad suure grupi suhtlusverbid, mis väljendavad lisaks ka seda, kuidas midagi öeldakse. Sageli kannavad need ka halvustavat hinnangut, nt *tänitama, näägutama, kriiskama, kiljuma, nähvama*. Samas on kõigile meessoost subjektidele eriomased verbid *käsutama* ja *käskima*, mis näitavad võimu. Ainult vanemlikul võimupositsioonil olevad subjektid *ema* ja *isa* omavad võrdselt sõna *käskima*.

Kollokatsioonid peegeldavad ühiskonnas kehtivaid võimusuhteid. Paremat juurdepääsu majanduslikele hüvedele väljendavad rahalist seisu näitavad adjektiivid ja rahakasutust väljendavad verbid. Adjektiivide analüüsist selgus, et kollokaat *rikas* on eriomane eelkõige meessoost isikuid tähistavatele sõnadele *mees, isa* ja *poeg.* Võimusuhted ei lähe aga ainult erinevate sugude vahelt. Kui võrrelda sõnu *laps* ja *inimene*, on viimasele väga iseloomulikud verbid *ostma* ja *maksma*, mis näitavad, et majanduslik võim kuulub täiskasvanutele. Vägivaldsust

ja kuritegelikku käitumist on samuti tõlgendatud kui võimu näitamist. Nii sõnade *mees*, *isa*, *poiss* kui *poeg* 30 sagedama eriomase kollokaatverbi seas esineb palju agressiivseid ja vägivaldseid tegevusi tähistavaid verbe. Nt sõna *mees* 30 sagedamast eriomasest verbist tähistasid 40% erinevaid agressiivseid tegevusi: *tulistama, võitlema, ähvardama, röövima, vägistama, pooma, tabama, vahistama, kaaperdama, sõdima, haavama* ja *purustama.*

Neljanda peatüki tulemused kinnitasid hüpoteesi, et sugudevahelised erinevused avalduvad inimest tähistavate sõnade kollokatsioonides. Inimest tähistavate sõnade kollokaatidena esinevad adjektiivid ja verbid tõid esile eelkõige stereotüüpse nägemuse naissoost ja meessoost inimestest. Samas ei tohi unustada, et erinevustele vaatamata on kõigil inimest tähistavatel sõnadel suur ühiste kollokaatide osa, mis avaldus nii kolmandas kui neljandas peatükis just kõige sagedamate kollokaatide võrdluses.

Raske on kindlaks teha, kas erinevused sõnade kollokatsioonisuhetes peegeldavad tegelikkust (kurjategijatest moodustavad enamiku mehed, mistõttu verbid *tapma* ja *röövima* esinevadki ootuspäraselt mitmeid kordi sagedamini koos mehi tähistavate sõnadega), ühiskonnas kehtivaid soorolle (nt sõnad *mees, poeg, poiss* ja *isa* esinevad sagedamini koos verbiga *ehitama* ja *naine* esineb koos verbiga *kuduma*) või selliseid stereotüüpe, mis tegelikkusele ei vasta (sõnad *naine*, *ema, tüdruk, tütar* esinevad sagedamini koos verbaalset väljendust tähistavate verbidega, kuigi uuringud ei kinnita, et naised oleksid verbaalselt aktiivsemad). Pigem on kollokatsioon abstraktsioon ja lihtsustus, mis näitab, kuidas tüüpiliselt inimesi mingis kultuuris mingil ajahetkel nähakse ja kujutatakse. Kollokatsioon kui sage koosesinemine on üldistus, mis ainult ei peegelda kuultuuris kehtivaid üldistusi ja stereotüüpset nägemust, vaid läbi kollokatsioonimustrite luuakse uusi stereotüüpe. Eelkõige luuakse neid läbi massimeedia. Ka käesoleva töö järeldused põhinevad korpustel, millest suurima osa moodustavad ajakirjandustekstid. Nendes pääseb omakorda mõjule peamiselt täiskasvanud ja enamasti haritud inimese keelekasutus. Võimalik, et ilukirjandustekstide või internetikeele suurema kaasatuse puhul uurimismaterjali oleks analüüsi tulemused erinenud. Siiski võib öelda, et väitekiri on esimeseks katseks tuua kollokatsioonimustrite põhjal välja inimest tähistavate sõnade omavahelisi seoseid ja peidetud tähendusi eesti keeles.

KIRJANDUS

Aavik, Johannes 1932. Keeleline käsiraamat. Uuenduste ning ametliku ja uuendusliku keele erinevuste tundmaõppimiseks. Tallinn, Tartu: Istandik.

Ahven, Andri; Kruusmaa, Kätlin-Chris 2013. Vangistus ja kriminaalhooldus – Kuritegevus Eestis 2012. Tallinn: Justiitsministeerium.

Apresjan, Jurij D. 1974. Regular Polysemy. – Linguistics 142, 5–32.

Arppe, Antti 2001. Corpus-Based Observations on the Use and Occurrence of Inflected Word Forms in Finnish – the Case of Synonyms. – Nordic and Baltic Morphology: Papers from a NorFA Course. Joensuu: Joensuun Yliopistopaino.

Arppe, Antti 2002. The usage patterns and selectional preference of synonyms in a morphologically rich language. – Proceedings of the 6th International Conference on Textual Data Statistical Analysis (JADT 2002). Vol. 1 s. 21–32. Rennes: Inria.

Baker, Paul 2006. Using corpora in discourse analysis. Bloomsbury Publishing.

Baker, Paul 2010. Will Ms ever be as frequent as Mr? A corpus-based comparison of gendered terms across four diachronic corpora of British English. – Gender and Language 1, 125–149.

Bartsch, Sabine 2004. Structural and Functional Properties of Collocations in English. A corpus study of lexical and pragmatic constraints on lexical co-occurrence. Tübingen: Gunter Narr Verlag.

Beauvoir, Simone de 1997. Teine sugupool. Tallinn: Vagabund.

Benson, Morton; Benson, Evelyn; Islon, Robert 1997. The BBI combinatory dictionary of English. Amsterdam, Philadelphia: John Benjamins.

Berlin, Brent; Kay, Paul 1969. Basic Color Terms: Their Universality and Evolution. Berkeley: University of California Press.

Biber, Douglas; Conrad, Susan; Reppen, Randi 1998. Corpus linguistics. Investigation language structure and use. Cambridge University Press.

Burgess, Adrienne 2000. Isarolli muutumine. Tallinn: Varrak.

Butler, Judith 1990. Gender trouble. Feminism and the subversion of identity. New York, London: Routledge.

Caldas-Coulthard, Carmen Rosa; Moon, Rosamund 2010. ‘Curvy, hunky, kinky’: Using corpora as tools for critical analysis. Discourse & Society 21/2, 99–133.

Church, Kenneth; Hanks, Patrick 1989. Word association norms, mutual information, and lexicography. – Proceedings of the 27th Conference of the Association for Computational Linguistics ACL'89. Association for Computational Linguistics, 76–83.

Cooper, William E.; Ross, John Robert 1975. World order. – Papers from the parasession on functionalism. Robin E. Grossman et al. (eds). Chicago, Illinois: University of Chicago, 63–111.

Coseriu, Eugenio 1967. Teoria del lenguaje y lingüistica general: cinco estudios. Madrid: Gredos.

Cowie, Anthony P. 1981. The treatment of collocations and idioms in learners' dictionaries. Applied linguistics 2(3), 223– 235.

Cruse, David Alan 2000. Meaning in Language. An Introduction to semantics and pragmatics. New York: Oxford University Press.

Cruse, David Alan 2001 [1986]. Lexical semantics. Cambridge: Cambridge University Press.

Crystal, David 2011. Internet linguistics: a student guide. London, New York: Routledge.

Dixon, Robert M. W. 1982. Where have all the adjectives gone?: and other essays in semantics and syntax. Berlin: Mouton Publishers.

Eesti etümoloogiasõnaraamat 2012. Iris Metsmägi, Meeli Sedrik, Sven-Erik Soosaar (toim). Tallinn: Eesti Keele Sihtasutus.

Eesti keele koondkorpus. Saadaval internetis aadressil <http://www.cl.ut.ee/ korpused/ segakorpus>. Vaadatud 15.07.2013.

Eesti keele põhisõnavara sõnastik 2014. Jelena Kallas, Mai Tiits, Maria Tuulik (toim). Tallinn: Eesti Keele Sihtasutus.

Ehala, Martin 2001. Eesti keele baassõnajärjest. – Keele kannul: pühendusteos Mati Erelti 60. sünnipäevaks. Reet Kasik (toim). Tartu: Tartu Ülikooli Kirjastus.

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

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

EKSS 1988 = Eesti kirjakeele seletussõnaraamat 1988–2007. Tallinn: Eesti Keele Sihtasutus.

EKSS = Eesti keele seletav sõnaraamat I-VI 2009. Tallinn: Eesti Keele Sihtasutus.

Engelberg, Mila 2002. The communication of gender in Finnish – Gender Across Languages. The linguistic representation of women and men 2. M. Hellinger, H. Bussmann (eds). Amsterdam, Philadelphia: John Benjamins Publishing Company.

Erelt, Mati 2005. Keeletüpoloogiast – Oma Keel 2005/1, 5–13.

Erelt, Mati; Erelt, Tiiu; Ross, Kristiina 2007. Eesti keele käsiraamat. Tallinn: Eesti Keele Sihtasutus.

Esser, Jürgen 1999. Collocation, colligation, semantic preference and semantic prosody: new developments in the study of syntagmatic word relations. – Words, Lexemes and Concepts: Approaches to The Lexicon. W. Falkner, H.-J. Schmid (eds). Tübingen: Gunter Narr Verlag, 155–166.

ESTMORF = Eesti keele morfoloogiline analüsaator. Saadaval internetis aadressil <http://www.filosoft.ee/html_morf_et/morfoutinfo.html>. Vaadatud 21.01.2014. Vt ka Kaalep, Heiki-Jaan 1998.

Evert, Stefan 2005. The Statistics of Word Coocurrences: Word Pairs and Collocations. PhD thesis. University of Stuttgart.

Firth, John Rupert 1957. Papers in linguistics, 1934–1951. Oxford: Oxford University Press.

Fodor, Jerry A.; Katz, Jerrold J. 1963. The Structure of a Semantic Theory – Language 39, 170–201.

Gale, William; Church, Kenneth; Yarowsky, David 1993. A method for disambiguating word senses in a large corpus. – Computers and the Humanities 26/1–2, 415–439.

Gesuato, Sara 2003. The company *women* and *men* keep: what collocations can reveal about culture. Dept. of Anglo-Germanic and Slavic languages and literatures. Saadaval internetis <http://ucrel.lancs.ac.uk/publications/cl2003/papers/gesuato.pdf> Vaadatud 10.10.2012.

Gómez, Pascual Cantos 2013. Statistical methods in language and linguistic research. Equinox.

Granger, Sylviane; Paquot, Magali 2008. Disentangling the phraseological web. – Phraseology: An Interdisciplinary Perspective. S. Granger, F. Meunier (toim.). John Benjamins Publishing, 27–49.

Grice, H. Paul 1991. Studies in the way of words. London: Harvard University Press.

Halliday, Michael Alexander Kirkwood 1966. Some notes on ‘deep’ grammar. – Journal of Linguistics 2/1, 57–67.

Halliday, Michael Alexander Kirkwood; Teubert, Wolfgang; Yallop, Colin; Čermáková, Anna 2004. Lexicology and corpus linguistics. An introduction. London, New York: Continuum.

Harris, Zellig 1946. From Morpheme to Utterance. – Language 22/3, 161–183.

Harris, Zellig 1954. Distributional structure. – Word. Journal of the linguistic circle of New York 10 /2–3, 146–162.

Harris, Zellig Sabbatai 1957. Co-occurrence and transformation in linguistic structure. – Language 33/3, 283–340.

Hawkes, Terence 2003 [1977]. Structuralism and Semiotics. Routledge: London, New York.

Hearst, Marti A. 1992. Automatic acquisition of hyponyms from large text corpora. – Proceedings of the 14th conference on Computational linguistics 2. Association for Computational Linguistics.

Heid, Ulrich 2008. Computational phraseology. An overview. – Phraseology. An interdisciplinary perspective. Amsterdam, Benjamins, 337–360.

Heinsoo, Heinike 2010. Adjektiivide tajumine ja õpetamine. Lähivõrdlusi. Lähivertailuja, 120–136.

Hjelmslev, Louis 2012. Sissejuhatus keeleteooria alustesse. Urmas Sutrop (koost). Eesti Keele Sihtasutus: Tallinn.

Hoey, Michael; Mahlberg, Michaela; Stubbs, Michael 2007. Text, discourse and corpora: Theory and analysis. London: Continuum.

Hunston, Susan; Francis, Gill. 2000. Pattern Grammar. Amsterdam: John Benjamins.

Huntoon, Alishia 2009. Gender Stereotypes. – Encyclopedia of gender and society 2. Jodi O'Brien (ed). Sage.

Hyde, Janet Shibley 2005. The Gender Similarities Hypothesis. – American psychologist 60 (6), 581–592.

Jaanits, Kadri 2004. Leksikaalsetest kollokatsioonidest soome ja eesti keeles. Magistritöö Tartu ülikooli läänemeresoome keelte õppetoolis.

Jakobson, Roman 1956. Two aspects of language and two types of aphasic disturbances. – Fundamentals of language. R. Jakobson, M. Halle (eds). The Hague: Mouton, 55–83.

Jantunen, Jarmo Harri 2001. What can corpus-linguistic analysis reveal about near synonyms. – Virittäjä 2, 170–192.

Jantunen, Jarmo Harri 2004. Synonymia ja käännössuomi. Korpusnäkökulma samamerkityksisyyden kontekstuaalisuuteen ja käännöskielen leksikaalisiin erityispiirteisiin. Joensuun yliopiston humanistisia julkaisuja 35.

Jantunen, Jarmo Harri 2009. „Minulla on aivan paljon rahaa" – Fraseologiset yksiköt suomen kielen opetuksessa. – Virittäjä 3, 356–381.

Jantunen, Jarmo Harri 2012. Akateemiset ja populaaritekstit korpusvertailussa: sanalistat, avainsanat ja fraseologiset yksiköt. – Genreanalyysi – tekstilajitutkimuksen käytäntöä. Vesa Heikkinen, Eero Voutilainen, Petri Lauerma, Ulla Tiililä, Mikko Lounela (toim). Helsinki: Kotimaisten kielten keskus.

Jantunen, Jarmo Harri; Brunni, Sisko 2012. Morfologinen priming ja fraseologia vieraan kielen oppimisessa: korpustutkimus oppijansuomesta. Lähivõrdlusi. Lähivertailuja 22, 71–100.

Jets, Katrin 2013. Tekstikorpusest kollokatsioonide tuvastaja. Magistritöö Tartu ülikooli arvutiteaduse instituudis.

Jütting, Johannes P.; Morrisson, Christian; Dayton-Johnson, Jeff; Dreschsler, Denis 2006. Measuring gender (in)equality: introducing the gender, institutions and development data base (GID). OECD Development Centre. Saadaval internetis <http://www.oecd.org/dev/36228820.pdf>. Vaadatud 12.02.2014.

Kaalep, Heiki-Jaan 1998. Tekstikorpuse abil loodud eesti keele morfoloogiaanalüsaator. – Keel ja Kirjandus 1, 22–29.

Kaalep, Heiki-Jaan; Muischnek, Kadri 2002a. Eesti kirjakeele sagedussõnastik. Tartu: Tartu Ülikooli Kirjastus.

Kaalep, Heiki-Jaan; Muischnek, Kadri 2002b. Püsiühendite leidmine teksti abil. – Tähendusepüüdja. Pühendusteos professor Haldur Õimu 60. sünnipäevaks. Renate Pajusalu, Tiit Hennoste (toim). Tartu Ülikooli üldkeeleteaduse õppetooli toimetised 3, Tartu, 185–195.

Kaalep, Heiki-Jaan; Muischnek, Kadri 2009. Eesti keele püsiühendid arvutilingvistikas: miks ja kuidas. – Eesti Rakenduslingvistika Ühingu Aastaraamat 5, 157–172.

Kallas, Jelena 2013. Eesti keele sisusõnade süntagmaatilised suhted korpus- ja õppeleksikograafias. Doktoritöö. Tallinn: Tallinna Ülikool.

Kallas, Jelena; Tuulik, Maria 2011. Eesti keele põhisõnavara sõnastik: ajalooline kontekst ja koostamispõhimõtted. – Eesti Rakenduslingvistika Ühingu Aastaraamat 7, 59–75.

Karlgren, Jussi; Sahlgren, Magnus 2001. From words to understanding. – Foundations of real-world intelligence. CSLI Publications. Y. Uesaka, P. Kanerva, H. Asoh (eds), 294–308.

Karlsson, Fred 2002. Üldkeeleteadus. Tallinn: Eesti Keele Sihtasutus.

Kasik, Reet 2002. Lingvistiline tekstianalüüs. – Teoreetiline keeleteadus Eestis. Renate Pajusalu, Ilona Tragel, Tiit Hennoste, Haldur Õim (toim). Tartu ülikooli üldkeeleteaduse õppetooli toimetised 4, Tartu, 74–88.

Kilgarriff, Adam 1992. Polysemy. PhD thesis. University of Sussex.

Kilgarriff, Adam 2005. Language is never, ever, ever, random. Corpus linguistics and linguistic theory, 1(2), 263–276.

Kilgarriff, Adam 2007. Googleology is bad science. Computational linguistics 33/1, 147–151.

Kilgarriff, Adam; Rundell, Michael 2002. Lexical profiling software and it's lexicographic applications: a case study. – Proceedings of EURALEX 2002. Copenhagen, 807–818.

Kilgarriff, Adam; Rychly, Pavel; Smrz, Pavel; Tugwell, David 2004. The Sketch Engine. – Proceedings of the 11th EURALEX international congress. G. Williams, S. Vessier (eds). Lorient: Université de Bretagne Sud, 105–115.

Kilgarriff, Adam; Kovar, Vojtech; Krek, Simon; Srdanovic, Irena; Tiberius, Carole 2010. A quantitative evaluation of word sketches. – Proceedings Euralex. Leeuwarden, The Netherlands.

Kirt, Riin 2009. Inimesega seotud hierarhiapuu Eesti Wordnetis. Bakalaureusetöö Tartu ülikooli eesti ja üldkeeleteaduse instituudis.

Klemm, Jürgen 2013. Tänapäeva eesti keele sugulussõnavara. Bakalaureusetöö Tartu ülikooli eesti ja üldkeeleteaduse instituudis.

Kont, Ilse; Pikamäe, Arno 1962. Sõnade tähenduste avamisest eesti kirjakeele seletavas sõnaraamatus. – Emakeele Seltsi aastaraamat 8, 122–131.

Krishnamurthy, Ramesh 2006. Collocations. – Encyclopedia of language and linguistics. Brown, Keith (ed). Oxford: Elsevier.

Lakoff, George 1987. Women, fire, and dangerous things. What categories reveal about the mind. Chicago: University of Chicago Press.

Langemets, Margit 2003. Polüseemia ja leksikograafia. – Emakeele Seltsi aastaraamat 49, 97–124.

Langemets, Margit 2009. Nimisõna süstemaatiline polüseemia eesti keeles ja selle esitus eesti keelevaras. Tallinn: Tallinna Ülikooli Kirjastus.

Langemets, Margit; Tiits, Mai; Valdre, Tiia; Voll, Piret 2010. In spe: üheköiteline eesti keele sõnaraamat. – Keel ja Kirjandus 11, 793–810.

Leacock, Claudia; Towell, Geoffrey; Voorhees, Ellen 1996. Towards building contextual representations of word senses using statistical models. – Corpus processing for lexical acquisition. B. Boguraev, J. Pustejovsky (eds), Cambridge, USA: MIT Press, 97–113.

Leech, Geoffrey 1991. The state of art in corpus linguistics. – English Corpus Linguistics. Karin Aijmer, Bengt Altenberg (eds). London, New York: Longman.

Lehrer, Adrienne 1974. Semantic Fields and Lexical Structure. Amsterdam, London: North Holland Publishing.

Liljeström, Marianne 2003. Soosüsteem – Võtmesõnad. Tallinn: Eesti Keele Sihtasutus.

Lindquist, Hans; Levin, Magnus 2008. Foot and mouth: The phrasal patterns of two frequent nouns. – Phraseology: An interdisciplinary perspective, 143–158.

Lindsalu, Pilleriin 2012. Kuritegevus Eestis 2011. aastal: kokkuvõte. – Kuritegevus Eestis 2011. Tallinn: Justiitsministeerium.

Lipka, Leonhard 1980. Methodology and representation in the study of lexical fields. – Perspektiven der lexikalischen Semantik. Beiträge zum Wuppertaler Semantikkolloquium vom 2.–3. Dezember 1977. Dieter Kastovsky (ed). Bonn: Bouvier Verlag Herbert Grundmann.

Lounsbury, Floyd G. 1956. A semantic analysis of the Pawnee kinship usage. – Language 32/1, 158–194.

Louw, Bill 1993. Irony in the Text or Insincerity in the Writer? The Diagnostic Potential of Semantic Prosodies – Text and Technology. In Honour of John Sinclair. Mona Baker, Gill Francis, Elena Tognini Bonelli (eds). Amsterdam: John Benjamins Publishing.

Lyons, John 1977. Semantics I–II. Cambridge: Cambridge Univesity Press.

Lyons, John 2002 [1981]. Language and Linguistics. An introduction. Cambridge University Press.

Marling, Raili 2011. Keel ja sugu. – Sissejuhatus soouuringutesse. Raili Marling (toim). Tartu: Tartu Ülikooli Kirjastus.

Mason, Oliver 2000. Parameters of collocation: The word in the centre of gravity. Language and Computers 30, 267–280.

Matthews, Peter 2001. A Short History of Structural Linguistics. Cambridge University Press.

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

Mel'čuk, Igor 1998. Collocations and Lexical Functions – Phraseology. Theory, Analysis and Applications. A. Cowie (ed). Oxford: Clarendon Press, 23–53.

Meunier, Fanny; Granger, Sylviane (eds) 2008. Phraseology in foreign language learning and teaching. John Benjamins Publishing.

Miller, Georg A.; Charles, Walter 1991. Contextual correlates of semantic similarity. – Language and Cognitive Processes 6/1, 1–28.

Mills, Sara 2002. Post-feminist text analysis. – Critical Discourse Analysis. Critical Concepts in Linguistics I. Michael Toolan (ed). London, New York: Routledge.

Muischnek, Kadri 2005. Eesti keele verbikesksed püsiühendid tekstikorpuses. Emakeele Seltsi aastaraamat 51, 80–106.

Murphy, Lynne M. 2006. Antonymy as lexical constructions: or, why paradigmatic construction is not an oxymoron. – Constructions SV 1/8.

Nesselhauf, Nadja 2003. The use of collocations by advanced learners of English and some implications for teaching. – Applied Linguistics 24, 223–42.

Nesselhauf, Nadja 2005. Collocations in a Learner Corpus. Benjamins.

Nida, Eugene Albert 1975. Componential Analysis of Meaning: An Introduction to Semantic Structures. The Hague: Mouton.

Nuolijärvi, Pirkko 1986. Kolmannen sukupolven kieli. Helsinkiin muuttaneiden suurten ikäluokkien eteläpohjalaisten ja pohjoissavolaisten kielellinen sopeutuminen. Helsinki: Suomalaisen kirjallisuuden seura.

Nuolijärvi, Pirkko 2005. „Se on tavallaan sellaista magiciä". Nykysuomalaisen kielimaailma. – Tahdon sanoa. Kirjoituksia kielen ja perinteen voimasta. Sirpa Huttunen, Pirkko Nuolijärvi (toim), 162–184.

Nuolijärvi, Pirkko; Tiittula, Liisa 2000. Televisiokeskustelun näyttämöllä: televisio-institutionaalisuus suomalaisessa ja saksalaisessa keskustelukulttuurissa. Helsinki: Suomalaisen Kirjallisuuden Seura.

Ogden C. K. 1932. Opposition. A Linguistic and Psychological Analysis. Bloomington, London: Indiana University Press.

Orav, Heili; Kerner, Kadri; Parm, Sirli 2011. Eesti Wordneti hetkeseisust. – Keel ja Kirjandus 2, 96–106.

Ots, Sander 2012. Statistikapõhise tarkvara loomine morfoloogiliste kollokatsioonide eraldamiseks eesti keele tekstidest. Bakalaureusetöö Tallinna Ülikooli informaatika instituudis.

Padó, Sebastian; Lapata, Mirella 2007. Dependency-based construction of semantic space models. Computational Linguistics 33/2, 161–199.

Pajusalu, Renate 2009. Sõna ja tähendus. Eesti Keele Sihtasutus.

Pajusalu, Renate; Tragel, Ilona; Veismann, Ann; Vija, Maigi 2004. Tuumsõnade semantikat ja pragmaatikat. Tartu Ülikooli üldkeeleteaduse õppetooli toimetised 5, Tartu.

Palmer, Harold E. 1949 [1938]. A Grammar of English words. London, New York, Toronto: Longman.

Panther, Klaus-Uwe; Thornburg, Linda L. 2012. Antonymy in language structure and use – Cognitive linguistics between universality and variation. Mario Brdar, Milena Žic Fuchs, Ida Raffaelli (eds). Newcastle, UK: Cambridge Scholars.

Pauwels, Anne 2003. Linguistic sexism and feminist linguistic activism. – The handbook of language and gender. Janet Holmes, Miriam Meyerhoff (eds). Oxford: Blackwell.

Pearce, Michael 2008. Investigating the collocational behaviour of *man* and *woman* in the BNC using Sketch Engine 1. – Corpora 3/1, 1–29.

Peterwagner, Reinhold 2005. What is the matter with communicative competence?: An analysis to encourage teachers of English to assess the very basis of their teaching. Wien: Lit Verlag.

Piits, Liisi 2004. Sõnade *mees, naine* ja *inimene* esinemisümbrusest ja stiilivärvingust. Magistritöö Tallinna Ülikoolis.

Piits, Liisi 2006. Sõnade *mees* ja *naine* kollokatsioonide võrdlemise võimalusi eesti, saksa ja inglise keele korpustes. Keel ja Arvuti. Tartu Ülikooli üldkeeleteaduse õppetooli toimetised 6, Tartu. Mare Koit, Renate Pajusalu, Haldur Õim (toim), 105–114.

Piits, Liisi 2007. Kas eesti *mees* on mees või inimene? – Ariadne Lõng: nais- ja meesuuringute ajakiri 2005/2006, 3–10.

Piits, Liisi 2011. Mis värvi on jõulud? – Värvinimede raamat. Mari Uusküla, Urmas Sutrop (toim). Tallinn: Eesti Keele Sihtasutus, 115–120.

Piits, Liisi 2013. Distributional Hypothesis: Words for ‘human being’ and their Estonian collocates. Trames: Journal of the Humanities and Social Sciences, 141–158.

Pilvre, Barbi 2011a. Meedia ja sugu. – Sissejuhatus soouuringutesse. Raili Marling (toim). Tartu: Tartu Ülikooli Kirjastus.

Pilvre, Barbi 2011b. Naiste meediarepresentatsioon Eesti ajakirjanduskultuuri ja ühiskonna kontekstis. Tartu: Tartu Ülikooli Kirjastus.

Puna, Kerli 2006. Soospetsiifilised isikunimetused sõnaraamatutes ja tekstides. Magistritöö Tartu ülikooli eesti keele õppetoolis.

Pustejovsky, James 1995. The Generative Lexicon. Language, Speech, and Communication Series. The MIT Press.

Renouf, Antoinette; Banerjee, Jayeeta 2007. Lexical repulsion between sense-related pairs. – International Journal of Corpus Linguistics 12 (3), 415–43.

Rosch, Eleanor 1973. Natural categories. – Cognitive psychology 3/4, 328–350.

Ross, Kristiina 1996. Naine eesti keele grammatikas. – Vikerkaar 11/12, 102–106.

Rubenstein, Herbert; Goodenough, John B. 1965. Contextual correlates of synonymy. – Computational Linguistics.

Rätsep, Huno 2002. Sõnaloo raamat. Tartu: Ilmamaa.

Rätsep, Huno 2009. Mul kogu küla sugulasi täis. Sugulussõnadest ja nende päritolust. – Oma Keel 1.

Sahlgren, Magnus 2006. The Word-Space Model. Using distributional analysis to represent syntagmatic and paradigmatic relations between words in high-dimensional vector spaces. A Dissertation submitted to Stockholm University in partial fulfillment of the requirements for the degree of Doctor of Philosophy. Saadaval internetis. <http://soda.swedish-ict.se/437/1/TheWordSpaceModel.pdf>. Vaadatud 21.01.2014

Sahlgren, Magnus 2008. The Distributional Hypothesis. – From context to meaning: Distributional models of the lexicon in linguistics and cognitive science (Special issue of the Italian Journal of Linguistics), Rivista di Linguistica 20/1.

Saussure, Ferdinand de 2000. Course in general Linguistics. Charles Bally and Albert Sechehaye (eds). London: Duckworth.

Saussure, Ferdinand de 2006. Writings in General Linguistics. Oxford: Oxford University Press.

Schütze, Hinrich 1992. Dimensions of meaning. – Proceedings of the 1992 ACM/IEEE Conference on Supercomputing. IEEE Computer Society Press, 787–796.

Schütze, Hinrich 1998. Automatic word sense discrimination. – Computational Linguistics 24/1, 97–123.

Scott, Mike 2004. WordSmith Tools version 4. Computer program. Oxford: Oxford University Press.

Siepmann, Dirk 2005. Collocation, colligation and encoding dictionaries. Part I: Lexicological aspects. – International Journal of Lexicography 18/4, 409–443.

Sinclair, John McHardy 1991. Corpus, Concordance, Collocation. Oxford University Press.

Sinclair, John McHardy 2003. Reading concordances. An Introduction. London: Longman.

Sinclair, John McHardy 2004. Trust the Text. Language, corpus and discourse. Roland Carter (ed). London, New York: Routledge.

Stefanowitsch, Anatol; Gries, Stefan Th. 2003. Collostructions: Investigating the interaction between words and constructions. International Journal of Corpus Linguistics 8/2, 209–243.

Stefanowitsch, Anatol; Gries, Stefan Th. 2009. Corpora and grammar – Corpus linguistics: an International handbook 2. A. Lüdeling, M. Kytö (eds). Berlin, New York: Mouton de Gruyter, 933–951.

Stewart, Dominic 2010. Semantic prosody. A critical evaluation. New York, London: Routledge.

Stubbs, Michael 1995. Collocations and semantic profiles: On the cause of the trouble with quantitative studies. – Functions of language 2/1, 23–55.

Stubbs, Michael 2001. Words and Phrases: Corpus Studies of Lexical Semantics. Oxford: Blackwell.

Stubbs, Michael 2009. Technology and phraseology: With notes on the history of corpus linguistics. – Exploring the lexis-grammar interface, 15–32.

Sutrop, Urmas 2000. Basic terms and basic vocabulary. – Estonian typological studies 4. Mati Erelt (toim). Tartu: Tartu Ülikool, 118–145.

Sutrop, Urmas 2002. Loetelukatse ja kognitiivse esiletuleku indeks. – Teoreetiline keeleteadus Eestis. Renate Pajusalu, Ilona Tragel, Tiit Hennoste, Haldur Õim (toim). Tartu: Tartu Ülikool, 240–254.

Sutrop; Urmas 2005. Temperatuurisõnad ja -skaalad. Emakeele Seltsi aastaraamat 50, 111–128.

Sutrop, Urmas 2011. Põhivärvinimed: ajalugu, teooria ja meetodid. – Värvinimede raamat. Mari Uusküla, Urmas Sutrop (toim). Tallinn: Eesti Keele Sihtasutus.

Talves, Kairi 2011. Soosotsioloogia. – Sissejuhatus soouuringutesse. Raili Marling (toim). Tartu: Tartu Ülikooli Kirjastus.

Tauli, Valter 1980. Eesti grammatika II. Lauseõpetus. Uppsala: Finsk-ugriska institutionen.

Taylor, John R. 1995. Linguistic Categorization: Prototypes in Linguistic Theory Oxford University Press.

Tognini-Bonelli, Elena 2001. Corpus linguistics at work 6. John Benjamins Publishing.

Trier, Jost 1931. Der deutsche Wortschatz im Sinnbezirk des Verstandes: die Geschichte eines sprachlichen Feldes. Heidelberg.

Uiboaed, Kristel 2010. Statistilised meetodid murdekorpuse ühendverbide tuvastamisel. – Eesti Rakenduslingvistika Ühingu Aastaraamat 6, 307–326.

Uiboaed, Kristel 2013. Kollostruktsioonilised meetodid ja konstruktsioonilise varieerumise tuvastamine. – Eesti ja soome-ugri keeleteaduse ajakiri. Journal of Estonian and Finno-Ugric Linguistics 4, 185–204.

Vainu, Vaike; Järviste, Liina; Biin, Helen 2010. Soolise võrdõiguslikkuse monitooring 2009. Sotsiaalministeeriumi toimetised 1, Tallinn.

Vanhatalo, Ulla 2005. Kyselytestit synonymian selvittämisessä. Sanastotietoutta kielenpuhujilta sähköiseen sanakirjaan. Väitöskirja. Helsinki: Yliopistopaino.

Vääri, Eduard 2013. Sugulusalane sõnavara läänemeresoome keeltes. Tallinn: Eesti Keele Sihtasutus.

Wang, Ying; Shaw, Philip 2008. Transfer and universality: collocation use in advanced Chinese and Swedish learner English. – *ICAME Journal* 32, 201–232.

Wortschatz = Leipzigi ülikooli kollokatsioonide otsingumootor. Saadaval internetis <http://corpora.informatik.uni-leipzig.de>. Vaadatud 22.01.2014.

Õim, Asta 1995. Antonüümisõnastik. Tallinn: Keele ja Kirjanduse Instituut.

Õim, Asta 2002. Leksikaalsete sünonüümide sisulisest küljest. – Nime murre. Tallinn: Eesti Keele Sihtasutus.

Õim, Haldur 1991. Lingvistilised korpused keeleuurimises. – Keel ja Kirjandus 5, 257–267.

Xiao, Richard; McEnery, Tony 2006. Collocation, Semantic Prosody, and Near Synonymy: A Cross-Linguistic Perspective. – Applied Linguistics 27 (1), 103– 9.

SUMMARY: COLLOCATIONS OF THE MOST FREQUENT ESTONIAN WORDS FOR 'HUMAN BEING'

Aim and structure

The present study is based on the structuralist assumption that language is a system of bound entities, in which the entities (e.g. words) do not have specific value in themselves, but only in comparison with other corresponding units. The meanings of the words cannot be defined without considering their relationships with other words of the language. The aim of this study is to examine the collocational relations of words used to refer to a human being with other words and the manifestation of their paradigmatic relations on the syntagmatic level. The study addresses the collocations of the Estonian words *ema* 'mother', *inimene* 'human being', *isa* 'father', *laps* 'child', *mees* 'man', *naine* 'woman', *poeg* 'son', *poiss* 'boy', *tüdruk* 'girl' and *tütar* 'daughter', aiming to find out if and how the meaning of a word is reflected in its collocational relations.

The current advance of corpora and corpus linguistics has boosted the investigation of syntagmatic relations of words, both in Estonia and beyond. Collocation is one of the phenomena of the syntagmatic level. Due to the different aims of the previous studies of the collocation there are now different schools, each having their own theoretical starting points and terminology. This study is based on the ideas of the British neo-Firthian school (Sinclair 2004, Stubbs 2001a) and the American structuralists (Harris 1954, 1957) to the effect that the distribution of a word reflects the meaning of the word. Thus the term *collocation* is used in a wider sense here, as based on the neo-Firthian tradition. The basic definition of the collocation, for this dissertation, relies on Stubbs (2001a) and Sinclair (2004): *Collocation is a lexical relationship between two frequently co-occurring words, as well as a word pair in such a relationship.*

In Estonia, collocations have previously been studied from the following points of view: language learning (Jaanits 2004, Heinsoo 2010), computational linguistics (Muischnek 2005; Kaalep, Muischnek 2009), mapping of dialect features (Uiboaed 2010), and lexicography (Kallas 2013). The present study is an attempt to fill a lacuna in Estonian linguistics by investigating the points of contact between the distributive behaviour of a word and its semantics. More specifically, the aim is to find out how the gender and age of the person referred to by the node word may influence the collocational relations of the word. The basic assumption is that every repeated co-occurrence carries a cultural allusion (Tognini-Bonelli 2001: 125) and a close examination of the respective collocates of words referring to opposite gender or age will eventually reveal the most stereotypical idea of a man or woman, boy or girl. The study will focus on the following research questions:

1. To what extent do the collocates of different words for 'human being' coincide and is there any correlation between this rate of coincidence and the values given to the semantic components of the words?

2. How are relations of the paradigmatic level reflected on the syntagmatic level?

3. What are the differences, if any, in the part of speech of the left vs right collocates?

4. How are differences in age and gender reflected in the collocates of words for 'human being'?

To answer these questions, a survey is given of the paradigmatic relations of the words for 'human being', component analysis (Fodor, Katz 1963, Nida 1975) is applied to describe the location of words in the semantic field of HUMAN BEING, and the collocational relations of words are thoroughly investigated using two complementary methods.

The present dissertation consists of four chapters and a conclusion. The introductory chapter discusses the concept and definitions of collocation, providing a brief survey of the history of collocation studies. The methods used in the study and the arguments for their selection are also presented.

Chapter 2 is meant to provide a theoretical background for studying the collocations of words referring to 'human being'. The semantic presentation of the words *ema* 'mother', *inimene* 'human being', *isa* 'father', *laps* 'child', *mees* 'man', *naine* 'woman', *poeg* 'son', *poiss* 'boy', *tüdruk* 'girl' and *tütar* 'daughter' in different dictionaries and in the Estonian Wordnet is described. Subsequently, the words for 'human being' are examined from the point of view of semantic field theory, while their semantic differences are investigated using component analysis. An overview is given of the main paradigmatic relations occurring between the words for 'human being', including different types of hyponymy, synonymy, and antonymy.

In Chapters 3 and 4, corpus data are analysed for the distribution of the words for 'human being', trying to find evidence of relations between the syntagmatic and paradigmatic levels of the Estonian language. Chapter 3 contains a separate analysis of left and right collocates. Coincidence of collocate clusters is analysed to find out what paradigmatic relations are reflected in collocational relations. The part-of-speech differences of the left vs right collocates are highlighted.

Chapter 4 deals with two lexico-grammatical constructions used for detecting collocations of the Estonian words for 'human being'. Antonym pairs are compared assuming that frequent collocational relations may reveal some covert or hidden meanings and the way how people are depicted in texts. Full lists of the collocates of the Estonian words for 'human being', which is the research material for the present study, are available in the Supplement.

The Conclusion summarizes the main results.

Material and methods

Two complementary methods and programs for corpus analysis are used. Chapter 3 describes the general distribution of the Estonian words for 'human being', having run the WordSmith Tools program on a lemmatized subcorpus (5 million words, journalistic texts) of the Balanced Corpus belonging to the Estonian Reference Corpus currently developed by the University of Tartu. As the subcorpus is relatively limited the collocations were retrieved without using either lexico-grammatical constructions or statistics of collocational strength; instead, a separate search was done for the left and right collocates, the thirty most frequent of which were subsequently subjected to comparative analysis. The collocates were drawn using a 3:3 collocate span, i.e. from the range of three words to the right and three words to the left of the node word. Considering the absolute co-occurrence frequency, a slightly higher weight was assigned to the collocates which were located nearer to the node word, assuming that the closer the collocate lies to the node, the more probably the two belong to one and the same syntactic phrase. As a result the obtained data were simple and transparent, while the list of collocates contained more pronouns and other high-frequency words, which would have missed researcher attention if collocation strength statistics had been used. A detailed analysis was performed on the 30 most frequent collocates of 10 words for 'human being'.

Chapter 4 describes how the collocates were retrieved using the Sketch Engine and the structural patterns devised by Jelena Kallas (2013) on the basis of phrase structure. Here the material comes from the Estonian Reference Corpus, which contains 203 million words. Although the data obtained this way depend on the structural patterns and they fail to offer a wider, purely frequency-based overview of all lexical collocates, they are more informative and better comparable with the data of other languages, because the use of grammatical constructions leads the researcher directly to the syntactic relations behind the collocational ones. The *adj_modifier* rules of the Sketch Engine are used to find the 30 most frequent modifying adjectives of each node word, while the *subject_of* rules help to find the verbs for which the node word is the subject. Verb analysis involves the 30 most frequent verbs of each node word and, in addition, the verbs distinctive to antonyms are compared. A collocate is regarded as distinctive to a member of an antonym pair if, in the corpus material, it co-occurs with one member of the pair at least four times more frequently than with the other.

A comparison of the frequency of the words for 'human being' in the Estonian Reference Corpus (ERC) and in the Balanced Corpus newspaper subcorpus (BCN) shows that their proportion in the two sources does not differ significantly. In both corpora the most frequent node word is *inimene* 'human being', similarly followed by *mees* 'man', *laps* 'child', *naine* 'woman' and *ema* 'mother'. Only the frequencies of *poiss* 'boy' and *isa* 'father' and of *poeg* 'son' and *tüdruk* 'girl', respectively, have the opposite ranking. The absolute frequencies of the ten most

frequent Estonian words for ‘human being’ occurring in the ERC and BCN are as follows:

	Freq. in ERC	Freq. in BCN
inimene ‘human’	522436	12277
mees ‘man’	307624	7293
laps ‘child’	245399	5583
naine ‘woman’	179449	4098
ema ‘mother’	68944	1504
poiss ‘boy’	59315	1221
isa ‘father’	57094	1330
tüdruk ‘girl’	44418	853
poeg ‘son’	40834	914
tütar ‘daughter’	28061	612

In both corpora words for a male person are more frequent than their female counterparts (except in the pair *ema* ‘mother’ and *isa* ‘father’). A comparison of the frequencies reveals an almost 20 times difference between the occurrences of *inimene* ‘human being’ and *tütar* ‘daughter’. In the process of analysis the differences are levelled to a certain extent, as the comparison addresses either the relative, not absolute frequencies of occurrence, or the rank position of the collocate among the other collocates of the node.

Results

Chapter 2 ascertains that the diagnostic, or semantically distinctive components of the most frequent words referring to humans are as follows: MALE, ADULT, KINSHIP/PARTNERSHIP, GENERATION OLDER THAN SELF, GENERATION YOUNGER THAN SELF. Chapters 3 and 4 investigate how the values given to those components are reflected in collocational relations.

In Chapter 3 the following hypotheses were tested:

1. There is a considerable coincidence among the collocate clusters of the Estonian words for ‘human being’.

2. The coincidence of the collocate clusters is in positive correlation with the number of semantic components with a similar value that are shared by the node words.

3. Paradigmatic relations can be detected in collocations.

4. The area where certain collocational relations can be detected (from left or right of the node) depends on the grammar of the particular language.

The study reported in Chapter 3 revealed that a large share of the left and right collocates of the Estonian ‘human’ semantic field coincided: an average of 43–57% of the 30 left collocates of all the analysed words coincided with the most

frequent collocates of other words for 'human being', while in the right collocates the coincidence was even higher, reaching up to 50–60%. Many collocates were shared by all ten node words. Eight words were observed among the 30 most frequent left collocates: *üks* 'one/a', *kaks* 'two', *kolm* 'three', *mina* 'I', *tema* 'he/she', *see* 'this', *olema* 'be' and *oma* '(one's) own'. Of the right collocates, the following seven were shared by all node words: *mina* 'I', *minema* 'go', *olema* 'be', *oma* '(one's) own', *saama* 'get', *see* 'it/this' and *tema* 'he/she'. Thus the hypothesis was confirmed that there is a considerable coincidence among the collocate clusters of the Estonian words for 'human being'.

Studying the possible correlations between the collocate clusters and intra-field lexical relations it was found that no semantic component separately taken (e.g. a similar value given to MALE) had a significant influence on cluster coincidence. And yet there was a tendency that the semantics of the node words with a larger number of shared collocates contained more components of a similar value. The manifestation of the tendency was nearly maximal in the left collocates. Thus the hypothesis was confirmed that the coincidence of collocate clusters is larger for node words with a greater number of semantic components of a similar value.

Also, Chapter 3 demonstrates that collocates are influenced by the syntactic structure of the sentence, notably, the part-of-speech distribution differs in the left and right collocates. The biggest difference was observed in the frequency of adjectives – 17% of the left collocates vs 2% of the right ones. Another big asymmetry was revealed in the location of verbs: 24% of the left collocates vs 43% (nearly twice as many) of the right ones. Thus it can be stated that although collocations are lexical syntagmatic relations, their occurrence largely depends on grammar, i.e. on syntax and word order patterns. Thus the hypothesis was confirmed that the area where this or that kind of collocational relations can be found depends on the grammar of the language.

As for paradigmatic relations, mainly those of antonymy were discovered in the collocations analysed. All antonomous word pairs such as *mees* 'man' – *naine* 'woman', *tütar* 'daughter' – *poeg* 'son', *tüdruk* 'girl' – *poiss* 'boy', *ema* 'mother' – *isa* 'father', the members of which differed in the value of the MALE component, had the greatest number of left collocates in common. This corroborates Sahlgren's suggestion that a comparison of the syntagmatic context, more specifically, collocates of words enables conclusions to be made about paradigmatic relations (in the present paper mainly about antonymy). If antonymy is regarded as a similarity relation – after all, only one of all the semantic components of two words gets an opposite value – there is absolutely no reason to protest Harris' distributional hypothesis, which argues that the semantic similarity of words is manifested in the coincidence of their contexts.

Chapter 3 led to the conclusion that if one overlooks the syntactic structure, it is difficult to make a difference between the verbs for which the node word is the subject and those for which it is an object. Therefore it was decided to proceed with the analysis of collocational relations (see Chapter 4) by means of certain

lexicogrammatical patterns, which give an idea of the possible syntactic relation between the node word and the collocate.

Chapter 4 tested the following two hypotheses:

1. The collocates of words for 'human being' reveal some gender differences.
2. The collocates of words for 'human being' reflect power relations in society.
3. The collocates of a word for 'human being' reveal the age of the typical referent of the node word.

The results of Chapter 4 confirmed the hypothesis that gender differences are indeed manifested in the collocates of the words for 'human being'. The verbs and adjectives occurring among the collocates illustrated the current Estonian stereotypes of male and female persons.

According to my analysis of collocational relations women are seen as more active in the private sphere, whereas men's preferable sphere of activity is public. A considerable part of the 30 verbs belonging to the word *ema* 'mother' denote routine housework, e.g. *küpsetama* 'bake', *keetma* 'cook', *pakkima* 'pack', *pesema* 'wash' and *askeldama* 'bustle', while fathers take care of more extraordinary tasks, suggested by the verbs *ehitama* 'build' and *meisterdama* 'make things'. This division of verbs is eloquent of the current division of labour in the Estonian community, where women normally do twice as much housework as men do (Talves 2011: 103). Although men's and women's contribution along the lines of paid work is nearly equal to-day, the role breakdown and stereotypes have survived.

In addition to the routine housework, Estonian women are usually expected to take care of communication with kindergarten and school, to play with children, help them and drive them wherever necessary (Vainu et al. 2010: 123). A considerable part of the collocate verbs distinctive to *ema* 'mother' refer to dealing with children and caring for them: *sünnitama* 'give birth', *imetama* 'breastfeed', *toitma* 'feed', *jootma* 'give something to drink', but also *keelitama* 'persuade', *tõrelema* 'scold', *tirima* 'drag', *raputama* 'shake' and *silitama* 'stroke, caress'. The following activities, however, are mainly associated with the father: *jalutama* 'have a walk (with)', *juhendama* 'instruct', *kiigutama* 'dandle', *patsutama* 'pat' and *meisterdama* 'make things'. Thus the collocations suggest that in Estonian culture mothers take care of the essential needs of a child, whereas the father's role is rather one of an entertainer.

The most frequent verb among the collocates of *isa* 'father', however, is *teenima* 'earn', which indicates that the primary contribution expected from a father is money earned for work. Further on, accumulated earnings may turn into a heritage: among the collocates of *isa* 'father' the verb *pärandama* 'bequeathe' occurs twelve times as frequently as among those of *ema* 'mother'.

The use of the word *töötav* 'working' indicates that in men their line of work is considered more important than in women. Although *töötav* 'working' was more frequent among the collocates of both *ema* 'mother' and *naine* 'woman' than in the case of male referents, there were differences in the use of the word. In the

case of women the phrase *töötav naine* ‘working woman’ was often opposed to *kodune naine* ‘housewife’, whereas in the case of men the word *töötav* ‘working’ was nearly always preceded by an attribute referring to job or occupation. This means that a homemaker’s position is more easily associated with women than with men, whereas in men the working vs stay-at-home opposition is practically inconceivable, which makes it more important to specify their job or office. Specifically, the typical idea of work means working out of home, excluding housework and raising children (Lakoff 1987: 80). This is also corroborated by Estonian collocations, as the corpus material contained the opposition of *kodune ema* ‘stay-at-home mother’ vs *töötav ema* ‘working mother’. The latter refers to a parent who works out of home, whereas the former does not, but of course she cannot enjoy her motherhood without a lot of baking, cooking, packing, washing, laundering etc.

Closer association of women with the private sphere is shown by the felt necessity to point out their marital status: the 30 most frequent modifiers of the word *naine* ‘woman; wife’ include *vallaline* ‘unmarried’, *üksik* ‘single’, *endine* ‘former’ and *uus* ‘new’. As for men, their marital status is not emphasized, at least no such words occur among the frequent collocates of *mees* ‘man; husband’. Instead, we find reference to social status in such collocates as *tähtis* ‘important’, *rikas* ‘rich’ and *vaene* ‘poor’. Social status is also important for father, as is proved by the collocates *kuulus* ‘famous’, *rikas* ‘rich’ and *suur* ‘big’ among the most frequent ones for *isa* ‘father’. The frequency of the first two among the collocates of *isa* ‘father’ surpasses their occurrence among the collocates of *ema* ‘mother’ by 20 to 30 times. For a woman it is considered necessary to point out her relations to men – ex-wife, new wife or a single woman. For a man it is considered important to define his rung of the social ladder – whether he is rich or poor, maybe fulfilling a significant role in the community.

An analysis of modifying adjectives revealed that in women it is considered important to describe their appearance, whereas for men it is rather personality traits that count. The words *ilus* ‘pretty’, *kaunis* ‘beautiful’, *kena* ‘good-looking’, which are relatively frequently used to refer to attractiveness, make up more than 5% of all adjectival attributes of *naine* ‘woman’, but less than 1% of the adjectival attributes of *mees* ‘man’. Thus, when describing the appearance of a woman beauty is many times more salient a feature than in the case of men. Nakedness is also a state mentioned many times more often in connection with women than with men, e.g. the word *paljas* ‘naked’ co-occurs with *naine* ‘woman’ over 4 times more often than with *mees* ‘man’.

The much-quoted sentence by Simone de Beauvoir (1997: 185) „one is not born, but rather becomes, a woman” is also confirmed by an analysis of the collocates of the words *tütar* ‘daughter’ and *poeg* ‘son’. Daughters are mostly described by their looks as little and pretty, while the adjectives for sons refer both to their appearance (*suur* ‘big’) and character, which, unlike that of the girls, is stereotypically headstrong, brave and faithful. This reveals the expectations

people tend to have for their offspring. For girls, beauty gets mentioned four times more often than for boys.

According to the analysis of distinctive verbs, all „female" words except *ema* 'mother' had one or two verbs among their collocates, which evidences the stereotypical idea of a woman as a (sex) object. For the word *tüdruk* 'girl', for example, the frequent distinctive collocates included such verbs as *maksma* 'cost', *värvima* 'use makeup', *ehtima* 'adorn'. The collocates of *naine* 'woman' included *värvima* 'use makeup' and *poseerima* 'pose'. The verbs *meeldima* 'be liked'and *esinema* 'perform' occurred among the collocates of *tütar* 'daughter'.

Also, the collocates distinctive to words denoting female persons include verbs referring to reproductive functions, such as *sünnitama* 'give birth', *viljastuma* 'be fertilized', *rasestuma* 'become pregnant' and adjectives of fertility, such as *rase* 'pregnant', *(last) ootav* 'expecting (a baby)', *imetav* 'breastfeeding'. No fertility verbs or adjectives were observed among the frequent collocates of words denoting male person.

According to the gender stereotype women are more emotional and verbally more able than men. Although no valid empirical proof has been produced to this effect and Janet Hyde (2005), for example, has demonstrated that the male-female differences in verbal abilities and behaviour are negligible, my analysis of communication verbs revealed that the verbs referring to the verbal expression of female persons and sounds uttered by them are far more diverse than those used in connection with male persons. For example, the 30 distinctive collocates of *tütar* 'daughter' include 13 verbs of communication or emotion, whereas *poeg* 'son' is the subject of three such verbs only. A large share of the collocates distinctive to *tüdruk* 'girl', *tütar* 'daughter', *naine* 'woman' and *ema* 'mother' is made up of such communication verbs that also convey some information on how something is uttered. Often such verbs have a derogatory connotation, e.g. *tänitama* 'rebuke', *näägutama* 'nag; twit', *kriiskama* 'scream', *kiljuma* 'shriek', *nähvama* 'retort'. All words denoting male, however, share the male-specific collocates *käsutama* 'command' and *käskima* 'order', which are both associated with power. Only *ema* 'mother' and *isa* 'father', who are both in the position of power, have a similar share of the collocate *käskima* 'order'.

Collocations reflect the power relations characteristic of the community. The adjectives of monetary status and verbs describing the use of money express access to benefits. My analysis of the adjectives showed that the collocate *rikas* 'rich' is mainly distinctive to *mees* 'man', *isa 'father'* and *poeg* 'son', which are all words with male referents. The watershed of power relations, however, does not always run between sexes. The fact that *ostma* 'buy' and *maksma* 'pay' are typical collocates of *inimene* 'human being', but not of *laps* 'child' indicates that ecomomic power belongs to the grown-ups.

Violence and criminal behaviour has also been interpreted as demonstration of power. The 30 collocates distinctive to *mees* 'man; husband', *isa* 'father', *poiss* 'boy' or *poeg* 'son' contain numerous verbs referring to aggressive or violent

activities. E.g., as many as 40% of the 30 most frequent collocates distinctive to *mees* 'man' designate aggressive activities: *tulistama* 'shoot', *võitlema* 'fight', *ähvardama* 'threaten', *röövima* 'rob', *vägistama* 'rape', *pooma* 'hang', *tabama* 'catch', *vahistama* 'arrest', *kaaperdama* 'hijack', *sõdima* 'make war', *haavama* 'wound' and *purustama* 'destroy'. The biggest group of the collocates distinctive to *poiss* 'boy' also consists of verbs of aggressive activities: *võitlema* 'fight', *tapma* 'kill', *lõhkuma* 'break', *kaklema* 'fight', *taguma* 'beat', *kiusama* 'bully', *röövima* 'rob', *tabama* 'catch' and *sundima* 'force'. Verbs of violence are also present among the collocates distinctive to the related pair of *isa* 'father' and *poeg* 'son'.

The frequent modifying adjectives *vaimne* 'spiritual' and *püha* 'holy' occurring among the collocates of *isa* 'father' indicate that the word often has a metaphorical sense of God, spiritual leader *etc*, whereas the role of *ema* 'mother' is almost invariably that of a parent. This difference between the collocational relations of *ema* 'mother' and *isa* 'father' reflects a patriarchal society, in which most of the clergy and spiritual leaders are male and even God is personalized as man.

It is difficult to determine whether the differences in the collocational relations of words reflect the reality (as most of the criminals are men the frequent co-occurrence of words for male persons with the verbs *tapma* 'kill' and *röövima* 'rob' is hardly surprising), gender roles (e.g. the words *mees* 'man', *poeg* 'son', *poiss* 'boy' and *isa* 'father' frequently co-occur with the verb *ehitama* 'build', while *naine* 'woman' co-occurs with the verb *kuduma* 'knit'), or stereotypes that do not correspond to reality (the words *naine* 'woman', *ema* 'mother', *tüdruk* 'girl', *tütar* 'daughter' frequently co-occur with verbs of verbal expression, but studies do not confirm that women are verbally more active than men). A collocation is rather an abstraction or simplification that shows how people are typically perceived in a culture. Collocation as a frequent co-occurrence is a generalization, which – besides reflecting the current generalizations and stereotypes of a culture – creates some new stereotypes by means of collocational patterns. After all, language enables the community not only to name things, but by naming, also to convey and develop attitudes.

ELULOOKIRJELDUS

Liisi Piits

Sünniaeg ja -koht: 13.08.1976, Tallinn
Kodakondsus:Eesti
Aadress: Eesti Keele Instituut
Roosikrantsi 6, Tallinn 10119
E-post: Liisi.Piits@eki.ee
Töökoht ja amet: Eesti Keele Instituut, teadur

Haridus

2000	Tallinna Pedagoogikaülikool, *baccalaureus artium* eesti keele ja kirjanduse erialal
2004	Tallinna Pedagoogikaülikool, *magister artium* eesti keele erialal
2005–2015	Tartu Ülikool, doktoriõpe (üldkeeleteadus)

Teenistuskäik

2004–2005	Rocca al Mare Kool; eesti keele ja kirjanduse õpetaja
2005–2008	Eesti Keele Instituut; erakorraline teadur
2008–	Eesti Keele Instituut; teadur

Väitekirjaga seotud publikatsioonid

Piits, Liisi (2013). Distributional Hypothesis: Words for ‘human being’ and their Estonian Collocates. Trames: Journal of the Humanities and Social Sciences, 141–158.

Piits, Liisi (2011). Mis värvi on jõulud? Uusküla, Mari; Sutrop, Urmas (Toim.). Värvinimede raamat (111–116). Tallinn: Eesti Keele Sihtasutus.

Piits, Liisi; Mihkla, Meelis (2011). Influences of contextual predictability and lexical prosody on Estonian word duration. In: Proceedings of the 17th International Congress of Phonetic Sciences: The 17th International Congress of Phonetic Sciences (ICPhS XVII), Hong Kong, China, August 17–21, 2011. (Toim.) Wai-Sum Lee & Eric Zee. Hong Kong: City University of Hong Kong, 2011, 1598–1601.

Piits, Liisi; Mihkla, Meelis; Kiissel, Indrek (2008). Effect of collocational strength on Estonian speech rate on the example of the verb olema ‘be’. The second Swedish Language Technology Conference (SLTC 2008), Stockholm, November 20-21, 2008. Stockholm 2008, 17–18.

Piits, Liisi (2007). Kas eesti mees on mees või inimene? Ariadne Lõng: nais- ja meesuuringute ajakiri, 2005/2006, 3–10.

Piits, Liisi (2006). Sõnade mees ja naine kollokatsioonide võrdlemise võimalusi eesti, saksa ja inglise keele korpustes. Tartu Ülikooli üldkeeleteaduse õppetooli toimetised 6, Tartu, 105–114.

CURRICULUM VITAE

Liisi Piits

Date and place of birth: 13.08.1976, Tallinn
Citizenship: Estonian
Address: Institute of the Estonian Language
Roosikrantsi 6, Tallinn 10119
E-mail: Liisi.Piits@eki.ee
Current position: Institute of the Estonian Language; Researcher

Education

2000	Tallinn Pedagogical University, 2000 BA Estonian language and literature
2004	Tallinn Pedagogical University, 2004 MA Estonian language
2005–2015	University of Tartu, PhD student (general linguistics)

Professional experience

2004–2005	Rocca al Mare Kool; Teacher
2005–2008	Eesti Keele Instituut; Extraordinary Researcher
2008–	Eesti Keele Instituut; Researcher

Publications

Piits, Liisi (2013). Distributional Hypothesis: Words for 'human being' and their Estonian Collocates. Trames: Journal of the Humanities and Social Sciences, 141–158.

Piits, Liisi (2011). Mis värvi on jõulud? Uusküla, Mari; Sutrop, Urmas (Toim.). Värvinimede raamat (111–116). Tallinn: Eesti Keele Sihtasutus.

Piits, Liisi; Mihkla, Meelis (2011). Influences of contextual predictability and lexical prosody on Estonian word duration. In: Proceedings of the 17th International Congress of Phonetic Sciences: The 17th International Congress of Phonetic Sciences (ICPhS XVII), Hong Kong, China, August 17–21, 2011. (Toim.) Wai-Sum Lee & Eric Zee. Hong Kong: City University of Hong Kong, 2011, 1598–1601.

Piits, Liisi; Mihkla, Meelis; Kiissel, Indrek (2008). Effect of collocational strength on Estonian speech rate on the example of the verb olema ‘be’. The second Swedish Language Technology Conference (SLTC 2008), Stockholm, November 20-21, 2008. Stockholm, 2008, 17–18.

Piits, Liisi (2007). Kas eesti mees on mees või inimene? Ariadne Lõng: nais- ja meesuuringute ajakiri, 2005/2006, 3–10.

Piits, Liisi (2006). Sõnade mees ja naine kollokatsioonide võrdlemise võimalusi eesti, saksa ja inglise keele korpustes. Tartu Ülikooli üldkeeleteaduse õppetooli toimetised 6, Tartu, 105–114.

DISSERTATIONES LINGUISTICAE UNIVERSITATIS TARTUENSIS

1. **Anna Verschik**. Estonian yiddish and its contacts with coterritorial languages. Tartu, 2000, 196 p.
2. **Silvi Tenjes**. Nonverbal means as regulators in communication: sociocultural perspectives. Tartu, 2001, 214 p.
3. **Ilona Tragel**. Eesti keele tuumverbid. Tartu, 2003, 196 lk.
4. Einar Meister. Promoting Estonian speech technology: from resources to prototypes. Tartu, 2003, 217 p.
5. **Ene Vainik**. Lexical knowledge of emotions: the structure, variability and semantics of the Estonian emotion vocabulary. Tartu, 2004, 166 p.
6. **Heili Orav**. Isiksuseomaduste sõnavara semantika eesti keeles. Tartu, 2006, 175 lk.
7. **Larissa Degel**. Intellektuaalsfäär intellektuaalseid võimeid tähistavate sõnade kasutuse põhjal eesti ja vene keeles. Tartu, 2007, 225 lk.
8. **Meelis Mihkla**. Kõne ajalise struktuuri modelleerimine eestikeelsele tekst-kõne sünteesile. Modelling the temporal stucture of speech for the Estonian text-to-speech synthesis. Tartu, 2007, 176 lk.
9. **Mari Uusküla**. Basic colour terms in Finno-Ugric and Slavonic languages: myths and facts. Tartu, 2008, 207 p.
10. **Petar Kehayov**. An Areal-Typological Perspective to Evidentiality: the Cases of the Balkan and Baltic Linguistic Areas. Tartu, 2008, 201 p.
11. **Ann Veismann**. Eesti keele kaas- ja määrsõnade semantika võimalusi. Tartu, 2009, 145 lk.
12. **Erki Luuk**. The noun/verb and predicate/argument structures. Tartu, 2009, 99 p.
13. **Andriela Rääbis**. Eesti telefonivestluste sissejuhatus: struktuur ja suhtlusfunktsioonid. Tartu, 2009, 196 lk.
14. **Liivi Hollman**. Basic color terms in Estonian Sign Language. Tartu, 2010, 144 p.
15. **Jane Klavan**. Evidence in linguistics: corpus-linguistic and experimental methods for studying grammatical synonymy. Tartu, 2012, 285 p.
16. **Krista Mihkels**. Keel, keha ja kaardikepp: õpetaja alagatatud parandussekventside multimodaalne analüüs. Tartu, 2013, 242 lk.
17. **Sirli Parm**. Eesti keele ajasõnade omandamine. Tartu, 2013, 190 lk.
18. **Rene Altrov**. The Creation of the Estonian Emotional Speech Corpus and the Perception of Emotions. Tartu, 2014, 145 p.
19. **Jingyi Gao**. Basic Color Terms in Chinese: Studies after the Evolutionary Theory of Basic Color Terms. Tartu, 2014, 248 p.
20. **Diana Maisla**. Eesti keele mineviku ajavormid vene emakeelega üliõpilaste kasutuses. Tartu, 2014, 149 lk.
21. **Kersten Lehismets**. Suomen kielen väylää ilmaisevien adpositioiden yli, läpi, kautta ja pitkin kognitiivista semantiikkaa. Tartu, 2014, 200 s.
22. **Ingrid Rummo**. A Case Study of the Communicative Abilities of a Subject with Mosaic Patau Syndrome. Tartu, 2015, 270 p.