

MARIA TUULIK

Adjektiivide süstemaatiline polüseemia
eesti keeles tajujadjektiivide näitel



MARIA TUULIK

Adjektiivide süstemaatiline polüseemia
eesti keeles tajujadjektiivide näitel



1632

TARTU ÜLIKOOL
kirjastus

Tartu Ülikooli humanitaarteaduste ja kunstide valdkond, eesti ja üldkeeleteaduse instituut

Väitekirja on filosoofiadoktori kraadi saamiseks kaitsmisele suunanud Tartu Ülikooli eesti ja üldkeeleteaduse instituudi nõukogu otsusega 13. oktoobril 2022.

Juhendajad: professor Renate Pajusalu (Tartu Ülikool)
dr Margit Langemets (Eesti Keele Instituut)

Oponent: PhD Ulla Vanhatalo (Helsingi ülikool)

Kaitsmise aeg ja koht: 16. detsember 2022 kell 14.15 Jakobi 2-438, Tartu ja Zoom

Doktoritöö valmimist on toetanud keeleteaduse, filosoofia ja semiootika doktori-
kool, rahastanud Euroopa Regionaalarengu Fond (Tartu Ülikooli ASTRA projekt
PER ASPERA).



Euroopa Liit
Euroopa
Regionaalarengu Fond



Eesti
tuleviku heaks

ISSN 1406-5657 (trükis)
ISBN 978-9916-27-096-7 (trükis)
ISSN 2806-237X (pdf)
ISBN 978-9916-27-097-4 (pdf)

Autoriõigus: Maria Tuulik, 2022

Tartu Ülikooli Kirjastus
www.tyk.ee

EESSÕNA

Semantika on valdkond, mis hakkas mind huvitama juba ülikooli päevil. Väga põnevalt ja kaasahaaravalt kõneles meile semantikast Renate Pajusalu, kellest saigi hiljem siinse väitekirja üks juhendajatest. Täpsemalt kujunes uurimisteema aga välja, kui asusin tööle Eesti Keele Instituudi sõnaraamatute osakonda. Leksikograafilises töös tuleb sõnasemantikaga seotud küsimusi pidevalt ette ja nii igapäevatöös kui ka teema valikul oli mulle eeskujuks ja nõuandjaks Margit Lange-mets, kelle palusin väitekirja juhendama. Tänan südamest Margitit ja Renatet suurepärase nõuannete, abivalmiduse, kannatlikkuse ja toetuse eest kõigil nendel aastatel. Aitäh, et te ei löönud mulle käega, kui mul jõud raugema kippus! Ilma teieta ei oleks seda tööd kunagi valminud.

Olen väga tänulik ka oma retsensentidele. Suur aitäh Tartu Ülikooli eelretsen-sendile Ann Veismannile, kelle põhjalik retsensioon aitas raampeatükki märksa paremaks ja terviklikumaks muuta. Tänan väga Ulla Vanhatalo väärtuslike ette-panekute ja sõbraliku tagasiside eest. Soojad tänusõnad Mari Uskülale, kelle asjatundlikke nõuandeid ja soovitusi hindan kõrgelt.

Doktoritöö kirjutamine on olnud pikk protsess ja mul on olnud õnn kogu selle aja töötada toetavas ja inspireerivas keskkonnas. Tänan häid kolleege ja kaas-doktorante Eesti keele Instituudist! Minu eriline tänu kuulub Ene Vainikule, kes on mind nii palju aidanud ja minuga kaasa mõelnud, et teda võiks pidada lausa mu kolmandaks juhendajaks. Väga väärtuslik on olnud ka lähedaste kolleegide – Kristina Koppeli ja Jelena Kallase – tugi, kellega läbi arutades muutus nii mõnigi mure väiksemaks ja sai nii mõnigi küsimus lahenduse.

Olen rõõmus ja tänulik, et peale toredate kolleegide EKI-st on mind abistanud reedeklubi eesotsas Helle Metslangiga. Reedeklubi on pakkunud võimalust oma kirjatükke vabamas, ent vägagi konstruktiivses õhustikus arutada ja Helle nakatav optimism on andnud jõudu tööga edasi liikuda ka siis, kui tulevik on tundunud tume.

Tänan Tartu Ülikooli meeldiva koostöö ja julgustava suhtumise eest! Amet-likes asjaajamistes on hindamatu abiline olnud Tiia Margus, kes leiab kiirelt lahendused kõigile probleemidele.

Viimaks, aga eriti palavalt, tänan oma perekonda ja sõpru, kes on olnud mulle vankumatult toeks ja rõõmuks sellel pikal teekonnal. Tänan oma ema ja isa, kelle peale saan alati kindel olla, ja elukaaslast Chrisi, kelle mõistev suhtumine ja iga-külgne abi on võimaldanud rahuliku südamega kirjatööle keskenduda.

Tallinnas 11. novembril 2022

SISUKORD

VÄITEKIRJA PUBLIKATSIOONID:	8
1. SISSEJUHATUS.....	9
1.1 Töö uurimisküsimused	10
1.2 Väitekirja ülesehitus ja publikatsioonide tutvustus	11
2. TAUST JA TEOREETILISED LÄHTEKOHAD.....	13
2.1 Distributiivne semantika.....	13
2.2 Süstemaatiline polüseemia	14
2.3 Polüseemia leksikograafias	16
2.4 Tajusõnad	17
2.5 Adjektiivi määratlemine	19
3. MEETOD JA ANALÜÜSI KÄIK	23
3.1 Semantilised tüübid	23
3.2 Kollokatsioonid, sõnavisandid ja tesaurus	24
3.3 Analüüsi käik.....	26
3.4 Meetodi piirangud	27
4. ALLIKAD JA MATERJAL.....	29
4.1 Korpused	29
4.2 Sõnastikud	31
4.3 Materjali kogumine	33
5. ÜLEVAADE UURITUD ADJEKTIIVIRÜHMADDEST	36
5.1 Rühmade suurus	36
5.2 Rühmade morfoloogiline kirjeldus.....	37
5.3 Rühmade sagedusandmete võrdlus	39
5.4 Rühmade polüseemsus	41
6. ÜLEVAADE POLÜSEEMIAMALLIDEST	42
6.1 Rühmade kattuvad polüseemiamallid	42
6.1.1 X-PSÜHH	43
6.1.2 Tajumallid.....	44
6.1.3 MATERJAL/AINE – X.....	45
6.1.4 X-NÄHTUS_LOODUS	46
6.1.5 X- IN/LOOM_OMADUS	46
6.2 Rühmade eriomased polüseemiamallid.....	47
6.3 Süstemaatiliste polüseemiamallide esitus sõnaraamatus	49
7. TULEMUSTE KOKKUVÕTE JA ARUTELU	52
7.1 Tajuadjektiivi süstemaatiline polüseemia.....	52
7.2 Adjektiivi korpusprofiil.....	55
7.3 Edasised uurimisvõimalused	57
SUMMARY	58

KIRJANDUS.....	66
LISA 1: Eesti adjektiivide semantilised tüübid.....	74
LISA 2: Värviadjektiivide polüseemiamallid.....	75
LISA 3. Temperatuuriadjektiivide polüseemiamallid.....	76
LISA 4. Kompimisadjektiivide polüseemiamallid.....	77
PUBLIKATSIOONID.....	79
ELULOOKIRJELDUS.....	178
CURRICULUM VITAE.....	179

VÄITEKIRJA PUBLIKATSIOONID

- P1 Tuulik, Maria (2014). Adjektiivide polüseemia korpuses ja sõnaraamatus. Eesti Rakenduslingvistika Ühingu aastaraamat, 10, 307–317.
- P2 Tuulik, Maria; Langemets, Margit (2016). Systematic Polysemy of Estonian Colour Adjectives. In: Margalitadze, Tinatin; Meladze, George (Ed.). *Lexicography and Linguistic Diversity. Proceedings of the XVII EURALEX International Congress* 702–708.
- P3 Tuulik, Maria (2020). Eesti temperatuuriadjektiivide polüseemiamallid. Eesti Rakenduslingvistika Ühingu aastaraamat. *Estonian Papers in Applied Linguistics*, 16, 223–240.
- P4 Tuulik, Maria (2023?). Eesti kompimisadjektiivide polüseemiamallid. Esitatud ajakirja *Keel ja Kirjandus*. Läbinud eelretsenseerimise.
- P5 Tuulik, Maria, Ene Vainik, Geda Paulsen, Ahti Lohk (2022). Kuidas ära tunda adjektiivi? Korpuskäitumise mustrite analüüs. Eesti Rakenduslingvistika Ühingu aastaraamat, 18, 279–302.

1. SISSEJUHATUS

Viimaste kümnendite jooksul on Euroopa leksikograafias toimunud suured muu- tused: sõnaraamatute formaat nihkub paberkujult elektrooniliseks¹ ning koos sellega on ka Eesti leksikograafia muutunud korpuspõhiseks ja liigub (pool)auto- maatsete lahenduste suunas (vt nt Koppel 2020, Kallas jt 2015). Arvutipõhise leksikograafia levikuga on kasvanud vajadus täpsema semantilise märgendamise ja süsteemsema esituse järele. Polüseemia on nii korpuslingvistikas kui ka leksikograafias jätkuvalt relevantne probleem. Eesti keele suurimas ja pidevalt täienevas EKI ühend sõnastikus on enamik elemente seotud konkreetse tähendusega, nt sünonüüme ühendatakse tähenduse, mitte märksõna tasandil ehk koostatakse kahepoolset (vt Tavast jt 2020). Seetõttu on süsteemsus tähendus- jaotustes olulisemal kohal, kui kunagi varem.

Siinne töö on esimene katse Eesti keeleteaduses selgitada tajuadjektiivide süste- maatilisi tähendustumustreid ehk polüseemiamalle. Polüseemiamallid pakuvad tuge leksikograafiale polüseemse sõna kirjeldamisel ning ka ainst keeleuurijale, kuna toovad välja seoseid eri tähendusvaldkondade vahel. Adjektiiv võib rõhu- tada erinevas kontekstis nimisõna erinevat omadust, näiteks *kare rätik* viitab puudutamisel mittesiledale, krobedale materjalile, *kare käitumine* karmile ja järsule olekule ja *kare hää* mittemeloodilisele, kähedale häälele. Uurides samasse semantilisse rühma kuuluvaid adjektiive ühiselt, on võimalik välja selgitada, milliste valdkondade vahel toimuvad süstemaatilised tähendusvaheldused ja seega, milliste teiste valdkondadega on uuritav semantiline rühm enim seotud. Täpsus- tan tajuadjektiivi tähendusvahelduste motivatsiooni töös kognitiivsete teooriate taustal (Lakoff ja Johnson 1980, Viberg 1983, Winter 2019 jt).

Väitekiri uurib eesti adjektiivide süstemaatilist polüseemiat kolme semantilise rühma näitel: värviadjektiivid, temperatuuriadjektiivid ja kompimisadjektiivid. Värvadjektiivide andmestikku kuulub 463 sõna (nt *punane, erkollane, tuha- karva, valkjās*), temperatuuriadjektiivide rühm moodustub 101 sõnast (nt *jahe, tulikuum, külmavõitu*) ja kompimisadjektiivide rühm 265 sõnast (nt *sile, terav, muhklik*). Andmestikud on komplekteeritud leksikograafiliste allikate ja korpuste abil. Iga sõna vähemalt üks tähendus võimaldab vastavalt väljendada kas 1) värvust, 2) temperatuuri, või 3) kombitavat omadust. Kuigi ka temperatuuri- adjektiive võib pidada kompimisadjektiivideks hulka, on nad siinses töös eraldi vaatluse all. Temperatuuri tajumisega on inimkehas seotud erinevad retseptorid võrreldes muu mehhaanilise tajuga (nt Zhang 2015) ning temperatuurisõnu on ka varem eesti tajusõnade uurimisel teistest kompimissõnadest lahus käsitletud (Sutrop 1998, Sutrop 2004). Väitekirja kompimisadjektiivide rühm keskendub (pinna)tekstuuri ja plastilisusega seotud omadustele, mida ei ole eesti keeles varem rühmana vaatluse alla võetud.

¹ 26 riiki hõlmav kasutajate uuring tõi välja, et 79,8% ükskeelsete sõnastike kasutajaid kasu- tab elektroonilist versiooni (Kosem jt 2019)

Töö üks teoreetiline lähtealus on distributiivse semantika hüpotees, mille kohaselt saab sõna tähendust tuletada tema kasutuskontekstist (Harris 1954, Firth 1957). Adjektiivide tähendus on teiste sõnaliikidega võrreldes eriti kontekstisidus ja sõltub substantiivist, mille täiendina teda kasutatakse (nt *soe vesi*, *soe naeratus*, *soe valgus*). Kõik väitekirja viis artiklit on korpuslingvistilised uurimused, kus vaatlen tegelikku keelekasutust ja konteksti suurimast uurimuse ajal saadaolevast eesti keele korpusest. Kasutan kontekstist süsteemsema ülevaate saamiseks korpuspäringuprogrammi Sketch Engine'i (Kilgarriff jt 2014, Kilgarriff jt 2004) tööriistu. Programm võimaldab vaadelda sõna kollokatsioone eri statistikute alusel, genereerida sõnavisandeid ja uurida sarnase kontekstiga sõnu tesauruse funktsiooniga. Polüseemia analüüsil on töös kasutusel kombineeritud meetod: kasutan visandeid ja tesaurust korpuskontekstist süsteemsema ülevaate saamiseks, kuid kasutusjuhtude analüüsil ja tähendusteks koondamisel tuginedes kvalitatiivsele analüüsile.

Tähenduste abstraherimisel ja polüseemiamallide väljaselgitamisel rakendan enda välja töötatud eesti adjektiivide semantilisi tüüpe. Substantiivsete tähenduste märgendamisel kasutan Margit Langemetsa (2010) eesti nimisõna süstemaatilise polüseemia analüüsil rakendatud semantilisi tüüpe. Süstemaatilise polüseemia määramisel tuginedes Langemetsa eeskujul Juri Apresjanile (1974), kelle järgi peab vähemalt kahe polüseemse sõna vähemalt kahe tähenduse vahel esinema sarnane tähendusstruktuur, et tähendustevahelist suhet saaks määrata süstemaatiliseks polüseemiaks. Tutvustan raampeatükis lähemalt väitekirja teoreetilisi lähte-kohti, uurimismeetodit ja materjali kogumist ning esitan üldistavalt uurimistöö tulemused. Väitekirja koostab koosneb raamosast ja viiest publikatsioonist (vt 1.2).

1.1 Töö uurimisküsimused

Doktoritöö põhieesmärk on tuvastada eesti tajuadjektiivide süstemaatilise polüseemia mallid, mis võimaldaks selgitada adjektiivisemantika seaduspärasid ja looks leksikograafidele teoreetilise aluse, mille põhjal oleks võimalik ühtlustada adjektiivide polüseemia esitust sõnaraamatus. Töö tegeleb ka adjektiivide kui sõnaliiki iseloomustavate parameetrite kaardistamisega korpuses. Väitekirja rakenduslik eesmärk on muuta sõnastikutööd hõlpsamaks ja süsteemsemaks ning aidata kaasa adjektiivide täpsemale semantilisele ja sõnaliigilisele esitusele leksikograafias. Rakenduslik väljund on analüüsis selgunud semantiliste tüüpide ja süstemaatilise polüseemia mallide lisamine Eesti Keele Instituudi sõnastikusüsteemi Ekilex (vt Tavast jt 2018, Langemets jt 2021), et neid oleks võimalik kasutada EKI ühendsõnastiku (Langemets jt 2022) koostamisel leksikograafiliste abivahenditena. Detailsemalt on töö uurimisküsimused järgmised (sulgudes esitan iga uurimisküsimuse järel publikatsiooni(d), kus vastavat teemat käsitlet):

1. Millised on 100 sagedasema adjektiivialusel põhilised eesti adjektiivide semantilised tüübid, mida saaks kasutada tähenduste märgendamisel? (P1)
2. Millised on värvi-, temperatuuri- ja kompimisrühma morfoloogilise struktuuri, sagedusandmete ja polüseemia seosed? (P2, P3, P4)
3. Kui rikkalikult on eesti keeles adjektiividega kaetud värvi-, temperatuuri- ja kompimisvaldkond? (P2, P3, P4)
4. Millised on semantiliste tüüpide abil korpuspõhiselt leitud värvi-, temperatuuri- ja kompimisrühma süstemaatilise polüseemia mallid? (P2, P3, P4)
5. Kuidas on polüseemiamallid leksikograafiliselt esitatud? (P2, P3, P4)
6. Millised on tajuadjektiivide tähenduse ülekande eelisvaldkonnad? (P2, P3, P4)
7. Milliste parameetrite abil oleks võimalik (prototüüpsele adjektiivile omistatavate tunnuste põhjal) adjektiive korpuses teistest sõnaliikidest eristada? (P5)

1.2 Väitekirja ülesehitus ja publikatsioonide tutvustus

Väitekirja koosneb raamosast ja viiest publikatsioonist (2014–2022). Raamosa esimeses peatükis annan ülevaate töö eesmärkidest. Raamosa teises, teoreetilise tausta peatükis tutvustan distributiivse semantika teooriat, süstemaatilise polüseemia käsitlemist ja polüseemia esitust leksikograafias. Annan ülevaate ka taju-sõnade eripärast ja liigitusest ning toon välja eesti adjektiiviklassiga seonduvad sõnaliigiprobleemid. Kolmandas, meetodi peatükis tutvustan semantilisi tüüpe analüüsivahendina, Sketch Engine'i tööriistu korpuskonteksti uurimisel ja peatun meetodi piirangutel. Neljandas osas "Allikad ja materjal" kirjeldan uurimusega seotud korpuseid, sõnastikke ja rühmade komplekteerimist. Viies osa annab ülevaate töös uuritud rühmade morfoloogilisest struktuurist, sagedusandmetest ja polüseemsusest. Kuuendas osas käsitlen kokkuvõtlikult tajuadjektiivide polüseemiamalle.

Artiklis "Adjektiivide polüseemia korpuses ja sõnaraamatus" (P1) uurin eesti adjektiivide polüseemiat koondkorpuse 100 sagedama adjektiivinäitel. Artikli esimeses pooles vaatlen valimi jagunemist polüseemia seisukohast ja selleks, et selgitada polüseemia ja struktuuri suhteid, analüüsin adjektiivide morfoloogilise struktuuri varieerumist. Teises pooles keskendun semantilistele tüüpidele, mille abil saaks polüseemsete adjektiivide tähendusnühkeid paremini kirjeldada. Tutvustan semantiliste tüüpide rakendusvõimalusi sõnaraamatute koostamisel ja pakun välja eesti adjektiivide peamised semantilised tüübid. Artiklis P1 esitatud semantiliste tüüpide loetelu ei ole ammendav, edasise analüüsi käigus tüübid täpsustusid.

Väitekirja teine artikkel "Systematic Polysemy of Estonian Colour Adjectives" (P2) käsitleb värviadjektiivide süstemaatilist polüseemiat. Artikli kaasautor on Margit Langemets. Uurime 463 värvisõna polüseemiat ja morfoloogilist struktuuri. Arvutame välja valimi juursõnade, tuletiste ja liitsõnade polüseemia indeksi, st keskmise tähenduste arvu sõna kohta. Süstemaatilise polüseemia mustrite

kaardistamisel kasutame eelmises artiklis välja töötatud adjektiivide semantilisi tüüpe. Värvisõnade rühmas tuvastame üheksa polüseemiamalli. Uurime korpusmaterjali põhjal mallide esilduvust ja arutame nende esituse üle “Eesti naaber sõnade sõnastikus” (2019). Olen panustanud artikli kõigi osade kirjutamisse.

Artiklis “Eesti temperatuuradjektiivide polüseemiamallid” (P3) annan ülevaate eesti temperatuuradjektiivide rühmast ja selgitan seaduspärasusi sõnade semantilises struktuuris. Eraldi pööran tähelepanu eesti põhitemperatuurisõnadele (Sutrop 2004) ja nendega korpuses (distributiivse tesauruse suhetes) seonduvate sõnade hinnangulisusele. Analüüsi tulemusena tuleb esile 10 polüseemiamalli. Artiklis arutlen levinumate polüseemiamallide võimaliku motivatsiooni üle. Peatun pikemalt temperatuuritähenduste vaheldustel psühholoogiliste omadustega ja teiste tajuvaldkondadega. Tajudega seotud polüseemiamallide juures kõrvutan enda tulemusi Åke Vibergi (1983) tajuhierarhia teooriaga. Kirjeldan ka polüseemiamallide leksikograafilist esitust ja toon näiteid esituste variatsioonist.

Neljanda artikli “Eesti kompimisadjektiivide polüseemiamallid” (P4) eesmärk on selgitada seaduspärasusi eesti kompimisadjektiivide semantilises struktuuris ja tuvastada korpuse põhjal valimi süstemaatilise polüseemia mallid. Käsitlen kompimisega seotud polüseemiamallide motivatsiooni kognitiivsete teooriate taustal. Kuna tegemist on kompimisadjektiivide rühma esmakordse käsitlusega eesti keeles, arutlen põhjalikumalt rühma piiritlemise raskuste üle ja annan valimist üksikasjaliku ülevaate. Kirjeldan põhilisi maailmaga suhestumise taktiilseid dimensioone, mis valimis esile tulid, ja arutlen valdkonna sagedamate sõnade üle. Lisaks uurin ühendsõnastikust kompimisadjektiivide polüseemiamallide esitust ning vaatlen mallide leksikograafilist esitusviisi.

Viies artikkel “Kuidas ära tunda adjektiivide? Korpuskäitumise mustrite analüüs” (P5) on kirjutatud koos Ene Vainiku, Geda Paulseni ja Ahti Lohuga. Artiklis uurime adjektiivide morfosüntaktilisi tunnuseid ja selgitame, kuidas on prototüüpsele adjektiivile omistatavad tunnused (nt ühildumine, võrrete moodustamine) adjektiiviklassile eriomased. Testime prototüüpse adjektiivide tunnuseid korpusandmestikus kuue parameetriga. Vaatluse all on 12 sõnarühma, millest igähte kuulub 10 sõna. Sõnavalikul arvestame adjektiiviklassi piiripealseid juhtumeid ja leksikograafilisi kitsaskohti. Olen kavandanud eksperimendi, tegelenud parameetrite väljatöötamisega, valinud testrühmad, läbi viinud koondanalüüsi ja kõrvalekaldeanalüüsi. Artiklis olen kirjutanud “Testsõnade valiku”, “Parameetrite”, “Koondanalüüsi” ja “Kõrvalekaldeanalüüsi” peatükid ning koostöö teiste autoritega “Kokkuvõtte ja arutluse” osa.

2. TAUST JA TEOREETILISED LÄHTEKOHAD

Teises peatükis tutvustan uurimuste teoreetilist tausta: distributiivset semantikat ja süstemaatilise polüseemia käsitlusi. Arutlen ka polüseemia esituse üle leksikograafias. Alapeatükis „Tajusõnad“ teen lühikese ülevaate tajusõnade käsitlusest kognitiivses lingvistikas. Kuna nii värvi-, temperatuuri- kui ka kompimisadjektiivid kuuluvad tajusõnavara hulka, siis sobib uurimuse tulemusi kõrvutada ka kognitiivsete teooriatega. Alapeatükis „Adjektiivi määratlemine“ tutvustan, kuidas olen uurimustes adjektiiviklassi piiritlenud.

2.1 Distributiivne semantika

Üks doktoritöö lähtealus on distributiivse semantika hüpotees, mille järgi kalduvad kaks sõna, mis esinevad sarnases kontekstis, väljendama sarnast tähendust (Harris 1954). Zellig Harris (1954: 155–156) kirjeldab tähendust distributsiooni funktsioonina: kui me leiame, et sõnad A ja B on tähenduselt kaugemad kui sõnad A ja C, siis sageli selgub, et ka A ja B distributsioonid on omavahel erinevamad kui sõnad A ja C, seega tähenduse erinevus korreleerub distributsiooni erinevusega.

John Rupert Firth (1957:11) arendas edasi ideed, mis käsitleb tähendust funktsioonina kontekstis ning sõnastas distributiivse semantika hüpoteesi ühe levinuma definitsiooni: sõna iseloomustavad tema kaasteelised² (*a word is characterized by the company it keeps*). Firthi tööd semantika ja kollokatsioonide vallas mõjutasid nii süsteemfunktsionaalset lingvistikat kui ka korpuslingvistika arengut (Halliday, Yallop 2007: 61). Teooria, mis näeb tähendust kui kasutust ja konteksti, on korpuslingvistika ajastul saanud uue mõõtme ja uued võimalused. Korpus võimaldab uurida loomulikke keelekasutust ja pakub uurijale mahukaimat keeleandmestikku. Siinses doktoritöös on kasutusel tänapäevased korpustööriistad, mis võimaldavad suure tekstikorpuse põhjal ülevaatlilikult uurida kontekstis sündivat tähendust.

Distributiivse semantika hüpoteesist on välja kasvanud statistiline semantika. Mudelite toel mõõdetakse tähenduste sarnasust nii, et distributiivne info teisedatakse vektoritesse, mille sarnasuse põhjal tehakse järeldusi tähenduse sarnasuse kohta (Burghard 1992). Vektorruumi mudelid esitavad sõnatähendusi ruumiliselt, tähenduselt sarnased sõnad asuvad üksteisele lähemal (Sahlgren 2006: 17–24). Gemma Boleda ja kolleegid (2012) on teinud katse määratleda distributiivse ja morfoloogilise info põhjal automaatselt katalaani polüseemsete adjektiivide semantilisi klasse. Eestis uurib adjektiivide polüseemiat distributiivse semantika mudelite abil Aimi Pikksaar³ (doktoriõpingute raames) ning sõna- ja tähendusvektoritega on eesti ühendverbe uurinud oma doktoritöös Eleri Aedmaa (2019).

² autori tõlge

³ https://www.etis.ee/CV/Aimi_Pikksaar

Siiski on statistiliste mudelite eristusvõime praegu veel piiratud: on tehtud katseid, kus eristatakse kahte tähendust (Heylen jt 2015, Hilpert ja Saavedra 2020, Pikksaar 2022), kuid ulatuslikuma polüseemia analüüsiks usaldusväärset mudelit ega automaatlahendust ei ole. Siinse töö eesmärk on uurida tajuadjektiivide (süstemaatilist) polüseemiat detailsemalt kui kahe tähenduse ulatuses ning seetõttu kasutan küll automaatseid abivahendeid korpuskontekstist parema ülevaate saamiseks (Sketch Engine'i sõnavisand, tesauruse ja sõnaloendi funktsioonid), ent konteksti tähendustesse paigutamisel ja nende tähenduste märgendamisel tuginen eelkõige siiski iseenda kui emakeelse kõneleja keeletajule.

2.2 Süstemaatiline polüseemia

Polüseemia probleemiga puutuvad kokku nii leksikograafid kui ka loomulikku keelt analüüsivad programmid. Kuigi sõnatähendust võib käsitleda erineval viisil (Trier 1931, Cruse 1986, Katz 1987, Goddard 2012 jt) või vaielda, kas see üldse eksisteerib (Hanks 2017), siis praktikud peavad igal juhul vastu võtma otsuse, kuidas sõna tähendusstruktuuri jagada ja kui mitme tähendusena seda esitada. Leksikograaf ja korpuslingvist Adam Kilgariff toob oma doktoritöös “Polüseemia” (orig Polysemy) (1995: 125) välja, et polüseemia võib esmasel kohtumisel meenutada mitmeid teisi keelenähtusi: homonüümiat, ellipsit, metafoori; süntaktilist variatsiooni või kollokatsiooni. Küsimusele “mida tähendab, kui sõnal on mitu tähendust?” saab vaadata kõigi nende nurkade alt.

Jurij Apresjani (1974) määratlusel saab polüseemiast rääkida siis, kui sõna kaks või enam tähendust on omavahel seotud – sealjuures ei pea aga olema ühisosa polüseemse sõna kõigi tähenduste vahel, vaid piisab, kui iga tähendus on seotud vähemalt ühega teistest tähendustest. Selles küsimuses, kas sõna üks tähendus on tema teise tähendusega seotud, ei pruugi uurijad (ega leksikograafid) aga alati sugugi üksmeelsed olla. Leksikograafid puutuvad traditsiooniliselt kokku ka polüseemia ja homonüümia eristusega, mis samuti nõuab tähenduste seotuse üle arutamist. Kui polüseemia puhul peaks iga tähendus vähemalt ühe teise sama sõna tähendusega seotud olema, siis homonüümia puhul langeb kokku eri sõnade vorm (kirjapildis või häälduses), nii et mitu tähendust esineb ühe, juhuslikult kokkulangevaks peetava vormi sees (Langemets 2010: 26). Ühest mõõdikut aga tähenduse seotuse üle otsustamiseks ei ole. Langemets (2010: 28–30) toob mitmeid näiteid Eesti leksikograafiast, kus varasemalt homonüümidenäidena esitatud sõnad (nt *taaler* kui ’münt’ või kui ’maa hindamise ühik’, *kann* kui ’lelu’ või kui ’lill’, *talent* kui ’vaimuanne’ või kui ’massiühik’) on hiljem kokku toodud polüseemse sõna tähendusteks. Ka Apresjan mõõnab (1974:13), et polüseemia ja homonüümia on suhtelised mõisted ja leiab ainult, et vähese ühisosaga tähendused on vähem homonüümsed kui need, millel igasugune ühisosa puudub.

Süstemaatilise ehk regulaarse polüseemia selgitamisel juhindun doktoritöös Apresjani käsitlusest (1974), mille kohaselt on sõna A polüseemia keeles regulaarne, kui keeles on vähemalt veel üks sõna B, mille tähendusi b_1 ja b_2 eristatakse üksteisest täpselt samal viisil nagu sõna A tähendusi a_1 ja a_2 ning kui a_1 ja b_1 ega a_2

ja b_j ei ole omavahel sünonüümid. Niisiis peab vähemalt kahe polüseemse sõna vähemalt kahe tähenduse vahel esinema sarnane tähendusstruktuur, et tähenduste vahelist suhet saaks määrata süstemaatiliseks polüseemiaks. Apresjan kasutab küll mõistet regulaarne polüseemia, ent siinses doktoritöös olen Langemetsa (2010) eeskujul valinud süstemaatilise polüseemia nimetuse. Neid kahte mõistet on teaduskirjanduses kasutatud sünonüümselt. James Pustejovskyl (1995) on sama nähtust kirjeldades kasutusel ka mõiste “loogiline polüseemia”.

Regulaarset polüseemiat juhivad Apresjani (1974) järgi protsessid, mis on produktiivsed, korrapärased ja ennustatavad – sarnaselt protsessidele, mis juhivad sõnaloomet. Ta näebki polüseemiat semantilise derivatsioonina. Apresjan toob ühe sellise protsessina välja metonüümse tähendusülekanne (nt *foot* 'jalg' → *foot of the mountain* 'mäejalam'), mida ta peab oma olemuselt regulaarsemaks kui metafoorset tähendusülekanne. Traugot ja Dasher (2002: 280) leiavad, et kuna keelekasutajatel on tendents kategoriseerida, kerkivad uued polüseemiad juba olemasolevatele kontseptuaalsetele struktuuridele.

Eesti nimisõnapolüseemias esineb Langemetsa doktoritöö (2010) põhjal metonüümset polüseemiat oluliselt rohkem kui metafoorset polüseemiat. Kui metafoorseks polüseemiaks määratleti analüüsis 100 tähendusülekanne paari, siis metonüümseks liigitus 380 juhtumit ehk ligi neli korda rohkem (Langemets 2010: 217). Anastasiya Lopukhina viis koos kolleegidega läbi eksperimendi (2018), kus osalejatel paluti polüseemseid sõnu (nii nende otseseid, metafoorseid kui ka metonüümseid tähendusi) sisaldavaid fraase grupeerida nii, et samad tähendused oleksid samas rühmas. Tulemuste põhjal järeldasid uurijad, et tähenduste talletamise viis mentaalses leksikonis erineb sõnaliigiti: nimisõnade ja verbide puhul talletatakse otseseid ja metonüümseid tähendusi mentaalses leksikonis koos ning metafoorseid tähendusi neist eraldi. Adjektiivide puhul aga kattusid metonüümset tähendused suurel määral nii otseste kui metafoorsete tähendustega. Siinses doktoritöö artiklites metafoorseid ja metonüümseid tähendusvaheldusi arvuliselt ei loetleta, kuid teen järelduste peatükis tähelepanekuid tajuadjektiivide tähendusülekanne eelismehhanismide kohta.

Leksikograafialähedaselt on uurinud polüseemiat James Pustejovsky, kes nimeatab loogiliseks polüseemiaks (sünonüümne süstemaatilise polüseemiaga) sellist komplementaarset mitmetähenduslikkust, kus ei toimu leksikaalse kategoria muutust ja sõna mitmel tähendusel on osaliselt kattuv, teineteisest sõltuv, ühine tähendus (1995: 28). Seega leian, et kuna süstemaatiline polüseemia on mono-seemiale lähemal kui mittesüstemaatiline polüseemia (vt ka nt Langemets 2009: 29), ei toimiks ilmselt distributiivse semantika automaatanalüüsi mudelid selle puhul hästi. Mida erinevamad on tähenduste kasutuskontekstid, seda paremini suudavad distributiivsed mudelid neid eristada. Kuna süstemaatilise polüseemia puhul on tähendustel suurem ühisosa ja ka rohkem kattuvaid kontekste, järeldan, et (praegu veel) peaks analüüsima süstemaatilist polüseemiat mitteautomaatsete meetoditega.

2.3 Polüseemia leksikograafias

Nii eesti kui inglise sõnastike ülevaated (Langemets 2010; Atkins, Rundell 2008) tõendavad, et enamik keeles leiduvaid sõnu on ühetähenduslikud. Nende tähenduste selgitamine ja esitus sõnastikes peaks olema seega kergem ülesanne, kuna leksikograaf ei pea pead murdma mitmetähenduslikkuse üle. Mida lähemale jõuame aga sageli kasutusel olevatele sõnadele, seda polüseemsemaks muutuvad sõnad (vt nt P1, P2). Kui võrrelda omavahel eesti keele suurt seletavat sõnaraamatut (=EKSS), millel on ligikaudu 140 000 märksõna, ja eesti põhisõnavara sõnastikku, mis sisaldab ligikaudu 5000 sagedasemat sõna, leiabki polüseemseid sõnu erineval määral. EKSSi 1. trüki 137 767 märksõnast on polüseemne ligikaudu iga kümnes sõna (kõigist märksõnadest on enam kui ühe põhitähendusega sõnu kokku 14 432 (10,5%)) (Langemets 2010: 254). Põhisõnavara sõnastiku (=PSV) 5392 märksõnast on polüseemsena esitatud 1296 sõna ehk ligi neljandik (24%) märksõnadest⁴. Kui arvestada juurde, et PSV on oma tähenduste esituses ka oluliselt lakoonilisem kui EKSS (PSV sihtrühm on keeleõppijad ja valitud on ainult tähendused, mis korpuses sageli esile tulevad (vt lähemalt Kallas, Tuulik 2011)), siis oleks võrdselt põhjaliku tähenduste esituse korral polüseemsuse erinevus ilmselt veelgi suurem. Kuna keele põhisõnavara kipub olema polüseemne, peavad leksikograafid oma töös neid sagedaid ja polüseemseid sõnu analüüsima ja kirjeldama. Et saada ülevaade adjektiivide süstemaatilise polüseemia võimalikust leksikograafilisest esitusest, uurin artiklites korpuse kõrval ka polüseemia-mallide esitust sõnastikus.

Paratamatult jäävad sõna semantilist analüüsi mõjutama uurija subjektiivsed otsused. Geart van der Meer (2006: 604) on öelnud „Pärast sajandeid praktilist leksikograafiat, ei ole ikka veel üksmeelt selles, kuidas jagada leksikaalse üksuse semantilist ruumi“. Ei ole üksmeelt uurijate vahel, mida sõnatähendused endast kujutavad; ei selles, kui laiad või kitsad need olema peaks; ega lõplikke juhiseid, kuidas teha kindlaks, kus üks tähendus lõppeb ja teine algab (Kilgarrieff 2006: 29). Kilgarrieff (1995) võtab kokku, et sõnastikutähendusena väljatoomisel lähtutakse sageli sellest, kas tähendus on piisavalt sage ja ebapiisavalt ennustatav. Kuid ka see määratlus jätab kõvasti tõlgendamisruumi. Käesolev töö on välja kasvanud leksikograafilistest küsimustest ja soovist tabada süsteemi eesti adjektiivipolüseemias, töö ei pretendeeri edasi arendama kognitiivse ja funktsionalistliku keeleteaduse polüseemiateooriaid.

Korpus võimaldab vaadelda loomulikku keelt kasutusel, uurida korraga suurt hulka suhtlussündmusi. Leksikaalsemantik Alan Cruse (1986: 53) kinnitab, et leksikaalsel üksusel võib olla erinev tähendus igas konkreetsetes erinevas kontekstis, milles ta esineb. Siiski on sõnastike töö olnud tähenduste esitusest üldistada. Leksikograafilises traditsioonis (Atkins ja Rundell 2008: 272–273) tuletavad leksikograafid lõpmatust arvust sõna üksikkasutustest lõpliku valiku leksikograafilisi seisukohti (*lexicographic utterance*), mis seletavad, kuidas see sõna panustab kõigi nende üksiksündmuste tähendusse.

⁴ Allikas on PSV andmebaas Eelexi sõnastikubaasis: <https://eelex.eki.ee/>

Sõnastikud esitavad polüseemset sõna enamasti nummerdatud tähenduste reana, kusjuures tähendused on esitatud põhitähenduste ja alltähendustena (nt EKI ühendsõnastik) või üksikult eraldiseisvate tähenduste loendina (nt PSV). Korpuslingvist ja leksikograaf Patrick Hanks leiab aga (2000: 205), et nummerdatud definitsioonide read sõnastikes aitavad kaasa sellele, et tekib moonutatud pilt tegelikult keelekasutuses toimuvast. Eesti tajuverbide polüseemia uurija Mariann Proos (2021: 14–15) on hästi kokku võtnud kognitiivse keeleteaduse seisukoha, mille kohaselt ei ole keelekasutajate peas eraldiseisvaid leksikaalseid kirjeid, millest igaüks esindab ainult ühte tähendust, vaid polüseemia on keeruline struktuur kohati kattuvaid ja ebamääraste piiridega kategooriaid. Seega tasub ettevaatusega suhtuda ka sinse töö tulemustesse. Tuleb silmas pidada, et analüüsiüksused on siiski abstraktsioonid. Ma ei väida, et alati on tegemist on kindla-piiriliste kognitiivselt eraldiseisvate tähendustega. Uurin rühmade semantikat pigem ülevaاتlikult, lootusega anda aimu süsteemist, mis võib just nende semantiliste rühmade puhul polüseemiat kujundada ning eesmärgiga pakkuda polüseemiamallide kujul tuge leksikograafidele, kes peavad sõnatähendusi eristades igal juhul otsuseid vastu võtma.

2.4 Tajusõnad

Doktoritöö artiklites on vaatluse all värvi-, temperatuuri- ja kompimisadjektiivid. Kogu doktoritöö valim on seotud tajudega ning seetõttu sobib tulemusi kõrvutada ka kognitiivsete teooriate ja neurouuringutega. Kognitiivne lingvistika näeb keelt kui fundamentaalselt kognitiivsetele protsessidele tuginevat nähtust (Langacker 1987, Johnson 1987, Lakoff ja Johnson 1980) ning kognitiivsete protsesside alla kuuluvad nii nägemismeel kui ka kompimise eest vastutav somatosensoorne meel.

Traditsiooniline mudel, mida tutvustas antiikaja filosoof Aristoteles, jaotab tajud viieks meeleks: nägemine, kuulmine, kompimine, maitsmine ja haistmine (nt Sorabji 1971). Mudeli järgi on iga meel seotud konkreetse meeelundiga: silmad nägemisega, kõrvad kuulumisega, nahk kompimisega, keel maitsmisega ja nina haistmisega (nt de Vignemont ja Massini 2015). Siiski pole traditsiooniline mudel ainus võimalik jaotusviis. Inglise keele tajusõnade uurija Bodo Winter nimetab viie meelega mudelit „kasulikuks fiktsiooniks“ – tohutuks üldistuseks, ent adekvaatseks alguspunktiks keele ja tajude uurimisel (2019: 11–16). Probleemidena toob Winter (2019: 11–16) välja näiteks küsimuse, millise meelega paigutada valu (mida töötlevad kompimisest erinevad ajustruktuurid) ja kuhu liigitub Jacobsoni organ, mis vastutab feromoonide taju eest. Probleemidele vaatamata leiab Winter, et tänapäeva lääne ühiskonnas on viie meelega mudel kultuuriliselt dominantseim ning võimaldab saada olulisi tulemusi ja teha tajusõnavara kohta üldistusi, isegi kui tegemist on tahtliku abstraktsiooniga (2019: 235–247).

Winter (2019: 79–97) pakub välja seisukoha, et lingvistilised kategooriad nagu „maitsesõnad“ või „kompimissõnad“ joonistuvad välja ainult siis, kui keelekõnelejad, kaasa arvatud lingvistid, räägivad keelest metalingvistilisel moel ning et tõeline keelekasutus ei pruugi olla struktureeritud selliste kategooriate ümber.

Kuna ka minu töö käsitleb selliseid kategooriaid, siis toon vastuargumendiks ajuuritud, mis on proovinud kaardistada, kuidas keelelist-semantilist informatsiooni ajus töödeldakse (Huth 2016, Damasio 1996, Lerner jt 2011). Huth ja kolleegid (2016) kasutasid sisendina jutustusi ja märgendasid sisendsõnu valdkondade kaupa (visuaalne, taktiline, numeraalne, lokatsiooniline, temporaalne jne). Tulemused näitasid, et semantilise informatsiooniga tegelevad ajukoore erinevad laiad piirkonnad, ent iga laiem piirkond sisaldab kindlaid alasid, mis on seotud konkreetset tüüpi semantilise informatsiooniga, nt numbrid, asukohad, puudutusega seotud omadused jne. Need ajukoore-kaardid on erinevatel inimestel (detailideni) sarnased ning annavad alust väita, et eri semantiliste valdkondade sõnade töötlemisega tegelevad erinevad ajupiirkonnad.

Läbi keeleajaloo on välja pakutud erinevaid meelte hierarhiaid (nt Ullmann 1945, Williams 1976, Viberg 1983), mis järjestavad meeli selle põhjal, kui sageli nende sõnavara kasutatakse selleks, et rääkida teistest meeltest. Hierarhia eelduseks on seisukoht, et hierarhias kõrgemal kohal olevad meeled on kaetud rikkalikuma sõnavaraga. Doktoritöö artiklites kasutasin võrdlusalusena Vibergi (verbide põhjal) välja pakutud hierarhiat (1983):

• NÄGEMINE>KUULMINE>KOMPIMINE>HAISTMINE/MAITSMINE

Vibergi hierarhia põhjal on kõrgemad meeled nägemine ja kuulmine, kompimine paikneb keskel ja haistmine/maitsmine koos kõige alumisel astmel. Siiski võib keeltevahelise universaalse tajusõnade hierarhia olemasolus kahelda. 20 erineva keele kõnelejaid hõlmav uurimus (Majid jt 2018) leidis, et keeled erinevad selles, milliseid tajuvaldkondi nad rikkalikumalt kodeerivad ja leitud mustrid lubasid järeldada, et leksikaalne tähistus oli kultuuriliselt mõjutatud.

Mitmed uurijad on täheldanud, et värvisõnavara ja kompimissõnavara ei ole keelte vahel universaalsed (vt nt värvisõnavara kohta Berlin ja Kay 1969, Kay jt 2009, Gibson jt 2017, Wierzbicka 2008; kompimissõnavara kohta Majid jt 2018). Üks uuemaid uurimusi (Josserand jt 2021) vaatleb keeli, milles on olemas eraldi sõna sinise värvi tähistamiseks. Eksisteerib ka keeli, kus esineb värvikategooria (ing k *grue*), mis ühendab sinise ja roheline ühe nimetaja alla. Uuriti 142 kogukonda ja leiti, et üks mõjur värvisõna väljakujunemisel on UV-kiirgusest tingitud muutused silma võrkkestal, mis mõjutavad sini-rohe värvitaju ja raskendavad nende värvide eristamist ning seetõttu on värvisõnavara vastavalt kohastunud. Keeltes esines eraldi sõna sinise tarbeks tõenäolisemalt ka siis, kui keelekogukond paiknes seisva veekogu (nt järve) lähistel. Niisiis tuleb värvi-, ja ka muude tajusõnade variatsiooni lahti mõtestades arvestada nii kultuuriliste, bioloogiliste kui ka füüsilise keskkonnaga seotud teguritega.

Tajusõnade uurimises on oluline asjaolu, et tajusõnu saab kasutada ka muudes domeenides peale nende põhilise domeeni, st teistes kontekstides kui sõna peamine tajuvaldkond, nt *kare hääl* (põhiliselt kompimisega seotud adjektiiv väljendamas hääleomadust) (Winter 2019: 67). Selliseid väljendusi on peetud sünestetilisteks metafoorideks (nt Strik Lievers 2016) või keeleliseks sünesteesiaks

(nt Whitney 1952). Sünesteesia (keeleteadusest väljaspool) on tajunähtus, mis puhul ühe meele stimuleerimine toob kaasa tahtmatud kogemused ka teises meeleelundis (nt Cytowic 2002), näiteks tekib helide kuulmine inimeses värviaistingu. Kognitiivses keeleteaduses defineerib Yu (2003: 20) sünesteetilist metafoori kui metafoori, mis katab mitut tajuvaldkonda. Winter (2019: 67–97) väidab aga, et sünesteetilised metafoorid pole ei tõeliselt sünesteetilised ega ka mitte metafoorid. Ta toob esile, et keeleväline sünesteesia on haruldane ja erk tajukogemus, kus vahelduvad modaalsused ei ole omavahel seotud, on idiosünkraatilised ja kogeja kontrolli alt väljas, samas kui sünteetilised metafoorid on üldlevinud, seotud meeled on korrelatsioonis keskkonnaga ja kogeja kontrolli all. Sünesteetiliste metafooride metafoorsuse vastuargumentideks toob Winter (2019: 95–97) kaks aspekti: esiteks pole põhjust eeldada sünesteetilist metafoori, kuna väljenduses kohtuvad tajud on tugevalt omavahel seotud (tajuassotsiatsioon) või siis pole põhjust eeldada sünesteetilist metafoori, kuna tajudevahelised kasutused järgivad lihtsalt nende sõnade hinnangulisi tähendusi (nt magus kui hea). Winteri järeldus ühtib Rakova (2003: 147) seisukohaga, et kõik sünesteetiliste adjektiivide tähendused on nende otsesed (mitte metafoorsed) tähendused. Võtmata seisukohta tähenduste metafoorsuse osas, vaatleb sinne doktoritöö valitud tajuadjektiivide süsteemseid tähendusmalle ja selgitab välja põhilised valdkonnavaieldused.

Eesti keeleteaduses on tajusõnu käsitletud erinevad autorid. Eesti keele tajusõnade ülevaate andis Urmas Sutrop oma doktoritöös “Tajusõnavara eesti keeles: tüpologia, ajalugu ja kognitsioon” (1998). Tajuverbe on põhjalikult uurinud Mariann Proos (2019, 2020a, 2020b, 2021). Anu Kalda on uurinud doktoritöö teemana tajumetafooride tõlkimist (nt Kalda 2020, Kalda 2021). Keele, taju ja mõtlemise omavahelist mõju on analüüsinud Mari Uusküla (nt Uusküla 2018, Uusküla 2019, Uusküla ja Bimler 2020). Täpsemalt on Uusküla uurinud, kuidas nimetatakse ja kategoriseeritakse värve erinevates keeltes (Uusküla 2008, Uusküla 2011, Uusküla 2014 jt). Eesti tajusõnadest ongi kõige rikkalikumalt ilmunud värvisõnade uurimusi (vt Uusküla ja Sutrop 2011, Rätsep 2018, Oja 2001, eesti viipekeele kohta Hollmann 2010). Kuid käsitletud on ka lõhnasõnu (Zurbuchen 2017, Sutrop 2001, Sünd 2020) ja maitsesõnu (Sutrop 1998, Vainik 2018). Tajusõnad on seotud ka emotsioonisõnavaraga, mida on eesti keele seisukohast uurinud Ene Vainik oma doktoritöös (2004) ja hilisemates uurimustes (Vainik 2017, Vainik 2018 jt). Konkreetsemalt adjektiivide semantika ja isikuomaduste sõnavaraga eesti keeles on tegelenud Heili Orav (2006). Sinne doktoritöö keskendub esimesena tajuadjektiivide polüseemiale ning käsitleb ka esimesena kompimisõnade rühma.

2.5 Adjektiivi määratlemine

Adjektiivirühmade moodustamisel ja uurimisel on vajalik piiritleda adjektiivi kui sõnaklassi. Suur osa tajusõnadest on eesti keeles adjektiivid (nt eesti temperatuurisõnade loetelukatses olid adjektiivid kõik üle nelja korra nimetatud temperatuurisõnad (Sutrop 2004: 116). Sarnaselt soome keelele (Juvonen ja Nikunlassi 2015)

võivad paljud eesti temperatuurisõnad siiski esineda nii adjektiivivi kui substantiivi rollis (*külm jook* vs *külma kartma*). Teoreetiliselt on substantiivne kasutus võimalik ka iga värvisõnaga (*punane kleit, kannab punast*), kuid siiski on korpuskasutust uurides sagedaim adjektiiv.

Uus eesti keele grammatika (Metslang jt: ilmumas) jagab sõnad süntaktiliselt iseseisvateks sõnadeks ja mitteiseseisvateks ehk abisõnadeks, semantiliselt täistähenduslikeks ja mittetäistähenduslikeks sõnadeks ning morfoloogiliselt muutuvateks ja muutumatuks sõnadeks. Adjektiivivid kuuluvad iseseisvate täistähenduslike sõnade hulka, väljendavad tüüpjuhul (asja) omadust, saavad lauses esineda ühilduva täiendina ja öeldistaitena ning muutuvad käändes, arvus ja võrdlustmes.

Eesti adjektiivisüntaksi uurija Mati Ereht (2017) on välja toonud, et adjektiive, nagu teisigi sõnaliike, iseloomustavad teatud morfoloogilised, süntaktilised ja semantilised omadused, millele võib lisanduda pragmaatiline aspekt. Adjektiivivi põhiline semantiline eesmärk on viidata omadusele. Sageli on adjektiivivid seotud võrdluskategooriaga – neist on võimalik moodustada kesk- ja ülivõrret, ning adjektiividele on tüüpiline ka gradatsiooni väljendus määrsõnade abil (*väga soe, ülimalt pehme*).

Morfoloogiauurija Ülle Viks (1977: 525) toob esile, et kuigi sõnaliiki iseloomustab mingi diferentsiaaltunnuste kimp, ei pea iga sõna puhul sugugi olema esindatud kõik vastava sõnaliigi diferentsiaaltunnused. Viksi käsitluses on sõnaliik üldjuhul mittehomoogeenne klass, mis koosneb mitmest natuke erinevast homoogeenest rühmast. Samas ei moodustagi sõnaliigid alati teravate ja selgete piiridega klasse – leidub sõnu, mis jäävad kahe (või rohkemagi) sõnaliigi vahele. Nendel sõnadel on osa tunnuseid omased ühele sõnaliigile, osa teisele.

Selline teoreetiliselt plaanis igati arusaadav üleminekuala võib aga osutuda probleemiks rakenduslingvistikas, nt korpusmärgendamises ja leksikograafias, kus iga sõna peab endale külge saama sõnaliigimärgendi. Samuti tuleb keelesüsteemi ebamäärasusega rinda pista keeleteaduslikes uurimustes, kus on vaja võtta seisukoht, millise sõna võib analüüsi kaasata kui konkreetse sõnaliigi esindaja. Kuna doktoritöö materjal pärineb nii korpusest kui ka sõnaraamatutest/andmebaasidest, siis olid vajalikud täpsemad piirid: mida ikkagi lugeda adjektiiviks ja kaasata uurimusse.

Eesti keele sõnaliigi seisukohast ambivalentsete vormide uuring (Vainik, Paulsen ja Lohk 2020) näitas, et adjektiiviklassil esineb piiriala verbidega (nt mineviku ja oleviku partitsiibid *avatud, armunud, tulev* ning supiinid *rääkimata, värvimata*); nimisõnadega (nt *vaimulik, rase, pull*), pronoomenitega (nn proadjektiivid nagu *niisugune, samane, teistsugune*) ja adverbidega (nt *kiivas, krussis, purjus*). Kõik need äärealad vajasis selgitamist ka 100 sagedasema adjektiivivi ning tajoadjektiivide piiritlemisel.

Põhiliselt vajasis erilist tähelepanu partitsiibid ja käändumatud adjektiivivid. Käändumatud omadussõnad valimis olid nt *karva, -laadi, -võitu, -verd*-osisega liitsõnad (morfoloogilise analüüsi kohta vt pt 5.2). Eriti palju esines käändumatuid adjektiive värvisõnade hulgas tänu *-karva* osisele (*kullakarva, hiirekarva, ihukarva, kahvatuvõitu, meekarva, taevakarva, tulekarva* jt), kuid neid leidis ka

temperatuurisõnade (*jahedavõitu, kuumavõitu, külmavõitu, leigevõitu*) ja kompimisõnade rühmas (*kõvavõitu*). Käändumatud adjektiivid kaasasin uurimisse nende semantilisele ja süntaktilisele rollile tuginedes.

Partitsiipe uurides võis korpuses täheldada adjektiivse kasutuse (st kui palju esines sõna korpuses adjektiivi rollis) erinevat määra – moodustus skaala. Skaala ühes otsas esines partitsiipe, mille puhul oli korpuses adjektiivne kasutus valdav (nt *erinev*), isegi sel määral, et verb oli peaaegu läbipaistmatu, nt adjektiiv *palav* (*palama*). Seejärel esines sõnu, millel leidis kasutust nii verbivormina kui adjektiivselt (*läinud, kehtiv, tulev, langenud*). *Kehtiva* puhul on täheldatav ka leksikograafilise esituse areng, kuna esimese uurimuse ajal (P1) see eraldi märksõnana seletavast sõnaraamatust puudus, kuid praeguseks (2022) on adjektiivina ühend sõnastikku kaasatud. Adjektiivide automaatsesse loenditesse korpuses sattus ka sõnu, mille puhul verbiline kasutus ületas adjektiivset (nt *mõeldud, toimuv*) ja välja tuli sorteerida ka selgeid verbivorme (*asuv, kuuluv, pandud, osanud, hakanud*).

Rohkem kui ühes sõnaliigis esinevatest sõnadest oli valimis enim sõnu, millel leidis nii adjektiivset kui substantiivset kasutust. Neid esines igas uuritud rühmas: 100 sagedasema adjektiivi hulgas nt *kohalik, vana, noor, tuttav*; värvide puhul teoreetiliselt kõik värvisõnad; temperatuurisõnade rühmas nt *soe, külm, kuum* ning kompimisadjektiividest nt *kuiv, märg, karvane*. Temperatuurisõnade substantivatsiooni on uurinud Silvi Vare (2006), kes leidis, et kalduvus substantivatsiooniks näib olevat seotud omaduse intensiivsuse astmega ja tähenduse üldise iseloomuga. *Soe* ja *külm* kui kahe erineva temperatuuripooluse kaks kõige neutraalsemat esindajat, on substantiveerunud kõige ulatuslikumalt, samas kui ebamäärasema sisuga liikmed *leige* ja *jahe* ei ole muutusele allunud.

Nii adjektiivina kui ka substantiivina esinevatel sõnadel varieerus tähenduste (ja vastavate sõnaliikide) sagedus korpuses suurel määral, nt *poliitilise* kasutus substantiivselt poliitvangi tähenduses moodustab sõna *poliitiline* üldkasutusest oluliselt väiksema osa kui sõna haige substantiivne kasutus *haige* üldsagedusest. Pöörasin substantiivse tähendusega sõnadele analüüsis eraldi tähelepanu ja kaasasin need ka süstemaatilise polüseemia mallidesse.

EKI sõnaliigi-töörühma uurimused (nt Paulsen jt 2019, Paulsen jt 2020) annavad ülevaate sõnaliikide rollist eesti leksikograafide töös. Leksikograafide uuringus (Paulsen, Vainik ja Tuulik 2019) toodi välja sõnaliigitusprobleeme kirjeldavaid näiteid, millest 26% moodustasid adjektiiviklassiga seotud sõnad. Erilise probleemkohana rõhutati küsimust, kuidas teha kindlaks, kas verbi partitsiibil on juba piisavalt adjektiivset kasutust, et oleks õigustatud sõna lisamine sõnastikku/andmebaasi ka adjektiivina. Sellest inspireeritult uurib doktoritöö viimane artikkel (P5), millised parameetrid võiksid korpuses aidata adjektiive teistest sõnaliikidest eristada, et leksikograafidel oleks tulevikus ebaselgete juhtumite puhul võimalik paremini hinnata sõna adjektiveerumise astet.

Artiklis (P5) katsetatud parameetrid aitaksid adjektiivirühmi piiritleda ka teistes doktoritöö uurimustes. Lähtusin parameetrite väljatöötamisel adjektiivile omastest tunnusjoontest, mis pärinevad nii teoreetilisest kirjandusest (Erelt 2017) kui ka praktilisest leksikograafilisest tööst. (Testitud parameetreid ja korpusuuringu tulemusi kirjeldan kokkuvõtvas osas.) Piiripealsete juhtumite puhul

kontrollisin 1. sõna kalduvust esineda lauses substantiivi atribuudi või predikaatiivina, 2. sõna ühildumist järgneva substantiiviga, 3. võrdlusastmete võimalikkust, 4. määrsõnalaiendite võimalikkust, 5. seda, kas sõna saab alustada lauset täiendi positsioonis substantiivi ees. Adjektiiviks määramisel arvestasin kõigis uurimustes nii semantilist aspekti – kas sõna viitab omadusele – kui ka süntaktilist aspekti – kas sõna esineb adjektiividele omastes süntaktilistes rollides (täiend, öeldistäide).

3. MEETOD JA ANALÜÜSI KÄIK

Artiklites selgitan tajuadjektiivide semantikat ja polüseemiat korpuskonteksti ja semantiliste tüüpide abil. Korpuskonteksti uurin korpuspäringuprogrammiga Sketch Engine (Kilgarriff jt 2014), mis võimaldab vaadelda sõna kollokatsioone eri statistikute alusel, genereerida sõnavisandeid ja uurida sarnase kontekstiga sõnu tesauruse funktsiooniga. Kuna distributiivse semantika automaatsed mudelid pole ulatuslikuma polüseemia eristuses veel kuigi täpsed (vt pt 2.1), siis on sinse väitekirja uurimustes kasutusel kombineeritud meetod: kasutan visandeid ja tesaurust korpuskontekstist süsteemsema ülevaate saamiseks, kuid kasutusjuhtude analüüsil ja tähendusteks koondamisel tuginen kvalitatiivsele analüüsile. Tähenduste abstraherimisel ja polüseemiamallide väljaselgitamisel rakendan semantilisi tüüpe, mida töötasin samuti välja doktoritöö käigus. Tutvustan peatükis põgusalt ka analüüsi käiku ja kombineeritud meetodi piiranguid.

3.1 Semantilised tüübid

Et tähendusi analüüsi tegemiseks formaliseerida, kasutasin semantilisi tüüpe. Semantilised tüübid on abstraktsed abivahendid sõnastikutöös ja korpusanalüüsis, mis võimaldavad tähendusi rühmitada ning polüseemia uurimisel paremini tuvastada süstemaatilisi tähendusvaheldusi. Nimisõna semantilised tüübid on näiteks TOIT, TAIM ja ESE (Langemets 2010: 307–318) ning adjektiivide semantilised tüübid nt OMADUS_VÄRV, OMADUS_AEG, OMADUS_KOHT ja OMADUS_MÕÕDE (P1: 314). Uurimuses (P1) kasutasin eesti keele adjektiivide semantiliste tüüpide väljatöötamisel eeskujuna Dixoni (1982) seitset olulisemat adjektiivide semantilist tüüpi, leksikaalse andmebaasi SIMPLE koostamisel väljatöötatud 18 tüüpi (Lenci jt 2000: 30–33) ning Hundsnurcheri ja Spletti (1982) uurimuse põhjal loodud GermaNeti 15 adjektiivide semantilist klassi.

Adjektiivide semantilise tüübistiku aluseks on minu enda uurimused (eeskätt P1, aga ka P2 ja P3). Substantiivsete tähenduste märgendamisel on kasutusel Margit Langemetsa eesti nimisõna semantilised tüübid (2010). Langemets (2009) kasutab substantiivide polüseemia kodeerimisel 80 semantilist tüüpi (liigisuhet). Artiklis P1 pakun välja 15 semantilist tüüpi spetsiaalselt eesti keele (polüseemsete) adjektiivide kirjeldamiseks, hilisemastes uurimustes (P2, P3) tüübid lisanduvad ja täpsustuvad.

Langemets (2009: 185) jõuab järeldusele, et kõikidesse semantilisesse tüüpi-desse kuuluvad sõnad on valmis uusi tähendusi vastu võtma, määrav on üldine kasutuse laienemine ja sagedus. Ta märgib, et suhted olemasolevate ja võimalike tähenduste vahel kalduvad järgima samu reegleid, mis kehtivad olemasolevate tähenduste vahel. Semantiliste tüüpide kodeerimine sõnastikus või sõnakogus on samm leksikaalsemantilise info formaalsema esituse suunas, mis võimaldab kirjeldada polüseemiat semantilise tüübi süstemaatilise tähendusnihe kaudu (ja seostada baasis süstemaatilisi malle vastava sõnatähendusega). Süstemaatilist

polüseemiat silmas pidades on märgistatud tähendusi semantiliste tüüpidega nt hollandi keele sõnastikusüsteemis (Vliet 2007) ja inglise keele leksikaalses andmebaasis CoreLex (Buitelaar 1998). Semantilisi tüüpe kasutatakse ka WordNeti tesaurus (nt Koeva jt 2018).

3.2 Kollokatsioonid, sõnavisandid ja tesaurus

Doktoritöö uurimustes kasutan korpuskontekstist kiirema ja süsteemsema ülevaate saamiseks Sketch Engine'i sõnavisandite funktsiooni. Sõnavisandite grammatika põhjal otsib programm statistika põhjal grammatilistes suhetes olevaid sõnapaare ja kolmikuid ning teeb uurija või leksikograafi jaoks sõna süntaktilisest ja kollokatiivsest käitumisest kompaktse kokkuvõtte (Kallas 2013: 26). Eesti keele sõnavisandite grammatika autor on Jelena Kallas (2013). Visandites esile toodud adjektiivikonstruktsioonide aluseks on olnud Mari Erelti (1979, 1986), Valter Tauli (1980), EKG (1993) ja Krista Kerge (2000) käsitlused (vt juurde Kallas 2013: 52–58). Kollokatsioonidena käsitleti (Kallas jt 2015 eeskujul) sisusõnade tähenduslikke ja statistiliselt esilduvaid kombinatsioone teiste leksikaalsete ja grammatiliste üksustega.

Adjektiivi sõnavisandis on võimalik vaadelda kollokatsioonipaare (nt sageduse või suhte tugevuse järjekorras): adjektiiv+substantiiv (mis on adjektiivi tähenduse selgitamisel üks olulisemaid suhteid), adverb+adjektiiv ning uuritava adjektiivile eelnevaid ja järgnevaid adjektiive (joonis 1). Kaardistatud on nii omadussõnalised, nimisõnalised kui ka määrsõnalised kollokaadid. Esile on toodud ka adjektiiviga seotud rinnastus- ja võrdlutarandid (*sinine nagu meri*) ning täiendi ja öeldistäite funktsioonis esinemised (joonis 2).

Constructions	nimisõnaga	määrsõnaga	eelneva omadussõnaga	järgneva omadussõnaga
nimetav ...	valgus sinise valguse	ülени ülени sinine	tume tume sinine	tribuline sinine tribuline
omastav ...	especially: beauty especially: e-commerce	silmi On silmi siniseid , on merekarva	especially: cars especially: e-commerce	kirju sinine kirju muu
osastav ...	taevas sinise taeva	kirkalt kirkalt sinine	hele hele sinine	sinine sinine sinine
seesütle	especially: Literature especially: fiction	koledalt ära suretatud ja koledalt sinine , mis paistab	taevakarva taevakarva sinine	lilla sinine lilla roosa
alalütle				
seestütle				
kaasütle	silmi			

Joonis 1. Väljavõte sõna *sinine* sõnavisandist I

ja/või	kui/nagu	teigusõna laiendina saavas	teigusõna_laiendina_nimetavas	... on "sinine"
punane ... punane , sinine	saks ... Sinine kui saks	siniseks_värvima ... siniseks	sinine_tulema ... tuleb sinine	taevas ... taevas on sinine
especially: DOAJ especially: e-commerce	taevas ... sinine nagu taevas	siniseks_tegema ... tribuunid siniseks	sinine_saama ... sinine	silmi ... silmi sinine
roheline ... roheline , sinine	meri ... sinine nagu meri	siniseks_muutama ... siniseks	sinine_meeldima ... meeldib sinine	värv ... värv on sinine
especially: Wikipedia 2021 especially: e-commerce		siniseks_minema ... siniseks läheb	sinine_valima ... sinine	especially: e-commerce
kollane ... sinine , kollane			sinine_paistma ... paistab sinine	lemmikvärv ... lemmikvärv on sinine
especially: e-commerce				silmaalune ...

Joonis 2. Väljavõte sõna *sinine* sõnavisandist II

Enim oli doktoritöö artiklites vaatluse all ühendkorpus 2019, mille sõnavisandite genereerimisel kasutab Sketch Engine sõnavisandite grammatika versiooni 1.8. (Uusim versioon on 2.0, mida rakendatakse ühendkorpuses 2021 (Koppel ja Kallas 2022)). Suhete otsingul on proovitud ette arvestada teatud morfoloogilise märgenduse vigadega ja versioon 1.8 kasutab vastavaid stoppsõnade loendeid (nt välistatakse adjektiivse kollokaadina lemmad *andev*, *paiknev*, *viibiv* jne, mis on korpuses adjektiivina märgendatud). Grammatika aluseks olevate regulaaravaldiste ja stoppsõnade täielike loenditega on võimalik tutvuda Sketch Engine'i programmis visandite funktsiooni aknas.

Kollokatsioonide leidmiseks ja nende omavahelise tugevuse arutamiseks kasutatavad korpuslingvistilised tööriistad erinevad statistilise seose mõõdikuid. Sketch Engine'i programm võimaldab korpusest kollokatsioone otsida erinevate statistikute põhjal: t-skoor, MI-skoor (vastastikuse informatsiooni väärtus) (vt lähemalt Church ja Hanks 1990), MI³-skoor (vt Oakes 1998), log-tõepära funktsioon (vt Dunning 1993), minimaalne tundlikkus (vt Pedersen 1998), Dice'i koeffitsient ja sellega seotud logDice statistik (Rychlý 2008) (eesti keeles vt välja toodud statistikute kohta Kallas 2013: 18–29). Sõnavisandite puhul kasutatakse sõnadevahelise seose tugevuse mõõtmiseks logDice'i üldist sõnavisandi skoori⁵. Seose tugevus võtab arvesse kollokatsiooni moodustavate sõnade sageduse korpuses, nt *kena maja* on nõrgem kollokatsioon, kuna nii *kena* kui *maja* võivad esineda koos paljude teiste sõnadega, ent *kahekorruseline maja* on tugevam kollokatsioon, kuna *kahekorruseline* esineb sageli koos *majaga* ja ei esine koos nii paljude teiste sõnadega. Sketch Engine'i kaasasutaja Pavel Rychlý (Rychlý 2008) toob valitud logDice'i skoori eelisena välja, et kui selle aluseks olev Dice skoor annab küll häid kollokaadikandidaate, ent tema väärtusteks on väga väikesed arvud, siis logDice'i puhul on see probleem kõrvaldatud ja väärtused jäävad mõistlikku vahemikku. Lisaks töötab valem hästi erineva suurusega korpuste peal.

Kasutasin semantiliste rühmade komplekteerimisel ja sõnade semantika uurimisel abivahendina ka Sketch Engine'i tesauruse funktsiooni, mis genereerib sarnase korpuskontekstiga⁶ sõnade loendi. Statistika põhjal toob programm esile sõnu, millel on sarnane grammatiline ja kollokatiivne käitumine. Joonisel 3 on esitatud ühe värviadjektiivi (*sinine*), ühe temperatuurijadjektiivi (*jäine*) ja ühe kompimisadjektiivi (*kare*) tesaurusepilt ühendkorpus 2019 põhjal. Joonisel on sõna järel kirjas tema korpussagedus, sõnad on järjestatud aga seose tugevuse alusel, st et eespool on otsitud märksõnaga kõige sarnasemalt käituvad sõnad. (Adjektiivid on korpuses märgendatud neljal moel: adjektiivid (*adjective positive*), keskvõrdevormid (*adjective comparative*), ülivõrdevormid (*adjective superlative*) ja käändumatud adjektiivid (*adjective indeclinable*)). Jooniselt nähtub, et sageli satuvad sõna tesaurusesse semantiliselt lähedased sõnad, näiteks värvide puhul teised värvid. Rychlý ja Kilgarriff (2007) toovad välja, et tesauruse tulemused

⁵ Valemiga saab tutvuda dokumendis Sketch Statistics, mis asub veebiaadressil: <https://www.sketchengine.eu/documentation/statistics-used-in-sketch-engine/>

⁶ Sõnade sarnasusskoori täpset arvutust näeb dokumendist Sketch Statistics, mis asub veebiaadressil: <https://www.sketchengine.eu/documentation/statistics-used-in-sketch-engine/>

paranevad korpuse suuruse kasvades. Suurema korpuse puhul on kindlam, et sarnased sõnad eristatakse mürast (sõnadest, mis lihtsalt juhtuvad jagama mõnda konteksti).

sinine as adjective positive 94,344x		jäine as adjective positive 9,823x		kare as adjective positive 9,385x	
Word	Frequency ?	Word	Frequency ?	Word	Frequency ?
1 kollane	97,863 ...	1 lumine	14,009 ...	1 krobeline	2,911 ...
2 roosa	47,001 ...	2 kivine	6,248 ...	2 niiske	39,255 ...
3 punane	222,874 ...	3 mudane	5,497 ...	3 sile	38,785 ...
4 hall	80,002 ...	4 karge	7,359 ...	4 karvane	21,465 ...
5 värviline	49,986 ...	5 kõle	5,929 ...	5 kuiv	92,900 ...
6 roheline	184,526 ...	6 libe	25,202 ...	6 libe	25,202 ...
7 pruun	43,710 ...	7 tuuline	12,095 ...	7 läikiv	17,783 ...
8 hele	50,642 ...	8 jahe	40,747 ...	8 kleepuv	5,858 ...
9 valge	295,692 ...	9 porine	6,695 ...	9 märg	58,482 ...
10 tume	75,650 ...	10 liivane	7,742 ...	10 jäme	20,980 ...

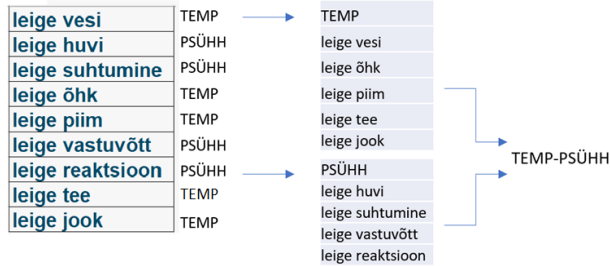
Joonis 3. Sõnade *sinine*, *jäine* ja *kare* distributiivne tesaurus Sketch Engine'is

3.3 Analüüsi käik

Analüüsi eesmärk on kontrollida sõna sobivust konkreetse semantilise rühma adjektiivide hulka, tuletada sõna korpuskontekstist sõna tähendus(ed) ja abstraherida tähendusi piisavalt, et erinevate sõnade tähendused või sama sõna erinevad tähendused oleksid omavahel paremini võrreldavad. Kirjeldan lühidalt analüüsi samme:

1. Otsida kandidaatsõnad (vt täpsemalt materjali kogumise peatükist), mille puhul võib eeldada, et sõnaga on võimalik väljendada (vähemalt ühte) omadust, mis osutab värvile, temperatuurile või kombitavale omadusele;
2. vastava tähenduse esinemise kontroll sõna korpuskonteksti põhjal (sõnavisanditest ja konkordantsidest) ja tähenduse märgendamine sobiva semantilise tüübiga (VÄRV, TEMP või PUUDUTUS);
3. sõna võimaliku polüseemia selgitamine (põhiliselt sõnavisandite ja konkordantside toel, vähemal määral ka distributiivse tesauruse abil); 4.1 kasutusjuhtude analüüs: sarnaste kasutusjuhtude koondamine; 4.2 iga tähenduse puhul eriomaste joonte selgitamine, mis seovad koondatud kasutusjuhte ning eristavad seda konkreetset tähendust teistest;
4. kõigi eristatud tähenduste semantiliste tüüpidega märgendamine. Nende tähenduste fikseerimine koos semantiliste tüüpide ja korpusnäidetega;
5. reeglipärade otsimine sõnade tähendusstruktuuris; polüseemiamallide tuvastamine ja kirjeldus.

Toon näite (joonis 4) temperatuurijadjektiivi *leige* kõige sagedamate A+S (adjektiiv+substantiiv) korpuskollokatsioonide analüüsist ja semantiliste tüüpidega märgendamisest. Joonis näitab, kuidas lisasin kollokatsioonide põhjal semantilised tüübid (vajadusel tutvusin lähemalt konkordantsidega). Koondasin kollokatsioonid tüüpide alla (TEMP=temperatuuriga seotud kasutus; PSÜHH=meeleolu, iseloomu ja vaimsete võimetega seotud kasutused). Fikseerisin sõnal *leige* tähendusvahelduse TEMP–PSÜHH. Kui mitme adjektiivi tähendusstruktuur väljendas sama tähendusvaheldust, moodustus polüseemiamall.



Joonis 4. Adjektiivi *leige* kollokatsioonide analüüs

3.4 Meetodi piirangud

Töös kasutatav polüseemiauurimise meetod on praktilise kallakuga ja välja kasvanud leksikograafilise töö kogemusest. Põhiline piirang semantiliste tüüpide kasutamisel on tähenduste koondamise, semantilise rühmitamise ehk semantilistesse tüüpidesse jagamise erinevad võimalused. Artiklites kirjeldatu on üks võimalik viis sõnatähendusi analüüsida, mis ei välista teistsugust interpretatsiooni. Näiteks soovi korral saaks ju semantilise tüübi OMADUS_TEMP jagada omakorda peenemaks: TEMP_SOE (millega saaks märgendada nt sõnu *põrgukuum*, *palav*, *unesoe*) ja TEMP_KÜLM (nt *kivikülm*, *jahe*, *jääne*) või vastupidi – koondada semantilised tüübid HÄÄL, LÖHN ja MAITSE kokku tüübiks OMADUS_TAJU.

Nagu teoreetiliste lähtepunktide peatükis (2.3) esile tõin, on igasugune tähendus, mis üritab hõlmata rohkem kui üht konkreetset suhtlussündmust, abstraktsioon. Semantilised tüübid on neid tähendusi omakorda koondavad abstraktsioonid. Kui tulemus on nõ abstraktsioonide abstraktsioon, võib tekkida küsimus, kui palju tuuma alles jääb, kui palju nihkub või läheb kaduma semantilist sisu. Kuid kuna siinse töö eesmärk on leida värvi-, temperatuuri ja kompimisadjektiivide rühmade polüseemia reeglipärasid, on praktiline just vastaval tasemel üldistus.

Korpusest tähendust uurides ei saa infot suhtlussituatsiooni välise konteksti kohta ega suhtlejate emotsionaalse oleku kohta (see ei pruugi olla alati tekstist tajutav). Korpusest ei näe ka seda, mis täpselt toimub inimese meeles, nt võivad tähendused siiski olla suhtlejate jaoks erinevalt liigendatud, isegi kui korpuskontekst tundub sarnane. Toon näiteks sõna *särk*, mida võib mõista kui üldmõistet,

mille alla paigutuvad T-särk, triiksärk, alussärk jt või siis kui sama tasandi sünonüümi, mispuhul särk=T-särk, särk= triiksärk, särk=alussärk jne.

Kuna uurisin sõnasemantikat korpuspäringusüsteemi abiga, siis tulid sõnavisandites, nt adjektiiv+substantiiv kollokatsioonides, esile sõnad, mis on korpuses vastavate sõnaliikidega märgendatud. See tähendab, et sõnaliigimärgenduse eksimatus on ülimalt oluline usaldusväärsete visandite saamisel. Eesti ühendkorpuste märgendamiseks kasutatava morfoanalüsaatori täpsus on (vähemalt) 97% (Kaalep ja Vaino, 2001). Kuid Sketch Engine'i tarkvaraarendaja Miloš Jakubiček (2021) on selgitanud, et nt 97% täpsus sõnetasandil tähendab ainult 50%-list täpsust lausetasandil, nii et pooled korpuslaused sisaldavad vähemalt üht valesti märgendatud sõna. Pidasin kollokatsioonide uurimisel korpusmärgendust kriitiliselt silmas, et analüüsi oleks kaasatud soovitud sõnaliigiga sõnad.

4. ALLIKAD JA MATERJAL

Neljas peatükk käsitleb doktoritöös kasutatud allikaid ja uurimismaterjali. Alapeatükis 4.1 tutvustan uurimuste aluseks olevaid tekstikorpuseid: Tartu Ülikooli koondkorpust ja Eesti ühendkorpuste sarja: 2013, 2017, 2019. Teises alapeatükis (4.2) selgitan, kuidas sõnastikumaterjal on doktoritöö uurimustesse kaasatud ja kirjeldan töö seisukohalt olulisemaid sõnastikke. Kolmandas alapeatükis (4.3) annan ülevaate tajuadjektiivide valimi komplekteerimisest ja eri rühmade materjali kogumise erinevatest protseduuridest.

4.1 Korpused

Dokoritöö artiklid on valminud enam kui kaheksa aasta jooksul ja see on olnud eesti korpuslingvistikas murranguline aeg. Viimase kümnendi jooksul on eesti keelele loodud ühendkorpuste sari (vt täpsemalt Koppel ja Kallas 2022), mis võimaldab keelt uurida suurema tekstikogu põhjal kui kunagi varem, ja kanda on kinnitanud korpuspõhine sõnastikekoostamine (vt nt Kallas jt 2015, Langemets jt 2018, Koppel 2020). Sellest tulenevalt on ka artiklite korpusanalüüsi alus ajas muutunud ja kasvanud. Olen kõigis doktoritöö uurimustes kasutanud artikli kirjutamise hetkel suurimat (ja värskemal) saadaolevat eesti keele korpust. Kõik kasutatud korpused on olnud kättesaadavad ja analüüsitavad Sketch Engine'i päringusüsteemis.

Esimene doktoritöö artikkel (P1), mis tegeles semantiliste tüüpide väljatöötamisega 100 sagedasema adjektiivil põhjal, kasutas aluseks Tartu Ülikooli koondkorpust⁷. Koondkorpus oli ühendkorpuste sarja sünnini (aastal 2013) suurim eestikeelne korpus ja ka esimene, milles oli võimalik Sketch Engine'i kaudu päringuid teha. Eesti keele koondkorpus koosneb terviktekstidest, mille hulka kuuluvad ilukirjandusteosed 1990. aastast, ajakirjanduslikud tekstid aastatest 1995–2008, 1,3 miljoni sõna ulatuses teadustekste, Eesti seadused ja Riigikogu stenogrammid ning ca 22 miljonit sõna uue meedia tekste. Korpuse koostajad (Muischnek jt 2011) toovad esile, et uue meedia korpus sisaldab omakorda nelja allkorpust: jututubade tekste, uudisgruppide tekste, foorumitekste ja kommentaaride tekste.

Järgnevad doktoritöö artiklid ammutasid materjali eesti keele ühendkorpustest. Teine artikkel (P2), mis käsitles värvisõnade polüseemiat, tugines kollokatsiooni-sõnastiku andmebaasi kaudu „Eesti keele ühendkorpus 2013“-le. Kolmas artikkel (P3 2020), mis uuris temperatuurisõnu, analüüsis „Eesti keele ühendkorpus 2017“ keelematerjali ning neljas (P4) ja viies artikkel (P5), mis uurisid vastavalt kompimisadjektiive ja adjektiivil määratlemist korpuses, kasutasid alusena „Eesti keele ühendkorpust 2019“.

⁷ <https://www.cl.ut.ee/korpused/segakorpus/index.php?lang=et>

Nüüdseks on ühendkorpuste sarjas kokku neli versiooni: ühendkorpused 2013, 2017, 2019 ja 2021. Koppel ja Kallas (2022) toovad välja, et 2013 a ühendkorpuse lähtetekstid olid lausestatud, morfoloogiliselt analüüsitud ja automaatselt ühestatud OÜ Filosoofi analüsaatori ESTMORF abil ning järgmiste versioonide märgendamiseks on kasutatud eesti keele töötlusprogrammi estNLTK teeke. Kogu korpuse sisu on klassifitseeritud žanridesse (blogid, foorumid, perioodika, veebikaubandus, entsüklopeedia, akadeemiline kirjutamine, kirjandus) ja teemadesse (kokku 24 teemat, nt loomad, teadus, sport, religioon), mis võimaldab teha täpsemat keeleanalüüsi. Ühendkorpuste sari on valminud Eesti Keele Instituudi ja tarkvarafirma Lexical Computing Ltd. koostööna.

Koppel ja Kallas (2022) toovad välja, et eesti keele ühendkorpus katab perioodi 1990datest kuni tänapäevani, mis tähendab, et tänapäeva eesti keelt on võimalik uurida umbes 30 aasta lõikes. Nad märgivad, et juba alates ühendkorpus 2013-st moodustavad suurima korpuse osa veebikorpused. Kui 2013. aastal kogutud veebikorpuse suurus oli umbkaudu 313 mln sõnet, mis moodustas 56% ühendkorpus 2013 mahust; siis 2017. aasta korpuses oli veebitekste 763 mln sõnet, mis moodustas 80% ühendkorpus 2017 mahust; 2019. aastal 615 mln sõnet, moodustades 87% ühendkorpus 2019 mahust; ning 2021. aastal 951 mln sõnet, mis moodustab lausa 92% ühendkorpus 2021 mahust. Koondkorpus on siiski jäänud alati ühendkorpuste osaks. Kallas jt (2015) leiavad, et leksikograafi vaatevinklist täiendavad koondkorpus ja veebikorpus teineteist hästi: esimene võimaldab analüüsida eelkõige kirjakeelt ja teine annab parema ülevaate just kõnekeelele lähedamast internetikeelest.

Kristina Koppel tõdeb oma näitelausete korpuspõhist automaattuvastust käsitlevas doktoritöös (2020: 68), et kuna ühendkorpus on suures osas just veebikorpus, sisaldab ta seetõttu veebikorpusele tüüpilisi probleeme, nagu näiteks masintõkelisi ja automaatselt genereeritud tekste. Automaatse märgendamisega kaasnevad ka lemmatiseerimise, morfoloogilise analüüsi ja lausepiiride tuvastamisega seotud vead.

Samas on korpuslingvist Adam Kilgarriff (2003) toonud veebi plussidena välja, et tegemist on äärmiselt suure keeleandmete allikaga, mis võimaldab seetõttu ka ulatuslikumaid uuringuid ning lisaks on veebi kasutamine ka tasuta ning koheselt võimalik. Suurem keeleandmestik võimaldab paremini leida infot ka harvemate sõnade kasutuse kohta. Ka ühendkorpuste kogumaht on iga versiooniga kasvanud. Kui ühendkorpus 2013 sisaldas u 563 miljonit sõna (Kallas jt 2015), siis ühendkorpus 2017⁸ sisaldas juba 1,1 miljardit sõna ja ühendkorpus 2019⁹ 1,5 miljardit sõna.

Doktoritöö viimane uurimus (P5) erineb eelnevatest selle poolest, et korpuspääringute tegemisel ei olnud kasutusel programm Sketch Engine, vaid ekstraheerisime huvipakkuvaid parameetreid ühendkorpusest 2019 programmeerimiskeeles

⁸ <https://metashare.ut.ee/repository/browse/eesti-keele-uhendkorpus-2017/b616ceda30ce11e8a6e4005056b40024880158b577154c01bd3d3fcfc9b762b3/>

⁹ <https://metashare.ut.ee/repository/browse/estonian-national-corpus-2019/cd9633fab22e11eaa6e4fa163e9d4547b71a2df64d1f43f1ac26dbd8508ea951/>

Python kirjutatud loogiliste avaldiste abil ning eraldasime korpusest testsõna esinemisjuhud, mil konkreetne parameeter kehtis. Kuna artiklis katsetasime adjektiivsete joonte väljaselgitamist korpuses nii adjektiiviklassi kui ka muudesse sõnaliikidesse kuuluvate sõnade peal, siis pidi testsõna enda sõnaliik olema korpuspäringul alati määramata, vaba.

4.2 Sõnastikud

Sõnastikumaterjal on olnud doktoritöö uurimustesse kaasatud mitmel moel. Esiteks kasutasin abiks märksõnastikke/sõnastikubaase, et leida vaatlusalusesse semantiliselt rühma kuuluvaid sõnu. Täielikult sõnastikubaasi-põhiselt on kokku pandud värvirühm, kuid ka temperatuurirühma ja kompimisadjektiivide koondamisel töötasin korpuse kõrval läbi sõnastikubaasi märksõnu. Teiseks uurisin korpuse kõrval sõnastikust polüseemiamallide esitust, et uurida, kas ja kui võrd süstemaatiliselt on polüseemiamallid sõnaraamatus esitatud, kuna üks doktoritöö eesmärke on teha järeldusi polüseemiamallide leksikograafilise esituse kohta. Kolmandaks kasutasin uurimustes korpusnäidetele täienduseks sõnastiku definit-sioonikatkeid ja näitelauseid tähenduste illustreerimisel.

Uurimust alustades tuli otsustada, millist leksikograafilist allikat aluseks valida. Eesmärk oli valida tähenduste esituse uurimiseks suurim, ülevaatlikum ja usaldusväärsem eesti keele sõnastik. Eesti leksikograafias on esil nii normiv kui ka kirjeldav suund ja kaks vastavat peamist esindajat on õigekeelsussõnaraamat (ÕS) ja seletav sõnaraamat (vt nt Langemets 2008). Kuna õigekeelsussõnastike traditsioonis on kombeks olnud anda tähendusi, kui üldse, siis piiratud mahu, osutus semantiliselt uurimise puhul paremaks valikuks deskriptiivne suund. Doktoritöö alguse ajal oli Eesti keele suurim ja põhjalikema tähenduste esitusega sõnaraamat „Eesti keele seletav sõnaraamat“ (EKSS), mis ilmus kuues köites aastal 2009 ja sisaldas ligi 150 000 märksõna. Kuid EKSS-i tugevus – selle põhjalikkus ja ulatuslikkus – oli saavutatud 50-aastase töö tulemusena, mis teistpidi muutis ta doktoritöö võrdlusalusena siiski kaheldavaks. Doktoritöö lähenemine on sünkroonne ja tähendused on ajas muutuvad. Eesmärk ei olnudki küll süstemaatiliselt polüseemiat „välja nõrutada“ sõnastikutähendustest, vaid ikkagi nõ elavast keelest, korpusest, kuid siiski oleks ka korpuspõhiste mallide kõrvale olnud hea uurida täiesti tänapäevast tähenduste esitust.

EKSS-i toimetustöödega paralleelselt hakati aga koostama Eesti keele ühekõitelist sõnaraamatut (EKS) (Langemets jt 2010). Uue sõnaraamatu põhialuseks jäi suure seletava sõnaraamatu andmebaas, kuid tähendusi ja näiteid ajakohastati tekstikorpuste ja interneti abil. Seetõttu sobib sõnaraamat allikaks, mida eesti keele semantiliselt analüüsil korpuse kõrval võrdlusmaterjaliks kasutada ja on ühtlasi ka parim leksikograafiline allikas, mille märksõnade seast leida sobivaid sõnu uurimiseks. EKS on mahukas eesti keele sõnavara sisaldavaid teosed ja sealjuures ka nüüdisaegsem, kuna tegemist on Eesti esimese seletava sõnaraamatuga, mille koostajad kasutasid aluseks keelekorpus (Langemets jt 2018).

Sõnastikualus on püsinud kõikide uurimuste lõikes peaausjalikult sama, kuigi selle vorm on ajas muutunud. Esimeste uurimuste aluseks oli EKS-i tööbaas ja viimastel juba EKI ühendsõnastik, mille esmane märksõnastik (120 000 sõna), tähendusjaotus, seletused ja kasutusnäited “Eesti keele sõnaraamatust 2019“ (=EKSist) pärinevad. Doktoritöö valimi sõnad pärinevad osaliselt niisiis EKI ühendsõnastiku andmebaasist, mida koostatakse, toimetatakse ja täiendatakse pidevalt. Ühendsõnastiku eesmärk on koondada erinevaid sõnastikke ja keeleandmeid, et kasutajal oleks võimalik saada kogu info kätte ühest kohast (Tavast jt 2020).

Eesti keele instituudi leksikograafidel on olnud viimase kümne aasta jooksul kaks tööbaasi: Eelex (vt lähemalt Jürviste jt 2011) ja Ekilex (vt lähemalt Tavast jt 2018). Praegu tehakse koostamis- ja toimetamistöid Ekilexi baasis ja kasutajale saavad uuendused nähtavaks järgmisel päeval Sõnaveebi portaalis (Sõnaveebi kohta vt Koppel jt 2019, Langemets jt 2021). Et leksikograafias toetab semantilise tüübi lisamine sõnatähendustele sama tähendusvälja sõnade ühtlasemat kirjeldust (ja lisaks võimaldab ka uurijatel sama rühma sõnu korraga välja otsida), siis on adjektiivide semantiliste tüüpide väljatöötamisel silmas peetud ka kasu andmebaasitöös.

Esimeses (P1) ja teises uurimuses (P2) kasutasin sõnaraamatuesituste uurimisel ühe allikana ühekõitelise eesti keele seletava sõnaraamatu (EKS-i) andmebaasi Eelexis, kus juba märgiti polüseemia ühtlasema esituse huvides sõnatähenduste juurde semantilisi tüüpe. Koostamisel olid kasutusel Margit Langemetsa (2010) substantiivide analüüsil tuvastatud tüübid. Esimeses uurimuses pakkusin välja semantilised tüübid eesti adjektiivide jaoks.

Teises uurimuses (P2) oli peale EKS-i olulisel kohal ka eesti naabersõnade sõnastik (vt lähemalt Kallas jt 2015), millele mõeldes tehti tähelepanekuid värvisõnade polüseemiamallide esituse kohta. Naabersõnade sõnastik ilmus aastal 2019, kuid selle materjal on integreeritud ka EKI ühendsõnastikku. Naabersõnade sõnastik sisaldab umbes 10 000 märksõna ja on mõeldud eeskätt edasijõudnud eesti keele õppijatele. Sõnastikku koostati poolautomaatselt ning kogu sõnastiku andmebaas oli genereeritud ühendkorpus 2013 põhjal sõnavisandite funktsiooni abil.

Kolmandas uurimuses (P3) kasutatud sõnastikuandmed olid saadud 2019. aastal veebisõnaraamatuna ilmunud “Eesti keele sõnaraamatust” (EKS), mis oli sel ajal ka juba Eesti Keele Instituudi ühendsõnastiku ning keeleportaali Sõnaveeb keskne andmestik. Neljanda uurimuse (P4) sõnastikumaterjal pärines Eesti Keele Instituudi ühendsõnastikust 2020.

Uuritud semantiliste rühmade morfoloogilise struktuuri analüüsis ehk sõnade juuradjektiivideks, tuletisteks ja liitsõnadeks jagamisel tuginesin erinevates doktoritöö uurimustes Silvi Vare sõnaperede sõnaraamatu (2012) võrguversioonile¹⁰. Sõnaperede sõnaraamat sisaldab 120 märksõna, mis on sõnadevaheliste moodustusseoste abil koondatud 9000 sõnapereks.

¹⁰ <http://www.eki.ee/dict/sp/index.cgi>

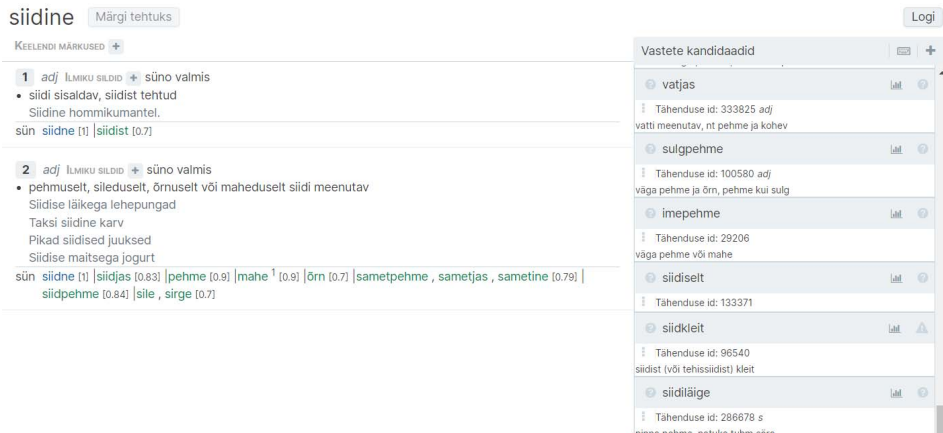
4.3 Materjali kogumine

Doktoritöö esimene uurimus (P1) keskendus eesti keele 100 sagedasemale adjektiivile, mistõttu sai uurimuseks vajaliku materjali korpusest välja pärida Sketch Engine'i sõnaloendi funktsiooniga. Järgnevad uurimused aga, mis käsitlesid konkreetseid semantilisi rühmi, vajasid teistmoodi lähenemist. Iga semantilise rühma süstemaatilise polüseemia uurimisele eelnes rühma sobivate sõnade kogumine. Värvadjektiivide, temperatuuradjektiivide ja kompimisadjektiivide rühma sõnad ei olnud uurijale võrdselt kättesaadavad, seega materjali kogumise peatükis tutvustan lühidalt iga rühma komplekteerimise eripära.

Kõige hõlpsam oli koondada värvadjektiivide rühma. Nimelt oli EKS-i tööbaasis adjektiivide (ja substantiivide) märgendamisel kasutusel semantiline tüüp VÄRV, mis võimaldas kõiki värve väljendavaid sõnu korraka välja pärida. Värvadjektiivide valimi moodustasidki sõnad, mille vähemalt üks tähendus on märgistatud semantilise tüübiga VÄRV. Kokku oli selliseid sõnu 463. Süstemaatilise polüseemia malle selgitasin uurimuses (P2) 115 polüseemse värvisõna põhjal. Ülejäänud valim oli monoseemne, kuid oli siiski kaasatud värvisõnade morfoloogilisse analüüsi.

Temperatuurisõnad andmebaasis eraldi märgendatud ei olnud. Temperatuuradjektiivide rühma kokku kogumiseks töötasin Eelexi baasis läbi enam kui 12 000 adjektiivi. Otsisin baasist sõnu, millel vähemalt ühe tähenduse definitsioonis oleks temperatuur tähendus- komponendina eksplitsiitselt välja toodud. Kõigepealt uurisin 3190 leidu, mis olid märgendatud semantilise tüübiga OMADUS_FÜÜS, seejärel analüüsisin semantilist tüüpi OMADUS_KVAL (6076 leidu) ja katus-tüüpi OMADUS (2819 leidu) kandvaid sõnu. Kokku sobis uurimusse 101 sõna (P3), mille hulgas oli polüseemseid sõnu 46.

Kompimisadjektiivide komplekteerimine oli kõige keerulisem ülesanne. Väljakutseks osutus juba rühma piiritlemine, kuna kompimisadjektiivid erinesid värvi- ja temperatuurisõnadest ka selle poolest, et eelnevaid uurimusi eesti keelele ei leidunud. Kasutasin valimi kokkupanekul erinevaid allikaid, programme ja leksikograafilisi abivahendeid (P4). Enamik valimi sõnadest pärineb "EKI ühendsõnastikust 2020", mida sõelusin ühendsõnastiku märksõnade hulgast välja sünonüümide tööriista abil (joonis 5). Sünonüümide lisamisel on abiks automaatselt genereeritud loend sünonüümikandidaate, mis on ekstraheeritud olemasolevatest sõnaraamatutest ning kasutatud on ka semantilist peegeldamist ja distributiivse sarnasuse arvutamist korpuse põhjal (vt Tavast jt 2020).



Joonis 5. Sünonüümide lisamise vaade Ekilexi tööbaasis

Kompimisadjektiivide leidmiseks korpusest kasutasin ka leksikograafilise tarkvara Sketch Engine sõnavisandite ja distributiivse tesauruse funktsiooni (Kilgarriff jt 2014). Sõnavisandid esitavad sõna kollokatsioone ja võimaldavad otsida kompimisadjektiive näiteks kompimiseks sobivate substantiivide kaudu (joonis 6) või teiste kompimisadjektiivide ja/või suhte abil (joonis 7). Lisaks oli valimi kogumisel kasulik Sketch Engine'i distributiivse tesauruse funktsioon, mis toob statistika abil esile sarnase grammatilise ja kollokatiivse käitumisega sõnu ja võimaldab seeläbi kaudselt vaadelda ka sarnase tähendusväljaga sõnu (joonis 8). Kompimisadjektiivide andmestikku kuulub töös 265 sõna, millest polüseemseid sõnu on 102.

omadussõnaga	
ümbritsev	...
olemasolev	...
tasane	...
kõva	...
külm	...
kindel	...
keskmine	...
vaba	...
kuiv	...
niiske	...
märg	...

Joonis 6. Väljavõte sõna *maapind* sõnavisandist

ja/või	
lumine	...
märg	...
jäine	...
sile	...
külm	...
pehme	...
mudane	...
porine	...
pime	...
järsk	...
kõva	...

Joonis 7. Väljavõte sõna *libe* sõnavisandist

THESAURUS

Estonian National Corpus 2019 (Estonian NC 2019)



kollane as adjective positive 97,863x

...

	Word	Frequency ?	
1	sinine	94,344	...
2	roosa	47,001	...
3	punane	222,874	...
4	pruun	43,710	...
5	roheline	184,526	...
6	värviline	49,986	...
7	hall	80,002	...
8	valge	295,692	...
9	hele	50,642	...
10	tume	75,650	...

Joonis 8. Sõna *kollane* tesaurusepilt Sketch Engine's

Kokkuvõtvalt võib öelda, et värvi-, temperatuuri- ja kompimisadjektiivide rühma moodustamisel oli põhiliseks allikaks sama (ajas täienev) sõnakogu (EKS, hilisema nimega EKI ühendsõnastik). Kasutasin värvirühma kokupanekul juba andmebaasis esinevat semantilist tüüpi VÄRV (andmebaasi vanad adjektiivitüübid ja minu pakutud uus tüübistik on esitatud P1-s). Temperatuurisõnad andmebaasis eraldi märgendatud ei olnud. Sõelusin Eelexi sõnastikubaasis läbi enam kui 12 000 EKS-i märksõna, et leida sõnu, mille vähemalt ühe tähenduse definitsioonis on temperatuur tähenduskomponendina eksplitsiitselt välja toodud (vt täpsemalt P3). Kompimisadjektiivide komplekteerimisel kasutasin lisaks ühendsõnastikule ka Sketch Engine'i sõnavisandeid ja tesaurust. Kõrvaldada omadust väljendava tähenduse olemasolu kontrollisin korpuse abiga.

5. ÜLEVADE UURITUD ADJEKTIIVIRÜHMADEST

Viiendas peatükis analüüsin ja kõrvutan uuritud rühmi suuruse, morfoloogiliste tunnuste, korpussageduse ja polüseemsuse seisukohast. Andmeid esitan võrdlevalt esmakordselt raampeatükis. Morfoloogia alusel jagan rühmad juursõnadeks, liitsõnadeks, tuletisteks ja kesksõnakujulisteks adjektiivideks. Annan ülevaate ka iga uuritud rühma (100 sagedasemat adjektiivi, värviadjektiivid, temperatuuriadjektiivid ja kompimisadjektiivid) põhilistest tuletusliidetest. Esitan jooniseid rühmade sagedusjaotuse kohta ja uurin, kuidas sagedus, morfoloogiline struktuur ja polüseemsus üksteist mõjutavad.

5.1 Rühmade suurus

Uuritud adjektiivide arv on esitatud semantiliste rühmade kaupa tabelis 1. Suurim uuritud rühm on värviadjektiivid 463 sõnaga, järgneb kompimisadjektiivide rühm 265 sõnaga ning väikseim semantiline rühm on temperatuuriadjektiivid 101 sõnaga. (Esimese uurimuse, P1 valim sisaldas korpuse põhjal esile tulnud 100 sagedasemat adjektiivi ja oli semantiliselt piiritlemata.) Doktoritöö värvirühma suurus jääb varem kahe katselise meetodiga saadud tulemuse keskele: Uusküla ja kolleegid (2012: 56) said eesti värvisõnade rühma suuruseks loetelukatsega 285 sõna ja nimetamiskatsega 638 sõna. Temperatuuriadjektiivide rühma suurus on sarnane Urmas Sutropi (2004: 116) läbi viidud loetelukatse tulemustega, kus keelejuhid nimetasid kokku 663 temperatuurisõna, mille hulgas oli erinevaid sõnu 92.

Tabel 1. Sõnade arv uuritud semantilistes rühmades

Semantiline rühm	Värvi-adjektiivid	Temperatuuri-adjektiivid	Kompimis-adjektiivid
Valimi suurus	463	101	265

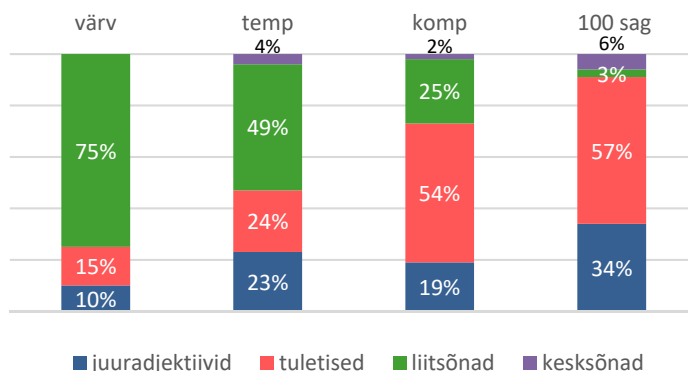
Rühmade suurust on huvitav kõrvutada ka meelte hierarhia seisukohast. Winter (2019: 342) pakub välja, et meelte hierhia aluseks võib olla tõik, et mõne meelega seotud omaduste väljendamiseks on keeles rohkem adjektiive kui mõne teise meele puhul. Siinse töö nägemismeelega seotud rühm (värvid) on tõesti adjektiivirikkam kui kompimismeelega seotud rühmad kokku (366 sõna, kui arvestada temperatuuri- ja kompimisadjektiive ühiselt). Lisaks on nägemisega seotud rühm kitsam ja ei sisalda näiteks objektide ega nähtuste kujuga seotud sõnavara (nt *kõrge*, *madal*) või üldisemaid hinnangulisi adjektiive (nt *ilus*, *inetu*). Kui eeldame, et valimi suurus peegeldab suuremat leksikaalset tähistatust keeles laiemalt, sobituvad tulemused meelte hierarhia (Viberg 1983) reastusega. Nägemise kui kõrgema meelega seonduv sõnavara on rikkalikum kui kompimismeelega seonduv. Lõplikke järeldusi siiski teha ei saa, sest kuigi rühmad on koostatud

põhilises osas sama, eesti suurima sõnakogu alusel, on kompimisadjektiivide piiritlemiseks erinevaid võimalusi (vt täpsemalt P4) ja samuti ei sisalda värvirühm kõiki nägemismeelega seotud adjektiive eesti keeles.

5.2 Rühmade morfoloogiline kirjeldus

Valimi morfoloogilisel jagamisel tuginesin Silvi Vare „Sõnaperede sõnaraamatule“ (Vare 2012). Sõnad, mis Sõnaperede sõnastikust puudusid, määrasin analoogia põhjal. Rühmad jagunesid juursõnadeks (nt *must, soe, krobe*), liitsõnadeks (nt *veripunane, ahjusoe, teraskõva*), tuletisteks (nt *kollane, pakasene, kühmuline*) ja kesksõnakujulisteks adjektiivideks (nt *põletav, torkiv*). Värvadjektiivide valimisse kesksõnakujulisi adjektiive ei kuulunud. Sõnad, mis lõppesid sufiksoidiga *-võitu* (nt *kuumavõitu, külmavõitu*), määrasin „Sõnaperede sõnaraamatu“ abiga liitsõnaks. Sellised sõnad on korpuses märgendatud küll hoopis adverbina, mitte adjektiivina, ent semantika ja süntaktilise käitumise põhjal analüüsisin nad käändumatuteks adjektiivideks ja kaasasin valimisse.

100 sagedasema adjektiivi ja värvirühma morfoloogilise ülevaate esitasin ka doktoritöö artiklites (P1, P2), kuid temperatuurirühma ja kompimirühma jaotust kirjeldan esmakordselt töö raamosas. Esitan valimi morfoloogilise struktuuri varieerumise järgneval protsent-graafikul (joonis 9).



Joonis 9. Rühmade jaotumine morfoloogilise struktuuri järgi

Eesti keele 100 sagedasema adjektiivi hulgas on juursõnu 34% (nt *uus, suur, vaba*) ja värvadjektiivide rühmas ainult 10% (nt *kirju, oranž, lilla*). Temperatuuradjektiivid (23%) ja kompimisadjektiivid (19%) paigutuvad nende kahe vahele (nt *leige, vilu, lämbe; kare, sile, libe*). On teada, et sageli kasutatavad sõnad nõ kuuluvad lühemaks (Zipf 1949) ja seega on 100 sagedama adjektiivi suurim juursõnade arv igati ootuspärane. On loogiline, et mida sagedam on valim, seda suurem on juursõnade osakaal. Suurem juursõnade osakaal lubab omakorda oletada, et temperatuurisõnad ja kompimissõnad esinevad rühmana keskmiselt sagedamini kui värvadjektiivid. Sagedusandmeid võrdlen järgmises peatükis.

Temperatuuradjektiivide andmestikus on juuradjektiive 23%. Sutropi (2004) välja selgitatud temperatuuri põhisõnad (*külm, soe, kuum ja jahe*) ja ülemineku-grupi (*palav, leige*) adjektiivid on kõik juursõnad. Põhitemperatuurisõnad ongi keeltes enamasti morfoloogiliselt liitsõnad ja omasõnad (Sutrop 2002). Ja ka teiste valdkondade põhisõnad (värvisõnad, maitseõnad, lõhnaõnad) jagavad sama tendentsi (Sutrop 2002: 39–41). Samas näiteks eesti põhivärvisõnade (põhivärvisõnade kohta vt lähemalt Sutrop 1995) seas on neli tuletist (kollane, punane, sinine, roheline).

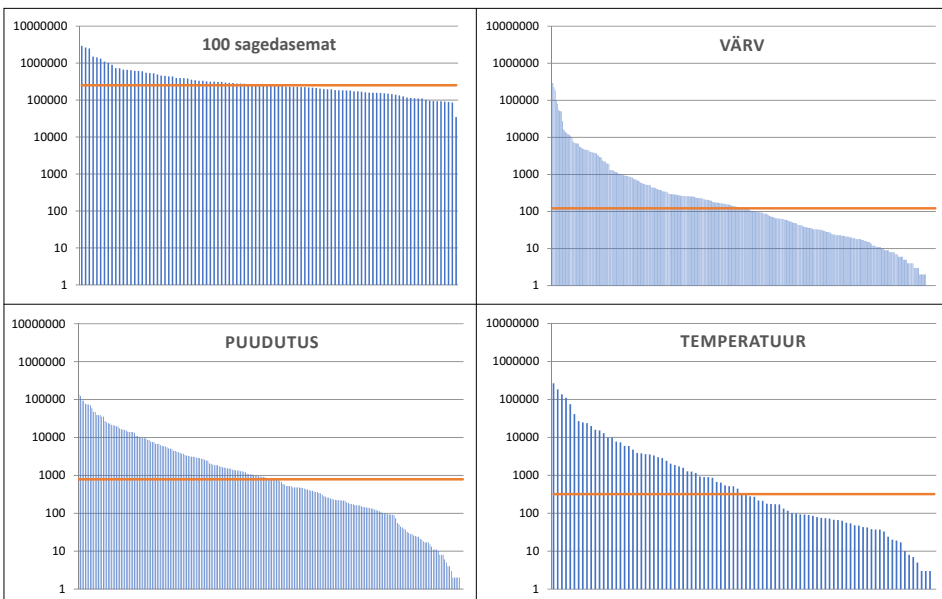
Liitsõnad moodustavad $\frac{3}{4}$ (75%) värvadjektiivide rühmast ja ligikaudu pool (49%) temperatuuradjektiivide rühmast ning on seega valdavaim morfoloogiline kuju, milles nende rühmade adjektiivid eesti keeles esinevad. Kuna paljud värvisõnad ja temperatuurisõnad põhinevad võrdlusel (nt *rebukollane, vaseroheline, kõrbekuum, kivikülm*), on selline jaotus loomulik. Kompimisrühma adjektiividest moodustasid liitsõnad väiksema osa, 25%, ja eesti 100 sagedasema adjektiivi hulgas on liitsõnu ainult 3.

Tuletised moodustavad kõige suurema osa (57%) 100 sagedasema adjektiivi rühmast ja peaaegu samal määral esineb neid kompimisadjektiivide rühmas (54%). Nimetatud rühmadest poole vähem (24%) esineb tuletisi temperatuuradjektiivide rühmas ja kõige väiksema osa (15%) moodustavad tuletised värvide rühmas (kuigi nende seas on ka põhivärvisõnu). Eesti 100 sagedasema adjektiivi hulgas on esil *ne*-tuletised (nt *tänane, tõsine, endine*), *lik*-tuletised (nt *ohtlik, vajalik, võimalik*) ja *line*-tuletised (nt *tavaline, poliitiline, tõeline*). Kompimisadjektiivide rühma iseloomustavad peamiselt *ne*-tuletised (nt *sametine, siidine*), *line*-tuletised (nt *sakmeline, sämbuline, hambuline*) ja *jas*-tuletised (nt *soomusjas, tokerjas, kleepjas*). Temperatuurirühmas on levinumad *ne*- (nt *jäine, põuane, leitsene*) ja *line*-tuletised (nt *tuline, pakaseline, arktiline*). Värvirühmas on sagedamad *ne*-tuletised (nt *roheline, kollane*), *jas*-tuletised (nt *valkjās, mustjās, kahkjās*), *kas*-tuletised (nt *punakas, pruunikas, sinakas*) ja *line*-tuletised (nt *värviline, roheline*).

Kasiku (2015: 347) järgi on eesti keeles omadussõnaliiteid võrreldes nimi-sõnaliidetega vähe, ent ometi on enamik omadussõnadest tuletised just kolme arvuka ja produktiivse tuletustüübi tõttu: *-ne*, *-line* ja *-lik* liited. Kasik põhjendab arvukust sellega, et need liited liituvad ka vabalt liitsõnadele ja võõrtüvedele ega ei ole seotud kindla tähenduskategooriga. *Lik*-tuletisi esines uuritud rühmadest märkimisväärselt ainult 100 sagedasema adjektiivi rühmas; semantilistes rühmades leidis ainult üksikuid näiteid. Küll aga esinesid kõigis kolmes rühmas (värvid, temperatuur, kompimine) *ne*- ja *line*-tuletised ning värv- ja kompimisrühmas *jas*-tuletised ning ainult värvirühmas *kas*-tuletised. *kas*- ja *jas*-tuletiste tähendus oleneb Kasiku (2015) põhjal tähismõiste kategooriast: asjamõiste kaudu omadust nimetavad *kas*-tuletised on possessiivse tähendusega (nt *andekas*), *jas*-tuletised võrdleva tähendusega (nt *munajas*); omadusmõistega seostudes modifitseerivad mõlemad liited omaduse astet (nt *pruunikas, mõrkjas*).

5.3 Rühmade sagedusandmete võrdlus

Et oleks võimalik võrrelda uuritud rühmadesse kuuluvate sõnade korpussagedust, pärsin kõigile rühmadele Sketch Engine'i abil uued sagedusandmed Eesti ühendkorpusest 2019. Rühmade sagedusjaotust illustreerib logaritmskaalaga joonis 10, mis kõrvutab doktoritöö semantilisi rühmi ja esitab 100 sagedasema adjektiivide sagedusjaotuse, et võrrelda semantiliste rühmade jaotust semantiliselt kitsendamatute adjektiivide rühma andmestikuga. 100 sagedasema adjektiivide sageduspilt pakub ka konteksti uuritud semantilistele rühmadele ja võimaldab vaadelda, kuidas rühmade sagedaimad sõnad suhestuvad eesti keele kõige sagedamate adjektiividega. Punase joonega on joonisel tähistatud rühma mediaansagedus.

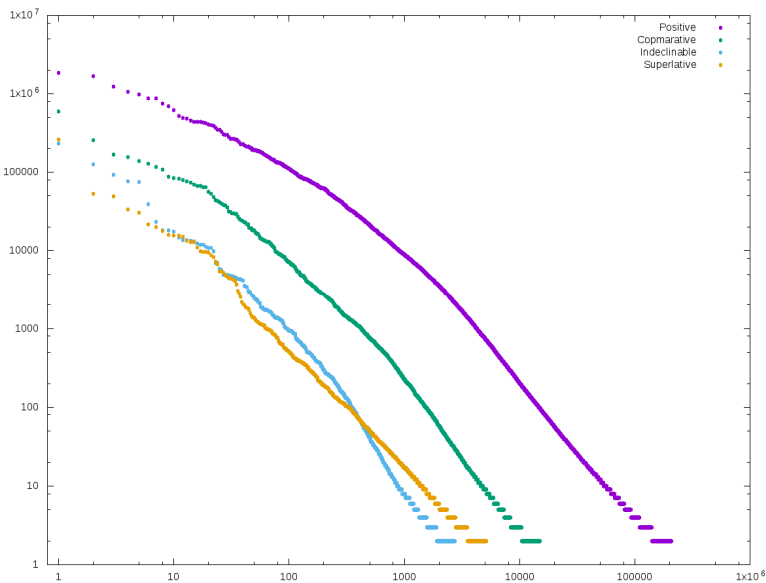


Joonis 10. Uuritud rühmade sagedusjaotus ühendkorpuses

Kõige madalama keskmise sagedusega uuritud rühm oli värviadjektiivide rühm (mediaan 122), mida iseloomustas ka suurim liitsõnade osakaal. Enamik madala sagedusega sõnu kõigis semantilistes rühmades olid liitsõnad. Rühma sageduse mediaanväärtuse poolest paigutus värvi- ja kompimisrühma vahele temperatuuradjektiivide rühm (mediaan 318) ning semantilistest rühmadest sagedaim oli kompimisrühm (mediaanväärtusega 789). (Eesti sagedaimate adjektiivide rühma mediaansagedus oli 261560.) Alla 100 esinemise korpuses oli kompimisadjektiivide rühmas 49 sõnal (nt *raudkõva*, *kidaline*, *paberjas*), temperatuuradjektiivide rühmas 34 sõnal (nt *põrgupalav*, *ihusoe*, *pakasene*) ja värviadjektiivide rühmas 171 sõnal (nt *porihall*, *tintmust*, *korallpunane*). Värvadjektiivide sagedusjaotuses on ka kõige märgatavam langustendents: sagedaimad värvid kuuluvad 100 sagedasema adjektiivide hulka (*must*, *valge*, *punane*, *roheline*), aga suure osa

rühmast moodustavad madala sagedusega sõnad. Teistes rühmades on distributsioon ühtlasem. Üle 10 000 esinemiskorra on korpusel 13 temperatuurisõnal, 18 värvisõnal ja 37 kompimissõnal.

Semantiliste rühmade sagedusjaotus järgib adjektiivide rühma sagedusjaotuse tendentse korpusel üldisemalt. Joonis 11 kujutab kõigi ühendkorpusel adjektiivide sagedusandmeid. Adjektiivid on korpusel märgendatud neljal moel: adjektiivid (*adjective positive*), keskvärdevormid (*adjective comparative*), ülivõrdevormid (*superlative*) ja käändumatud adjektiivid (*indeclinable*). Adjektiivide distributsioon korpusel järgib Zipfi seadust (1936) – on väiksem osa kõrge sagedusega adjektiive ja ülejäänute sagedus väheneb järk-järguliselt. Logaritmskaalaga joonis (11) illustreerib adjektiivisageduse variatsiooni. Sagedusjaotuse lõpuosas on palju adjektiive ühekordse esinemisega (*hapax legomena*). 63%-l korpusel adjektiividest on ainult üks esinemine, 12% ainult kaks esinemist ja 5% kolm esinemist. Kõige sagedasemana tulevad korpusel esile *adjective positive* märgendiga sõnad, sageduselt teisena keskvärdevormid (*adjective comparative*) ning kolmandana neljandana ülivõrdevormid (*adjective superlative*) ja käändumatute adjektiivide rühm (*adjective indeclinable*). Kahe viimase rühma sagedusjaotus korpusel on lähedane.



Joonis 11. Adjektiivide (*positive, comparative, indeclinable, superlative*) sagedusjaotus ühendkorpusel

5.4 Rühmade polüseemsus

Alapeatükis annan lühikese ülevaate uuritud rühmade polüseemsust. (100 sagedasema adjektiivi ja värviadjektiivide rühma andmed on avaldatud ka artiklites P1 ja P2.) Tabel 2 kõrvutab värvi-, temperatuuri ja kompimisadjektiivide valimit ning 100 sagedasema adjektiivi rühma: polüseemsete sõnade hulk on esitatud nii arvuliselt kui ka protsendina valimist.

Tabel 2. Polüseemsete sõnade arv uuritud rühmades

	VÄRV	TEMPERATUUR	KOMPIMINE	100 sagedasemat
Rühma suurus	463	101	265	100
Polüseemseid sõnu rühmas	115 (25%)	46 (46%)	102 (39%)	80 (80%)

Kõige väiksem polüseemsete sõnade osakaal on värviadjektiivide rühmas (25%) ning kõige kõrgem 100 sagedasema adjektiivi rühmas. Semantilistest rühmadest on kõrgeim temperatuurirühma polüseemsete sõnade osakaal (46%), kuid lähedane on ka kompimisrühma tulemus (39%). Samas tuleb silmas pidada, et värviadjektiivide rühm on sõnade arvult suurim ja temperatuuradjektiivide rühm väikseim. Tulemused kinnitavad, et sagedasti esinev sõna kaldub olema polüseemne (Zipf 1945) ja rühma polüseemsuse tase on korrelatsioonis ka selle morfoloogiliste tunnustega. Mida polüseemsem rühm, seda suurem juursõnade ja väiksem liitsõnade osakaal.

6. ÜLEVAADE POLÜSEEMIAMALLIDEST

Kuues peatükk esitab kokkuvõtlikult värviadjektiivide, temperatuuriadjektiivide ja kompimisadjektiivide polüseemiamallid, mida käsitlesid vastavalt P2, P3 ja P4. Esimene alapeatükk annab ülevaate süstemaatilistest tähendusvaheldustest, mis esinesid vähemalt kahes uuritud rühmas. Seejärel tutvustan levinumaid ühiseid malle lähemalt. Alapeatükis 6.2 käsitlen rühmade eriomaseid polüseemiamalle ehk malle, mis seonduvad analüüsis ainult ühe uuritud semantilise rühmaga. Alapeatükk 6.3 keskendub polüseemiamallide esitusele sõnaraamatus.

6.1 Rühmade kattuvad polüseemiamallid

Tabel 3 annab ülevaate süstemaatilistest tähendusvaheldustest, mis esinesid vähemalt kahel uuritud rühmal. Kuna doktoritöös uuritud semantilised rühmad on kõik tajusõnade rühmad, on ka rühmade tähendusvaheldustes rohkem kattuvust, kui oleks kaugemate valdkonna sõnadel. X tähistab tabelis doktoritöös uuritud sõnarühmi: värvi-, temperatuuri või kompimisadjektiive. A viitab adjektiivsele tähendusele ja S substantiivsele tähendusele. Tärn tähistab vastava polüseemiamalli puudumist rühmas.

Tabel 3. Valimi ühised polüseemiamallid

Mall	Värv	Temp.	Komp.	Malli kirjeldus
X-PSÜHH	9	24	43	Uuritud tajusõnal esines ka meeleolu, suhtumist või iseloomu väljendav tähendus, nt <u>must</u> kass => <u>must</u> meeleolu, <u>soe</u> saun => <u>soe</u> vastuvõtt, <u>terav</u> serv => <u>teravad</u> suhted
X-HÄÄL	9	6	18	Uuritud tajusõnal esines ka heli, häält iseloomustav tähendus, nt <u>heledad</u> seinad=> <u>hele</u> lauluhääl, <u>käre</u> pakane => <u>käre</u> mehehääl, <u>kõva</u> voodi => <u>kõva</u> lärm
MATERJAL/ AINE – X	16	1	11	Uuritud tajusõnal esines ka teatud materjalist/ainest koosnemist või sellega kaetud olemist väljendav tähendus, nt <u>pigine</u> korsten => <u>pigine</u> pimedus, <u>jäine</u> rada => <u>jäine</u> tuul, <u>siidine</u> hommikumantel => <u>siidised</u> põsed
X-NÄHTUS_ LOODUS (A, S)	4	14	10	Uuritud tajusõnal esines ka ilma, loodusnähtust, keskkonna olekut iseloomustav tähendus, nt <u>valge</u> kleit => <u>väljas</u> juba <u>valge</u> , <u>külm</u> vesi => ei kannata <u>külma</u> , <u>kuiv</u> pinnas => terve nädala pidas <u>kuiva</u>

Tabel 3. Järg

Mall	Värv	Temp.	Komp.	Malli kirjeldus
X-MAITSE	*	3	11	Uuritud tajusõnal esines ka maitset iseloomustav tähendus, nt <u>tuline</u> <u>radiaator</u> => <u>tulise maitsega karri</u> , <u>kuiv pesu</u> => <u>kuiva maitsega siider</u>
X-VÄRVUS	*	6	3	Uuritud tajusõnal esines ka värvust või valgust iseloomustav tähendus, nt <u>soe õhtu</u> => <u>soojad toonid</u> , <u>metalne pind</u> => <u>metalne läige</u>
X – LOOM_ OMADUS (S)	9	*	6	Uuritud tajusõnal esines ka sama omadust kandvat looma tähistav substantiivne tähendus, nt <u>kõrb hobune</u> => <u>kõrvid lähenesid</u> , <u>lühikarvaline kass</u> => <u>lühikarvaliste näitus</u>
X – IN_ OMADUS (S)	11	*	3	Uuritud tajusõnal esines ka sama omadust kandvat inimest tähistav substantiivne tähendus, nt <u>blond neiu</u> => <u>talle meeldivad blondid</u> , <u>karvane kasukas</u> => <u>karvaste pidu</u>

6.1.1 X-PSÜHH

Kõigil kolmel rühmal esinesid regulaarsed tähendusvaheldused psühholoogilise omadusega (semantiline tüüp PSÜHH). Enim esines psühholoogilisele omadusele ülekannet kombitavalt omaduselt (43 sõna tähendusstruktuuris). Kõige sagedasem oli psühholoogilise omadusega seotud mall ka temperatuuriadjektiivide rühmas (24 sõna). Nii temperatuuriadjektiivid kui kompimisadjektiivid kuuluvad taktiilsesse valdkonda, nii et on igati ootuspärane, et nende põhilistes polüseemia-mallides on kattuvust. Värvisõnade puhul oli malli esinemine tagasihoidlikum – psühholoogilist tähendust kandis 9 värviadjektiivi.

Nii värvi- kui temperatuurisõnu kasutatakse tunde intensiivsuse väljendamiseks. Malli VÄRV-PSÜHH näitlikustab adjektiiv *must* 'sünge, morn, rõõmutu' millega rõhutatakse lootusetust ja süngust: *must mure*, *must tulevik*, *mustad mõtted*, *must meeoleolu*. TEMPERATUUR-PSÜHH mallist sobivad intensiivistava tähenduse näideteks kollokatsioonid *palav armastus*, *tuline viha*, *põletav kirg*.

Kompimisadjektiivide rühmas (mall PUUDUTUS-PSÜHH) võis täheldada meeldiva puudutusaistingu vaheldumist positiivsete vaimsete omadustega ning ebameeldivat puudutuskogemuse vaheldumist vaimselt ebameeldiva kogemusega. Näiteks adjektiividega *ogaline* ja *okkiline* iseloomustatavad esemed ja olendid (*kaktus*, *roos*, *siil* jt) võivad meid torkida ja meile haiget teha ning nende sõnadega saab väljendada ka psühholoogilisi omadusi, mis on meile rasked taluda ja ebameeldivad (*okkiline teekond*, *okkiline iseloom*). Ka temperatuuriadjektiivide rühmas ilmnes soojema ehk kogemuslikult meeldivama temperatuuriga seotud tähenduse vaheldumine positiivsete omadustega (*soe naeratus*, *palav armastus*) ja külma temperatuuri vaheldumine negatiivse emotsionaalse kogemusega (*külm vastuvõtt*, *jahedad suhted*, *jäine pilk*).

Lakoffi ja Johnsoni (2003) [1980] järgi suhestuvad kõik kognitiivsed nähtused (sealhulgas keele moodustamine ja mõistmine) erinevate kehaliste, afektiivsete ja tajuga seotud protsessidega. Ka süstemaatilise polüseemia uurimine loob pildi mõtlemise muistest ja tajutähenduste vaheldumiste motivatsioon langeb osaliselt kokku mõistemetafooride omaga – abstraktsemate tähenduste väljendamiseks ammutatakse ainest kehalisest kogemusest.

6.1.2 Tajumallid

Teistest tajuvaheldustest esines vähemalt kahel uuritud rühmal süstemaatilisi tähendusvaheldusi hääle-, maitse- ja värvifääriga. Hääle- ja helivaheldustega leidis süstemaatilise polüseemia malle lausa kõigil kolmel rühmal. Enim süstemaatilisi seoseid esines kompimise ja hääletähenduse vahel, malli PUUDUTUS–HÄÄL kandis 18 sõna valimist. Värvirühma analüüsis ilmnes hääletähendus 9 sõna tähendusstruktuuris ja temperatuurirühmas 6 sõna juures.

Ka helivahelduse puhul (mall PUUDUTUS–HÄÄL) esines tähendusülekanne meeldiva puudutusaistingu ja meeldiva hääle vahel ning seda nii temperatuurikompimisadjektiivide rühmas. *Soe, pehme, siidine, sametine* hääli kirjeldavad kõik nauditavat, meeldivat hääli ja ka samade sõnadega osutatakse ka aistingule, mis naha vastas mõnus. Seevastu *käre, kalk* ja *kare hääli* pole sugugi nii meeldivad kuulata ega pole taktiilselt meeldivad ka näiteks *käre talveilm, kalk kivi* või *kare materjal*. Värvirühmas oli seos meeldivuse ja eba-meeldivuse vahel ebaselgem. Määravam oli pigem värv ja hääle intensiivsus või siis just vastupidi – vähene intensiivsus. *Hõbedane* ja *hele* viitasid kõrgele ja kõlavale häälele, ja samas ka värvuse/valguse mõttes tugevat valgust kiirgavale, läikivale värvitoonile. *Akvareliga* viidati õrnadele, õhulistele värvitoonidele ja ka mahedale, õrnale kõlale. *Matt* ja *tuhm* jällegi osutasid säratule, läiketule värvile ning ka tuhmile ja kõlatule häälele.

Kohati oli keeruline eristada, kas seos on ikka hääletähenduse ja uuritud tajutähenduse (kompimine, temperatuur, värv) vahel või pigem on teiseks osapooleks juba sõna psühholoogiline tähendus, nt *kalk* viitab südametule inimesele, seega võiks eeldada, et *kalk hääli* on sellise inimese omadus; samuti osutab ergas nii erk-sale toonile, vaimult virgele loomusele kui ka väljendusrikkale, ilmekale häälele.

Maitsevaheldustega esines tähendusvaheldusi temperatuur- ja kompimisadjektiivide rühmas, kuid mitte värvisõnade juures. Temperatuurirühmas ilmnes süstemaatiline mall TEMPERAATUUR–MAITSE küll ainult 3 sõna tähendusstruktuuris (nt *tuline pann=>tuline karri*), kuid kompimisõnade rühmas esines mall PUUDUTUS–MAITSE 11 sõna juures (*kuiva maitsega siider, metalne järelmaitse, siidise maitsega jäätis*). Kompimisadjektiividel oli korpusel maitseõnadena rohkem kasutust ja täpsemaid tähendusi erialasõnavaras, alkoholsete jookide kirjeldamisel (*metalne, nahkne, siidine, mullane* jt). Huvitav küsimus on veel, kui võrd iseloomustab nt *sametine, sametjas, sametpehme maitse* tegelikult maitset ja kui palju mõjutab hinnangu andmisel ka konkreetse toidu tekstuur (nt *pehme, vahune*), kuid siiski võib eeldada, et viidatakse ka õrnale, mahedale maitsele.

Värvi-, värvuse omadust väljendati süsteemselt nii temperatuuri kui kompimisõnadega. Rohkem esines värvuse väljendamist temperatuurisõnade abil (6 sõna valimist, nt *soojad toonid, külm valgus*), kuid mõnel juhul oli värvuse tähendus lisandunud ka kompimisõnale (3 sõna juures, nt *vahane nahatoon*). Värvusega vaheldumist esines teistel uuritud rühmadel vähe, nii et siin ei kinnita süstemaatilise polüseemia uurimine tajuhierarhia loogikat, kust väljendusvahendeid laenatakse kõrgema meele poolt.

Küll aga kinnitavad süstemaatilise polüseemia mallid metafooriuurimisi (nt Williams 1976, Shen ja Cohen 1998, Shopin 2017), näidates tugevat sidet kompimismeel ja kuulmismeel vahel ning ka kompimismeel ja maitsemeele vahel. Nägemisvaldkonnal olid võrdselt tugevad süstemaatilised seosed kompimisvaldkonnaga (kui arvestada kokku temperatuuri- ja kompimisadjektiividega seotud mallid) ning kuulmismeeliga. Maitsemeelega seotud tähendustega nägemisvaldkonnal uuritud värvisõnade ulatuses süstemaatilist seost ei olnud.

6.1.3 MATERJAL/AINE – X

Semantiline tüüp MATERJAL/AINE pärineb Eesti nimisõnade semantilisest tüübistikust (Langemets 2010) ja tähistab tähendusi, mis väljendavad mingi materjali või ainega kaetud olemist / koos olemist või sellest tehtud olemist. Mall MATERJAL/AINE – X oli kõige produktiivsem mall värviadjektiivide rühmas (16 esinemist) ja kõrge sagedusega (11 sõna struktuuris) ka kompimisadjektiivide rühmas. Temperatuurisõnadel selline regulaarne tähendusvaheldus puudus.

Nii värviadjektiivide kui kompimisadjektiivide rühmas vaheldusid tajutähendused regulaarselt adjektiivsete ainetähendustega (MATERJAL/AINE (A)) ja ka substantiivsete ainetähendustega MATERJAL/AINE (S). Adjektiivse tähendusvahelduse näide värvirühmast on sõna *süsine*, mis võib tähendada „sõega kaetud, söestunud“ ja „täiesti must, väga tume“. Sarnased vaheldused esinevad ka nt adjektiividel *pigine, tindine, vahane, pronksine* ja *kuldne*. Substantiivse tähendusvahelduse korral tähistab sõna nii teatud materjali või ainet (substantiiv) kui ka sellega seostatavat värvitooni, näiteks vask kui metall ja ka kui punakas värv. Siia sobivad veel näideteks *purpur, ultramariin, vask, mahagon, neon* ja *rooste*.

Mallis MATERJAL/AINE – PUUDUTUS esineb regulaarne tulenemissuhe, kus teatud materjali/ainet tähistavale omadussõnale on liitunud selle materjaliga seonduv kompimisaistingu tähendus, mida omakorda saab kanda juba teistele (teisest materjalist) objektidele. Kollokatsioonides *sametine ülikond, sametine kate, sametine kangas* tähistab *sametine* sametist tehtud esemeid, kuid kollokatsioonides *sametine nahk, sametised huuled* juba sameti pehmust meenutavat aistingut. Samuti viitavad *takused püksid, takune riie* ja *takune nõör* konkreetsele takust tehtud materjalile, *takused juuksed* ja *takune habe* aga taku-taolisele pulstile, sassis olekule.

6.1.4 X-NÄHTUS_LOODUS

NÄHTUS_LOODUS on laiem semantiline tüüp, mis pärineb eesti nimisõnade tüübistikust (Langemets 2010) ja tähistab ilma, loodusnähtust, looduse olekut. Tüübiga NÄHTUS_LOODUS esines süstemaatilisi vaheldusi kõigil kolmel uuritud rühmal. Temperatuuradjektiivide rühmas oli selliseid vaheldusi 14 sõnal ja kompimisadjektiivide rühmas 10 sõnal. Värvisõnadel on teoreetiliselt universaalne elliptiline kasutus, kus värvisõnaga saab viidata mistahes vastavat värvi esemele/nähtusele: *taevas on halli, puudes on punast*, kuid siiski pole selline tähendusvaheldus enamasti nii süstemaatiline, et seda sõnastikes esile tuua või et korpusanalüüsil paistaksid konkreetse üht liiki entiteetid, mida ühise tähenduse alla koondada. Ka temperatuurisõnade puhul on võimalus looduse olekule viidata peaaegu universaalne, kuid samuti lähtusin siin väljatoomisel tähenduse korpus-esinemisest.

Kompimisõnade uurimuses ilmsid regulaarsed tähendusvaheldused nii adjektiivsete kui substantiivsete NÄHTUS_LOODUS tähendustega. Mallis vaheldus konkreetne kombitav tähendus, nt *kuivad riided* (PUUDUTUS), abstraktsema kirjeldusega, nt *kuiv suvi* (NÄHTUS_LOODUS A) ja ka substantiivse tähendusega, nt *nädal aega pidas kuiva* (NÄHTUS_LOODUS S). Langemets (2010: 159–161) on välja toonud, et sõnad, mis osutavad omadust ja ühtlasi ka omaduse kandjat, esindavad sõnaliigi vaheldumisega seotud metonüümsel ülekandel põhinevat süstemaatilist polüseemiat.

6.1.5 X- IN/LOOM_OMADUS

Semantilised tüübid IN_OMADUS ja LOOM_OMADUS koondavad substantiivseid tähendusi, mis osutavad vastavat omadust kandvale inimesele või loomale, näiteks *brünett* kui 'tumepruunide juustega' (*brünett neiu*) ja ka kui 'tumepruunide juustega inimene' (*veetlev Brünett*) või *võik* kui 'kollane' (*võik täkk*) ja ka kui 'kollane hobune' (*võik sõi kaera*). Nii X – LOOM_OMADUS kui X – IN_OMADUS malli kandsid uuritud rühmadest värviadjektiivid ja kompimisadjektiivid, temperatuuradjektiividel neid ei ilmnenu. Rohkem esines nimeetatud tähendusvaheldusi värvisõnade rühmas: malli X – LOOM_OMADUS 9 korral ja malli X – IN_OMADUS 11 sõna juures (kompimisrühmas vastavalt 6 ja 3 sõna tähendusstruktuuris).

Värvisõnade IN_OMADUS vahelduse puhul viidati enamasti kas inimese juuksevärvile või nahavärvile (*blond, šatään, Brünett, must, valge, kollane, punane, värviline*). LOOM_OMADUS tüübiga sai märgendada valdavalt hobusele viitavaid värvisõnu (*kõrb, raudjas, hiir, võik, kimmel*). Hobustele osutavad värvisõnad olid küll üldiselt madala korpussagedusega ja seega sõltub selle malli väljatoomise otsus sõnastikus kindlasti sihtrühmast, näiteks õppijasõnastikus võiksid need pigem kõrvale jääda.

Kuigi temperatuuradjektiivid ja kompimisadjektiivid on mõlemad seotud kompimismeelega ning värviadjektiivid nägemismeelega, siis ei tähendanud see, et ainult temperatuuri ja kompimise mallid omavahel kattuksid. IN/LOOM_

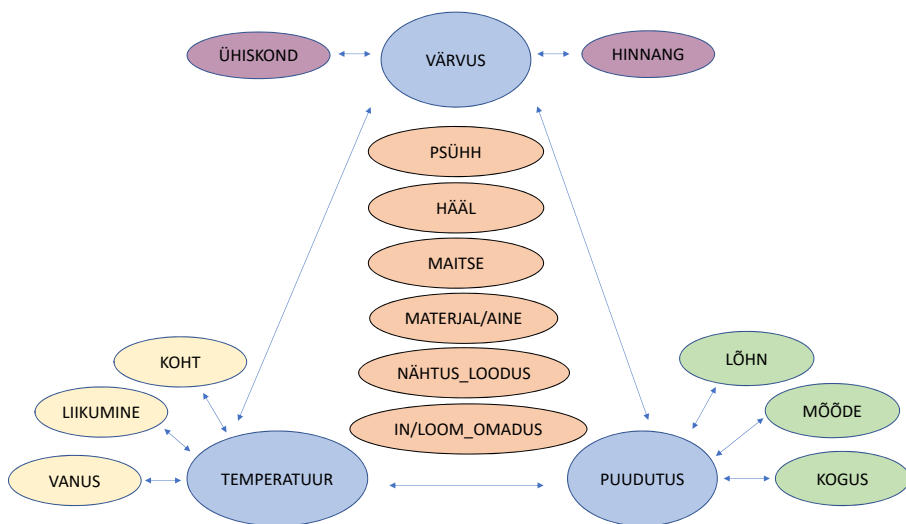
OMADUS tüüpi tähendustega vaheldusid süstemaatiliselt just värvi- ja kompimisadjektiivid, nii et inimesele või loomale viidatakse värvi või kompimisega seotud omaduse kaudu, mitte temperatuuriga. Temperatuur omadusena ei ole enamasti peale vaadates selgeks tehtav (nt radiaatorit vaadates me ei tea, kui külm või soe see on), seega on väljajäämine loogiline. Nägemismeel võib samas kompimissõnade töötlemisel ja mõistmisel sageli olla kompimismeele kõrval toetavas rollis või isegi esmane, näiteks saame hinnata looma *karvaseks* ilma teda puudutamata ja *karvane* pikajuukselisele inimesele viitavalt on ilmselt enamasti vaatluse teel selgeks tehtud.

Kokkuvõtvalt võib öelda, et temperatuuriadjektiivide ja kompimisadjektiivide rühmas oli süstemaatilise polüseemia mallides mitmeid kattuvusi. Kõige produktiivsem mall mõlemas rühmas oli X-PSÜHH ehk temperatuuri või kombitava omaduse vaheldumine psühholoogilist omadust tähistava tähendusega. Mall esines mõlemas rühmas kahesuunaliselt – sõnadel, mille esmatähendus osutas temperatuurile või kombitavale omadusele, oli ka psühholoogilise omaduse tähendus (nt *leige vastuvõtt*, *kivikõva iseloom*) ning sõnadel, mille esmatähendus oli psühholoogiline, esines temperatuuri või kombitavat omadust kirjeldav tähendus (nt *karm kliima*, *karmi karvaga koer*). Selliseid sõnu, kus psühholoogiline omadus oleks esmane ja kompimistähendus kaasneks, esines valimis vähem, kui sõnu, mis olid esmase psühholoogilise tähendusega ja teisese temperatuuritähendusega. Läbipõimunud semantika annab tunnistust valdkondadevahelisest tugevast sidemest. Temperatuuri- ja kompimisadjektiivide kokkukuuluvust näitavad ka mõlema rühma vaheldused hääle ja värvivaldkonna tähendustega.

6.2 Rühmade eriomased polüseemiamallid

Joonis 12 esitab ülevaate värvi-, temperatuuri- ja kompimiserühma süstemaatilise polüseemiast semantiliste tüüpide abil. Kolmiku keskel paiknevad semantilised tüübid, millega esines süstemaatilise polüseemia malle vähemalt kahel uuritud rühmadest (täpsemaks eristuseks vaata tabel 3). Lisaks on põhirühmad joonisel nooltega ühenduses semantiliste tüüpidega, millel leidis süstemaatilisi tähendusvaheldusi ainult ühendatud rühmaga. Kolm põhirühma on nooltega ühendatud ka omavahel, sest nende vahel esines samuti süstemaatilist polüseemiat.

Värviadjektiivide rühmal esines mall VÄRV-HINNANG, kus värvitähendus vaheldub hinnanguga seotud omadusega, näiteks hall kui „igav“, kahvatu kui „mitte muljet avaldav“ ja kuldne kui „väga hea, väärtuslik“. Teiste rühmade puhul sellist malli esile ei kerkinud, kuid tuli ette, et analüüsil oli keeruline eristada semantilise tüübiga HINNANG märgendatavaid tähendusi PSÜHH tüübi tähendustest. Näiteks kuidas märgendada *libeda* tähendust kollokatsioonides *libe sell*, *libe jutt*? Kuigi *libe* kannab siin konkreetseid psühholoogilisi omadusi (lipitsev, kaval, osav oma eesmärkide saavutamisel), kannab see endas ka negatiivset hinnangut. Lipitsev olek on midagi, milles suhtes võiks eeldada ühiskonnas laiemalt negatiivset hoiakut ja seega pole sõnavalik neutraalne.



Joonis 12. Ülevaade uuritud rühmade süstemaatilise polüseemiast

Teiseks eriomadeks polüseemiamalliks värviadjektiivide rühmas oli kolmikmall OMADUS_VÄRV – OMADUS_ÜHISKOND – IN_ROLL, kus poliitilise liikumise või ühiskondliku mõtlemise suunaga seotakse mingi värv, mis hakkab tähistama ka seda liikumist/suunda ennast ja ühtlasi ka vastavat mõtteviisi pooldavast inimest (*punane, roheline, valge, roosa, punaroheline, veripunane*). Näiteks punane kui kommunistlik, roheline kui ökoloogilist mõtteviisi pooldav inimene, valge kui parempoolsete kodanlike vaadetega isik, roosa kui vasakpoolsusesse kalduv ja punaroheline kui nii ökoloogilist ja ka vasakpoolset mõtteviisi pooldav inimene.

Temperatuuriadjektiivide rühmas oli ainsana regulaarne vaheldussuhe kohatähendustega. Mallis KOHT–TEMPERATUUR koondab semantiline tüüp KOHT mingi koha või piirkonnaga seotud omadust väljendavaid tähendusi (*arktiline, boreaalne, atlantiline* jt). Omadussõna, mis tähistab teatud geograafilisest piirkonnast pärinemist või sinna kuulumist, on hakanud tähistama ka vastava piirkonnaga seostatavaid kliimaolusid ja temperatuuri, näiteks tähistab sõna *arktiline* Arktikast pärinemist (*arktilised linnud, arktiline õhumass*) ja ka väga külma ja karmi ilma (*jõulude paiku oli täiesti arktiline lumetorm*).

Teised ainuomased temperatuurirühma süstemaatilise polüseemia mallid olid TEMPERATUUR–LIIKUMINE ja TEMPERATUUR–VANUS. LIIKUMINE märgib liikumise ja tempoga seotud omadusi ning kodeerib vaheldumist temperatuuritähenduse ning tempotähenduse vahel. Valimis esines mall rohkem soojaga seotud sõnade juures, nt palavikuline (*palavikuline nahk, palavikuline sagimine*), kuid esines ka kiiret liikumist iseloomustav tähendus külmaga seotud sõna juures käre (*käre talv, käre jõgi*). Semantiline tüüp VANUS koondab kellegi või millegi ea, olemasolu ajaga seotud tähendusi. Mallis TEMPERATUUR–

VANUS vaheldub temperatuuritähendus eaga seotud tähendusega, nt sõnas ahjusoe (ahjusoojad pirukad, ahjusoe raamat).

Kompimisadjektiivide eriomased mallid olid PUUDUTUS – LÕHN, PUUDUTUS–MÕÕDE ja PUUDUTUS–KOGUS. Mallis PUUDUTUS – LÕHN vaheldub kompimistähendus lõhna iseloomustava tähendusega, nt *terav nuga => terava lõhnaga gaas, mullane kartul => mullase lõhnaga seemed*. PUUDUTUS–MÕÕDE mallis kannavad sõnad koos kompimistähendusega ka ruumilist omadust tähistavat tähendust, nt *paks puder => paksud seinad või rasvane laud => rasvaste tähtedega tekst*. Lisaks leidis kompimisadjektiividel vaheldust PUUDUTUS–KOGUS, kus kompimistähendus esines koos hulgatähendusega, näiteks *kõva iste => kõva rahasumma*.

Eesti nägemis- ja kompimisvaldkonna sõnade süstemaatilise polüseemia analüüsi tulemused võimaldavad tulevikus paremini diskuteerida ka kultuurispetsiifilisuse ja universaalse kognitiivse ajendi suhete üle. Kindlasti paistab tulemustest üldinimlike kehaliste kogemuste olulisus, kuid kõrvutades tulemusi teiste keelte polüseemiauurimistega, oleks võimalik ka kultuurikesksem lähenemine.

6.3 Süstemaatiliste polüseemiamallide esitus sõnaraamatus

Polüseemiamallide esitus sõnaraamatus (vt pt 4.2) varieerus kõigis kolmes rühmas. Loomulikult sisaldasidki polüseemiamallid eri tüüpi tähendusvaheldusi, mida ei pruugi olla otstarbekas sõnastikus samal viisil esitada (näiteks vaheldust adjektiivsete tähenduste vahel või vaheldust adjektiivse ja substantiivse tähenduse vahel). Ometi leidis erinevat esitust ka sama malli sama sõnaliigiga osaliste vahel. Siinses peatükis on vaatluse alla võetud psühholoogilise omadusega seotud mall, kuna seda leidis kõigis kolmes rühmas rikkalikult. Enamasti olid mallis kodeeritud kaks tähendust eraldi põhitähendustena esitatud (malli X – PSÜHH näited 1, 2, 3). Välja on toodud ainult malli kandvate tähenduste definitsioonid, uurimisel lisatud semantiline tüüp ja üks kasutusnäide tähenduse kohta.

(1) okkaline

- 1 (torkavate) okaste, ogade või asteldega (kaetud) [PUUDUTUS] *Okkaline kala*
- 2 pahur ja tõrges [PSÜHH] *Tüdruk oli võõraste vastu okkaline*

(2) külm

- 1 madala temperatuuriga, vähese soojusega või hoopis soojuseta [TEMPERATUUR] *Täna on väljas külm ilm*
- 2 (inimese olemuse kohta:) ilma südamlikkuse ja soojuseta, tundetu või ükskõikne [PSÜHH] *Meid võeti vastu külma viisakusega*

(3) **roosa**

- 1 kibuvitsaõie värvi, kahvatupunane [VÄRVUS] *Roosa pesu*
- 2 (liiga) optimistlik, rõõmus ja muretu [PSÜHH] *Ole roosa ja rahulik!*

Malli kodeeritud tähendused võisid olla esitatud ka põhitähenduse ja alltähendusena (näited 4, 5, 6). Alltähenduse positsioonis esitatakse tavaliselt seotumaid tähendusi, mille puhul on märgata otsesemat tulenemissuhet ja Langemetsa (2010: 50) järgi on alltähendusena esitamine süstemaatiline polüseemia puhul sageli eelistatum esitusviis.

(4) **kõva**

- 1 (materjali, asjade vm kohta:) koostiselt tihe, survele vastupidav, mitte painduv [PUUDUTUS] *Hiina traditsiooni järgi peab voodi olema kõva*
- 1.2 (inimese kohta:) iseloomult, käitumiselt järeleandmatu, mingile mõjule mittealustuv [PSÜHH] *Ole kõva, ära anna alla!*

(5) **jäine**

- 1 külm nagu jää, väga külm [TEMPERATUUR] *Jäist tuult trotsides rühiti edasi*
- 1.1 (tunnete, psüühilise sfääri kohta:) äärmiselt ükskõikne, kalk [PSÜHH] *Külaline võeti vastu jäise viisakusega*

(6) **erk**

- 1 vaimult virge, terane ja (uuele) vastuvõtlik, kiiresti reageeriv, mitte uimane ega loid [PSÜHH] *Ta on heas vormis ja vaimselt erk*
- 1.1 (värvi kohta:) puhta, silmatorkava värvitooniga [VÄRVUS] *Huulepulk ei tohiks olla liiga erk*

Toodud näited annavad tunnistust, et leksikograafide taju võib erineda sama malli osaliste tähendusvahekorra tajumisel. Tähendusi võidakse pidada eraldiseisvateks üksusteks või teineteisest tulenevaks. Kolmas leksikograafiline esitusviis oli malli osaliste koondamine ühise tähenduse alla (näide 7). Selles näites on tähenduste kirjelduses väga olulised ka näitelauseid. Viimaks ilmnes korpusanalüüsi tulemuste ja leksikograafilise esituse kõrvutamisel ka sõnu, mille puhul ühendsõnastik kompimisadjektiivi psühholoogilist tähendust eraldi esile ei toonud (ei tähenduse ega kasutusnäidete kujul), kuid mis uuritud korpusnäidete põhjal malli siiski kandsid (näide 8).

(7) **teraskõva**

- 1 (hrl materjali või isikuomaduste kohta:) väga kõva, väga tugev, ülimalt vastupidav [PUUDUTUS] [PSÜHH] *Lapimaa jää on teraskõva. Tal on teraskõvad närvid*

(8) **limane**

limaga kaetud või lima sisaldav [PUUDUTUS] *Angerja nahk on limane*
[PSÜHH] korpusnäide: *Kaasajal baarides räuskavad mehed on aga küünilised, empaatiatundeta limased tegelased.*

Malli produktiivsust demonstreeriski ootuspäraselt paremini korpus. Väiksema sagedusega sõnade juures võis olla sõnastikus esitatud üks malli osapooltest.

Polüseemiamalli konkreetse parima leksikograafilise esituse valimisel mängivad rolli sõnastiku sihtrühm, sõnasagedus, vastava tähenduse sagedus ja muud tegurid. Samas näitab sama malli erinev esitus sageduse poolest võrreldavatel sõnadel, et polüseemiamallid võiksid süsteemsemat esitust abistada, kui leksikograafil oleks võimalus potentsiaalsete süstemaatilise polüseemia mallidega koostamise ajal tutvuda. Semantilise tüübi lisamine sõnatähendusele (käändsõnade juures) on ühendsõnastikus juba praegu kohustuslik element ja ehk hõlbustaks tööd polüseemsete sõnadega see, kui süsteem pakuks leksikograafile välja konkreetse semantilise tüübiga seotud võimalikke polüseemiamalle.

7. TULEMUSTE KOKKUVÕTE JA ARUTELU

Väitekirja eesmärk on uurida eesti tajuadjektiivide süstemaatilist polüseemiat kolme semantilise rühma näitel – selgitada semantiliste tüüpide abil korpuspõhiselt välja värvi-, temperatuuri- ja kompimisrühma polüseemiamallid. Töö tegeleb ka adjektiivirühma sõnaliigilise piiritlemisega ja uurib, millised adjektiivitunnused on korpuses kaardistatavad viisil, mis aitaks leksikograafil paremini adjektiiviklassi piiripealsete juhtumite puhul sõnaliiki määrata. Nii süstemaatilise polüseemia mallide selgitamise kui ka adjektiivitunnuste sõnaliigi korpusprofili loomise rakenduslik siht on hõlbustada sõnastikutööd ning aidata kaasa adjektiivitäpsemale esitusele leksikograafias.

7.1 Tajuadjektiivide süstemaatiline polüseemia

Alapeatükis 7.1 käsitlen järjest vastuseid uurimisküsimustele 1–6 (vt ptk 1.1). Et välja selgitada süstemaatilist polüseemiat, rakendasin adjektiivitunnuste korpuskonteksti analüüsil semantilisi tüüpe. Semantilised tüübid on abstraktsed abivahendid sõnastikutöös ja korpusanalüüsis, mis võimaldavad tähendusi rühmadesse koondata. Adjektiivitunnuste tüüpide arv võib erinevates käsitlustes suuresti varieeruda. Uurimuses kasutasin (P1) eeskujuna ja võrdlusmaterjalina Dixoni (1982) seitset olulisemat adjektiivitunnust semantilist tüüpi, leksikaalse andmebaasi SIMPLE koostamisel väljatöötatud 18 tüüpi (Lenci jt 2000: 30–33) ning Hundsnurcheri ja Spletti (1982) uurimuse põhjal loodud GermaNeti 15 adjektiivitunnust klassi, millel on omakorda 78 alaliiki. 100 eesti sagedama adjektiivitunnuste tähendusi analüüsid pakkusin (esimeses artiklis, P1) välja 15 semantilist tüüpi spetsiaalselt eesti keele (polüseemsete) adjektiivitunnuste kirjeldamiseks. Hilisemates uurimustes (P2, P3, P4) tüübitäiendasin ja täpsustasin, nt lisasin tüübid HÄÄL, MAITSE, LÕHN ja PUUDUTUS. Välja töötatud tüübid võimaldavad märgendada tähendusi adjektiivitunnuste polüseemia uurimisel ja ka adjektiivitunnuste täpsemat semantilist märgendamist sõnaraamatutöös.

Uuritud rühmade morfoloogiliste tunnuste, korpusageduse ja polüseemsete võrdlus kinnitab, et polüseemia eelistab sagedat sõna (vt ka Zipf 1945). Polüseemsete korrelatsioonis ka rühmade morfoloogilise struktuuriga – mida polüseemsem rühm, seda suurem on juursõnade ja väiksem liitsõnade osakaal. Uuritud rühmadest oli enim harvemaid sõnu (alla 100 esinemise korpuses) värviadjektiivitunnuste rühmas. Liitsõnad moodustasidki 75% värviadjektiivitunnuste rühmast ja ligikaudu pool (49%) temperatuuriadjektiivitunnuste rühmast. Seega on liitsõna põhiline morfoloogiline kuju, milles nende rühmade adjektiivitunnuste eesti keeles esinevad. Kuna paljud värvisõnad ja temperatuuriõnad põhinevad võrdlusel (nt *rebu-kollane*, *vaseroheline*, *kõrbekuum*, *kivikülm*), on selline jaotus loomulik.

Suurim uuritud semantiline rühm on värviadjektiivitunnuste 463 sõnaga, järgneb kompimisadjektiivitunnuste rühm 265 sõnaga ning väiksem rühm on temperatuuriadjektiivitunnuste 101 sõnaga.

Rühmade suurust on huvitav vaadelda ka meelte hierarhia seisukohast. Winter (2019: 342) pakub välja, et meelte hierhia aluseks võib olla tõik, et mõne meelega seotud omaduste väljendamiseks on keeles rohkem adjektiive kui mõne teise meele puhul. Siinse töö nägemismeelega seotud rühm (värvid) on tõesti adjektiivirikkam kui kompimismeelega seotud rühmad kokku (366 sõna, kui arvestada temperatuuri- ja kompimisadjektiive ühiselt). Lisaks on nägemisega seotud rühm kitsam ja ei sisalda näiteks objektide ega nähtuste kujuga seotud sõnavara (nt *kõrge, madal*) või üldisemaid hinnangulisi adjektiive (nt *ilus, inetu*). Kui eeldame, et valimi suurus peegeldab suuremat leksikaalset tähistatust keeles laiemalt, sobituvad tulemused meelte hierarhia (Viberg 1983) reastusega, mille kohaselt nägemise kui kõrgema meelega seonduv sõnavara on keeles rikkalikum kui kompimismeelega seonduv. Lõplikke järeldusi siiski teha ei saa, sest kuigi rühmad on koostatud põhilises osas sama, eesti suurima sõnakogu alusel, on kompimisadjektiivide piiritlemiseks erinevaid võimalusi (vt täpsemalt P4) ja samuti ei sisalda värvirühm kõiki nägemismeelega seotud adjektiive eesti keeles.

Kõik doktoritöös uuritud semantilised rühmad on tajusõnade rühmad ja seetõttu on ka rühmade tähendusvaheldustes (ja polüseemiamallides) tõenäoliselt rohkem kattuvust, kui oleks üksteisest kaugemate valdkonna sõnadel. Samuti mõjutab polüseemiamallide sarnasust, et temperatuuriadjektiivid kuuluvad kompimisadjektiividega üldistatult samasse tajuvaldkonda – mõlemad on seotud taktiilise tajuga. Kõigil kolmel rühmal esinesid regulaarsed tähendusvaheldused psühholoogilise omadusega (semantiline tüüp PSÜHH). Enim esines psühholoogilisele omadusele ülekannet kombitavalt omaduselt (43 sõna tähendusstruktuuris). Kõige sagedasem oli psühholoogilise omadusega seotud mall ka temperatuuriadjektiivide rühmas (24 sõna). Värvisõnade rühmas esines malli vähem – psühholoogilist tähendust kandis 9 värviadjektiivi.

Nii värvi- kui temperatuurisõnu kasutatakse tunnete intensiivsuse väljendamiseks (nt lootusetuse rõhutus: *must mure, must meeoleu*; kire rõhutus: *palav armastus, tuline viha, põletav kirg*). Mõlemas rühmas võis täheldada ka meeldiva puudutusaistingu vaheldumist pigem positiivsete vaimsete omadustega (nt *soe tuba, soe vastuvõtt; pehme padi, pehme loomuga*) ning ebameeldivat puudutuskogemuse vaheldumist vaimselt ebameeldiva kogemusega (nt *külm ilm, külm pilk; okkiline roos, okkiline iseloom*). Siiski leidis tajusõnadel ka näiteid tähendusvaheldustest, mis ei kinnita omaduse kognitiivse meeldivuse või ebameeldivuse järjepidevust. Mitmed kombitavat omadust väljendavad adjektiivid, millel esineb polüseemiamall PUUDUTUS–PSÜHH, ei ole kompimiskogemuse põhjal üheselt liigitatavad positiivseks või negatiivseks (nt *karvane loom, karvane tunne; tasane maapind, tasane loomuga*).

Lakoffi ja Johnsoni (1980) järgi põhineb keele moodustamine ja mõistmine erinevatel kehalistel, afektiivsetel ja tajuga seotud protsessidel. Abstraktsemate ja keerulisemate nähtuste väljendamiseks kasutatakse võrdkujusid kogemuslikult tuttavamatest ja kehapõhisematest valdkondadest. Süstemaatilise polüseemia uurimine osutab tugevale seosele kompimisvaldkonna ja tunnete valdkonna vahel. Kehaga/nahaga/suuga vahetult kogetav on ülimalt sobilik lähtevaldkonnaks, mille mõistetes räägitakse abstraktsematest nähtustest nagu psühholoogilised nähtused.

Tugevat seost valdkondade vahel kinnitas temperatuurirühmas ka kahepool-selt läbipõimunud semantika. Esmatähendusega temperatuurisõnadest kandsid malli kõik temperatuuri põhisõnad (nt *külm suhtumine, soe inimene, kuumad tunded, jahedad suhted*). Aga valimis leidis ka mitmeid peamiselt psühholoogi-lise tähendusega kasutusel olevaid sõnu, millel esines temperatuuritähendus (nt *kõhe, leebe, karm*). Kui temperatuuridjektiivide rühmas esines vaheldumine psühholoogilise omadusega kahesuunaliselt, siis kompimisadjektiivide rühmas on valdav tähenduse liikumine konkreetsest kombitavast omadusest abstraktsema psühholoogilise omaduse poole.

Uuemad uurimused lubavad oletada, et temperatuuriga seonduvad suhtumised keeles ei pruugi olla universaalsed, vaid sõltuvad ka keekekasutajate kultuurist ja keskkonnast (Majid jt 2018, Koptjevskaja-Tamm 2015). Uurisin (P3) eesti temperatuuri põhisõnade seondumist positiivse või negatiivse hinnanguga Sketch Engine'i tesauruse funktsiooni abil, mis võimaldas vaadelda korpusest sarnase kontekstiga sõnu. Ootuspäraselt ilmneseid külma puhul negatiivset hinnangut väljendavad adjektiivid (nt *igav, kole*) ja sooja puhul positiivse hinnanguga sõnad (nt *mõnus, ilus, kena, meeldiv*). Jaheda tugevamate tesauruseseste hulgas hinnangulisi omadussõnu ei esinenud. Põhisõna *kuum*, mis võiks viidata juba liig-sele soojusele ja omada ühiseid kollokatsioone ka negatiivset hinnangut väljen-davate omadussõnadega, seostus tugevamate tesaurusesuhete põhjal samuti ainult positiivse hinnanguga (nt *mõnus, kena*). Korpusanalüüs viitab eestlaste soojalembusele ja lubab oletada, et Eestis kui külma kliimaga piirkonnas on seos soojuse ja positiivsete emotsioonide vahel eriti tugev. Siiski oleks kindlamaks väitmiseks vajalik võrdlev uurimus. Eesti emotsioonisõnade uurija Ene Vainik (2016: 285) toob välja, et seda erinevates keeltes ja kultuurides ilmnevat püsiseost kiindumuse ja soojuse vahel on seletatud universaalse nn primaarmetafoorina ehk seosena, mis kujuneb meil kõigil imikueas kehalisest süles hoituse kogemusest.

Teistest tajualdkondadest esines vähemalt kahel uuritud rühmal süstemaati-lisi tähendusvaheldusi hääle-, maitse- ja värvivaldkonnaga. Hääle- ja helivald-konnaga leidis süstemaatilise polüseemia malle lausa kõigil kolmel rühmal. Enim süstemaatilisi seoseid esines kompimise ja hääletähenduse vahel, malli PUUDUTUS–HÄÄL kandis 18 sõna valimist. Värvirühma analüüsis ilmnese hääletähendus 9 sõna tähendusstruktuuris ja temperatuurirühmas 6 sõna juures. Maitsevaldkonnaga esines tähendusvaheldusi temperatuuri- ja kompimis-adjektiivide rühmas, kuid mitte värvisõnade juures.

Puudutuse ja heliga seotud tähenduste puhul (mallid PUUDUTUS–HÄÄL, TEMP–HÄÄL) võis samuti märgata tähendusülekannet meeldiva puudutus-aistingu ja meeldiva hääle vahel ning seda nii temperatuuri- kui ka kompimis-adjektiivide rühmas. *Soe, pehme, siidine, sametine* hääled kirjeldavad kõik naudi-tavat, meeldivat häält ja ka samade sõnadega osutatakse ka aistingule, mis naha vastas mõnus. Seevastu *käre, kalk* ja *kare hääled* pole sugugi nii meeldivad kuulata ega taktiilselt pole meeldivad ka *käre talveilm, kalk kivi* või *kare materjal*. Värvirühmas oli seos meeldivuse ja ebameeldivuse vahel ebaselgem. Määravam oli pigem värvi ja hääle intensiivsus või siis just vastupidi – vähene intensiivsus. *Hõbedane* ja *hele* viitasid kõrgele ja kõlavale häälele, ja samas ka värvuse/

valguse mõttes tugevat valgust kiirgavale, läikivale värvitoonile. *Akvarelsega* viidati õrnadele, õhulistele värvitoonidele ja ka mahedale, õrnale kõlale. *Matt* ja *tuhm* jällegi osutasid säratule, läiketule värvile ning ka tuhmile ja kõlatule häälele.

Williams (1976) on inglise keele sünesteetilisi adjektiive uurides välja toonud, et kõige rohkem teistest tajualadest pärit sõnu kasutatakse helide iseloomustamiseks, samas kui kompimisvaldkond on kõige sagedamini kasutusel allikvaldkonnana. Shen ja Cohen (1998) rõhutavad, et ka adjektiivide sünesteetilistes kasutustes on allikvaldkond sihtvaldkonnast konkreetsem ja seetõttu ligipääsetavam, seega võetakse täiend pigem kompimisalast ja harvem nägemis- või kuulmisvaldkonnast. Süstemaatilise polüseemia uurimise tulemused toetavad sünesteetilise metafoori uurimisel täheldatud tendentsi.

Värvusega vaheldumist esines teistel uuritud rühmadel üsna vähe, nii et siin ei kinnita süstemaatilise polüseemia uurimine Vibergi (1983) tajuhierarhia loogikat, mispuhul väljendusvahendeid laenatakse kõrgema, eeldatavasti sõnavaraliselt paremini kaetud, meele poolt. Temperatuuri- ja kompimisrühma sõnadel esines ka regulaarseid tähendusvaheldusi nii nägemise kui kuulmise valdkonna tähendustega (*soe valgus, jääne hää, siidine hää*). „Kõrgema kategooria“ tajude (nägemine-kuulmine) väljendamiseks on laenatud „madalama kategooria“, kompimisvaldkonna sõnu.

Kokku avaldus värviadjektiivide rühmas 9 polüseemiamalli, temperatuuriadjektiivide rühmas 10 polüseemiamalli ja kompimisadjektiivide rühmas 13 polüseemiamalli. Vähemalt kahes uuritud rühmas esinesid süstemaatilised vaheldused tüüpidega MATERJAL/AINE, NÄHTUS_LOODUS ja substantiivsete tähendustega, mis kandsid tüüpi IN/LOOM_OMADUS. Enamik polüseemiamalle moodustusid paarisuhetest, ent avaldus ka kolmiksuhteid: VÄRV – ÜHISKOND – IN_ROLL, PUUDUTUS–HÄÄL–MAITSE.

Polüseemiamallide esitus sõnaraamatus varieerus kõigis kolmes rühmas. Mall võis olla esitatud: 1) kahe eraldi põhitähendusena 2) põhitähenduse ja alltähendus(t)ena; 3) ühise koondtähendusena; 4) süstemaatilist polüseemiat polnud esitatud. Hindasin esitamata jätmiseks olukorda, kus tähendus sõnastikust puudus, ent korpuses arvestatava sagedusega esildus. Analüüs näitas, et leksikograafide taju võib erineda sama malli osaliste tähendusvahekorra tajumisel ning seetõttu võiks süsteemsemale esitusele kaasa aidata tutvumine vastava semantilise tüübiga seotud polüseemiamallidega.

7.2 Adjektiivi korpusprofiil

Alapeatükis 7.2 käsitlen vastust uurimisküsimusele 7 (vt ptk 1.1). Nii nagu adjektiivi semantika pole alati samal viisil kirjeldatud, pole ka adjektiivikategooria ise keeltes ühtlaselt esindatud ega selgelt piiritletud: on keeli, kus omadussõnu on piiratud arvul või pole adjektiivide klassi üldse (Dixon 1982: 1–7). Kõik semantilised adjektiivid pole keeles grammatiliselt adjektiivid. Eesti keeles on keeruline piiri tõmmata näiteks partitsiipide või käändumatute adjektiivide puhul. Väitekirja adjektiivirühmade kindlaks määramisel vajab selgitamist adjektiivi-verbi

piir (nt *põletav, torkiv, jahtunud*), adjektiiv-substantiivi piiriala (nt *must, soe, kuiv*) ning adjektiiv-adverbi piir (nt *kullakarva, jahedavõitu, kõvavõitu*). Töö uurib, millised on keeletehnoloogia võimalused hõlbustada sõnaliigi üle otsustamist adjektiiv piirialade puhul ja kuidas võiks korpuse andmete põhjal eristada adjektiive teistest sõnaliikidest.

Leksikograafide uuringus (Paulsen, Vainik ja Tuulik 2019) töid leksikograafid välja sõnaliigitusprobleeme kirjeldavaid näiteid, millest 26% moodustasid adjektiiviklassiga seotud sõnad. Põhilise murekohana nimetasid leksikograafid vajadust hinnata partitsiipide adjektiveerumise astet (nt sõnades *muteerunud, müüiv, kiusav*). Väitekiri (P5) pakub prototüüpsele adjektiivile omistatavate tunnuste põhjal välja parameetrid, mille abil saaks korpusest kontrollida, kuivõrd on sõnal adjektiivset kasutust ja mis aitaksid leksikograafil otsustada, kas lisada sõna sõnastikku/andmebaasi adjektiivina. Uurimus on aluseks leksikograafilise raken-duse loomisele.

Töö tulemusena välja pakutud parameetrite loomisel on lähtunud adjektiivile omastest tunnusjoontest teoreetilisest kirjandusest ja ka praktilisest leksikograafilisest tööst. Atribuudiparameeter [test_S] tugineb adjektiiv omadusele esineda atribuudi funktsioonis ja mõõdab testsõna esinemisi substantiivi ees. Võrdparameeter mõõdab sõnast moodustatud keskvõrde sagedust korpuses. Ühildumisparameeter [test_S_ÜHILD] lähtub adjektiivide omadusest ühilduda substantiiviga arvus ja käändes ning kontrollib testsõna ühildumist, kui sellele järgneb substantiiv. Lausealguseparameeter [LA_test_S] selgitab, kuivõrd alustab testsõna lauset täiendi positsioonis substantiivi ees. Adverbiparameeter [D_test] tugineb adjektiivide korpussuhtele D+A (nt *liiga soe, väga pehme*) ja mõõdab, kui sageli esineb testsõna ees korpuses adverb. Neljakohaline parameeter [V_?_test_S] selgitab, kui sageli esineb testsõna mustris V_?_test_S (järjend pärineb käsitsi morfoloogiliselt ühestatud korpuse analüüsist).

Adjektiveerumata partitsiipide rühma eristas prototüüpsete adjektiivide rühmast kõige paremini kolm parameetrit: lausealguseparameeter 90% täpsusega, adverbiparameeter 80% täpsusega ja keskvõrde parameeter 100% täpsusega (100% täpsus tähendab, et kõigi uuritud partitsiipide tulemused jäid kõrvalekaldeanalüüsis väljapoole adjektiivide järgi sätitud piirmäära.) Siit nähtub, et verbiühendi osana ei alusta partitsiip lauset substantiivi ees, kuid adjektiivilaadne partitsiip võib seda teha.

Prototüüpsed adjektiivide väärtused ületasid kõiki teisi uuritud sõnarühmi (teised sõnaliigid ja adjektiiv piirirühmad) keskvõrdparameetri ja adverbiparameetriga mõõtes. Adjektiivirühma piirväärtuste arvesse võtmisel (kõrvalekaldeanalüüs) osutusid parimateks eristajateks prototüüpsete adjektiivide ja teiste sõnaliikide vahel: lause alguse parameeter (eristas ~78% sõnadest), keskvõrde parameeter (eristas 100% teistest sõnaliikidest sõnu) ja ühildumisparameeter (eristas ~59% sõnadest, sealhulgas substantiive, adverbe ja verbe 100%-liselt). Tulemused annavad lootust luua rakendus, mis aitaks leksikograafidel hinnata adjektiiviklassiga piirneva keelendi käitumist tekstis ning saada tuge sõnaliigi-otsustuste langetamisel.

7.3 Edasised uurimisvõimalused

Adjektiivivi korpusprofiili lävendeid saaks edaspidi täpsustada, kui keskenduda leksikograafidele probleeme tekitavale partitsiipide rühmale. Oleks võimalik katsetada erinevaid sarnasusmeetodeid (eukleidiline kaugus, koosinussarnasus jt), et selgitada välja, milline meetod annaks parima tulemuse leksikograafilise töö seisukohast. Testida võiks ka uusi adjektiivide rühma iseloomustavaid parameetreid, mida artikkel P5 ei käsitlenud, näiteks predikatiiviparameetrit või adjektiiviviga seonduvaid tuletuslõppe.

Doktoritöös rakendatud meetodiga (korpuse ja adjektiivivi semantiliste tüüpide abil) on võimalik uurida ka eesti keele maitse-, lõhna-, ja kuulmisvaldkonda kuuluvate adjektiivide polüseemiat. Eesti tajuadjektiivide süstemaatilise polüseemia analüüsi tulemused võimaldavad tulevikus paremini diskuteerida ka kultuuri-spetsiifilisuse ja universaalse kognitiivse ajendi suhete üle. Siinse doktoritöö tulemused näitavad üldinimlike kehaliste kogemuste olulisust, kuid kõrvutades tulemusi teiste keelte polüseemiauurimistega, oleks võimalik ka kultuurikesksem lähenemine.

SUMMARY

Systematic Polysemy of Estonian Adjectives on the Example of Perception Adjectives

The main objective of the doctoral thesis is to identify the patterns of systematic polysemy in Estonian perception adjectives, which would make it possible to analyse the patterns of adjective semantics and create a theoretical basis for lexicographers that could be used to harmonise the presentation of adjective polysemy in dictionaries. The thesis also tackles mapping the parameters to characterise the adjective word class in corpora. The practical aim of the thesis is to make lexicographic work easier and to enable a more systemic semantic presentation for adjectives. Also, the thesis contributes to a better part-of-speech analysis of the ambiguous cases connected to adjective word class in corpora¹¹.

Introduction

In recent decades, major changes have been taking place in European lexicography: the format of dictionaries is shifting from paper to electronic (a survey of users covering 26 countries pointed out that 79.8% of users of monolingual dictionaries use an electronic version (Kosem et al. 2019)) and with it, Estonian lexicography has also changed from corpus-based and is moving towards (semi-)automatic solutions (see, for example, Koppel 2020, Kallas et al. 2015). As a result of the above, the need for more precise semantic tagging and more systematic presentation has increased. Polysemy continues to be a relevant problem in both corpus linguistics and lexicography. In the largest and constantly updated dictionary of the Estonian language, The EKI Combined Dictionary, most elements are related to a specific meaning, e.g. synonyms are combined at the meaning-level, not keyword-level, i.e. the synonym layer of the dictionary is compiled bilaterally (see Tavast et al. 2020). Therefore, systemicity is more important than ever in the word sense division.

This thesis is the first attempt in Estonian linguistics to study the systematic meaning patterns of perception adjectives. Polysemy patterns would provide support for the lexicographer in describing a polysemous word, as well as material for the linguistic researcher of the subject, as they show regular semantic shifts between different domains. The adjective can emphasise a different characteristic of a noun in different contexts, for example, *kare rätik* ‘coarse towel’ refers to a material that is non-smooth, rough to the touch, *kare käitumine* ‘coarse behaviour’ to a harsh and abrupt conduct, and *kare hääl* ‘coarse voice’ to a non-melodic, hoarse voice. By studying the adjectives belonging to the same semantic group together, it is possible to find out between which areas there are systematic differences of meaning and thus which other areas the semantic group being studied

¹¹ In Estonian, the adjectives’ morphological profile is characterised by inflection (in case and in number), gradation, and derivation.

is most closely related to. I will also specify the motivation of meaning shifts in the context of cognitive theories (Lakoff and Johnson 1980, Viberg 1983, Winter 2019, etc.). The thesis examines the systematic polysemy of Estonian adjectives on the example of three semantic groups: colour adjectives, temperature adjectives and tactile adjectives.

Structure of the thesis

The thesis consists of a framework chapter and five scientific articles. In the second, theoretical background part of the framework chapter, the theory of distributive semantics and the concept of systematic polysemy are presented and the presentation of polysemy in lexicography is discussed. Also, an overview of the distinctive features and classification of perception adjectives is given and the classification problems of Estonian adjectives are highlighted. In the third part, describing the method, I will introduce the use of semantic types as an analytical tool, as well as Sketch Engine tools for corpus study, and will also describe the limitations of the method. The fourth part on sources and material describes the corpora and dictionaries used for the research, as well as compiling the word groups. The fifth part provides an overview of the morphological composition, frequency data and polysemy of the word groups examined in the thesis. In the sixth part, I will introduce and summarise the polysemy patterns of perception adjectives.

Goals of the Thesis

- Giving an overview of the wide variety of words used to describe the domains of colour, temperature and touch.
- Explaining the connections in morphology, frequency data and polysemy between the word groups in the categories of colour, temperature, and touch.
- On the basis of 100 common adjectives, proposing the basic semantic types of Estonian adjectives that could be used to tag meanings.
- Using semantic types to identify patterns of systematic polysemy in the colour, temperature, and tactile adjective groups in corpora.
- Applying the semantic types and systematic polysemy patterns revealed as a result of the analysis in the Ekilex dictionary system of the Estonian Language Institute (see Tavast et al. 2018, Langemets et al. 2021) in the compilation of the EKI Combined Dictionary, (see Tavast et al. 2020) as lexicographical aids.
- Improving the knowledge of the most prevalent domains for meaning shifts in perception adjectives.
- Proposing parameters by which it is possible (based on the characteristics attributed to a prototypical adjective) to distinguish adjectives from other word classes in a corpus, with the aim of creating an automatic word classification tool for the lexicographer to assess cases where it is difficult to distinguish whether the word should be classified as an adjective.

Background and Method

One theoretical basis for the work is the distributional semantics hypothesis, according to which the meaning of a word can be derived from its context of use (Harris 1954, Firth 1957). The meaning of an adjective is especially contextual compared to other word classes and depends on the noun for which it is used as an attribute (e.g. *soe vesi* 'warm water', *soe naeratus* 'warm smile', *soe valgus* 'warm light'). All five articles of the thesis are corpus linguistic studies, where I examine actual language use and context, always using the largest available Estonian language corpus at the time of the study. For a more systemic overview of the context I use Sketch Engine (Kilgarriff et al. 2014, Kilgarriff et al. 2004) tools. The program allows to observe the collocations of a word based on different statisticians, to generate word sketches and study words with a similar context with the function of the thesaurus.

The distinctive character of the statistical models for distributional semantics remains limited: tests have been carried out in which two meanings are distinguished (Heylen et al. 2015, Hilpert and Saavedra 2020), but there is no reliable model or automatic solution for the analysis of more extensive polysemy. Thus, a combined method is used for this thesis: I use both sketches and a thesaurus to get a more systematic overview of the corpus context, but rely on qualitative analysis when analysing cases of use and condensing those into meanings.

In abstracting meanings and identifying polysemy patterns, I apply the semantic types (P1) of the Estonian adjective I have previously developed. When marking the noun-meanings, I use the semantic types applied to the analysis of the systematic polysemy of the Estonian noun by Margit Langemets (2010). In defining systematic polysemy, the dissertation follows the example of Langemets who used the definition by Yuri Apresjan (1974), according to whom a similar meaning structure must exist between at least two meanings of at least two polysemous words in order for the relationship between meanings to be determined as systematic polysemy.

The number of semantic types of the adjective can vary greatly, based on the approach used. In the study, I used the seven most important semantic types of Dixon (1982), the 18 types developed in the compilation of the lexical database SIMPLE (Lenci et al. 2000: 30–33) as an example and reference material, and GermaNet's 15 adjective semantic classes, which in turn have 78 subtypes, created on the basis of a study by Hundsnurher and Splett (1982). Analysing the meanings of the 100 most frequent adjectives in Estonian, I propose 15 semantic types specifically for describing the (polysemous) adjectives of the Estonian language. In later studies (P2, P3, P4) the types were updated and specified, e.g. the types VOICE, TASTE, SMELL and TOUCH were added. The developed types make it possible to tag meanings in the study of adjective polysemy, as well as allow for more accurate semantic tagging of adjective meanings in dictionary work.

The polysemy research method used in the thesis is one of a practical inclination and has grown out of the experience of lexicographical work. The main

limitation in the use of semantic types is to do with the different possibilities of aggregating meanings, semantic grouping or dividing them into semantic types. Any meaning that tries to encompass more than one specific communication event, is an abstraction. Semantic types, in turn, are abstractions that encompass these meanings. With this, the question arises of how much of the core meaning remains, the extent of shifts in meaning or the amount of semantic content lost. However, since the aim of this thesis is to find the polysemy rules for groups of colour, temperature and tactile adjectives, a generalisation is necessary.

Results and discussion

Since the semantic groups studied in the doctoral thesis are all groups of perception words, there is also expected to be more overlap in the meaning shifts with these words than one would find with words that are further away from each other in terms of subject field. It also affects the similarity of polysemy patterns that in general, temperature adjectives belong to the same area of perception as the tactile adjectives – both of which are related to tactile perception. All three groups had regular shifts of meaning with psychological characteristics (semantic type PSYCH). The most common shift to the psychological characteristic was from the tactile characteristic (43 words from the meaning structure). The pattern associated with psychological characteristics was also the most common in the group of temperature adjectives (24 words). In the group of colour words, the pattern was less present – 9 colour adjectives had a psychological meaning.

Both colour and temperature words are used to express the intensity of feelings (e.g., the oppression of hopelessness: *must mure* 'black worry', *must tulevik* 'black future', *mustad mõtted* 'black thoughts'; emphasizing passion: *palav armastus* 'hot love', *tuline viha* 'fiery anger', *põletav kirg* 'burning passion'. In both groups, shifts could be observed from pleasant tactile sensations to rather positive mental characteristics (e.g. *soe tuba* 'warm room', *soe vastuvõtt* 'warm reception'; *pehme padi* 'soft pillow', *pehme loomuga* 'soft-natured') and unpleasant alternation of touch experience with mentally unpleasant experience (e.g. *külm ilm* 'cold weather', *külm pilk* 'cold gaze'; *okkiline roos* 'thorny rose', *okkiline iseloom* 'thorny character'). However, there were also examples of meaning shifts in perceptual words that do not confirm the consistency of cognitive agreeableness or unpleasantness of a characteristic. A number of adjectives expressing tactile properties with the polysemy pattern TOUCH-PSYCH are not unambiguously categorised as positive or negative based on tactile experience (e.g. *karvane loom* 'furry animal', *karvane tunne* 'hairy feeling' (meaning 'unpleasant suspicion'); *tasane maapind* 'level ground', *tasase loomuga* 'level nature').

According to Lakoff and Johnson (1980), the formation and understanding of language is based on various bodily, affective, and perceptual processes. To express more abstract and complex phenomena, equivalences from more experientially familiar and body-based domains are used. The study of systematic polysemy points to a strong link between the domains of tactile touch and feelings. All that is directly experienced via the body/skin/mouth is extremely suitable as a starting

point for concepts to express more abstract phenomena such as psychological phenomena.

In the temperature group, the strong link was also illustrated by bilaterally intertwined semantics. Of the words with their first meaning in the domain of temperature, all the basic words for expressing temperature (e.g., *külm suhtumine* 'cold attitude', *soe inimene* 'warm person', *kuumad tunded* 'hot feelings', *jahedad suhted* 'cooled relationships') followed this pattern. Still, the sample included several words that are mainly used in a psychological meaning, that also showed a temperature-related meaning (*kõhe* 'chilling', *leebe* 'mild', *karm* 'harsh'). While in the temperature adjective group, the shift of meaning toward psychological characteristics occurs in both directions, in the tactile adjective group, the most prevalent meaning shift is one from a specific touchable characteristic toward a more abstract psychological one.

Newer studies allow for the presumption that the temperature-related attitudes in language are not universal, but depend also on the culture and environment of those using the language (e.g. Majid et al. 2018, Koptjevskaja-Tamm 2015). I studied the connection of basic Estonian words for temperature with either a positive or negative connotations with the thesaurus function of Sketch Engine that allows to examine words that have the most similar context in a corpus. As expected, for words that express cold weather, adjectives expressing a negative attitude emerged (e.g. *igav* 'boring', *kole* 'ugly'), whereas for warmth, there were adjectives linked to a positive attitude (e.g. *mõnus* 'nice', *ilus* 'beautiful', *kena* 'lovely', *meeldiv* 'pleasant'). As for coolness, the thesaurus offered no strong connections to attitude-related adjectives. According to the stronger links in the thesaurus function, the basic word *kuum* 'hot', that could refer to warmth that is already excessive and has common collocations with adjectives expressing a negative attitude, is still linked to only positive attitudes (e.g. *mõnus* 'nice', *kena* 'lovely'). Analysing the corpus, one will deduce that Estonians appreciate warmth, presuming that in Estonia as an area with a particularly cold climate, the link between warmth and positive emotions is especially strong. Still, a comparative study would be necessary to make this claim. Researcher of Estonian emotion terms, Ene Vainik (2016: 285) points out that the permanent connection between affection and warmth in different languages and cultures has been explained as an universal primary metaphor: it is a connection that we all make as infants, when we experience being physically held.

In other domains of perception, at least two groups exhibited systematic shifts of meaning with the domains of voice, taste and colour. For the domains of voice and sound, there were patterns of systematic polysemy in all three groups. The largest amount of systematic connections were found between tactile touch and voice; the pattern TOUCH–VOICE was present for 18 words out of the sample. In the colour group, analysis revealed a voice-related meaning in the meaning structure of 9 words; in the temperature group, in 6.

When it comes to voice, there were also meaning shifts between pleasant tactile sensations and pleasant voices, as noted in both temperature and tactile adjective groups. *Soe* 'warm', *pehme* 'soft', *siidine* 'silky', *sametine* 'velvety'

voice all refer to an enjoyable, pleasant voice; the same words are used to describe a sensation that is pleasant for the skin. In contrast, a voice that is *käre* 'harsh', *kalk* 'callous' ja *kare* 'rough' are not at all as pleasant to listen to, similar to how *käre talveilm* 'harsh winter weather', *kalk kivi* 'rough stone' or *kare materjal* 'rough material' are also rather unpleasant. In the colour group, the link between pleasantness and unpleasantness was less clear. The determining factor for this group was the intensity of the colour or voice, or to be more specific, lack of intensity. *Hõbedane* 'silver' ja *hele* 'bright' referred to a high and resonant voice, while for colours, both of these words also referred to a shade that radiates intense light and has a glossy sheen. *Akvarelne* 'watercolour, pastel' was used to refer to light, airy shades, as well as to soft, gentle sounds. *Matt* 'matte' and *tuhm* 'dull' were used to refer to a shineless, matte colour, as well as to a dull, toneless voice. As for taste, there were meaning shifts in the groups of temperature and tactile adjectives, but not in the colour adjective group.

Williams (1976) has pointed out that for synaesthetic adjectives, sound is the domain that most commonly uses words from other domains of perception, whereas the tactile domain is most often used as a source domain. Shen and Cohen (1998) emphasise that in the synaesthetic use of adjectives, the source domain is more specific than the target domain, which makes it more accessible; thus, the complement word is more commonly from the tactile domain, less commonly from the visual auditory domain. The results of studies of synaesthetic polysemy support the tendency observed when examining synaesthetic metaphors.

In other examined groups, shifts of meaning with colours occurred quite rarely, which means that in this case, the study of synaesthetic polysemy does not confirm Viberg's (1983) logic of the hierarchy of the senses, in which case the means of expression are borrowed from a sense that ranks higher in the hierarchy. The words of the temperature and touch group also had regular shifts of meaning with both the visual and auditory fields (*soe valgus* 'warm light', *jäine hääl* 'icy voice', *siidine hääl* 'silky voice'). To express perceptions of the higher category (vision-hearing), one borrows the words of a lower category, the field of touch.

In total, there were 9 polysemy patterns in the colour adjective group, 10 in the temperature adjective group, and 13 in the tactile adjective group. In at least two of the groups studied, there were systematic shifts of MATERIAL/SUBSTANCE, PHENOMENON_NATURE and noun meanings of type IN/ANIMAL_PROPERTY. Most polysemous patterns were formed by double, but also by triple connections: COLOUR -SOCIETY – IN_ROLL, TOUCH -SOUND-TASTE .

The presentation of polysemous patterns in the dictionary varied across all three groups. The pattern was represented as: 1) two distinct primary meanings; 2) primary meaning and submeaning(s); 3) a common aggregate meaning; 4) no systematic polysemy was presented. When there was a meaning that was not listed in the dictionary, but was used frequently enough in the corpus, I classified it as not having been presented. The analysis showed that the perception of lexicographers may differ in the perception of the meanings of the participants in the

same pattern; therefore, acquaintance with polysemy patterns associated with the respective semantic type could contribute to a more systematic representation.

Just as the semantics of the adjective are not always described in the same way, the adjective class itself is not evenly represented in different languages, nor is it always clearly defined: there are languages in which the adjectives are limited in number or there is no class of adjectives at all (Dixon 1982: 1–7). Not all semantic adjectives are grammatical adjectives in English. In Estonian, it is difficult to draw a line, for example, in the case of participles or non-declinable adjectives. The adjective-verb boundary (e.g. *põletav* 'burning', *torkiv* 'prickling', *jahtunud* 'cooled off'), the adjective-noun boundary (e.g. *must* 'black', *soe* 'warm', *kuiv* 'dry'), and the adjective-adverb boundary (e.g. *kullakarva* 'gold-tinted', *jahedavõitu* 'coolish', *kõvavõitu* 'somewhat firm') needed clarification when determining the adjective groups for the dissertation. The thesis explores the possibilities of language technology to facilitate deciding on word classes in more ambiguous cases, and how adjectives could be distinguished from other word types based on corpus data.

In a study of lexicographers (Paulsen, Vainik and Tuulik 2019), lexicographers pointed out examples of word classification problems, 26% of which were words related to the adjective class. As a main concern, the lexicographers cited the need to assess the degree of adjectivisation of participles (e.g. in words *muteerunud* 'mutated', *müüiv* 'selling', *kiusav* 'bullying'). The thesis (P5) proposes, on the basis of the attributes of the prototypical adjective, parameters for verifying the adjective's use in the corpus and for helping the lexicographer to decide whether to add the word to the dictionary/database as an adjective. The study is groundwork for creating a lexicographic application.

The parameters proposed as a result of the work are based on adjective characteristics from theoretical literature as well as practical lexicographic work. The attribute parameter [test_S] relies on the adjective's tendency to occur in the attribute function and measures the occurrences of the test word in front of the noun. The comparative form parameter measures the frequency of the comparative formed by the word in the corpus. The agreement parameter [test_S_ÜHILD] is based on the adjective's ability to match a noun in number and case, and verifies the concordance of the test word when followed by the noun. The sentence beginning parameter [LA_test_S] explains the extent to which a test word starts a sentence in a complement position in front of a noun. The adverb parameter [D_test] relies on the D+A corpus relation of adjectives (e.g. *liiga soe* 'too warm', *väga pehme* 'very soft') and measures how often the test word is preceded by an adverb in the corpus. The four-digit parameter [V_?_test_S] explains how often the test word occurs in the pattern V_?_test_S (the sequence comes from the analysis of the corpus integrated morphologically by hand, see P5 for more information).

The group of non-adjective participles was best distinguished from the group of prototypical adjectives by three parameters: the beginning-of-sentence parameter (with 90% accuracy), the adverb parameter (with 80% accuracy), and the comparative parameter (with 100% accuracy) (100% accuracy means that the

results of all the partitions studied were outside the adjective threshold set in the deviation analysis.) Thus, as part of the verb compound, the participle does not start a sentence in front of the noun, but an adjective-like participle may do so.

The prototypical adjective values exceeded all the other word groups studied (other word classes and groups borderlining adjectives), measured by the comparative form parameter and the adverb parameter. When considering the limit values of an adjective group (deviation analysis), the best distinctions between prototypical adjectives and other word classes were found to be the following: the beginning of a sentence parameter (distinguished ~78% of words), the comparative form parameter (distinguished 100% of other word classes), and the concordance parameter (distinguished ~59% of words, out of which nouns, adverbs, and verbs to 100%). The results give hope for creating an application that would help lexicographers assess the behaviour of the words that could be added to a dictionary as an adjective and to support them when deciding on word part-of-speech classification.

The thresholds of the adjective corpus profile could be further refined by focusing on the participle group, which causes problems for lexicographers. It would be possible to test different measures of similarity (Euclidean distance, cosine similarity, etc.) to find out which method would give the best result from the point of view of lexicographic work. It would also be possible to test new parameters characterizing the adjective word class which were not discussed in the article P5, for example the predicative parameter or derivation word-endings related to adjectives.

With the method applied in the doctoral thesis (using the corpus and adjective semantic types), it is also possible to study the polysemy of Estonian adjectives pertaining to the sense of taste, smell, and hearing. The results of the systematic polysemy analysis of Estonian perceptual adjectives will enable better future discussions of the relationship between cultural-specific and universal cognitive motives. The results of this doctoral thesis highlight the importance of universal human bodily experiences, though by comparing the results with polysemy studies of other languages, a more culturally oriented approach would also be possible in the future.

KIRJANDUS

- Aedmaa, Eleri 2019. Detecting Compositionality of Estonian Particle Verbs with Statistical and Linguistic Methods. Doktoritöö, Tartu Ülikool, Tartu Ülikool, Humanitaarteaduste ja kunstide valdkond.
- Apresjan, Jurij 1974. Regular Polysemy. – *Linguistics*, 142. Hague–Paris: Mouton & Co, 5–32.
- Atkins, B.T. Sue, Michael Rundell 2008. *The Oxford guide to practical lexicography*. Oxford: Oxford University Press.
- Berlin, Brent, Paul Kay 1969. *Basic Color Terms: Their Universality and Evolution*. Berkeley, Los Angeles: University of California Press, 1969.
- Boleda, Gemma, Sabine Schulte im Walde, Toni Badia 2012. Modeling Regular Polysemy: A Study on the Semantic Classification of Catalan Adjectives. – *Computational Linguistics*, Vol 38 (3), 575–616.
- Buitelaar, Paul 1998. *CoreLex: Systematic Polysemy and Underspecification*. Doktoritöö. Brandeis University, Computer Science.
- Church, Kenneth Ward, Patrick Hanks 1990. Word Association Norms, Mutual Information, and Lexicography, in *Computational Linguistics*, 16(1):22–29.
- Cruse, David Alan 1986. *Lexical Semantics*. Cambridge: Cambridge University Press.
- Cruse, David Alan 2000. Contextual variability. – *Meaning in Language*. Oxford: Oxford University Press.
- Cytowic Richard E. 2002. *Synesthesia: A Union of the Senses*. Second Edition. Cambridge, Massachusetts: MIT Press.
- Damasio Hanna, Thomas J. Grabowski, Daniel Tranel, Richard D. Hichwa, Antonio R. Damasio 1996. A neural basis for lexical retrieval. – *Nature*, 380: 499–505.
- de Vignemont, F., O. Massin 2015. Touch. – *The Oxford Handbook of philosophy of perception*. Oxford: Oxford University Press, 294–313.
- Dixon, Robert M.W. 1982. Where have all the adjectives gone? And other essays in syntax and semantics. – *Janua linguarum. Series Maior*, 107. Berlin: Walter de Gruyter.
- Dunning, Ted 1993. Accurate Methods for the Statistics of Surprise and Coincidence. – *Computational Linguistics* 19:1.
- EKG = Eesti keele grammatika II. Süntaks. 1993 Erelt, M., Kasik, R., Metslang, H., Rajandi, H., Ross, K., Saari, H., Tael, K., Vare, S. 1993. Tallinn: Eesti Teaduste Akadeemia Keele ja Kirjanduse Instituut.
- EKSS = Eesti keele seletav sõnaraamat 1–6. 2009. (“Eesti kirjakeele seletussõnaraamat” 2., täien- datud ja parandatud trükk.) [The Explanatory Dictionary of Estonian.] Toimetanud Margit Langemets, Mai Tiits, Tiia Valdre, Leidi Veskis, Ülle Viks, Piret Voll. Eesti Keele Instituut. Tallinn: Eesti Keele Sihtasutus.
- Erelt, Mati 1979. Predikatiivne adjektiiv (lausemallid). – *Ars grammatica*, Tallinn: Eesti Raamat.
- Erelt, Mati 1986. Eesti adjektiivisüntaks. Tallinn: Valgus.
- Erelt, Mati 2017. Sissejuhatus süntaksisse. – Eesti keele süntaks. Eesti keele varamu III. Tartu: Tartu Ülikooli Kirjastus, 537–564.
- Fillmore, Charles J., Beryl Ts Atkins 200. Describing polysemy: the case of ‘crawl’. – *Polysemy: Theoretical and Computational Approaches*. Oxford: Oxford University Press, 91–110.
- Firth, John Rupert 1957. *A Synopsis of Linguistic Theory, 1930–1955*. – *Studies in linguistic analysis*. Oxford: Blackwell.

- Frawley, William 1992. *Linguistic Semantics*. Hillsdale, NJ, Hove and London: Lawrence Erlbaum
- Gibson, Edward, Richard Landley Jones Futrell, Julia Jara-Ettinger, Kyle Adam Mahowald 2017. Color naming across languages reflects color use. – *Proceedings of the National Academy of Sciences* 114(40):201619666.
- Goddard, Cliff 2012. Semantic Primes, Semantic Molecules, Semantic Templates: Key Concepts in the NSM Approach to Lexical Typology. – *Linguistics*, 50(3), 711–743.
- Halliday, M.A.K., Colin Yallop 2007. *Lexicology: A Short Introduction*. London Continuum International Publishing Group
- Hanks, Patrick 2000. Do Word Meanings Exist?. – *Computers and the Humanities*, 34(1/2), 205–215.
- Hanks, Patrick 2017. Mechanisms of Meaning. – *International Conference on Computational and Corpus-Based Phraseology*.
- Harris, Zellig 1954. Distributional Structure. – *Word*, 10(2–3), 146–162.
- Heylen, Kris, Thomas Wierbaert, Dirk Speelman, Dirk Geeraerts 2015. Monitoring polysemy: Word space models as a tool for large-scale lexical semantic analysis. – *Lingua*, 157, 153–172.
- Hilpert, Martin, David Correia Saavedra 2020. Using token-based semantic vector spaces for corpus-linguistic analyses: From practical applications to tests of theoretical claims. – *Corpus Linguistics and Linguistic Theory*, 16(2), 393–424.
- Hundsnurher, Franz, Jochen Splett 1982. *Semantik der Adjektive des Deutschen: Analyse der semantischen Relationen*. Westdeutscher Verlag.
- Huth, Alexander G., Wendy A. De Heer, Thomas L. Griffiths, Frédéric E. Theunissen, Jack L. Gallant 2016. Natural speech reveals the semantic maps that tile human cerebral cortex. – *Nature*, 532, 453–458.
- Johnson, Mark. 1987. *The body in the mind: The bodily basis of meaning, imagination, and reason*. Chicago: University of Chicago Press.
- Josserand, Mathilde, Emma Meeussen, Asifa Majid, Dan Dediu 2021. Environment and culture shape both the colour lexicon and the genetics of colour perception. – *Scientific Reports* 11, 19095. <https://doi.org/10.1038/s41598-021-98550-3>
- Juvonen, Päivi, Ahti Nikunlassi 2015. Temperature adjectives in Finnish. – *The Linguistics of Temperature. Typological Studies in Language* 107. Amsterdam: John Benjamins, 491–536.
- Jürviste, Madis, Jelena Kallas, Margit Langemets, Maria Tuulik, Ülle Viks 2011. Extending the functions of the EELex dictionary writing system using the example of the Basic Estonian Dictionary. *Electronic Lexicography in the 21st Century New Applications for New Users. Proceedings of the eLex 2011 Conference*, 10–12 November 2011, Bled, Slovenia. Ljubljana: Trojina, Institute for Applied Slovenian Studies, 106–112.
- Kaalep, Heiki-Jaan; Vaino, Tarmo 2001. Complete morphological analysis in the linguist's toolbox. – *Congressus Nonus Internationalis Fenno-Ugristarum. Pars V. Dissertationes sectionum: linguistica. II*. Tartu, 9–16.
- Kalda, Anu 2020. Translating perception metaphors. A process-based approach. – *Philologia Estonica Tallinnensis*, 5, 109–133. <https://doi.org/10.22601/PET.2020.05.04>
- Kalda, Anu 2021. Translating perception metaphors: Linguistic, cultural and social implications. – *Taikomoji Kalbotyra*, 16, 86–109. <https://doi.org/10.15388/Taikalbot.2021.16.6>
- Kallas, Jelena, Maria Tuulik 2011. Eesti keele põhisõnavara sõnastik: ajalooline kontekst ja koostamispõhimõtted. – *Eesti Rakenduslingvistika Ühingu aastaraamat*, 7, 59–75.

- Kallas, Jelena, Maria Tuulik, Madis Jürviste 2012. Leksikograafilise tarkvara Sketch Engine eesti keele moodul. – Eesti ja soome-ugri keeleteaduse ajakiri 3–2, 57–77.
- Kallas, Jelena 2013. Eesti keele sisusõnade süntagmaatilised suhted korpus- ja õppeleksikograafias. Doktoritöö, Tallinna Ülikool.
- Kallas, Jelena, Kristina Koppel, Maria Tuulik 2015. Korpusleksikograafia uued võimalused eesti keele kollokatsioonisõnastiku näitel. – Eesti Rakenduslingvistika Ühingu aastaraamat, 11, 75–94.
- Kallas, Jelena, Kristina Koppel, Geda Paulsen, Maria Tuulik 2019. Eesti keele naaber sõnad 2019. Sõnaveeb. DOI: 10.15155/3-00-0000-0000-0000-0823EL.
- Kasik, Reet 2015. Sõnamoodustus. Tartu: Tartu Ülikooli Kirjastus.
- Katz, S. 1987. Estimation of probabilities from sparse data for the language model component of a speech recognizer. – IEEE Transactions on Acoustics, Speech, and Signal Processing, 35(3), 400–401.
- Kay, Paul, Brent Berlin, Luisa Maffi, William R. Merrifield, Richard Cook 2009. The World Color Survey. Stanford: CSLI.
- Kerge, Krista 2000. Eesti süntaks võõrkeeleõppe praktikule: käsiraamat. Tallinn: TEA Kirjastus.
- Kilgarriff, Adam 1992. Polysemy. Doktoritöö, University of Sussex, UK.
- Kilgarriff, Adam, Gergory Grefenstette 2003. Web as Corpus. – Computational Linguistics 29(3): 333–347.
- Kilgarriff, Adam; Rychly, Pavel; Smrz, Pavel; Tugwell, David 2004. The Sketch Engine. – Proceedings of the 11th EURALEX International Congress (Lorient, 6–10 July 2004). Lorient: Université de Bretagne-Sud, 105–117.
- Kilgarriff, Adam, Vit Baisa, Jan Bušta, Miloš Jakubíček, Vojtěch Kovář, Jan Michelfeit, Pavel Rychlý, Vit Suchomel 2014. The Sketch Engine: ten years on. – Lexicography 1(1), 7–36.
- Koeva, Svetla, Cvetana Dimitrova, Valentina Stefanova, Dimitar Hristov 2018. Mapping WordNet Concepts with CPA Ontology. – Proceedings of the 9th Global Wordnet Conference, 69–76.
- Koppel, Kristina, Arvi Tavast, Margit Langemets, Jelena Kallas 2019. Aggregating dictionaries into the language portal Sõnaveeb: issues with and without a solution. Proceedings of the eLex 2019 Conference, 1–3 October 2019, Sintra, Portugal. Brno: Lexical Computing CZ, s.r.o., 434–452.
- Koppel, Kristina 2020. Näitelausete korpuspõhine automaattuvastus eesti keele õppe sõnastikele. Doktoritöö, Tartu Ülikool, Humanitaarteaduste ja kunstide valdkond.
- Koppel, Kristina, Jelena Kallas 2022. Eesti keele ühendkorpuste sari 2013–2021: mahukaim eestikeelsete digitekstide kogu. – Eesti Rakenduslingvistika Ühingu aastaraamat, 18, 207–228.
- Koptjevskaja-Tamm, Maria 2015. Introducing “The Linguistics of Temperature”. – Maria Koptjevskaja-Tamm (Ed.), *Typological Studies in Language* 107, 1–40.
- Kosem, I., R. Lew, C. Müller-Spitzer, M.R. Silveira, S. Wolfer, A. Dorn, A. Gurrutxaga, K. Ceberio, E. Etxeberria, M.A. Lefer, D. Geeraerts, K.Š. Despot, T. Stojanov, N. Ljubešić, M. Škrabal, B. Štěpánková, V. Vodrážková, H. Lorentzen, L. Trap-Jensen, J. Kallas, M. Tuulik, K. Koppel, M. Langemets, T. Heinonen, I. Thomas, T. Margilitadze, S. Markantonatou, V. Giouli, C. Mulhall, I. Kernerman, Y. Ben-Moshe, T. Sadan, A. Abel, M.N. Curcio, L. Tanturovska, B. Nikovska, C. Tiberius, O. Grønvik, M. Hovdenak, S. Berg-Olsen, K.E. Karlsen, C.E.S. Ore, M. Biesaga, T.Z. Kuhn, J. Silvestre, E.I. Tamba, G. Haja, M.R. Clim, M.I. Patrascu, T. Tasovac, S. Petrović, Š.A. Holdt, C.V. Riveiro, M.J.D. Vázquez, E. Volodina, I.

- Pilán, E. Sköldberg, L. Holmer, H. Nesi 2019. The Image of the Monolingual Dictionary Across Europe. Results of the European Survey of Dictionary use and Culture. – *International Journal of Lexicography*, 32 (1), 92–114.
- Lakoff, George, Mark Johnson 1980. *Metaphors We Live By*. Chicago: University of Chicago Press.
- Langacker, Ronald W. 1987. *Foundations of Cognitive Grammar: Theoretical Prerequisites*, Volume 1. Stanford: Stanford University Press.
- Langemets, Margit 2008. „Eesti kirjakeele seletussõnaraamatu“ viiskümmend aastat. – *Emakeele Seltsi aastaraamat*, 181–199.
- Langemets, Margit 2009. *Nimisõna süstemaatiline polüseemia eesti keeles ja selle esitus keelevaras*. Doktoritöö, Tallinna Ülikool.
- Langemets, Margit 2010. *Nimisõna süstemaatiline polüseemia eesti keeles ja selle esitus keelevaras*. Tallinn: Eesti Keele Sihtasutus.
- Langemets, Margit, Mai Tiits, Tiia Valdre, Piret Voll 2010. In spe: ühekõiteline eesti keele sõnaraamat. – *Keel ja Kirjandus*, 11, 793–810.
- Langemets, Margit, Udo Uibo, Mai Tiits, Tiia Valdre, Piret Voll 2018. Eesti keel uues kuues. *Eesti keele sõnaraamat 2018*. – *Keel ja Kirjandus*, 12, 942–958.
- Langemets, Margit, Kristina Koppel, Jelena Kallas, Arvi Tavast 2021. Sõnastikukogust keeleportaaliks. – *Keel ja Kirjandus*, 8–9, 755–770. <https://doi.org/10.54013/kk764a6>
- Langemets jt 2022: Langemets, Margit; Hein, Indrek; Jürviste, Madis; Kallas, Jelena; Kiisla, Olga; Koppel, Kristina; Leemets, Tiina; Mäearu, Sirje; Paet, Tiina; Päll, Peeter; Raadik, Maire; Risberg, Lydia; Tammik, Hanna; Tavast, Arvi; Tiits, Mai; Tsepelina, Katrin; Tuulik, Maria; Valdre, Tiia; Viks, Ülle; Sai, Edgar; Smirnova, Alina; Tubin, Valentina (2022). *EKI ühendsõnastik 2022*. <https://doi.org/10.15155/3-00-0000-0000-08C0AL>.
- Lenci, Alessandro, Nuria Bel, Federica Busa, Nicoletta Calzolari, Elisabetta Gola, Monica Mona-chini, Antoine Ogonowski, Yvonne Peters, Wim Peters, Nilda Ruimy, Marta Villegas, Antonio Zampolli 2000. SIMPLE: A General Framework for the Development of Multilingual Lexicons. – *International Journal of Lexicography*, 13(4), 249–263.
- Lerner Yulia, Christopher J. Honey, Lauren J. Silbert, Uri Hasson 2011. Topographic mapping of a hierarchy of temporal receptive windows using a narrated story. – *J Neurosci*. 31(8):2906–15.
- Lievers, Francesca Strik 2016. Synaesthetic metaphors in translation. – *Linguistic Studies and Essays*, 54(1), 43–70.
- Lopukhina, Anastasiya, Anna Laurinavichyute, Konstantin Lopukhin, Olga Dragoy 2018. The mental representation of polysemy across word classes. – *Frontiers in psychology* 9, 192
- Majid, A, S. Roberts, L. Cilissen, K. Emmorey, B. Nicodemus, L. O’Grady, B. Woll, B. LeLan, H. de Sousa, B. Cansler, S. Shayan, C. de Vos, G. Senft, N. Enfield, R. Razak, S. Fedden, S. Tufvesson, M. Dingemanse, O. Ozturk, P. Brown, C. Hill, O. Le Guen, V. Hirtzel, R. van Gijn, M. Sicoli, S. Levinson 2018. Differential coding of perception in the world's languages. – *Proceedings of the National Academy of Sciences of the United States of America*, 115(45), 11369–11376.
- Meer, Geart van der 2006. It’s about Time: On Coherence and Simplicity in Dictionary Entries. – *English Studies* 87(5), 602–616.
- Metslang, Helle, Mati Ereht, Tiit Hennoste, Reet Kasik, Pire Teras, Annika Viht, Eva Liina Asu, Külli Habicht, Liina Lindström, Pärtel Lippus, Renate Pajusalu, Helen

- Plado, Andriela Rääbis, Ann Veismann (ilmumas). Eesti grammatika. Tartu: Tartu Ülikooli kirjastus.
- Jakubiček, Miloš 2021. Morphology is an open problem of NLP. Ettekanne sõnaliikide töötoas „Quantitative answers to qualitative questions? The challenge of ambiguity in corpus data“. Eesti Keele Instituut. Tallinn (<https://www.youtube.com/watch?v=Fluqe7FZiBo&list=PLfDCKnTgUT4GJffKuCxqW8lc3lwSGxOD1&index=3>)
- Muischnek, Kadri, Heiki-Jaan Kaalep, Raul Sirel 2011. Korpuslingvistiline lähenemine eesti internetikeele automaatsele morfoloogilisele analüüsile. – Eesti Rakenduslingvistika Ühingu aastaraamat, 111–127. Tallinn: Eesti Rakenduslingvistika Ühing. <https://doi.org/10.5128/ERYa7.07>
- Oakes, Michael P 1998. Statistics for Corpus Linguistics. Edinburgh: Edinburgh University Press.
- Oja, Vilja 2001. Värvuste tähistamisest eesti murretes. Doktoritöö, Tartu Ülikool.
- Orav, Heili 2006. Isiksuseomaduste sõnavara semantika eesti keeles. Doktoritöö. Tartu Ülikool.
- Paulsen, Geda, Ene Vainik, Maria Tuulik, Ahti Lohk 2019. The Lexicographer’s Voice: Word Classes in the Digital Era. Electronic lexicography in the 21st century. Proceedings of the eLex 2019 Conference, 1–3 October 2019, Sintra, Portugal. Brno: Lexical Computing CZ, s.r.o., 319–337.
- Paulsen, Geda; Ene Vainik, Maria Tuulik 2020. Sõnaliik leksikograafi tööalual: sõnaliikide roll tänapäeva leksikograafias. – Eesti Rakenduslingvistika Ühingu aastaraamat, 16, 177–202. <https://doi.org/10.5128/ERYa16.11>
- Pedersen, Ted 1998. Proceedings of the Fifteenth National Conference on Artificial Intelligence and Tenth Innovative Applications of Artificial Intelligence Conference, Madison, Wisconsin: AAAI Press / The MIT Press.
- Pikksaar, Aimi 2022. Empiirilised andmed semantikas: kui juhuslik peaks olema polüseemia uurimisel kasutatav korpusvalim? Ettekanne. Eesti Rakenduslingvistika Ühingu konverents 16.–17.juuni 2022, Tallinn. Slaidid veebiaadressil: https://www.rakenduslingvistika.ee/wp-content/uploads/2022/07/Aimi_Pikksaar_Empiirilised-andmed-semantikas.pdf
- Proos, Mariann 2019. Polysemy of the Estonian perception verb „nägema“ ‘to see’. In: Laura J. Speed, Carolyn O’Meara, Lila San Roque, Asifa Majid (Ed.). Perception Metaphors (231–252). John Benjamins. <https://doi.org/10.1075/celcr.19.12pro>.
- Proos, Mariann 2020a. Usage patterns of Estonian experience perception verbs: a random forest approach. SKY Journal of Linguistics, 33, 67–101.
- Proos, Mariann 2020b. Feeling your neighbour: an experimental approach to the polysemy of tundma ‘to feel’ in Estonian. Language and Cognition, 12 (2), 282–309. <https://doi.org/10.1017/langcog.2019.44>.
- Proos, Mariann 2021. Meaning and usage of Estonian Experience Perception. Doktoritöö. Tartu Ülikool, Humanitaarteaduste ja kunstide valdkond.
- PSV = Eesti keele põhisõnavara sõnastik 2014. Kallas, Jelena, Mai Tiits, Maria Tuulik, Kristina Koppel, Madis Jürviste. Eesti Keele Instituut. Tallinn: Eesti Keele Sihtasutus
- Pustejovsky, James 1995. The Generative Lexicon. – Language, Speech, and Communication.
- Rakova, Marina 2003. The Extent of the Literal: Metaphor, Polysemy and Theories of Concepts. New York: Palgrave Macmillan.
- Rieger, Burghard B. 1992. On Distributed Representations in Word Semantics. – ICSI Berkeley 12–1991.

- Rätsep, Kaidi 2018. Colour terms in Turkish, Estonian and Russian: how many basic blue terms are there?. Doktoritöö, Tartu Ülikool.
- Rychlý, Pavel, Adam Kilgarriff 2007. An efficient algorithm for building a distributional thesaurus (and other Sketch Engine developments). – Proceedings of the 45th Annual Meeting of the ACL on Interactive Poster and Demonstration Sessions, Prague: Association for Computational Linguistics, 41–44.
- Rychlý, Pavel 2008. A Lexicographer-Friendly Association Score. – Proceedings of 2nd workshop on recent advances in Slavonic natural languages RASLAN 2008. Brno: Masaryk University, 6–9.
- Sahlgren, Magnus 2006. The Word-Space Model. Doctoral thesis. Stockholm University.
- Shen, Yeshayahu, Michal Cohen 1998. How come silence is sweet but sweetness is not silent: a cognitive account of directionality in Poetic synaesthesia. – Language and Literature, Volume 7, Issue 2, 123–140.
- Sorabji, Richard 1971. Aristotle on Demarcating the Five Senses. – The Philosophical Review, 80(1), 55–79.
- Sutrop, Urmas 1995. Eesti keele põhivärvinimed. Keel ja Kirjandus, 12, 797–808.
- Sutrop, Urmas 1998. The Vocabulary of Sense Perception in Estonian: Typology, History, and Cognition. Doctoral Thesis. University of Konstanz, Germany.
- Sutrop, Urmas 1999. Temperature terms in the Baltic area. Mati Ereit. – Estonian: Typological Studies, 3, 185–203. Tartu: Tartu Ülikool. (Tartu Ülikooli eesti keele õppetooli toimetised; 11).
- Sutrop, Urmas 2002. The Vocabulary of Sense Perception in Estonian: Structure and History. Frankfurt am Main: Peter Lang
- Sutrop, Urmas 2004. Temperatuurisõnad ja skaalad [‘Temperature terms and scales’]. – Emakeele Seltsi aastaraamat, 50, 111–128.
- Tavast, Arvi, Margit Langemets, Jelena Kallas, Kristina Koppel 2018. Unified Data Modelling for Presenting Lexical Data: The Case of EKILEX. – Proceedings of the XVIII EURALEX International Congress: Lexicography in Global Contexts, Ljubljana, 17–21 July 2018. Ljubljana University Press, Faculty of Arts, 749–761.
- Tavast, Arvi, Kristina Koppel, Margit Langemets, Jelena Kallas, Jelena 2020. Towards the superdictionary: layers, tools and unidirectional meaning relations. Proceedings of the XIX EURALEX Congress: Lexicography for Inclusion, Vol. I. Ed. Gavriilidou, Z, Mitsiaki, M, Fliatouras, A. Alexandroupolis, Greece: Democritus University of Thrace, 215–223.
- Traugott, Elizabeth Closs, Richard B. Dasher 2001. Regularity in Semantic Change. Cambridge: Cambridge University Press.
- Trier, Jost 1931. Der deutsche Wortschatz im Sinnbezirk des Verstandes, Doktoritöö, Bonn.
- Ullmann, Stephen 1957. The Principles Of Semantics. A Linguistic Approach To Meaning. Oxford: Basil Blackwell.
- Uusküla, Mari 2008. The basic colour terms of Czech. – Trames Journal of the Humanities and Social Sciences, 12 (1), 3–28.
- Uusküla, Mari 2011. Terms for red in Central Europe: An areal phenomenon in Hungarian and Czech. – New Directions in Colour Studies, Amsterdam: John Benjamins, 147–156.
- Uusküla, Mari, Liivi Hollmann, Urmas Sutrop 2012. Basic colour terms in five Finno-Ugric languages and Estonian Sign Language: A comparative study. – Journal of Estonian and Finno-Ugric Linguistics, 3 (1), 47–86.
- Uusküla, Mari, Urmas Sutrop 2011. Värvinimede raamat. Tallinn: Eesti Keele Sihtasutus.
- Uusküla, Mari 2014. Linguistic categorisation of blue in Standard Italian. – Colour Studies: A Broad Spectrum, Amsterdam/Philadelphia: John Benjamins. 67–78.

- Uusküla, Mari 2018. Loetelukatse ja nimeandmiskatse psühholingvistika meetodina. Elised ja kitsaskohad. – Keel ja Kirjandus, 8–9, 683–696. <https://doi.org/10.54013/kk730a5>
- Uusküla, Mari 2019. Translation of Colour Terms: An Empirical Approach toward word-translation from English into Estonian. – Journal of Estonian and Finno-Ugric Linguistics, 10 (2), 69–84. <https://doi.org/10.12697/jeful.2019.10.2.05>
- Uusküla, Mari, David L. Bimler 2020. When does ‘bright’ mean ‘prototypical’? Color term modifiers in eight European languages, examined with color-survey data. – Journal of the Optical Society of America. A, Optics, image science, and vision, 37 (5), A305–A312.
- Vainik, Ene, 2004. Lexical knowledge of emotions: the structure, variability and semantics of the Estonian emotion vocabulary. Doktoritöö, Tartu Ülikool, Filosoofia-teaduskond.
- Vainik, Ene 2017. Tunnetest, piltlikult. – Mäetagused, 69, 9–38. <https://doi.org/10.7592/MT2017.69.vainik>
- Vainik, Ene 2018. Emotion meets taste. Taste-motivated emotion terms in Estonian. – Folklore: Electronic Journal of Folklore, 71 (1), 129–155. <https://doi.org/10.7592/FEJF2018.71.vainik>
- Vainik, Ene, Geda Paulsen, Ahti Lohk 2020. A typology of lexical ambiforms in Estonian. Proceedings of the XIX EURALEX Congress: Lexicography for Inclusion, Vol. 1. Ed. Gavriilidou, Z.; Mitsiaki, M.; Fliatouras, A. Alexandroupolis, Greece: Democritus University of Thrace, 119–130.
- Vare, Silvi 2006. Adjektiivide substantivatsioonist ühe tähendusrühma näitel. – Keele ehe. Tartu: Tartu Ülikool, 30, 205–222.
- Vare, Silvi 2012. Eesti keele sõnapered. – Tänapäeva eesti keele sõnavara struktuuri-analüüs I–II. Tallinn: Eesti Keele Sihtasutus.
- Viberg, Åke 1983. The verbs of perception: A typological study. Explanations for Language Universals. Toim. Butterworth, B; Comrie, B; Dahl, Ö Berlin: Mouton de Gruyter, 123–162.
- Viks, Ülle 1977. Sõnaliik kui niisugune. – Keel ja Kirjandus 9, 521–525.
- Vliet, Hennie van der 2007. The Referentiebestand Nederlands as a multi-purpose lexical database. – International Journal of Lexicography, 3, Vol. 20, 239–258.
- Wierzbicka, Anna 2008. Why There Are No ‘Colour Universals’ in Language and Thought. – The Journal of the Royal Anthropological Institute, Vol. 14, No 2, 407–425
- Williams, Joseph M. 1976. Synaesthetic adjectives: A Possible Law of Semantic Change. – Language, Volume 52, Issue 2, 461–478.
- Winter, Bodo 2019. Sensory linguistics: Language, perception and metaphor. Amsterdam: John Benjamins, xiv, 289.
- Yu, Ning 2003. Synesthetic metaphor: A cognitive perspective. – Journal of Literary Semantics, 32, 19–34.
- Zhang, Xiuming 2015. Molecular sensors and modulators of thermoreception. – Channels, Volume 9, Issue 2, 73–81.
- Zipf, George Kingsley 1935. The Psycho-Biology of Language. Cambridge (Max Bunker).
- Zipf, George Kingsley 1945. The meaning-frequency relationship of words. – Journal of General Psychology, 3, 251–256. <http://dx.doi.org/10.1080/00221309.1945.10544509>
- Zipf, George Kingsley 1949. Human behavior and the principle of least effort. Addison-Wesley Press.

Võrguviited

- Eesti keele koondkorpus. cl.ut.ee/korpused/segakorpus/index.php?lang=et (20.06.2022).
- Eesti keele ühendkorpus 2017. <https://metashare.ut.ee/repository/browse/eesti-keele-uhendkorpus-2017/b616ceda30ce11e8a6e4005056b40024880158b577154c01bd3d3fcfc9b762b3/> (20.06. 2022).
- Eesti keele ühendkorpus 2019. <https://metashare.ut.ee/repository/browse/estonian-national-corporus-2019/cd9633fab22e11eaa6e4fa163e9d4547b71a2df64d1f43f1ac26dbd8508ea951/> (20.06.2022).
- Ekilex: Eesti Keele Instituudi sõnastiku- ja terminibaasisüsteem. <https://doi.org/10.15155/3-00-0000-0000-0000-0823CL> (20.06. 2022).
- estNLTK: Open source tools for Estonian natural language processing. <https://estnlk.github.io/estnlk/1.2/index.html#> (20.06.2022).
- PSV andmebaas Eelexi sõnastikubaasis: <https://eelex.eki.ee/> (20.06.2022).
- PSV=Eesti keele põhisõnavara sõnastik (elektrooniline versioon) <http://www.eki.ee/dict/sp/index.cgi> (20.06.2022).
- Sketch Engine. <https://www.sketchengine.eu/> (20.06.2022).
- Sketch Statistics. <https://www.sketchengine.eu/documentation/statistics-used-in-sketch-engine/> (20.06. 2022).
- Sõnaveeb. <https://sonaveeb.ee/> (20.06.2022).

LISA 1: Eesti adjektiivide semantilised tüübid

Semantiline tüüp	Kirjeldus	Näited valimist
OMADUS_FÜÜS	füüsilised omadused, mis ei kuulu kitsamatesse tüüpidesse	<i>tugev, nõrk, puhas</i>
OMADUS_MÕÕDE	ruumilised omadused	<i>pikk, lai, lühike, kõrge</i>
OMADUS_LIIKUMINE	liikumise ja tempoga seotud omadused	<i>kiire, aeglane</i>
OMADUS_VÄRV	värvusega seotud omadused	<i>must, valge, punane, roheline</i>
OMADUS_MASS	massiga seotud omadused	<i>raske, kerge</i>
OMADUS_TEMPERAATUUR	temperatuuriga seotud omadused	<i>külm, soe</i>
OMADUS_AEG	ajaga seotud omadused	<i>eilne, tänane, käesolev, praegune, endine</i>
OMADUS_KOHT	mingi kohaga seotud omadused	<i>sealne, kohalik</i>
OMADUS_VANUS	kellegi või millegi eaga, olemasolu ajaga seotud omadused	<i>vana (ka tähenduses 'kulunud'), noor, uus, värske</i>
OMADUS_KOGUS	hulgaga seotud omadused	<i>suur (tähenduses 'arvukas'), vähene</i>
OMADUS_ÜHISKOND	sotsiaalsete suhetega seotud omadused	<i>poliitiline, rahvuslik, sotsiaalne, avalik, riiklik</i>
OMADUS_RÕHUTAMINE	intensiivsust, suurust või väiksust rõhutavad omadused	<i>täielik, viimane (nt käitub nagu viimane narr)</i>
OMADUS_HINNANG	hinnangu ja sealhulgas modaalsusega seotud omadused	<i>õige, huvitav, ohtlik, sobiv, halb, võimalik</i>
OMADUS_PSÜHH	meeleolu, iseloomu ja vaimsete võimetega seotud omadused	<i>tõsine, tark*, rõõmus*</i>
OMADUS_SUHTELISUS	omadused, mida defineerib võrdlus või suhe millegi muuga	<i>sarnane, keskmine, viimane</i>

LISA 2: Värviajeksiivide polüseemiamallid

Mall	Sõnade arv	Näited valimist	Malli kirjeldus
OMADUS_ VÄRV- VÄRV (A+S)	Võimalik kõigi värvi-sõnade puhul	<i>kollane, roheline, punane, sinine</i>	Värviajeksiiv tähistab ka sama värvi substantiivset kasutust, näiteks „kaskedes on kollast“, „ta kannab punast“.
VÄRV- MATERJAL/AINE (S+S)	20	<i>purpur, ultramariin, vask, mahagon, neon, rooste</i>	Sõna tähistab nii teatud materjali või ainet (noun) kui ka sellega seostatavat värvitooni, näiteks <i>vask</i> kui metall ja ka kui punakas värv.
OMADUS_ VÄRV- OMADUS_ MATERJAL/AINE (A+A)	16	<i>pigine, tindine, vahane, süsine, pronksine</i>	Ajeksiiv väljendab mingi aine või materjaliga kaetud olemist/ koos olemist või sellest tehtud olemist ja ka vastavat värvi. Näiteks <i>süsine</i> kui „sõega kaetud, söestunud“ ja „täiesti must, väga tume“.
OMADUS_ VÄRV – OMADUS_ HINNANG (A+A)	13	<i>hall, kahvatu, värvitu, kuldne, roosiline, veretu</i>	Värvitähendus vaheldub hinnanguga seotud omadusega. Näiteks <i>hall</i> kui „igav“ ja <i>kahvatu</i> kui „mitte muljet avaldav“.
OMADUS_ VÄRV – INIMENE_ OMADUS (A+S)	11	<i>blond, šatään, brünett, must, valge, kollane, punane, värviline</i>	Sõna tähistab nii värvi kui ka teatud omadusega inimest (noun). Näiteks juuksevärve tähistavate sõnade puhul nagu <i>blond, šatään</i> ja <i>brünett</i> tähistab sama sõna ka vastava juuksevärviga inimest.
OMADUS_ VÄRV – LOOM_ OMADUS (A+S)	9	<i>kõrb, raudjas, hiir, võik, kimmel</i>	Sõna tähistab nii teatud värvitooni kui ka vastavat värvi looma (noun), näiteks <i>kõrb</i> kui „punakaspruun“ ja ka kui „punakaspruun hobune“.
OMADUS_ VÄRV – OMADUS_ PSÜHH (A+A)	9	<i>must, süngmust, roosa, terasene, roheline, erk</i>	Värvitähendus vaheldub meeleolu, iseloomu või vaimsete võimetega seotud omadusega. Näiteks <i>must</i> kui „masendav“ ja <i>roosa</i> kui „rõõmus/muretu“.
OMADUS_ VÄRV – OMADUS_ HÄÄL (A+A)	9	<i>hele, tuhm, hõbedane, akvarelne, matt</i>	Värvitähendus vaheldub hääle või heli omadusega, näiteks <i>hõbedane</i> kui „kõrge, heledalt kõlav“.
OMADUS_ VÄRV – OMADUS_ ÜHISKOND – IN_ ROLL (A+A+S)	6	<i>punane, roheline, valge, roosa, punaroheline, veripunane</i>	Poliitilise liikumise või ühiskondliku mõtlemise suunaga seotakse mingi värv, mis hakkab tähistama ka seda liikumist/suunda ennast ja ühtlasi ka vastavat mõtteviisi pooldavat inimest. Näiteks <i>punane</i> kui kommunistlik ja <i>roheline</i> kui ökoloogilist mõtteviisi pooldav inimene.

LISA 3. Temperatuuradjektiivide polüseemiamallid

Mall	Sõnu valimis	Näited	Malli kirjeldus
TEMPERATUUR- PSÜHH	24	<i>külm, soe, jääne, leige, kuum</i>	Temperatuuritähendus vaheldub meeleolu, suhtumist või iseloomu väljendava omadusega
TEMPERATUUR- NÄHTUS_FÜÜS	14	<i>vilu, toasoe, palav, suvesoe</i>	Adjektiivne temperatuuritähendus vaheldub substantiivitähendusega, mis tähistab vastavat füüsikalist nähtust
TEMPERATUUR- KOOSNEMINE	6	<i>jääne, krõbe, pehmeke, kalk</i>	Temperatuuritähendus vaheldub millegi koostist, ehitust väljendava tähendusega
TEMPERATUUR- HÄÄL	6	<i>käre, kalk, soe, sume</i>	Temperatuuritähendus vaheldub heli, häält iseloomustava tähendusega
TEMPERATUUR- VÄRVUS	6	<i>külm, soe, sume, kalk, mahe</i>	Temperatuuritähendus vaheldub värvust või valgust iseloomustava tähendusega
KOHT- TEMPERATUUR	5	<i>arktiline, boreaalne, atlantiline</i>	Adjektiiv, mis tähistab teatud geograafilisest piirkonnast pärinemist või sinna kuulumist, on hakanud tähistama ka vastava piirkonnaga seostuvaid kliimaolusid ja temperatuuri
TEMPERATUUR- LIIKUMINE	5	<i>palavikuline, palav, käre</i>	Temperatuuritähendus vaheldub liikumist ja tempot iseloomustava tähendusega
TEMPERATUUR- PUUDUTUS	4	<i>pehme, karm</i>	Temperatuuritähendus vaheldub tähendusega, mis iseloomustab puudutamisel tekkivat aistingut
TEMPERATUUR- MAITSE	3	<i>tuline, mahe, pehme</i>	Temperatuuritähendus vaheldub maitset iseloomustava tähendusega
TEMPERATUUR- VANUS	2	<i>ahjusoe, soe</i>	Temperatuuritähendus vaheldub millegi või kellegi iga või olemasolu aega iseloomustava tähendusega

LISA 4. Kompimisadjektiivide polüseemiamallid

Mall	Sõnu valimis*	Näited	Malli kirjeldus
PUUDUTUS- PSÜHH	43	<i>pehme, okkaline, libe, kuiv, nüri, jäik</i>	Kompimistäendus => meeleolu, suhtumist või iseloomu väljendav tähendus, nt <i>okkaline roos/okkaline iseloom</i>
PUUDUTUS- HÄÄL	18	<i>kõva, kare, sametine, karune</i>	Kompimistäendus => heli, häält iseloomustav tähendus, nt <i>sametine nahk/sametine hääl</i>
PUUDUTUS- MAITSE	11	<i>terav, õline, sametpehme, ümar</i>	Kompimistäendus => maitset iseloomustav tähendus, nt <i>teravad nurgad/terav karri</i>
PUUDUTUS – VÄRVUS	3	<i>vahane, metalne</i>	Kompimistäendus => värvust või valgust iseloomustav tähendus, nt <i>metalne pind/ metalne läige</i>
PUUDUTUS – LÕHN	4	<i>terav, õline, mullane</i>	Kompimistäendus => lõhna iseloomustav tähendus, <i>terav nuga/ terava lõhnaga gaas</i>
PUUDUTUS- TEMPERATUUR	4	<i>karm, pehme</i>	Kompimistäendus => temperatuuritähendus, <i>pehme voodi/ pehme suveõhtu</i>
MATERJAL/ AINE (A)- PUUDUTUS	11	<i>siidine, sametine, takune</i>	Teatud materjalist/ainest koosnemist väljendav tähendus esineb koos tähendusega, mis markeerib selle materjali puudutamisel tekkivat puuteaistingut ja on omakorda kantav ka muust materjalist objektidele, nt <i>sametine ülikond/ sametine nahk</i>
PUUDUTUS- MATERJAL/ AINE (S)	4	<i>villane, puuvillane, linane</i>	Teatud materjaliga seotud kompimistäendus esineb koos substantiivse tähendusega, mis tähistab sellest materjalist eset, nt <i>villased sokid/ kannan villast</i>
PUUDUTUS- NÄHTUS_ LOODUS (A)	10	<i>kuiv, niiske, märg, tatine, lumine, ligane</i>	Kompimistäendus => ilma, loodusnähtust iseloomustav adjektiivne tähendus, nt <i>tatine taskurätik/ tatine talv</i>
PUUDUTUS- NÄHTUS_ LOODUS (S)	4	<i>kuiv, märg</i>	Kompimistäendus => ilma, loodusnähtust iseloomustav substantiivne tähendus, nt <i>märg pesu/ mantel peab märga</i>
PUUDUTUS- IN/LOOM_OMA DUS (S)	6	<i>karvane, siledakarvaline</i>	Kompimistäendus => sama omadust kandvat inimest või looma tähistav substantiivne tähendus, nt <i>karvane elukas/ karvaste pidu</i>
PUUDUTUS- MÕÖDE	4	<i>rasvane, paks</i>	Kompimistäendus => ruumilist omadust tähistav tähendus, nt <i>paks puder/ paksud seinad</i>
PUUDUTUS- KOGUS	3	<i>kõva, paks, tihe</i>	Kompimistäendus => hulgatehendus, nt <i>kõva iste/ kõva rahasumma</i>

PUBLIKATSIOONID

ELULOOKIRJELDUS

Nimi: Maria Tuulik
Sünniaeg: 10.06.1985
Kodakondus: Eesti
E-post: maria.tuulik@eki.ee

Hariduskäik

2013–2022 Tartu Ülikool, doktoriõpe, eesti ja soome-ugri keeleteadus, üldkeeleteadus
2007–2010 Tartu Ülikool, magistratuur, eesti ja soome-ugri keeleteadus, õppekavajärgseks spetsialiseerumiseks üldkeeleteadus
2004–2007 Tartu Ülikool, bakalaureuseõpe, eesti ja soome-ugri keeleteadus, õppekavajärgseks spetsialiseerumiseks eesti keel, kõrvalerialad: suhtekorraldus ja kommunikatsiooniteadused
2001–2004 Tallinna Arte Gümnaasium, reaalklass
1994–2002 Nõmme Muusikakool, klaveri eriala
1992–2001 Tallinna 49. Keskool, kunstikallakuga klass

Teenistuskäik

2016–... Eesti Keele Instituut, leksikograaf-nooremteadur
2010–2016 Eesti Keele Instituut, leksikograaf
01.01.2007–31.12.2008 Tartu Ülikool, kaastöö TEKsauruse projektile, mõistete koostaja

Teadusorganisatsiooniline tegevus:

2013–... Eesti Rakenduslingvistika Ühingu liige
2011–... Emakeele Seltsi liige
2017–2021 keeleõppe ja rahvahanke ühendamise Euroopa võrgustiku (enet-Collect) juhtkomitee liige
2014–2017 Euroopa elektroonilise leksikograafia (ENeL) võrgustiku juhtkomitee liige

CURRICULUM VITAE

Name: Maria Tuulik
Date of birth: 10.06.1985
Citizenship: Estonian
E-mail: maria.tuulik@eki.ee

Education

2013–2022 University of Tartu, Estonian and Finno-Ugric Linguistics (general linguistics), PhD
2007–2010 University of Tartu, Estonian and Finno-Ugric Linguistics (general linguistics), MA
2004–2007 University of Tartu, Estonian and Finno-Ugric Linguistics BA, secondary degree in Public Relations and Communication Sciences
2001–2004 Tallinn Arte Gymnasium
1994–2002 Nõmme Music School, piano
1992–2001 Tallinn 49. Highschool

Professional employment

2016–... Institute of the Estonian Language, lexicographer-junior researcher
2010–2016 Institute of the Estonian Language, lexicographer
01.01.2007–31.12.2008 University of Tartu, collaboration on TEKsaurus project

Membership in professional organisations:

2013–... Member of the Estonian Association of Applied Linguistics
2011–... Member of the Mother Tongue Society
2017–2021 Member of the management committee of the European Network for Combining Language Learning with Crowdsourcing Techniques (enetCollect)
2014–2017 Member of the management committee of the European Network of e-Lexicography

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: socio-cultural perspectives. Tartu, 2001, 214 p.
3. **Iлона 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 structure 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 algatatud parandussekventsides 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 lk.

22. **Ingrid Rummo.** A Case Study of the Communicative Abilities of a Subject with Mosaic Patau Syndrome. Tartu, 2015, 270 p.
23. **Liisi Piits.** Sagedamate inimest tähistavate sõnade kollokatsioonid eesti keeles. Tartu, 2015, 164 lk.
24. **Marri Amon.** Initial and final detachments in spoken Estonian: a study in the framework of Information Structuring. Tartu, 2015, 216 p.
25. **Miina Norvik.** Future time reference devices in Livonian in a Finnic context. Tartu, 2015, 228 p.
26. **Reeli Torn-Leesik.** An investigation of voice constructions in Estonian. Tartu, 2015, 240 p.
27. **Siiri Pärkson.** Dialoogist dialoogsüsteemini: partneri algatatud parandused. Tartu, 2016, 314 lk.
28. **Djuddah A. J. Leijen.** Advancing writing research: an investigation of the effects of web-based peer review on second language writing. Tartu, 2016, 172 p.
29. **Piia Taremaa.** Attention meets language: a corpus study on the expression of motion in Estonian. Tartu, 2017, 333 p.
30. **Liina Tammekänd.** Narratological analysis of Võru-Estonian bilingualism. Tartu, 2017, 217 p.
31. **Eva Ingerpuu-Rümmel.** Teachers and learners constructing meaning in the foreign language classrooms: A study of multimodal communication in Estonian and French classes. Tartu, 2018, 218 p.
32. **Kaidi Rätsep.** Colour terms in Turkish, Estonian and Russian: How many basic blue terms are there? Tartu, 2018, 181 p.
33. **Kirsi Laanesoo.** Polüfunktsionaalsed küsilauseid eesti argivestluses. Tartu, 2018, 176 lk.
34. **Maria Reile.** Estonian demonstratives in exophoric use: an experimental approach. Tartu, 2019, 240 p.
35. **Helen Türk.** Consonantal quantity systems in Estonian and Inari Saami. Tartu, 2019, 149 p.
36. **Andra Rumm.** Avatud küsimused ja nende vastused eesti suulises argivestluses. Tartu, 2019, 217 lk.
37. **Eleri Aedmaa.** Detecting Compositionality of Estonian Particle Verbs with Statistical and Linguistic Methods. Tartu, 2019, 271 p.
38. **Kristina Koppel.** Näitelausete korpuspõhine automaattuvastus eesti keele õppesõnastikele. Tartu, 2020, 249 lk.
39. **Ilze Tälberga.** On the equivalents of the Latvian verbal prefixes in Estonian. Tartu, 2020, 210 p.
40. **Roger M. A. Yallop.** The affect and effect of asynchronous written artefacts (cover letters, drafts, and feedback letters) within L2 English doctorate writing groups. Tartu, 2020, 312 p.
41. **Mariann Proos.** Meaning and usage of Estonian experience perception verbs. Tartu, 2021, 175 p.
42. **Helen Hint.** From full phrase to zero: a multifactorial, form-specific and crosslinguistic analysis of Estonian referential system. Tartu, 2021, 271 p.

43. **Anton Malmi.** The production of Estonian palatalization by Estonian and Russian speakers. Tartu, 2022, 180 p.
44. **Mari Aigro.** In any case? Estonian spatial cases as argument markers. Tartu, 2022, 223 p.