Tartu Ülikool
Filosoofiateaduskond
Üldkeeleteaduse õppetool
Arvutuslingvistika eriala

Kadri Kerner

Sõnatähendused tekstides ja tesauruses ühestajate erimeelsuste põhjal

Bakalaureusetöö

Juhendaja Kadri Vider, M.A.

Tartu 2004

Sisukord

1. Sissejuhatus

Uus eesti keele strateegia[1] näeb muuhulgas ette ka keeletehnoloogilise toe arendamist sellisele tasemele, et eesti keel oleks võimeline funktsioneerima tänapäeva infoühiskonnas. Keele tehnoloogiline tugi hõlmab elektroonilisi keeleressursse, keeletöötlustarkvara ja keeletehnoloogilisi rakendussüsteeme. Keeleressursid on elektroonilised andmekogud (tekstid, sõnastikud, andmebaasid), mida kasutatakse keeletarkvara väljatöötamiseks. (Muischnek jt 2003: 7)

Keeleressursiks on ka semantiline andmebaas, mis on leksikaalsete andmebaaside alaliigiks. Selliste andmebaaside olemasolu paljudes keeltes võib viia mitmete heade tulemusteni, nt automaatsed tõlkesõnastikud, intelligentsed info-otsisüsteemid. (Muischnek jt 2003: 29)

Alates 1997. aastast on Tartu Ülikooli arvutuslingvistika uurimisrühmas loodud sõnasemantikal põhinevat leksikaalset andmebaasi – TEKsaurust ehk Tartu Ülikooli eesti keele tesaurust[2]. Tänapäeval muutub järjest olulisemaks automaatne sõnatähenduste ühestamine ning olulise osa sellest moodustab sõnatähenduste ühestamine (Kahusk, Kaljurand 2002: 185). Sõnatähenduste ühestamise ajaloost ja meetoditest annab ülevaate esimene peatükk.

Sõnatähenduste ühestamist peetakse loomuliku keele töötlemise juures üheks kesksemaks probleemiks (Ide, Veronis 1998). Tänaseks on eesti keelele loodud automaatne morfoloogiline analüüs ja süntees ning süntaktiline analüüs, kuid täielikku semantilist automaatset analüüsi veel ei ole. Semantilise ühestamise aktuaalsusest annavad märku SENSEVAL-1 ja SENSEVAL-2 projektid[3] ning eesti keele strateegias mainitud leksikaal-semantilise andmebaasi, TEKsauruse suurendamine.

[1] http://www.eki.ee/keelenoukogu/strateegia.html
[2] http://www.cl.ut.ee/ee/ressursid/teksaurus.html
[3] http://www.senseval.org/

Automaatse sõnatähenduste ühestamise süsteemi loomisel on vajalik tekstis teatud sõnad käsitsi ühestada, luua semantiliselt ühtlustatud treeningkorpus ning seejärel samale ühestatud tekstile rakendada automaatne ühestamissüsteem. Käsitsi ühestamine toimub TEKsauruse tähendusnumbrite põhjal. TEKsaurusest ja käsitsi ühestamisest räägib lähemalt teine peatükk.

Praeguseks on käsitsi ühestatud 43 teksti (~40 000 ühestatavat sõna ja ~100 000 tekstisõna) ning tulevikus on plaanis käsitsi ühestatud korpust suurendada. See näitab, et käsitsi märgendamise täiustamise probleem on küllaltki vajalik.

Töö eesmärgiks on uurida käsitsiühestajate arvamuste erinevust – milliste sõnade tähenduste osas on erinevused suured ning kas tekivad mingid tähenduste klastrid.. Uurimise alla on võetud 74 sõna, millest 24 on substantiivid ja 50 verbid. Sõnade valikul on lähtutud selle esinemissagedusest nii erinevuste alamkorpuses kui ka ühestatud korpuses. Töö metoodika, sh sõnade valik, on esitatud eraldi peatükina.

Sõnatähenduste ühestamine pole ka inimesele lihtne ülesanne (Vider jt 2000). Käsitsi ühestamine viitab ka konkreetsetele probleemidele TEKsauruses. Eriti näitavad problemaatilisi kohti käsitsiühestajate eriarvamused ning seepärast ongi töö praktilises osas, neljandas peatükis, analüüsi aluseks võetud käsitsiühestajate erinevuste failid. Töö käigus püütakse välja tuua TEKsauruse puudusi, mida esitatakse olemasolu korral iga sõnaartikli juures eraldi.

2. Sõnatähenduste ühestamine

Sõnatähenduste ühestamine (ingl *word sense disambiguation*) kuulub leksikaalse semantika (ehk sõnasemantika) uurimisvaldkonda. On lekseeme, millel on rohkem kui üks tähendus, sõnad on mitmetähenduslikud ehk polüseemsed (Karlsson 2002: 243). Näiteks on eesti keeles polüseemne sõna *aeg*; TEKsaurusest tehtud päring mainitud polüseemsele sõnale *aeg* näitab, et erinevaid tähendusi on kuus.

Arvutirakendused, mis käsitlevad loomuliku keele tekste, peavad toime tulema muuhulgas ka polüseemiaga. Arvutilingvistika uurimisobjektiks on see, kuidas ühestada tähendusi automaatselt. Polüseemia võib tuleneda regulaarsest tähenduse laiendamisest, seda võib põhjustada kontekst või võib polüseemia olla tingitud metafoorist ning metonüümiast. (Ravin, Leacock 2002: 23)

Sõnatähenduste ühestamise vastu on huvi tuntud juba alates 1950ndatest aastatest. Öeldakse, et sõnatähenduste ühestamine on vahendava ülesandega, seda vajavad näiteks:

- masintõlge – tõlkida tuleks õige tähendus;
- info-otsing, mille puhul oleks vajalik eemaldada need tähendused, mis pole vastaval otsingul vajalikud;
- grammatiline analüüs – lauseliikmete määramiseks juhul, kui sõna tähenduse teadmine kindlas kontekstis on vajalik;
- kõnetöötlus (ühtemoodi kõlavate sõnade eristamine) jt.

Sõnatähenduste ühestamine hõlmab endas kahte sammu:

1) kindlaks teha kõik erinevad tähendused kindlas kontekstis. Selle alamülesande lahendamisel kasutatakse:
    - elektroonilist sõnastikku, mis esitab sõnade kõikvõimalikud tähendused;
    - kategooriate, omaduste gruppi või seoseis olevaid sõnu (nt sünonüüme nagu tesauruses);
    - tõlke- või ülekandesõnastikku, mis hõlmab endas teise keele tõlkeid.

2) Teha kindlaks, milline tähendus on kõige sobivam (lähtutakse sellest, et mingi sõna tähendus sõltub tema kontekstist). Püütakse leida meetodit, mis kirjeldaks vastavust sõna võimalik kontekstide ja võimalike tähenduste vahel. Kasulikku andmestikku mingi sõna seostamiseks tema tähendusega saab, kui kasutada keeleväliseid teadmiste allikaid: leksikaalsed, entsüklopeedilised ressursid (Ide, Veronis 1998).

Sõnatähenduse täpse definitsiooni üle vaieldakse palju. Adam Kilgarriff kirjutab oma artiklis "*I don't believe in word senses*" (Kilgarriff 1997), et sõnatähenduste ühestamine eeldab erinevate tähenduste olemasolu ning erinevad tähendused on teada-tuntud komistuskohaks. Artikli autor väidab, et mõistel *sõnatähendus* puuduvad põhialused.

Sõnadel on tihti rohkem kui üks tähendus, mõnikord on need erinevused suured, mõnikord väiksemad ning mingi sõna kindla tähenduse saab uurides tema konteksti. Inimesele pole see kerge ülesanne, arvutile ka mitte. Näiteks sellised laused: *poiss hüppas pangalt vette; auto jäi panga ees seisma ja kolm maskides meest tulid välja*. Inimene saab tänu kontekstile öelda, millist *panka* mõlemas lauses silmas on peetud, arvuti sellest ilmselt aru ei saa. See on küll eesti keelekäsitluses väidetavalt homonüümia, kuid võõrkeelsed uurijad toovad vastava valvenäite polüseemia seletamiseks.

2.1. Sõnatähenduste ühestamise erinevad meetodid ajaloo vältel

2.1.1. Varajane masintõlge

Masintõlke uurimise alal (1950ndatel) tehti esimesed sammud automaatse STÜ suunas. Sellesse perioodi jääb palju põhilisi olulisi uurimusi, kuid suuremate ressursside puudumisel jäid ideed testimata. Keskenduti peamiselt spetsiifiliste sõnaraamatute arendamisele, kus igal sõnal oli vastavalt valdkonnale ainult üks kindel tähendus. Näiteks matemaatika valdkonnas oli sõnale *triangle* (ee kolmnurk) ainult üks, matemaatikasse sobiv definitsioon; *triangle* (ee triangel) kui muusikariist oli välja jäetud. M. Mastermani sõnatähenduste ühestamist Roget tesauruse põhjal (1957) peetakse varajaseks statistiliseks lähenemiseks, mida on jätkanud ka hilisemad uurijad. (Ide, Veronis 1998)

2.1.2. Tehisintellekti meetodid

1. Sümbolmeetodid (ingl *Symbolic methods)*

Semantilised võrgustikud. 1962. aastal arendas M. Masterman välja kogumi, kus esines 100 primitiivmõiste tüüpi. Töö tulemuseks oli semantiline võrgustik (ingl *semantic network)*, milles mõisted olid omavahel seotud semantiliste suhetega. Mõiste (ingl *concept*) on sõlm (ingl *node)* ja semantilisi suhteid märgitakse kaartega (ingl *arcs).* 1961–1969 töötas semantiliste võrgustikega edasi R. Quillian. Ta valmistas võrgustiku, mis sisaldas ühenduslülisid sõnade ja mõistete vahel. Ühenduslülid märgendati erinevate semantiliste suhetega. (Ide, Veronis 1998)

Võrgustikud ja freimid. 1976–1978 kasutas P. Hayes semantiliste võrgustike ja freimide kombinatsiooni, kus sõlmed näitasid substantiivii tähendust (ingl *noun sense)* ja ühenduslülid verbi tähendust (ingl *verb sense),* freimideks olid suhted IS-A (klassi-liigi suhe) ja PART-OF (osa-terviku suhe). Nii oli võimalik käsitleda homonüüme, aga mitte polüseemiat. 1987. aastal kasutas G. Hirst samuti võrgustikke ja freime ning eelmainitud R. Quilliani kõige lühema tee põhimõtet. Ta tutvustas mõistet "polaroidsed sõnad" (ingl *polaroid words*) – mehhanism, mis süntaktilise informatsiooni põhjal ja semantiliste suhete põhjal järjest eemaldab sobimatuid tähendusi; lõpuks peaks ideaalis jääma vaid üks, õige tähendus. Kuid see mehhanism ei tule tihti toime metafoorsete sõnadega ning eemaldab kõik tähendused. (Ide, Veronis 1998)

***Case based* lähenemisviisid.** Aastatel 1968–1975 loodud Y. A. Wilksi "eelistava semantika" (ingl *preference semantics)* reeglid kehtestavad teatud piirangud, mis põhinevad semantilistel tunnustel (nt elus-eluta). Ingliskeelses näitelauses *My car drinks gasoline* (ee minu auto võtab palju bensiini) on piirangud/kitsendused sõnal *drink* (ee jooma), mis eeldab elusat subjekti, kuid lubab ka elutut. Hiljem on tõestatud, et ka see meetod ei suuda toime tulla polüseemiaga. (Ide, Veronis 1998)

Ontoloogilised seletused. 1988. aastal valminud K. Dahlgreni süsteem sisaldab endas tähendusi ühestavat komponenti, mis kasutab mitmeid informatsiooni tüüpe: kinnisühendeid (ingl *fixed phrases*), süntaktilist informatsiooni ja tervemõistuslikke

argumentatsioone/seletusi (ingl *commonsense reasoning*); viimast kasutatakse vaid siis, kui kaks esimest meetodit ei saa ühestamisega hakkama.

2. Ühenduslikud meetodid (ingl *connectional methods)*

Nende meetoditega hakati tegelema psühholingvistikas 1960ndatel ja 1970ndatel. Olulisemad uurimused: 1971. aastal D. Meyeri ja R. Schvaneveldti "Semantiline vermimine" (ingl *Semantic priming)*. Idee seisneb selles, et me saame järgnevatest sõnadest aru selle põhjal, mida me oleme juba kuulnud.

"Leviv käivitumismudel" (ingl *Spreading activation model)*, kus mingi mõiste aktiveerub ja laieneb edasi temaga seotud sõnadele; aktivatsioon nõrgeneb edasi jaotudes.

1983. aastal kasutasid G. Cottrell ja S. Small neuraalseid võrgustikke sarnaselt R. Quillianile (sõlm on mõiste). (Ide, Veronis 1998)

Tehisintellekti põhiste meetodite kriitikaks on peetud seda, et ühestamist on testitud piiratud kontekstis (sageli ainult ühe lause piires). Ning seetõttu on raske kindlaks määrata tegelikus keelekasutuses. (Ide, Veronis 1998)

2.1.3. Teadmistel põhinevad meetodid

Neid meetodeid hakati kasutama 1980ndatel, kui said kättesaadavaks suured leksikaalsed ressursid (elektroonilised sõnastikud, tesaurused, leksikonid).

Elektroonilised sõnaraamatud. R. Amsler ja A. Michiels proovisid automaatselt nõrutada leksikaalseid ja semantilisi teadmusbaase (ingl *knowledge base*) elektroonilistest sõnaraamatutest. 1986. aastal püüdis Lesk luua sõnaraamatust teadmusbaasi. Iga sõna tähendus oli vastavuses "signatuuriga". Signatuurid koosnesid nn sõnade kottidest (ingl *bag of words*), mida on kasutatud sõnatähenduse definitsioonides (Stephen Wan 1999). 1990. aastal arvutas Y. Wilks sõnade definitsioonide koosesinemise sagedust, ning 1990. aastal J. Veronis ja N. Ide lõid neuraalse võrgustiku, kus iga sõna on ühendatud oma tähendustega, mis on ühendatud ka sõnadega nende definitsioonides (või signatuurides) ja need omakorda vastavate sõnade tähendustega jne. Mitmed erinevad autorid (R. Krovetz, W. B. Croft jt) on

proovinud tulemuste parandamiseks kasutada *Longman Dictionary of Contemporary English* (LDOCE) elektroonilises versioonis täiendavaid informatsioonivälju. Lisainformatsioon on sarnane tesaurusele, kuid tesaurus on paremini struktureeritud. (Ide, Veronis 1998)

Elektroonilise sõnaraamatu kasutamine ongi üheks enimlevinud ühestamismeetodiks. Tavalisim on see, et sõna tähenduse leidmiseks võrreldakse tema konteksti sõnaraamatus esinevate näitelausetega ja sarnaseim valitakse välja. Sõnastike viga on see, et nad annavad küll põhjalikku leksikaalset informatsiooni, kuid neis puudub vajalik pragmaatiline informatsioon ning sageli on uurimused liiga ühest sõnastikust sõltuvad. Elektroonilised sõnastikud võivad olla väga detailsed, kuid siiski on seal teatud puudusi, millega arvuti hakkama ei saa. (Ide, Veronis 1998)

Tesaurused, mis esitavad sõnadevahelisi seoseid, enamasti just sünonüümiat. Vaieldakse pidevalt hierarhias ülemistel kohtadel olevate üldmõistete üle, kuid tesaurused pakuvad siiski rikkalikku sõnade võrgustikku, mis võib olla potentsiaalselt kasulik keeletöötluses. Ka tesaurused on loodud inimestele kasutamiseks, seepärast pole sealsed sõnadevahelised suhted arvuti jaoks just kõige paremal viisil esitatud. 1985. aastal kasutas A. Patrick Roget' tesaurust, et ühestada verbide tähendusi. Ta uuris, kuidas on seotud lähedased sünonüümid vastava sõnaga. 1992. aastal jaotas D. Yarowsky sõnaklassid tavalisemate kategooriate alusel ning kasutas entsüklopeediat, leidmaks signatuure; edasi vaatles, missugused signatuurid esinevad kõige rohkem antud sõnaklassiga. (Ide, Veronis 1998)

Arvutileksikonid. 1980ndatel hakati looma (algul käsitsi) suuremahulisi teadmisbaase, näiteks WordNet. Semantiliste leksikonide loomisel on kaks põhilist lähenemist:

1) loetletavad leksikonid (ingl *enumerative)*, kus tähendused on esitatud eksplitsiidselt;
2) generatiivsed leksikonid, kus semantiline informatsioon (seostatuna antud sõnadega) ei ole täpsustatud ning kasutatakse generatiivseid reegleid, et tuletada täpset tähenduse informatsiooni.

Loetletavaks leksikoniks on WordNet – praegu kõige tuntum ja kasutatavam ressurss inglise keele sõnatähenduste ühestamisel. WordNet kombineerib mitmeid erinevaid

ressursse, mida on ühestamise juures eelnevalt kasutatud: igal sõnatähendusel on definitsioon; esitatud on sünohulgad (ingl *synset)*; mõisted on organiseeritud hierarhiasse ning WordNetis on esitatud sõnadevahelised suhted: hüponüümia, antonüümia ja meronüümia. WordNetti on rakendatud ka sõnatähenduste ühestamisel info-otsingu jaoks. (Ide, Veronis 1998)

1993. aastal arvutas M. Sussna sõnatähenduste vahelist semantilist distantsi mõõtes sõnade semantilist seotust WordNetis. Suhete tulemus (skoor) arvutatakse tähenduse konteksti sõnade vahel ning ühestatuks saab see tähendus, mille tulemus on kõrgem. (Ide, Veronis 1998)

Ometigi öeldakse, et WordNet pole kõige parem ressurss sõnatähenduste ühestamisel, sest sõna erinevad tähendused on liiga "peeneks" tehtud. See pole veel selgeks saanud, kui eristatud peavad sõnade tähendused ühestamise jaoks üldse olema. (Ide, Veronis 1998)

2.1.4. Korpusepõhised meetodid

Korpusepõhine lähenemine ühestamisele on andnud tulemusi statistika ja tõenäosuse kasutamise tõttu. Ühestamisel saab kasutada ka semantiliselt käsitsi ühestatud korpust, kuid häda on selles, et korpus ei kata ilmselt kõiki tähendusi, vaid ainult neid, mis esinevad sagedamini. Luuakse mingi algoritm, mida treenitakse korpusel ja mis peaks oskama automaatselt sõna tähendusi eristada. On püütud kasutada ka Markovi peitmudelit (ingl *Hidden Markov Model)*, mis on efektiivne olnud süntaktilisel märgendamisel. Ühestatud korpuse loomine on paraku aga aeganõudev ja kallis töö. Mõned uurimused:

- 1973. aastal demonstreeris S. Weiss, et ühestamisreegleid saab õppida käsitsi semantiliselt ühestatud korpusest.
- 1990ndatel hakati kasutama valvega õppimise meetodit (ingl *supervised learning approaches)*; töötab semantiliselt märgendatud korpusel;
- On püütud automaatselt ühestada treeningkorpust kasutades andmestiku usaldusväärsust kontrollivaid meetodeid (ingl *bootstrapping)*. 1991. aastal lõi M. Hearst algoritmi, mis on seotud ka statistilise informatsiooniga.

- Silumismeetodit (ingl *smoothing)* kasutatakse harvaesinevate sündmuste/sõnade probleemi jaoks. Püütakse kindlustada seda, et harvaesinevad sõnad ei saaks tõenäosuseks nulli.
- Klassidel põhinev meetod püüab kõik ühte kategooriasse kuuluvad sõnad klassidesse jaotada. Kohati on see meetod hea, kuid tekib teatav informatsioonikadu, sest tegelikult ei käitu kõik sarnasesse klassi kuuluvad sõnad ühtemoodi.
- Sarnasusel põhineval meetodil on umbes sama idee, kuid seal ei panda sõnu mingitesse kindlatesse gruppidesse. Igal sõnal on potentsiaalne sarnaste sõnade kogum. Sõnad loetakse sarnasteks, kui nad esinevad sarnases kontekstis, kontekstid on sarnased, kui nad sisaldavad sarnaseid sõnu. Treeningkorpuses on ainult mõned näited iga sõna kohta, mitte nagu teiste meetodite puhul, kus vajalikud on sajad näited. (Ide, Veronis 1998)

Meetodite hindamine. Kuna kõik meetodid ja uurimused on erinevad – põhinevad erinevatel ressurssidel ja erinevatel tingimustel – on neid raske omavahel võrrelda. Piiratud tähendustega on kindla spetsiifikaga tekstid; suurem hulk tähendusi on üldisematel tekstidel. Hinnata saab ilmselt selle järgi, kui palju on sõnatähenduste uurimused andnud mingisse kindlasse rakendusse, näiteks masintõlkesse või info-otsingusse. Kõige rohkem on püütud liita sõnatähenduste ühestamise meetodeid info-otsingusse, kuid uurimused näitavad, et kui ühestamine pole piisavalt täpne, siis info-otsingu tulemused pigem vähenevad. (Ide, Veronis 1998)

Viimasel ajal tehtud uurimused hõlmavad endas tüüpiliselt vähe sõnu (enamasti nimisõnu) ja tähendused on laiemalt eristatud. Tänu suurematele ressurssidele ja arendatud statistilistele vahenditele, on uurimustulemused muidugi paranenud; arvatakse, et palju kasulikku saab leida ka leksikaalse semantika alalt. Praegune probleem paistab seisnevat selles, et erinevaid tähendusi püütakse leida tavalistest sõnaraamatutest, mis ei ole loodud siiski esitama tähendust vastavas kontekstis. (Ide, Veronis 1998)

2.1.5. "Üks tähendus ühe teksti kohta"

W. Gale, K. Church ja D. Yarowsky leidsid sõnatähenduste ühestamise süsteemi väljatöötamisel, et sõna tähendus sõltub ühest kindlast tekstist. Seda nimetatakse diskursuse efektiks ja see tähendab, et kui sõna esineb kaks või enam korda ühes tekstis, siis tema tähendus on tavaliselt üks ja seesama kogu teksti piires (inglise keeles kuni 98%). Eelnevalt mainitud autorid avastasid, et sõnatähendustel on kombeks esineda koos kindlate klastritena (ingl *cluster*). Raske on leida kaht või enamat ühe sõna tähendust ühe teksti seest. Selle hüpoteesi tõestamiseks tehti erinevaid teste (nii ükskeelsete kui ka kakskeelsete ressurssidega) substantiividega ning jõuti järeldusele, et piisavalt suure ressursi olemasolu korral on diskursuse efekt tõepoolest tugev. (Gale, Church, Yarowsky 1992)

Sama tendents ilmneb ka eesti keele nimisõnade puhul ning harvaesinevate verbide puhul. Näiteks võiks tuua sõna *tüdruk*, mis esineb teksti läbivalt ainult ühes tähenduses (ühe tähendusnumbriga). Verbidest näiteks sõnad *ärgitama, õngitsema.*

On kaks võimalikku rakendust, kuidas ära kasutada tähelepanekut "üks tähendus ühe teksti kohta":

1) lisakitsendusena ühestamisalgoritmides;
2) ei pea ühe teksti sees sõnu eraldi ühestama, vaid saab ühe korraga antud sõnale tähendusnumbri anda.

Muidugi võib tekkida vigu, kuna see tendents pole sajaprotsendiline, aga suuremate tekstihulkade puhul on see suureks abiks. (Gale, Church, Yarowsky 1992)

Seda tendentsi, küll pisut modifitseeritult, on rakendatud SENSEVAL-2 võistlusel (rahvusvaheline projekt, kus hinnati ühestamissüsteeme) ning mõningad tulemused ja ühestamissüsteemid on ka paranenud.

Teisedki uurijad on seda fenomeni uurinud – on ka vastuväiteid leitud. Eelkõige see, et tekst peab olema äärmiselt teemakeskne ning et lisaks tuleks arvestada ka tendentsi "üks tähendus ühe kollokatsiooni kohta". Fenomen ei kehti siiski adjektiivide puhul ja sagedaste verbide puhul nagu *pidama, saama, andma* jt.

2.1.6. Tähenduserinevuste uurimine

Mitmesus on väga tavaline, eriti sagedamini esinevate sõnade seas. Sõnatähenduste ühestamise probleemiks on luua mõistlikud tähenduste erinevused – mitte liiga üle-

eristatud (kui kahe tähenduse vahel on väga ähmane piir) ega liiga ala-eristatud. (Chklovski, Mihalcea 2003)

WordNetis on palju polüseemseid sõnu – umbes 20%-l sõnadel on rohkem kui üks tähendus ning polüseemsel sõnal on keskmiselt kolm erinevat tähendust. Liiga üle-eristatud tähenduste puhul on käsitsimärgendajatel õiget tähendust raske valida. On tehtud palju teste ja katseid, kus inimene ei oskagi üleüldse polüseemiat kahtlustada. (See on loomulik, sest inimene saab tähendusest aru tänu kontekstile (Chklovski, Mihalcea 2003)). Keeletehnoloogilised rakendused ei vaja väga detailset eristust; samas ei ole võimalik ka väga palju üldistusi teha, s.t sarnaseid tähendusi grupeerida. (Tomuri 2001)

Kõrge polüseemia tase tekib sageli leksikograafide käe all, sest nemad peavad eristama sõna kõiki võimalikke tähendusi (Peters, Peters, Vossen 1998).

Sarnaseid sõnatähendusi on püütud grupeerida klastritesse, et vähendada mitmesust – nii on tehtud EuroWordnetis ning on saavutatud seeläbi täpsemad tulemused ja väiksem müra osakaal. (Peters, Peters, Vossen 1998)

Sarnaseid tähendusi leitakse sagedus- ja korratusmaatriksite moodustamise abil. Printsiip: tähendus A on sarnane tähendusele B kui A on sageli märgendatud kui B. Seejärel rakendatakse tavaliselt klaster-meetodit, mille abil sarnased tähendused grupeeritakse. Selline automaatne grupeerimine kattub täiesti ka inimekspertide arvamusega. Nii on saadud suhteliselt head tulemused ning tõestatud, et ka erimeelsuste uurimisest võib kasu olla. (Chlkovsky, Mihalova 2003)

Viidi läbi eksperiment, milles paluti inimestel (mitte-leksikograafidel) sõnatähendusi internetis ühestada WordNeti tähendusnumbrite põhjal[4]. Selgus, et WordNetis liiga üle-eristatud tähendused tekitasid segadust. Need tähendused, mille suhtes oldi ühel arvamusel, koguti korpusesse ning erimeelsustele rakendati eelmainitud klaster-meetodit, mis automaatselt moodustas laiemad tähendusgrupid. Näiteks sõna, millel WordNetis oli algupäraselt kuus tähendust jaotati nii, et tähendused 1,2,6 moodustasid omaette tähendusklastri ja 3,4 ning 5 veel omaette klastrid. (Chlkovsky, Mihalova 2003)

[4] http://teach-computers.org/

3. TEKsaurus ja käsitsi sõnatähenduste ühestamine

3.1. Tartu Ülikooli eesti keele tesaurus ehk TEKsaurus

Eesti üldkeele tesaurust ehk eesti WordNet'i on koostatud aastast 1997. Tesauruse koostajaiks on professor Haldur Õimu juhtimisel olnud arvutuslingvistika uurimisrühma liikmed Neeme Kahusk, Leho Paldre, Heili Orav ja Kadri Vider. Eesti keele põhisõnavara tähendused peaks olema enamuses kaetud, kuigi mõisteid on umbes 100 000 ringis. TEKsauruses on substantiivid, verbid ja adjektiivid (~300). (Lõpparuanne 2000–2002)

Eesti üldkeele tesauruse loomist on toetanud Eesti Teadusfond ja Eesti Informaatikakeskus sihtprogrammis "Eesti keeletehnoloogia", samuti riikliku sihtprogrammi "Eesti keel ja rahvuskultuur" keeletehnoloogia allprojekt. (Vider jt 2000)

Eeskujuks võeti Wordneti ja Eurowordneti[5] põhimõtted. (WordNeti ja EuroWordneti erinevus on see, et Princetoni Ülikoolis loodud Wordnet on ükskeelne, EuroWordnet aga mitmekeelne).

Eesti keele tesaurus on keelelistel teadmistel põhinev, sest põhineb tekstikorpusel ja olemasolevatel sõnaraamatutel (elektrooniliselt olemasolevad):

- Eesti Kirjakeele Seletussõnaraamat
- KeeleWebi[6] kaudu on kasutatud Filosofti tesaurust, Asta Õimu Sünonüümisõnastikku, Antonüümisõnastikku ja Fraseoloogiasõnaraamatut.

Eesti keele wordnet-tüüpi tesaurust koostati üldisematest sõnadest ning mõistetest lähtudes: esimeseks etapiks oli baasmõistete leidmine ja kirjeldamine Eesti kirjakeele seletussõnaraamatu (EKSS) seletuste sagedusloendi põhjal. Baasmõisted moodustavad keele wordneti tuuma ning esindavad keele peamist semantilist jaotust. Järgmine etapp oli moodustada sünohulgad, mille moodustamisel võeti arvesse ka sõna esinemissagedust eesti keele tekstikorpuses. Eesti keele tesauruse loomisel on kõige

[5] http://www.illc.uva.nl/EuroWordNet/
[6] http://ee.www.ee/

rohkem keskendutud sünonüümiale. Sünohulgad on loodud EKSSi ja sünonüümisõnastike leksikaalse info ümberstruktureerimisega. (Vider 2001)

Seletused on püütud koostada nii, et nad kehtiksid terve sünohulga suhtes. Seletused ei saa olla vastuolulised või teineteist välistavad. Mõnikord on kasutatud ka vastandust või antonüümiat. Sünohulk võib olla TEKsauruses ka üheliikmeline, mis erineb sünonüümisõnastiku sünonüümireast. (Vider 2001)

Eesti keele tesauruse loomisel kasutati ka hüperonüümiasuhet ning ideaalis peaks igal sõnal olema üks hüperonüüm. Täiuslikkuse huvides on tehtud mööndus ideaalile ning osa sõnu võib kuuluda ka mitmesse hierarhiasse. (Vider 2001)

Teisi semantilisi suhteid (antonüümiat, osa-terviku suhteid, osalus-rollisuhteid, põhjussuhteid, osasündmuse suhteid) on kasutatud ebaregulaarselt (Vider 2001).

TEKsauruse koostamiseks kasutatakse EuroWordNeti sisestusliidest Polaris, mis on Lernout ja Hauspie loodud[7].

TEKsauruse päringunäide järgmisel leheküljel:

[7] http://www.cl.ut.ee/ee/ressursid/teksauruse_struktuur.html

Päring eesti keele tesaurusest

Päringuaknasse võib sisestada algvormis nimi-, tegu- või omadussõnu. Täpsemalt vaata tesauruse lehelt.

Otsi Unusta ära

Päringusõna sisaldavatest tesauruse kirjetest kuvatakse:

1. veerus - päringusõna sisaldavate sünohulkade ülemmõisted ehk hüperonüümid, kui hüperonüümiasuhe eksisteerib;
2. veerus - sünohulk, milles igale sõnale järgneb tema tähenduse number ja päringusõna on rasvases kirjas;
3. veerus - seletus, mis peaks ideaaljuhul kehtima kõigi sünohulga liikmete kohta;
4. veerus - näide või näited ühe või mitme erineva sünohulga liikme kasutuse kohta selles tähenduses.

Kaeba sisu üle Kadri Viderile ja vormi üle Neeme Kahuskile

keel

Hüperonüüm(id)	Sünohulk	Seletus	Näide
keel 3	**keel 2**, inimkeel 1	kirja- või kõnekeel suhtlemisvahendina, vastandatud nn. tehiskeelele	Tõlgib signaalid inimkeelde.
rihm 1	**keel 6**, piug 1, piitsarihm 1	nöör v. rihm piitsavarre otsas	
organ 2, elund 1	**keel 4**	toitu haarata, segada, maitsta ja neelata aitav ning häälitsemisel osalev liikuv elund suuõõne põhjas	
kommunikatsioon 1, suhtlus 1	kõne 2, kõnekeel 1, **keel 1**	inimese olulisim suhtlemisvahend, mõtete ning tunnete vahendaja	Kõne areng on seotud mõtlemise arenguga.
suhtlus 2, suhtlemine 1, lävimine 1, kommunikatsioon 2	jutt 1, kõne 3, **keel 5**	verbaalne suhtlemine, see, mida keegi räägib, ütleb, jutustab	Tema kõne oli segane Oli kuulda laia venekeelset kõnet.
kommunikatsioon 1, suhtlus 1	**keel 3**	häälikuid v. kokkuleppelisi sümboleid kasutav süstemaatiline suhtlusvahend	Külalise keel oli kohalolijate jaoks võõras.
asjandus 1, vahend 2	**keel 7**, pillikeel 1	üle kahe tugialuse pinguldatud häälestatav heliallikas	Viiulil katkes keel. Tõmbab poognaga üle keelte.

3.2. Käsitsi sõnatähenduste ühestamine

Eesti keele puhul on valitud Eesti üldkeele tesaurus ehk TEKsaurus abistavaks sõnastikuks. Tähendusnumbrite põhjal on ühestatud substantiive, põhiverbe ja modaalverbe (Vider jt 2000).

Enne semantilise analüüsi tegemist on vaja teha morfoloogiline analüüs ning leida igale sõnavormile õige lemma. Morfoloogiliselt ühestatud korpuse maht on juba üle 400 000 tekstisõna ning seda ühestatud korpust kasutataksegi sõnatähenduste semantilisel ühestamisel (Lõpparuanne 2000–2002). Morfoloogilise analüüsi teeb programm ESTMORF[8]. Tekstid, mida ühestatakse, on võetud Eesti Kirjakeele korpusest[9]. Sõnavormi õige lemma taga on tunnused, märgitud plussmärgiga. Kui tunnust/käändelõppu/pöördetunnust ei lisandu või on tegu muutumatu sõnaga, siis lisatakse plussmärgi taha 0. // // märkide vahel on morfoloogiline info.

@ märgile järgnev info on lisatud automaatselt programmi *semyhe* poolt. Viimase kooloni järel olev number näitab, mitu tähendust on antud sõnal TEKsauruses.

Iga teksti ühestas korraga kaks inimest; // ja @-märgi vahele paneb käsitsiühestaja TEKsauruses esineva tähendusnumbri, mis tema arvates õigeim on (vt ka käsitsimärgendaja juhendit, Lisa 2). Kui sõna TEKsauruses puudub, siis märgitakse // ja @ märgi vahele automaatselt 0 ning sel juhul pole ühestajal antud sõna vaja ühestada. Eelfiltrina on toiminud programm *semyhe*, mis lisab @-märgile järgneva info. (H.Orav, K.Vider 2002).

```
Ülejärgmisel
ülejärgmine+l //_A_ pos sg ad //
päeval
päev+l //_S_ com sg ad // 6 @ päev:1711:8
oli
ole+i //_V_ aux indic impf ps3 sg ps af //
õnn
õnn+0 //_S_ com sg nom // 1 @ õnn:492#8698#8938:3
```

Näide ühestatud tekstist.

[8] http://www.ee/eks/ready/estmorf.html
[9] http://www.cl.ut.ee/ee/corpusb/tykk.html

Kui sõna esineb TEKsauruses, aga mitte vastavas tähenduses, siis märgiti +1 ning lisati eraldi kommentaaride faili uus tähendus.

```
'vabandama' vestluse alustamisel on midagi muud.
'võtta või jätta' kõlab pigem kahtlusena. Äkki võiks kokku panna?
'silm' ei pruugi olla ainult nägemisorgan, nt uksesilm
'vaatama' pole siin päris see
'sõna' pisut abstaktsem. Pole ju mõeldud konkreetset keeleüksust
'pöördumine' pisut teises tähenduses.
'Draamateater' on siin kui konkreetne institutsioon
'esseist' siin on morfoloogiline analüüs valesti
'passima' PUUDU tähenduses vaatama
'kukkuma' ükski neist hästi ei sobi. välja kukkuma käib äkki kokku?
'kukkuma' tähenduses hakkama võiks olla..
'tükk' PUUDU tähenduses asi või tegu
'venitama' PUUDU tähenduses sõnu, häälikuid venitama
'peo' morfoloogine analüüs on valesti
'aadress' pole ikka asukoht..
'tulu'PUUDU mittemajanduslikus mõttes
'väljapääs' PUUDU tähenduses koht, kust välja saab. Sissepääsu vastand
'üllatus' ei pruugi alati rõõmu valmistada
'teene' PUUDU tähenduses täna pälviv tegu
'ülal pidama' PUUDU tähenduses käituma
'kujund' PUUDU tähenduses kirjanduslik/metafoorne kujund
'vormistama' PUUDU tähenduses loodud/vormitud...
```

Näide kommentaarifailist.

Juhul kui kaks ühestajat on tähenduste suhtes eriarvamusel, lepitakse omavahel kokku üheainsa tähenduse suhtes. Moodustatakse erinevuste fail, mis näitab vaieldava sõna lähikonteksti paari sõnaga. Igal sõnal peaks olema ainult üks tähendus, seega valitaksegi välja ainult üks tähendus. Praegu on erinevalt ühestatud umbes 20% juhtudest.

```
@@ -541,11 +541,11 @@
 et
     et+0 //_J_ sub //
 ei
     ei+0 //_V_ aux neg //
 saanud
-    saa+nud //_V_ main indic impf ps neg // 2 @ saama:4256:12
+    saa+nud //_V_ main indic impf ps neg // 10 @ saama:4256:12

 sõnagi
     sõna+gi //_S_ com sg part // 1 @ sõna:323:1
 suust
     suu+st //_S_ com sg el // 1 @ suu:1277:2
 -
@@ -555,11 +555,11 @@
```

Näide erinevuste failist.

3.3. Eesti keele semantilise ühestaja loomine

Semantilise ühestamise programmi olemasolu on eeldus keeletehnoloogilistele rakendustele, kus on oluline sõnatähenduste tuvastamine (nt masintõlge, info-otsing jt). Sellise programmi loomine on pisut keerulisem kui näiteks süntaktilise või ühestamise puhul, sest semantika jaoks puuduvad formaliseeritud käsitlused. (Lõpparuanne 2000–2002)

Semantilise ühestamise programme on püütud välja töötada juba 40 aastat; viimaste aastate jooksul on hakatud kasutama suuri tekstikorpusi, mis on vähemalt morfoloogiliselt ühestatud ning ka semantilisi andmebaase (WordNet). Eesti keele jaoks on semantilise ühestamise programmi väljatöötamiseks kasutatud morfoloogiliselt ühestatud korpust ja eesti keele semantilist andmebaasi, mis töötati välja suures osas EuroWordNeti raames. Eesmärgiks oli välja töötada semantilise ühestamise programm, mis tuvastaks tekstisõnade tähendused vastavalt semantilisele andmebaasile. Selle raames oli vaja täita kaks allülesannet:

1. Semantilise ühestamise programmi koostamine.
2. Programmi koostamiseks vajaliku käsitsi ühestatud korpuse loomine, mille najal programmi testida ja arendada.

Projekti käigus tegeldi samaaegselt kolme ülesandega:

1. käsitsi semantiliselt ühestatud korpuse loomine ja kasvatamine vähemalt 50 000 sõnani;
2. ühestamise aluseks oleva eesti keele tesauruse tähendussüsteemi täiendamine;
3. semantilise ühestamise programmi loomine ja selle esimese variandi täiustamine

Projekti käigus osaleti rahvusvahelises projektis SENSEVAL-2, mille võistlusel võrreldi erinevaid ühestamisprogramme. SENSEVALi tulemuste analüüsi põhjal täiustati ka loodud programmi *semyhe*. (Lõpparuanne 2000–2002)

SENSEVAL-2 (ingl *Evaluation of Word Sense Disambiguation Systems*) oli teine rahvusvaheline projekt, kus hinnati sõnatähenduste ühestamise süsteeme, toimus juunis 2001. aastal. SENSEVAL projekt asutati 1997.aastal, et tuua kokku uurijaid ja lahendada sõnatähenduste ühestamise probleem. Eesmärgiks oli hinnata sõnatähenduste ühestamise algoritme ja süsteemide nõrku ning tugevaid kohti. SENSEVAL ei püüa välja selgitada võitjat, vaid kutsub üles avastama teaduslikke aspekte. (Edmonds, Cotton 2001)

SENSEVAL-2 võistluse eesmärgiks oli tuua uusi keeli osalema ning kokku osales 12 riiki kolme erineva süsteemitüübiga, mis olid alljärgnevad:

1) valitud sõnade ülesanne (ingl *lexical sample task*), mille valisid baski, inglise, itaalia, jaapani, korea, hispaania ja rootsi keel;

2) kõikide sõnade ülesanne (ingl *all-words task)*, mille valisid tsehhi, hollandi, inglise ja eesti keel;

3) tõlkimise ülesanne, mille valis jaapani keel. (Edmonds, Cotton 2001)

Kõikide sõnade ülesande puhul tuleb märgendada peaaegu kõik sisusõnad. Jooksva teksti näitel. Valitud sõnade ülesande puhul selekteeritakse esiteks leksikonist sõnade näited; ning süsteem peab seejärel märgendama sõnade näidete juhte lühikeses tekstilõigus. Tõlkimisülesanne – kus sõnatähendus on defineeritud vastavalt tõlke eristusele. Kokku esitati 93 süsteemi kolmekümne nelja erineva uurimisgrupi poolt.

Kõigepealt kasutasid meeskonnad etteantud vastustega treeningkorpusi ning nende põhjal täiustati ja kohandati oma süsteeme. Seejärel rakendati süsteeme testkorpusel. Eesti keele ülesannet lahendasid kaks süsteemi: *semyhe* (TÜ-s loodud) ja *JHU_Estonian*, loodud John Hopkinsi Ülikoolis. Mõlemad süsteemid lahendasid ülesannet enam-vähem võrdselt. (Lõpparuanne 2000–2002)

Eesti keele valitud ülesande kõrval taheti muuhulgas testida ka koostatud tesauruse sobivust tekstitähenduste ühestamiseks. Eesti meeskonda kuulusid Heili Orav, Neeme Kahusk ja prof.Haldur Õim. (Lõpparuanne 2000–2002)

Programmi ***semyhe***[10] teostas Kaarel Kaljurand programmeerimiskeeles Perl. Sisendtekst peab olema morfoloogiliselt analüüsitud ning väljundile lisatakse semantiline analüüs. *semyhe* põhineb eesti WordNeti hüponüümide ja hüperonüümide hierarhial ja on mõeldud ühestamaks kõiki substantiive ja verbe. s*emyhe* leiab sõna tähenduse eesti WordNeti hüponüümia/hüperonüümi puust. Praegune versioon peaks leidma igale morfoloogiliselt ühestatud sõnale ühe vaste. Morfoloogilise analüüsi teeb programm ESTMORF. s*emyhe* kasutab oma tesauruse kuju. Väljundis märgitakse morfoloogilise informatsiooni kõrvale sõnade semantika kirjeldus – tähendusnumbrid. (Lõpparuanne 2000–2002)

[10] http://www.psych.ut.ee/~kaarel/eng_semyhe/

Sõnatähenduste ühestamise algoritm on sama nii substantiividele kui verbidele. Ühestamine toimub kahes jaos: kõigepealt ühestatakse substantiivid ja seejärel verbid või vastupidises järjekorras.

4. Tähenduserinevused sõnatähendustes

4.1. Uurimuse metoodika

Analüüsi aluseks on võetud erinevuste failid ajavahemikust 2000. aasta september kuni 2004. aasta veebruar. Kokku oli 39 erinevuste faili. Igas failis oli keskmiselt 180 erimeelsuste rida (kokku ~10 000 rida).

Tähendusklaster on ühetüübiliste tähenduste kobar, mille moodustavad käsitsiühestajate eriarvamustest saadud sõnatähenduste paarid

Erinevuste alamkorpusesse on sisestatud iga fail eraldi, sest on arvestatud fenomeni "üks tähendus ühe teksti kohta" (vt ka ptk 2.1.5.). Kui kaks ühestajat on olnud eriarvamusel, siis on erinevuste failis läbivalt kaks tähendust ning seega on antud lõputöös mõttekas analüüsida sõnade esinemist ühe teksti piires.

```
piirama:925:9
 tüdruku
-    tüdruk+0 //_S_ com sg gen // 1 @ tüdruk:2436:3
+    tüdruk+0 //_S_ com sg gen // 3 @ tüdruk:2436:3
 sisse
-    sisse+0 //_D_ //
.....
.....
.....
tüdruku
-    tüdruk+0 //_S_ com sg gen // 1 @ tüdruk:2436:3
+    tüdruk+0 //_S_ com sg gen // 3 @ tüdruk:2436:3
 käest
     käsi+st //_S_ com sg el // 1 @ käsi:1215:5
 kinni
@@ -1212,7 +1212,7 @@
```

Näide: kuidas üks kindel tähendus esineb ühe teksti kohta.

Sisestatud on lekseemid, mitte sõnavormid. Kasutatud on andmetöötlusprogrammi R[11]. Seejärel on kõik sõnad koos vaieldavate tähendusnumbritega sisestatud ühte faili (vt Lisa 3. CD), milles kokku on 4659 sõna koos võimalike tähenduserinevustega. Järgnevalt on moodustatud sõnade esinemissageduste tabel, kokku 983 lekseemi (vt

[11] http://www.r-project.org/

Lisa 3. CD). Erinevuste alamkorpuses suure esinemissagedusega ning ühestatud korpuses sageli esinevate sõnade seast on valitud 50 verbi ja 24 substantiivi.

Iga sõnaartikli juurde on moodustatud tabel, milles on esitatud sõna kõik võimalikud tähendusklastrid (tabeli päises märgitud ***sõna+täh***) ning esinemissagedus erinevuste alamkorpuses (tabelis märgitud ***sagedus***).

Iga sõnaartikli juures on ka joonis, milles on esitatud sõna kõik tähendused, sünohulgad (sulgudes ja kaldkirjas) ning hüperonüüm(id). Hüperonüümiasuhe on esitatud märgiga '=>'. Vajadusel on joonisele kantud ka seletus ja/või näide. Hüperonüümid, sünohulgad, seletused ja näited on võetud TEKsaurusest. Analüüsitud on tähendusklastreid, mis esinevad rohkem kui viis korda erinevuste alamkorpuses. Väiksema sagedusega tähendusklastrid on loetud juhuslikeks või vähemtähtsateks.

Välja on toodud:

- Autohüponüümia suhted. Autohüponüümia viitab sõnadele, mille tähendused paaruvad nii hüperonüümi kui hüponüümina. Üks sõna tähendus on teisele sama sõna tähendusele hüperonüümiks. (Peters, Peters 1998)
- Ko-hüponüümia (ingl *sisters*), mille puhul on sõnatähendustel üks ja seesama hüperonüüm. (Peters, Peters 1998)
- Osaline ko-hüponüümia, mille puhul sõnatähenduste mingi hulk hüperonüüme kattuvad.

Kõik TEKsaurusest puuduvad näitelaused on leitud ühestatud korpusest.

4.2. Verbid

OLEMA

olema8
=>olema1 (*käima24*)
olema4 (*olemas olema1, olelema2, eksisteerima2)*
=>olema3 (*viibima1*)
=>olema5 (*püsima1, jääma2*)
=> olema2 (*paigal püsima1, jääma2)*
=>olema7 (*asuma2, asetsema1, paiknema1*)
sobima 4, kokku sobima 1, vastavuses olema 1, sobiv olema 1, vastama 2
=>olema6 (*võrdne olema1, ühtima2, võrduma1***)**
olema9; seletus: *kellegi või millegi valduses, omanduses, käsutuses, kellegagi või millegagi ühtekuuluvana*

Joonis 1.

sõna+täh	***sagedus***
olema;4,8	197
olema;4,9	63
olema;8,9	50
olema;2,8	33
olema;4,7	32
olema;3,7	19
olema;3,4	17
olema;3,8	15
olema;5,8	15
olema;7,8	15
olema;6,8	12
olema;2,9	9
olema;2,4	8
olema;1,8	5
olema;7,9	5
olema;6,9	4
olema;4,5	3
olema;1,4	2
olema;3,9	2
olema;4,6	2
olema;5,7	2
olema;1,9	1
olema;2,+1	1
olema;2,3	1
olema;4,+1	1
olema;5,6	1
olema;5,9	1
olema;6,7	1
olema;8,+1	1

Tabel 1.

olema4, *olema8* ja *olema9* on tipmised mõisted ja seega peab neid tähendusi selgelt eristama; arvestada tuleb ka seda, et verb *olema* esineb kõige sagedamini.

olema8 esineb ühestatud korpuses kõige rohkem, ta on ka kõige üldisema tähendusega; järgnevad *olema4* ja *olema9*.

olema5, *olema3* ja *olema7* on ko-hüponüümid, s.t et neil on samad hüperonüümid. *olema5* on *olema2* hüperonüüm ja *olema8* ning *olema1* vahel esineb autohüponüümia.

Tipmiste tähenduste märgendamisel tuleks lisaks süntaksile arvestada ka sõna konkreetse tähenduse sagedusi. *olema4*, *olema8* ja *olema9* eristamiseks tuleb määrata lausetüüp:

- *olema4* on eksistentsiaallauses; väljendab subjekti üldist, antud hetkel või minevikus eksisteerimise fakti; nt: *nüüd on seal maju nagu seeni*, kus lause algul on aega ja kohta väljendavad määrused, grammatiline subjekt on lause lõpul;
- *olema8* on predikatiivlauseis, iseseisva tähenduseta või väga üldise tähendusega olles iseloomustab koos predikatiivi või adverbiaaliga subjekti omadust, olemust, tunnust, kuuluvust jne; nt: *sõprussuhted on mõnikord tormilise loomuga*, kus omadussõnaliseks öeldistäiteks on nimisõnafraas kaasaütlevas;
- *olema9* on possesiivlauseis; kellegi või millegi valduses, omanduses, käsutuses, kellegagi või millegagi ühtekuuluvana; nt: *neil olid kõverad tiivad*, kus lausealguline tegevussubjekt on määrus (*neil*) ning grammatiline subjekt on lause lõpul (*tiivad*).

olema4 korral on võimalik asendada sõna *olema* sõnaga *eksisteerima*; *olema8* esineb ühestatud korpuses kõige sagedamini (vt Lisa 1), seega tuleks seda tähendusnumbrit kahtluse korral eelistada.

olema4 ja *olema9* tähendused on TEKsauruses selgesti eristatavad, kuid tekstis ei tule ilmselt iga kord erinevus välja. Näiteks lauses *isegi Kristuse näol on mingi pide*.

olema8 ja *olema9* eristamiseks tuleb määrata lausetüüp; kahtluse korral peaks olema eelistatud *olema8*. Mõnedes kohtades võib tekkida raskusi lausetüübi määramisel, nt: *temal olevat väga tähtis jutt; töölkäijatel on korraline puhkus*.

Ühestatud korpuses esineb *olema2* suhteliselt harva, seega võrdluses *olema8*-ga tuleb eelistada *olema8*-t. *olema2* valikul peaks proovima asendada sõna *olema* sünohulgas olevate sõnadega. *olema4* ja *olema3* ning *olema4* ja *olema7* korral peaks märgendamisel valima hüperonüümi (antud juhul olema4), sest hüperonüüm on üldisema tähendusega ja esineb ühestatud korpuses sagedamini (vt Lisa 1). *olema3* ja *olema7* eristamise raskus on tingitud sellest, et tegemist on ko-hüponüümidega; nende tähenduste eristamiseks tuleb vaadelda sünohulka, seletusi ning näitelauseid, et valida sobivaim tähendus kindlas kontekstis.

Suhteliselt sagedalt esinevad erimeelused *olema8* eristamisel *olema3*-st, *olema5*-st ning *olema7*-st. Sõna *olema* tähendused 3, 5, 7 on olema4 hüponüümid ning seega esinevad

eksistentsiaallauses. *olema6* ja *olema8* eristamist hõlbustaks TEKsauruses puuduv *olema6* näide (ühestatud korpusest näitelause: *tema püütud heeringad olid kui täismehe käsivarred*). *olema2* ja *olema9* tähenduserinevus on TEKsauruses selge. *olema2* ja *olema4* korral tuleks eelistada hüperonüümi (antud juhul *olema4*).

SAAMA

kätte saama1, saavutama1, võitma1
=>saama4 (*omandama2*)
=> saama1 (*pärima3, omandama3*)
tunda saama 2, läbi elama 2, kogema 3 **=> saama2**
=> saama5 (*hankima2, nõutama1, soetama1, muretsema1*)
teisenema1, muutuma1
=>saama7
=>saama3 (*jääma4, muutuma3, minema8*)
saama8 (*tulema1*)

saama10 (*võima1*); seletus: *võimeline olema, võimalik olema, väljendab tegevuse võimalikkust ja subjekti võimelisust v. võimalust selleks*

saama12; seletus**:** *esineb sisult 1. isikut (v. impersonaali) esindavates passiivilausetes (olevikus; lihtminevikus)*
korda minema1, õnnestuma1
=>saama6
võima2, tohtima1
=>saama11
jääma4, muutuma3, minema8, saama3
=>saama9 (*jõudma3*)

Joonis 2.

sõna+täh	***sagedus***
saama;10,11	24
saama;10,9	9
saama;10,2	8
saama;2,4	7
saama;3,8	6
saama;10,6	5
saama;6,9	5
saama;10,3	4
saama;10,7	4
saama;2,8	4
saama;2,9	4
saama;3,7	4
saama;3,9	4
saama;4,5	4

saama;2,3	3
saama;6,8	3
saama;7,8	3
saama;1,3	2
saama;1,41	2
saama;10,12	2
saama;11,1	2
saama;2,+1	2
saama;8,12	2
saama;9,11	2
saama;1,6	1
saama;1,9	1
saama;10,+1	1
saama;10,1	1
saama;10,5	1

saama;10,8	1
saama;11,2	1
saama;11,4	1
saama;11,5	1
saama;12,+1	1
saama;2,6	1
saama;2,7	1
saama;4,6	1
saama;4,7	1
saama;5,6	1
saama;5,9	1
saama;6,7	1
saama;7,9	1
saama;8,11	1
saama;9,2	1

Tabel 2.

saama4 on autohüponüümia suhtes *saama1*-ga, *saama2*-ga ja *saama5*-ga.

saama3 ja *saama7* on ko-hüponüümid.

saama8, *saama10*, *saama12* on tipmised mõisted.

Erinevuste seas on kõige sagedasemad *saama10* ja *saama11*, mis on seletatav sellega, et TEKsauruses esitatud seletused kattuvad osaliselt: mõlema seletuses on fraas *võimalik olema. saama10* esineb ühestatud korpuses sagedamini kui *saama11* (vt Lisa 1). Keeruline on valida sobivat tähendust nt selliseis lauseis: *kevadel ei saa enam loomi Kraakovi mäele saata; midagi ei saa teha vastu naise tahtmist*.

saama10 ja *saama9* on TEKsauruses tähendusvahega, ilmselt pole tekstis tähenduste erinevus mitte nii selgelt tajutav. Nt lauseis: *saavad kui tahavad; siin saavad kõik kõikjale sisse; ega soldatite kartusel saanud jääalust kalapüüki ära jätta; Peeter ei saanud lauset lõpetada, sest...*

Kui märgendamisel kahelda *saama2* ja *saama4* vahel, tuleks eelistada hüperonüümi: *saama4*, kuigi tähendused on tajutavalt sarnased ning esinevad ühestatud korpuses enam-vähem võrdselt (vt Lisa 1).

saama3 ja *saama8* tähenduserinevus TEKsauruses on selge. Mõnes lauses, nt: *uueks asupaigaks sai ühiselamu; vastuseks saime, et...*, võib tekkida raskusi sobiva tähendusnumbri valimisel, sest *saama3* näitelausetes on rõhutatud saava käände olemasolu (*tahan saada näitlejaks; tuju läks halvaks; sain tigedaks*).

saama2 ja *saama10* eristamisel on võimalik arvestada, et *saama10* esineb sagedamini ning on seega eelistatum (vt Lisa 1).

TULEMA

tulema1 (*saama8*)

tekkima1
=>tulema2

liikuma3
=>tulema3 (*kohale jõudma1, jõudma2, pärale jõudma1, saabuma3*)
kulgema2, liikuma2, siirduma2
=>tulema4

evima1, omama2, vajama2, vaja olema1, tarvis olema1
=>tulema5 (*pidama1*)

Joonis 3.

sõna+täh	*sagedus*
tulema;3,4	32
tulema;1,4	16
tulema;1,2	11
tulema;2,3	8
tulema;2,4	6
tulema;1,3	4
tulema;1,5	3
tulema;2,5	3
tulema;4,+1	3
tulema;1,+1	2
tulema;1,1	2
tulema;3,5	2
tulema;2,+1	1
tulema;3,+1	1
tulema;4,5	1
tulema;5,+1	1

Tabel 3.

tulema1 on tipmine mõiste.

tulema3 ja *tulema4* hüperonüüm on sama sõna, aga erinev tähendusnumber. *tulema3* ja *tulema4* eristamiseks: *tulema4* on üldisem, ilma sihtpunktita liikumine.

tulema1 ja *tulema4* eristamiseks: *tulema1* peab olema võimalik asendada sõnaga *saama*.

PIDAMA

pidama7; seletus: *osutab mingi asjaolu esinemise v. millegi toimumise tõenäosusele v. võimalikkusele*

pidama8; seletus: *osutab millelegi, mis oleks võinud juhtuda v. peaaegu oleks juhtunud*

säilitama1, talletama1, hoidma6
=>**pidama3** (*hoidma2*)
=>**pidama9**
=>**pidama12** (*kinni pidama5, kinni hoidma3*)
=>**pidama13**

oletama1, uskuma1, arvama2, mõtlema3
=>**pidama2** (lugema1, arvama5)
=>**pidama4** (*välja tegema1, hoolima1, arvestama4*)

evima1, omama2, vajama2, vaja olema1, tarvis olema1
=>**pidama1** (*tulema*5)
vajama2, vaja olema1, tarvis olema1
=>**pidama5**

rääkima6
=>**pidama14**

katsuma2, püüdma5, proovima4, üritama1
=>**pidama11** (*talitsema2, ohjeldama2, takistama3, tagasi hoidma2*)

kavandama1, planeerima1, plaanima1
=>**pidama6** (*kavatsema1, plaanitsema1, mõtlema4*)

meelt lahutama2, lõbutsema1
=>**pidama10** (*tähistama3, pühitsema1*)

Joonis 4.

sõna+täh	***sagedus***
pidama;1,5	35
pidama;5,7	8
pidama;1,7	7
pidama;1,11	4
pidama;1,4	4
pidama;1,2	3
pidama;1,6	3
pidama;2,4	3
pidama;4,5	3
pidama;1,3	2
pidama;2,5	2
pidama;5,6	2
pidama;5,8	2
pidama;6,7	2
pidama;1,+1	1
pidama;10,+1	1
pidama;11,1	1
pidama;11,12	1
pidama;11,2	1
pidama;11,3	1
pidama;11,5	1
pidama;11,7	1
pidama;4,+1	1
pidama;4,7	1

Tabel 4.

pidama7 ja *pidama8* on tipmised mõisted.

pidama3 on *pidama9*, *pidama12* ja *pidama13* suhtes hüperonüüm. *pidama9*, *pidama12* ja *pidama13* on ko-hüponüümid.

pidama2 ja *pidama4* on autohüponüümia suhtes (*pidama2* on üheks hüperonüümiks).

pidama1 ja *pidama5* on osalised ko-hüponüümid.

pidama1 ja *pidama5* hüperonüümid kattuvad osaliselt ning seega võib tähenduste eristamine olla segadusttekitav. TEKsauruses on tähendus eristatav, mitte-leksikograafi taustaga inimesel on ilmselt mõnel juhul tekstis raske vahet teha, nt lauses: *pidid oma koolipõlve ülbuse eest kallilt maksma*.

pidama5 ja *pidama7* tähendused on TEKsauruses eristatavad, samuti ka *pidama1* ja *pidama7*.

TEGEMA

tegema5 (*sooritama4, teostam 4*)
=>tegema3 (*tegutsema3, toimima4*)

muutma2
=>tegema6 (*tingima3, tekitama3, põhjustama1*)
=>tegema7 (*looma5, esile kutsuma2, tekitama4*)

meeli ärritama1
=>tegema4 (*häält tegema1, kõlama2*)

looma2, valmistama1
=>tegema8 (*valmistama3*)

Joonis 5.

sõna+täh	*sagedus*
tegema;3,5	22
tegema;5,7	12
tegema;6,7	10
tegema;5,6	9
tegema;6,8	6
tegema;5,8	5
tegema;6,5	5
tegema;3,6	4
tegema;5,3	4
tegema;6,+1	3
tegema;3,7	2
tegema;3,8	2
tegema;1,2	1
tegema;1,5	1
tegema;4,6	1
tegema;4,8	1

Tabel 5.

tegema5 on tipmine mõiste.

tegema3 ja *tegema5* vahel on autohüponüümia suhe; *tegema5* on ülemmõiste *tegema3* suhtes. Ühestatud korpuses esineb tegema5 sagedamini, seega märgendamisel tuleks eelistada seda.

tegema6 ja *tegema7* on autohüponüümia suhtes (*tegema6* on üheks hüperonüümiks); märgendamisel eelistada *tegema6*.

tegema5 ja *tegema7* on TEKsauruses selgelt eristatava vahega; märgendamisel tuleks eelistada *tegema5*, sest see esineb ühestatud korpuses märksa sagedamini (vt Lisa 1).

tegema6 ja *tegema8* eristamist aitaks TEKsaurusesse *tegema8* näite lisamine (ühestatud korpusest: *plastiliinist loomi tegema*).

MINEMA

minema6 (*kuluma2*)
arenema1
=>minema9 (*jooksma6, edenema3, sujuma1, laabuma1*)
=>**minema17** (*kulgema1, käima9*)

suunduma1, tüürima1, siirduma1
=>minema3
=>**minema4** (*lahkuma1*)
sobima4, kokku sobima1, vastavuses olema1, sobiv olema1, vastama2
=>minema11 (*hakkama10, kõlbama1, sobima3*)
=>**minema12** (*mahtuma2*)

jooksma6, edenema3, sujuma1, minema9, laabuma1
=>minema14 (*müüma1, müübima2, kaubaks minema1*)

vajama2, vaja olema1, tarvis olema1
=>minema13 (*kuluma3*)

kulgema2, liikuma2, siirduma2
=>minema1

minema16 (*etenduma2*)
teisenema1, muutuma1
=>minema8 (*jääma4, muutuma3, saama3*)

asuma2, asetsema1, paiknema1, olema7
=>minema15 (*viima3*)

eemalduma1, kaugenema1
=>minema2 (*lahkuma3, ära minema2*)

tekkima1
=>minema10 (*kujunema1, juhtuma2*)

kaotsi minema1, kaduma6
=>minema5 (*haihtuma2*)

edenema1, edasi liikuma1, edasi minema2
=>minema7 (*kuluma1, kaduma9, mööduma1*)

Joonis 6.

sõna+täh	***sagedus***
minema;3,4	13
minema;1,3	11
minema;13,5	7
minema;1,4	5
minema;10,3	2
minema;13,3	2
minema;2,3	2
minema;2,4	2
minema;3,8	2
minema;4,8	2
minema;1,2	1

minema;1,8	1
minema;10,8	1
minema;12,3	1
minema;13,1	1
minema;14,5	1
minema;3,+1	1
minema;4,5	1
minema;4,9	1
minema;7,9	1
minema;8,9	1
minema;9,17	1

Tabel 6.

minema6 ja *minema16* on tipmised mõisted.

minema3 ja *minema4* on ko-hüponüümid, mille tõttu võib neid tekstis olla raske eristada; eelistatud *minema3* (vt Lisa 1); *minema3* üldisem, *minema4* puhul lahkutakse.

minema1-l ja *minema3*-l on sama sõna hüperonüümiks, aga erineva tähendusega ning tähenduste eristamisel oleks eelistatum *minema3*, mis esineb ühestatud korpuses palju sagedamini. Nii TEKsauruses kui tekstis on tähendusvahe olemas, kuid pealiskaudsel süvenemisel antud tähendustesse ei ole tõepoolest võimalik alati vahet teha. Muidugi esineb ka üksikuid lauseid, kus on peaaegu võimatu üht ja ainsat tähendust valida; nt: *...kui Joona isa mõne aja pärast puhkama läks...*

minema11 ja *minema12* on ko-hüponüümid.

minema9 ja *minema17* vahel on autohüponüümia suhe (*minema9* on üheks hüperonüümiks).

minema5 ja *minema13* on TEKsauruses selge tähendusvahega.

minema1 ja *minema4* üks hüperonüüm on sama, aga erineva tähendusnumbriga.

Minema9 on *minema14* üheks hüperonüümiks.

JÄÄMA

eksisteerima2, olemas olema1, olema4, olelema2
=>**jääma1** (*püsima2, olema5*)
=>**jääma2** (*paigal püsima1, olema2*)
säilima1, seisma1
=>**jääma3**

teisenema1, muutuma1
=>**jääma4** (*muutuma3, minema8, saama3*)

olema8
=>**jääma5**

Joonis 7.

sõna+täh	***sagedus***
jääma;1,5	16
jääma;1,2	9
jääma;2,5	7
jääma;4,5	7
jääma;1,3	4
jääma;1,4	4
jääma;3,4	4
jääma;3,5	3
jääma;2,3	1
jääma;5,+1	1

Tabel 7.

jääma1 ja *jääma2* on autohüponüümia suhtes (*jääma1* on hüperonüüm *jääma2* suhtes); tähenduste märgendamisel tuleks eelistada *jääma1*.

jääma1 ja *jääma5* on sarnase tähendusega ja esinevad ühestatud korpuses suhteliselt võrdselt. TEKsauruses on vahe eristatav, tekstis on võimalik määrata tähendusnumbrit 1, kuid antud töö autori keeletaju ei suuda eristada täpselt *jääma5*. Nt lauses *mõistmatult jäi ta särava ja kumava Elleni otsa vaatama* või *jäid teda uskuma ka või!* Mõlemas lauses on sõna *jääma* saanud tähendusnumbri 5.

Sõna *jääma* tähendustest on täiesti selgelt tajutav tähendus *jääma4*-l ja *jääma3*-l; teised tähendused paistavad olema tekstis raskesti eristatavad, sest erinevuste korpusest ilmneb, et peaaegu kõik tähendused kombineeruvad (nt *jääma5* kombineerub kõikide teiste tähendustega, mis viitab osaliselt sellele, et see tähendus on märgendajatele ebaselge).

HAKKAMA

hakkama5 (*algama2, pihta hakkama1, peale hakkama1*)
=>**hakkama2** (*asuma4*)

kinnituma2, fikseeruma1
=>ha**kkama6** (*kinni jääma2, kinni võtm 2*)
=>**hakkama9** (*nakkama1*)
jääma4, muutuma3, minema8, saama3
=>**hakkama3**
=>**hakkama4**

mõju avaldama4, stimuleerima1, mõjuma1, mõjustama3, toimima1
=>**hakkama7**

võima1, saama10
=>**hakkama8**

sobima4, kokku sobima1, vastavuses olema1, sobiv olema1, vastama2
=>**hakkama10** (*kõlbama1, minema11, sobima3*)

Joonis 8.

sõna+täh	***sagedus***
hakkama;2,3	17
hakkama;2,5	10
hakkama;2,6	9
hakkama;3,5	3
hakkama;10,2	2
hakkama;2,4	2
hakkama;1,3	1
hakkama;2,8	1
hakkama;4,5	1

Tabel 8.

hakkama5 on tipmine mõiste.

hakkama5 ja *hakkama2* on autohüponüümia suhtes (*hakkama5* on üks hüperonüümidest).

hakkama6 ja *hakkama9* on autohüponüümia suhtes (*hakkama6* on üks hüperonüümidest).

hakkama3 ja *hakkama4* on ko-hüponüümid.

hakkama2 ja *hakkama3* eristus TEKsauruses on selge, kuna *hakkama2* esineb märgatavalt sagedamini, tuleks eelistada märgendamisel seda (vt Lisa 1).

Sõna *hakkama* tähendusnumbritest esineb kõige sagedamini *hakkama2* (hakkama+*ma*-infinitiiv, vt ka Lisa 2).

VÕTMA

võtma1; seletus: *midagi enda kätte saama, omastama*

vajama1, nõudma3, tahtma2
=>**võtma2** (*nõudma6, neelama5*)

suhtlema1, lävima1
=>**võtma3** (*närima2, kraaksuma2, kärisema3*)

muutma2
=>**võtma4**; (seletus puudub)

ära võtma2, eemaldama2, kõrvaldama1
=>**võtma5** (*tõstma4, ammutama1*)

mõtlema 1
=>v**õtma6** (*mõistma4, käsitama1)*

liigutama2
=>**võtma7** (*aset võtma2, asuma5*)

panema4
=>v**õtma8**

Joonis 9.

sõna+täh	*sagedus*
võtma;1,4	15
võtma;1,5	4
võtma;1,3	3
võtma;3,+1	3
võtma;3,6	3
võtma;1,7	2
võtma;4,7	2
võtma;1,+1	1
võtma;1,6	1
võtma;1,8	1
võtma;2,5	1
võtma;3,4	1
võtma;4,6	1
võtma;4,8	1
võtma;5,+1	1
võtma;5,7	1
võtma;6,+1	1
võtma;8,+1	1

Tabel 9.

võtma1 on tipmine mõiste.

võtma1 ja *võtma4* on väga sarnased, esinevad ka ühestatud korpuses enam-vähem võrdselt (vt Lisa 1). Tundub, et *võtma1* seletus peaks käima hoopis *võtma4* alla ning *võtma1* ei ole mitte konkreetselt kätte võtma, vaid abstraktsem. Praegu on ühestatud korpuses märgendatud tähendusnumbriga 1 sellised laused: *võtke siis mind kah juba ligi; ma tahtsin võtta auto* ning tähendusnumbriga 4 sellised laused: *ta võtab taskust kirja; võtab sigareti; läksin toru võtma*.

MÕTLEMA

mõtlema1; seletus: *ajutegevuse abil omadusi ja seoseid leidma, järeldusi ja otsustusi tegema*
=>**mõtlema6** (*kujutlema1, ette kujutama1*)

mõtlema3 (ol*etama1, uskuma1, arvama2*)

teavet andma1
=>**mõtlema2** (*silmas pidama2*)

kindlaks määrama1, kinnistama1, fikseerima1, määrama1
=>**mõtlema5** (*määrama6, ette nägema1*)

kavandama1, planeerima1, plaanima1
=>**mõtlema4** (*kavatsema1, pidama6, plaanitsema1*)

Joonis 10.

sõna+täh	*sagedus*
mõtlema;1,3	16
mõtlema;1,6	15
mõtlema;1,2	5
mõtlema;1,4	4
mõtlema;3,6	3
mõtlema;2,3	1

Tabel 10.

mõtlema1 ja *mõtlema3* on tipmised mõisted.

mõtlema1 ja *mõtlema6* on autohüponüümia suhtes (*mõtlema6* on hüperonüüm).

TEKsauruses on *mõtlema1* ja *mõtlema3* tähendused eristatavad; kahtluse korral on *mõtlema1* eelistatud, sest esineb ühestatud korpuses sagedamini. Mõnes lauses on sõna *mõtlema* tähendust raske määrata, nt: *nüüd mõtlen ma, et võib-olla oli see...* (ühestatud korpuses märgendatud kui *mõtlema3*) vrd *...ja mõtles, et elus tuleb kõik ikka väikese hilinemisega kätte* (märgendatud tähendusnumbriga 1).

Samuti on eristatav *mõtlema1* ja *mõtlema6* vahe; märgendamisel tuleks eelistada *mõtlema1*, sest viimatimainitu on hüperonüüm ning esineb ühestatud korpuses ka sagedamini.

Ilmneb, et *mõtlema1* kui tipmine mõiste on märgendajatele raske tajuda, kuivõrd *mõtlema* tähenduses 1 on võimeline kombineeruma peaaegu kõigi teiste *mõtlema* tähendustega.

TEADMA

teadma1 (*teadlik olema1*)

võima1, saama10
=>**teadma2** (*oskam2*)

olema8
=>**teadma3** (*tundma4*)

Joonis 11.

sõna+täh	***sagedus***
teadma;1,3	22
teadma;1,2	11
teadma;1,+1	4
teadma;2,3	3

Tabel 11.

teadma1 on tipmine mõiste.

teadma1-le on seletust vaja, siis on ehk selgem, aga TEKsauruses on *teadma1* ja *teadma3* tähendused eristatavad. Nagu on näha erinevuste failidest, on *teadma1* kõige raskemini eristatav tähendus, sest see kombineerub ka ülejäänud kahe tähendusega.

Üksikutes lausetes võib tekkida probleeme *teadma1* ja *teadma3* valikul, nt: *mida me õieti teame teistest inimestest*.

Märgendamisel tuleks eelistada *teadma1*, sest selles tähenduses esineb sõna *teadma* kõige sagedamini (vt Lisa 1).

RÄÄKIMA

suhtlema 1, lävima 1
=>**rääkima1** (*kõnelema 1*)
valdama 2, oskama 1, mõistma 3, asja tundma 1,kõnelema 1, **rääkima 1**
=>**rääkima3**

edasi andma 1, edastama 1, väljendama 4
=>**rääkima2** (*kõnelema 2*)
väljendama 1, teatama 1, avaldama 3, edasi andma 1, edastama 1, väljendama 4
=>r**ääkima4** (*kõnelema 3*)

informeerima 1, teada andma 1, teatama 2
=>**rääkima5**

esinema 2
=>**rääkima6**

Joonis 12.

sõna+täh	***sagedus***
rääkima;4,5	8
rääkima;1,5	7
rääkima;2,5	7
rääkima;1,2	3
rääkima;1,4	2
rääkima;2,4	2
rääkima;3,4	2
rääkima;3,5	2
rääkima;5,6	2
rääkima;1,3	1
rääkima;2,6	1
rääkima;3,6	1

Tabel 12.

rääkima1 ja *rääkima3* on autohüponüümia suhtes (*rääkima1* on üheks hüperonüümiks). *rääkima4* ja *rääkima5* üheks hüperonüümiks on sama sõna, aga erineva tähendusnumbriga. Tähendused on eristatavad nii TEKsauruses kui ka tekstis, kuid selleks, et tähenduste erinevusest aru saada, peaks olema rohkem näiteid, rohkem kogemusi tähenduste eristamise vallas.

Antud sõna tähendusnumbrite vahel tekib tihe klaster – kõik tähendused kombineeruvad kõigiga. Sellest võib järeldada, et sõna *rääkima* tähendused on liiga peeneks tehtud, üle-eristatud.

ANDMA

teavet andma 1
=>**andma6**
=>**andma13** (*esitama 7*)

liigutama 2
=>**andma1** (*ulatama 2*)
=>**andma2** (*loovutama 2*)

võimaldama 2, lubama 2
=>**andma3** (*pakkuma 2*)

tingima 3, tegema 6, tekitama 3, põhjustama 1
=>**andma4** (*tekitama 2*)

valmistama 2, tootma 3, produtseerima 3
=>**andma5** (*produtseerima 2, tootma 2*)

väljendama 1, teatama 1, avaldama 3
=>**andma7**

peksma 2, lööma 1
=>**andma8** (*virutama 2, panema 8, äigama 1*)

korraldama 3
=>**andma9** (*korraldama4*)
tegema 5, sooritama 4, teostama 4
=>**andma10**

kindlaks määrama 1, kinnistama 1, fikseerima 1, määrama 1
=>**andma11** (*omistama 1*)

pingutama 3, vaeva nägema 1
=>**andma12**

Joonis 13.

Sõna+täh	*sagedus*
andma;3,4	7
andma;6,7	4
andma;1,2	3
andma;10,3	3
andma;1,3	2
andma;10,11	2
andma;2,3	2
andma;4,+1	2
andma;1,12	1
andma;1,7	1
andma;10,2	1
andma;10,6	1
andma;11,12	1
andma;2,6	1
andma;2,7	1
andma;3,9	1
andma;4,6	1

Tabel 13.

andma1 ja *andma2* on ko-hüponüümid.

andma6 ja *andma13* on ko-hüponüümid.

TEKsauruses on *andma3* ja *andma4* selgelt erineva tähendusega, ühestatud korpuses esinevad nad suhteliselt võrdselt. Andma3-e puhul on võimalik asendada sünohulgas oleva sõnaga *pakkuma*.

andma6-l ja *andma7*-l on küllalt sarnane näitelause; *andma6*: *Ajalehed, raadio, televisioon annavad informatsiooni* vrd *andma7*: *Ajalehed annavad olulist infot*.

Kuigi antud sõnal on palju erinevaid tähendusi, s.t tal on kõrge polüseemiatase, ei teki erinevuste failidest olulisi tähendusklastreid, mis annab märku sellest, et TEKsauruses on tähendused selgelt eristatud.

KÄIMA

kulgema 2, liikuma 2, siirduma 2
=>**käima2**
=>**käima4** (*kurseerima 1, sõitma 1*)

liikuma 3
=>**käima5** (*kukkuma 2*)
=>**käima25** (*liikuma1)*

minema 1
=>käima1 (*sammuma 1, kõndima 1, astuma 2*)

kandma 7
=>**käima3**

üle minema 1, ületama 2
=>**käima6** (*läbistama2*)

levima 2
=>**käima7** (*levima 1, liikvel olema 1, ringi liikuma 1*)

tulema 3, kohale jõudma 1, jõudma 2, pärale jõudma 1, saabuma 3
=>**käima8**

jooksma 6, edenema 3, sujuma 1, minema 9, laabuma 1
=>**käima9** (*kulgema 1, minema 17*)

talitlema 1
=>**käima10** (*töötama 2, , funktsioneerima 1*)

tegutsema 3, toimima 4, tegema 3
=>**käima11** (*tegutsema 1, toimima 2*)

toimuma 2, aset leidma 1, juhtuma 3, sündima 2
=>**käima12** (*toimuma 1*)

tegema 5, sooritama 4, teostama 4
=>**käima13**

armatsema 1, armastama 4
=>**käima14** (*kurameerima 1*)

kestma 2, vastu pidama 1
=>**käima15**

teisenduma 1, transformeeruma 1, muunduma 1
=>**käima16** (*käärima 2*)

mõju avaldama 4, stimuleerima 1, mõjuma 1, mõjustama 3, toimima 1
=>**käima17**

puutuma 1, juurde kuuluma 1, seotud olema 1
=>**käima18**

sobima 4, kokku sobima 1, vastavuses olema 1, sobiv olema 1, vastama 2
=>**käima19** (*kuuluma 4*)

edenema 1, edasi liikuma 1, edasi minema 2
=>**käima21**

ihkama 1, himustama 1, ihaldama 1
=>**käima22**

olema 8
=>**käima24** (*olema1*)

Joonis 14.

Sõna+täh	*sagedus*
käima;1,2	5
käima;2,24	3
käima;1,21	2
käima;1,24	2
käima;1,25	2
käima;11,25	2
käima;19,25	2
käima;1,23	1
käima;1,29	1
käima;11,12	1
käima;18,19	1
käima;19,24	1
käima;21,24	1
käima;21,25	1
käima;6,13	1
käima;8,17	1
käima;8,25	1
käima;9,24	1
käima;9,25	1

Tabel 14.

käima2 ja *käima4* on ko-hüponüümid.

käima5 ja *käima25* on ko-hüponüümid.

Sõna *käima* tähenduserinevused on TEKsauruses ja ka tekstis suhteliselt selged; ei esine ko-hüponüüme ega autohüponüümiat. Enim kaheldud tähendused *on käima1* ja *käima2*; nende eristamisel tuleks silmas pidada TEKsauruse seletusi ning, et *käima1* on jalgsi (jalgade peal) edasi liikuma; *käima2* on üldisem. Tekstis tulevad tähendusvahed hoolikamal vaatlemisel välja.

KÜSIMA

suhtlema 1, lävima 1
=>**küsima1** (*küsitlema 2, nõudma 2, pärima 1*)
=>**küsima4;** seletus: *küsimust esitama*
=>**küsima5** (uuri*ma 5, pärima 2*)
=>**küsima6** (*intervjueerima 1, küsitlema 1, usutlema 1*)

taotlust esitama 1, taotlema 3
=>**küsima7** (*paluma 3*)

kontrollima 3
=>**küsima8**

edasi andma 1, edastama 1, väljendama 4
=>**küsima2** (*nõutama 2, paluma 2)*

Joonis 15.

sõna+täh	***sagedus***
küsima;1,4	17
küsima;1,5	4
küsima;4,5	3
küsima;4,7	3
küsima;2,4	2
küsima;1,2	1
küsima;7,+1	1

Tabel 15.

küsima4 on autohüponüüm *küsima6* ja *küsima5* suhtes.

küsima1 ja *küsima4* on ko-hüponüümid.

küsima4 kasutatakse otsese kõne saatelausena.

Kõige segasema tähendusega on *küsima4* ilmselt just selle tõttu, et seletusest on puudu asjaolu, et antud tähendust kasutatakse muuhulgas ka otsese kõne saatelause.

Märgendamisel tuleks eelistada *küsima4*, sest see tähendus esineb ühestatud korpuses kõige sagedamini (vt Lisa 1). TEKsaurusest on puudu *küsima1* näide: *vabandas ja küsis, kas ma olen kuulnud taksojuhi mõrvast.*

JÕUDMA

jääma 4, muutuma 3, minema 8, saama 3
=>**jõudma3** (*saama 9*), seletus: *seisundisse, olekusse, olukorda; kohta, olukorda, seisundisse, asendisse*

liikuma 3
=>**jõudma2** (*tulema 3, kohale jõudma 1, pärale jõudma 1, saabuma 3*)
seletus: *liikumise tulemusena sihtpunkti jõudma*.

suutma 1
=>**jõudma4** (*jaksama 1*)

Joonis 16.

sõna+täh	***sagedus***
jõudma;2,3	19
jõudma;3,+1	3
jõudma;1,3	2
jõudma;3,4	2
jõudma;1,4	1
jõudma;2,4	1
jõudma;4,+1	1

Tabel 16.

jõudma2 ja *jõudma3* tähendused on sarnased; TEKsauruses on tähenduserinevus mingil määral olemas, kuid tekstis ei ole sageli võimalik vahet teha (sageli võivad *jõudma3* tähendused esineda ka *jõudma2* all). Nt lauseis: *kiri jõudis pärale* (ühestatud korpuses tähendusnumbriga 3); *Kill jõudis üles ja lõngutas akent* (tähendusnumbriga 3). Samuti esinevad *jõudma2* ja *jõudma3* tekstis peaaegu võrdselt.

TUNDMA

tundma1 (*tajuma 4, aistima 1*)

tundma3 (*kogema 1, tunnetama 3*)

olema 8
=>**tundma4** (*teadma 3*)

Joonis 17.

sõna+täh	***sagedus***
tundma;1,3	20
tundma;1,4	3
tundma;3,4	3
tundma;1,+1	1

Tabel 17.

tundma1 ja *tundma3* on tipmised mõisted.

TEKsauruses on selged tähendusvahed, ilmselt ei tule tähenduserinevused mõnikord tekstis välja, nt lauseis: *tunnen, et ta mõtleb minule, tunneb end vanana, tunnen õhus midagi pahelist*. Ei ole võimalik öelda, kas on tegemist aistimise või kogemisega. Eristamist hõlbustaks TEKsaurusesse *tundma3* näitelause lisamine (ühestatud korpusest: *tundma tüdimust; tundma end tähtsana*).

LEIDMA

kogema 2, tunda saama 1, läbi elama 1
=>**leidma5** *(pälvima 2, osaks saama 1)* näitelause: *Ettepanek leidis kohe pooldajaid.*
=>**leidma6** *(saavutama 2)* näitelause: *Ettepanek leidis kohe pooldajaid*
mõtlema 1
=>**leidma2** (*märkama 4, avastama 2*)
leidma1 (*avastama 1*)

oletama 1, uskuma 1, arvama 2, mõtlema 3
otsusele jõudma 2, otsustama 2
=>**leidma4** (*arvama 8, arvamusele jõudma 1*)

otsusele jõudma 2, otsustama 2
=>**leidma8** (*järeldust tegema 1, järeldama 1*)

leidma3 (*välja valima 1*)

mahti olema 1, mahti saama 1, mahtima 1
=>**leidma7** (*aega jaguma 1*)

Joonis 18.

sõna+täh	*sagedus*
leidma;1,2	6
leidma;1,3	5
leidma;3,6	3
leidma;1,5	2
leidma;1,4	1
leidma;1,6	1
leidma;1,7	1
leidma;1,8	1
leidma;2,4	1
leidma;3,4	1
leidma;4,5	1
leidma;4,7	1

Tabel 18.

leidma1 ja *leidma3* on tipmised mõisted.

leidma5 ja *leidma6* on autohüponüümia suhtes (*leidma5* on hüperonüüm).

Leidma4 ja leidma8 on osalised hüponüümid.

leidma5 ja *leidma6* näitelaused on identsed (vt joonis 18).

leidma1 ja *leidma2* eristus on raske sellepärast, et mõlema tähenduse sünohulgas on sõna *avastama* (erineva tähendusnumbriga). TEKsauruses on tähendused eristatud ning tähenduserinevused tulevad ka tekstis välja (näiteks tähendusnumbriga 1 on lauseis: *leidsin ühe asja; ei leidnud kinnast kusagilt*; tähendusnumbriga 2 on lauseis: *ei leidnud viga endast; ta leiab end seismast jälle siin*) ; võib-olla peaks *leidma2* saama seletuse, mis oleks märgendamisel abiks.

ÜTLEMA

väljendama 1, teatama 1, avaldama 3
=>**ütlema2** (*lausuma 2, sõnama 1*)

taotlust esitama 1, taotlema 3
=>**ütlema3** (*käskima 4, käsku andma 1*)

Joonis 19.

sõna+täh	*sagedus*
ütlema;2,3	9
ütlema;2,+1	5
ütlema;2,4	4
ütlema;3,4	3
ütlema;1,2	2

Tabel 19.

TEKsauruses on vahed eristatavad; täpsem seletus märgendamisjuhendis aitab vahet teha. *ütlema2* ja *ütlema3* eristamiseks: *ütlema2* esineb otseses kõnes; eelistatum on *ütlema2*, sest see esineb ühestatud korpuses mitmeid kordi sagedamini (vt Lisa 1). Paremini saaks neid eristada, kui lisada TEKsaurusesse *ütlema2* näitelause (ühestatud korpusest: *tahtis midagi öelda; mõtles ning ütles: "Tere").*
Mingil märgendamisperioodil on eksisteerinud TEKsauruses ka *ütlema4*.
Suur +1 hulk viitab tõenäoliselt vajadusele luua *ütlema2* kõrvale mingi uus tähendus, mis seda modifitseeriks.

ARVAMA

arvama2 (*oletama 1, uskuma 1, mõtlema 3*)
=>arvama4 (*oletama 4, kartma 2*)
=>arvama1 (*oletama 2, tõenäoliseks pidama 1, eeldama 2*)
=>arvama5 (*lugema 1, arvama 5, pidama 2*)
otsusele jõudma 2, otsustama 2 **=>arvama8** (*leidma 4, arvamusele jõudma 1*)

kindlaks määrama 1, kinnistama 1, fikseerima 1, määrama 1
=>arvama3 (*loendama 1, lugema 2*)

silmas pidama 3, arvestama 1, arvesse võtma 1
=>arvama6 (*arvestama 2, arvesse võtma 2*)

mõtlema 1
=>arvama7 (*mõistatama 1, ära arvama 1*)

Joonis 20.

sõna+täh	*sagedus*
arvama;1,2	9
arvama;2,8	6
arvama;2,6	3
arvama;1,3	1
arvama;2,+1	1
arvama;2,5	1
arvama;3,6	1

Tabel 20.

arvama2 on tipmine mõiste.

arvama2 hüponüümid on *arvama4*, *arvama1*, *arvama8*.

arvama4 ja *arvama1* ja *arvama5* on ko-hüponüümid (*arvama8* on eelmainitute osaline ko-hüponüüm).

arvama1 ja a*rvama2* eristamiseks: eelistada *arvama2*; see on hüperonüüm ning esineb ühestatud korpuses sagedamini. TEKsaurusest on puudu *arvama1* näide: *ma poleks arvanudki, et tänapäeval nii viletsates tingimustes elatakse; ei osanud arvatagi, et...*

arvama2 ja *arvama8* eristamisel eelistada *arvama2* (*arvama2* on hüperonüüm ja esineb sagedamini).

SEISMA

püsima 2, olema 5, jääma 1
=>**seisma1** (*säilima 1*)
=>**seisma5** (*laagerdama 1*)

paigal püsima 1, olema 2, jääma 2
=>**seisma2** (*püsti olema 1*)

eksisteerima 2, olemas olema 1, olema 4, olelema 2
=>**seisma3** (*seisnema 1*)

loobuma 2, pooleli jätma 1, katkestama 2
=>**seisma4** (*seisma jääma 2, toppama 2, stoppama 1*)

õigustama 2, eest seisma 1, kostma 1, kaitsma 2
=>**seisma6**

olema 8
=>**seisma7**

Joonis 21.

sõna+täh	*sagedus*
seisma;1,2	3
seisma;2,3	3
seisma;2,7	3
seisma;3,+1	3
seisma;2,6	2
seisma;4,6	2
seisma;6,+1	2
seisma;1,+1	1
seisma;1,6	1
seisma;2,4	1
seisma;3,4	1

Tabel 21.

seisma1 ja *seisma5* on ko-hüponüümid.

Märgendamisel tuleks eelistada *seisma2*, sest esineb ühestatud korpuses kõige sagedamini (vt Lisa 1). TEKsauruses on puudu näitelaused; *seisma1*: *see seisis ammusest ajast asutuse kontos*, *seisma4: jätsin auto seisma*, *seisma2: turistid seisavad kohviku järjekorras*, *seisma5* ei esinenud ühestatud korpuses kordagi.

Erimeelsuste korpuse põhjal saab öelda, et antud sõna tähendused on TEKsauruses selgelt eristatud, sest ei teki märkimisväärseid tähendusklastreid.

ELAMA

eksisteerima 2, olemas olema 1, olema 4, olelema 2

=>**elama1** (*elus olema 1*)

=>**elama4** (*toituma 2, elatuma 1)*

=>**elama2;** seletus: *päev-päevalt kulgevast eluprotsessist osa võtma (hrl. täpsustatult elamise laadi, viisi, tingimuste vms. poolest)*

asuma 2, asetsema 1, paiknema 1, olema 7

=>**elama3** (*elunema 1, asuma 3, elutsema 1*)

Joonis 22.

sõna+täh	*sagedus*
elama;1,3	11
elama;1,2	7
elama;2,3	3

Tabel 22.

elama1 ja *elama2* on ko-hüponüümid, mistõttu võib nende tähenduste eristamine olla komplitseeritud.

elama1 ja *elama4* on autohüponüümia suhtes (*elama1* on üheks hüperonüümiks).

TEKsauruses on *elama1* ja *elama3* selge tähendusvahega; tekstis tekivad mõningad juhud, kus on raske tähendust eristada, nt lauses: *seda lootust, et isa veel kuskil elaks, enam pole*. Selles lauses võib tekkida küsimus, kas *isa elab* (vastandina surnud olemisele) või kas *isa elab kuskil* (elutseb ja asub kuskil).

OSKAMA

võima 1, saama 10
=>**oskama1** (*valdama 2, mõistma 3, asja tundma 1*)
=>**oskama2** (*teadma 2*)

Joonis 23.

sõna+täh	***sagedus***
oskama;1,2	19
oskama;2,+1	2

Tabel 23.

oskama1 ja *oskama2* on ko-hüponüümid.

TEKsauruse tähendusvahe on selge, mis ei tule tekstis välja, nt lauseis: *...et oskavad nendega ringi käia; sõimata oskab värvikalt*. TEKsauruses puudub *oskama1* näide (ühestatud korpusest: *oskab lugeda*).

SÕITMA

kulgema 2, liikuma 2, siirduma 2
=>**sõitma1** (*kurseerima 1, käima 4*)
=>**sõitma3;** seletus: *sõidutatud saama*
=>**sõitma4;** seletus: *liikuma kindlas suunas v. kindla esmärgiga (ka piltl)*
minema 1
=>**sõitma2**

Joonis 24.

sõna+täh	*sagedus*
sõitma;2,4	10
sõitma;1,2	3
sõitma;2,3	3
sõitma;3,4	3
sõitma;1,4	1

Tabel 24.

sõitma1, *sõitma3* ja *sõitma4* on ko-hüponüümid.

Sõna *sõitma* tähendustest peaks eelistatuim olema *sõitma2*, sest see esineb ühestatud korpuses kõige sagedamini (vt Lisa 1).

Puudu on TEKsaurusest *sõitma3* näide (ühestatud korpusest: *siis sõitiski lift üles*).

JÄTMA

jätma 1; seletus: *(edasi) olla laskma, mingis kohas, seisundis, asendis, olukorras, tegevuses, millegi jaoks, mingi ajani.*
=>**jätma6** *(hülgama 1, maha jätma 2)*

muutma 2
=>**jätma2;** seletus*: mingisuguseks muutma v. muutuda laskma, mingisse olukorda v. seisundisse minna laskma v. siirma.*
=>**jätma3** (*järele jätma 1*)

loovutama 2, andma 2
=>**jätma4** (*maha jätma 1*)

lõpetama 2, lõpule viima 1
=>**jätma5** (*järele jätma 2, lõpetama 3*)

Joonis 25.

sõna+täh	*sagedus*
jätma;1,2	8
jätma;1,3	6
jätma;1,4	3
jätma;1,6	3
jätma;2,3	3
jätma;1,5	2
jätma;2,4	1
jätma;3,5	1
jätma;3,6	1

Tabel 25.

jätma1 on tipmine mõiste.

jätma1 ja *jätma6* on autohüponüümia suhtes (*jätma1* on hüperonüüm).

jätma2 ja *jätma3* on autohüponüümia suhtes (*jätma2* on hüperonüüm).

TEKsauruses on *jätma1* ja *jätma2* eristatavad, nende eristamine tekstis on keerulisem (ühestatud korpuses esinevad nad suhteliselt sarnaselt, seega eelistada üht või teist ei saa). Antud töö autor ei ole võimeline tajuma erinevust lausetes: (sulgudes on sõna *jätma* tähendusnumber, mis on ühestatud korpuses) *no jäta*(1) *piim lastele*; *kinnitamata jäetud* (1); *teda võidi ka põranda alla jätta* (2); *jätsin* (2) *auto seisma*; *ega sellepärast saa tegusid tegemata jätta* (1); *...ei saa kalapüüki ära jätta* (2). Erinevuste failist selgub, et *jätma1* on kõige problemaatilisem tähendus, sest ta kombineerub kõigi teiste tähendustega.

KADUMA

teisenema 1, muutuma 1
=>**kaduma2** (*otsa saama 2, lakkama 2, lõppema 1*)
=>**kaduma3** (*surema 1, kustuma 2, koolma* 1)
=>**kaduma6** (*kaotsi minema 1*)
=>**kaduma7;** seletus: *(mingi takistava teguri tõttu) lakkama nähtav olemast*
=>**kaduma10** (*vaibuma 2, hääbuma 2, kustuma 3, sumbuma 1*)
minema 4, lahkuma 1
=>**kaduma8;** seletus: *kusagilt, hrl. kellegi eest, juurest, nähtavalt, silmapiirilt (kiiresti) ära minema, lahkuma*

edenema 1, edasi liikuma 1, edasi minema 2
=>**kaduma9** (*kuluma 1, minema 7, mööduma 1*)

Joonis 26.

sõna+täh	***sagedus***
kaduma;7,8	3
kaduma;10,2	2
kaduma;2,7	2
kaduma;2,8	2
kaduma;8,9	2
kaduma;1,2	1
kaduma;1,6	1
kaduma;10,9	1
kaduma;2,6	1
kaduma;2,9	1
kaduma;6,7	1
kaduma;6,8	1
kaduma;7,9	1

Tabel 26.

kaduma3, *kaduma6* ja *kaduma10* hüperonüümiks on *kaduma2* (autohüponüümia suhe). *kaduma6* on *kaduma7* suhtes hüperonüüm.

TEKsauruses on tähendusvahed tajutavad; erinevuste failides ei teki märkimisväärseid tähendusklastreid hoolimata sellest, et esineb autohüponüümiat ning ko-hüponüüme.

TEKsauruses on puudu *kaduma3* näide, kuid seda tähendust pole ühestatud korpuses kordagi esinenud.

LÖÖMA

tegema 5, sooritama 4
=>**lööma1** *(peksma 2)*

tegutsema 3, toimima 4, tegema 3
=>**lööma2** *(taguma 2)*

puudutama 1, puutuma 2
=>**lööma3** *(põrkama 1)*

kätte saama 1, saavutama 1, võitma 1
=>**lööma4** *(alistama 1, võitma 2, võitu saama 1, peale jääma 1*)
katma 2
=>**lööma5**

jääma 4, muutuma 3, minema 8, saama 3
=>**lööma6**

tegema 4, häält tegema 1, kõlama 2
=>**lööma7**

Joonis 27.

sõna+täh	***sagedus***
lööma;1,2	9
lööma;2,3	2
lööma;2,6	2
lööma;6,7	2
lööma;1,+1	1
lööma;2,+1	1
lööma;2,4	1
lööma;2,5	1

Tabel 27.

Kõige suurem tähendusklaster tekib *lööma1* ja *lööma2* vahel. *Lööma1* on elusolendile liiga tegema ning *lööma2* pole vägivaldne (see tuleks ka käsitsimärgendaja juhendisse

lisada või seletustes ära märkida). Ühestatud korpuses on mõned märgendid justkui sobimatud: *Kill polnud ju löönud (2); kas Kill lõi või mitte (1); Joorik lõi mu pikali (2); aga see, et ta mind lõi (2); võib-olla tema ei löönudki seda meest üldse (1).*

TEKsauruses on puudu *lööma1* näide (ühestatud korpusest: *tema ei löönud seda meest*).

LUGEMA

lugema 3; seletus: *kirjutatut häälikulises kõnes esitama v. mingist märgisüsteemist aru saama, lahti mõtestada oskama*

oletama 1, uskuma 1, arvama 2, mõtlema 3
=>**lugema1** (*arvama 5, pidama 2*)

kindlaks määrama 1, kinnistama 1, fikseerima 1, määrama 1
=>**lugema2** (*loendama 1, arvama 3*)

Joonis 28.

sõna+täh	***sagedus***
lugema;1,3	9
lugema;1,2	4
lugema;3,+1	3
lugema;2,+1	1

Tabel 28.

lugema3 on tipmine mõiste.

lugema3 esineb ühestatud korpuses kõige sagedamini, seega tuleks märgendamisel eelistada seda (vt Lisa 1). Tähenduserinevused TEKsauruses on selged; puudu on *lugema3* näide (ühestatud korpusest: *ta loeb liiga palju*).

KIRJUTAMA

kandma 9

=>kirjutama1

informeerima 1, teada andma 1, teatama 2

=>kirjutama2

tegutsema 3, toimima 4, tegema 3

=>kirjutama3 (*kirja saatma 1*)

looma 2

=>kirjutama4

teatavaks tegema 1, avalikustama 1

=>kirjutama5

sisse kandma 1, registreerima 3

=>kirjutama6 (*arvele kandma 1*)

kohustama 1, määrama 4, käsundama 1, nimetama 4, panema 11

=>kirjutama7 (*välja kirjutama 1*)

Joonis 29.

sõna+täh	***sagedus***
kirjutama;4,5	4
kirjutama;1,2	2
kirjutama;1,4	2
kirjutama;1,7	2
kirjutama;2,3	2
kirjutama;2,4	2
kirjutama;1,3	1
kirjutama;3,4	1

Tabel 29.

Sõna *kirjutama* puhul ei esine autohüponüümiat, ko-hüponüüme, mis teeb tähenduse eristamise kergemaks. TEKsauruse tähendused väljenduvad ka tekstis.

MUUTUMA

muutuma 1 (*teisenema 1*); seletus: *oluliste tunnuste osas teistsuguseks v. täiesti teiseks saama v. minema*

=>muutuma3 (*jääma 4, minema 8, saama 3*); seletus: *kellekski, millekski v. mingisuguseks oma seisundit, olekut v. asendit muutma, senisest erinevaks muutuma, kellegi, millegi sarnaseks kujunema*

uut omadust v. tunnust omandama, mingisuguseks muutuma.
jääma 4, **muutuma 3**, minema 8, saama 3 **=>muutuma2**

Joonis 30.

sõna+täh	***sagedus***
muutuma;1,3	12
muutuma;1,2	2
muutuma;2,3	2

Tabel 30.

muutuma1 on hüperonüüm *muutma3* suhtes. Siin ei kehti ilmselt tähelepanek, et eelistatum on hüperonüüm, sest ühestatud korpuses esineb *muutuma3* sagedamini. *muutuma3* esineb sageli koos saava käändega (*muutus alandlikuks*). Tekstis on mõnikord raske vahet teha; ilmselt pole see iga kord mõttekaski. Nt lauses: *hallid, madalad pilved olid muutunud valgete äärtega rünkadeks* (3), kuigi võib ka oletada, et pilved muutusid täielikult ja just oma oluliste tunnuste osas.
muutuma3 on üheks *muutuma2* hüperonüümiks, kuid nende tähenduste osas on erimeelsused väga väikesed.

ISTUMA

liikuma 3
=>istuma1 (*istet võtma 1*)
sobima 4, kokku sobima 1, vastavuses olema 1, sobiv olema 1, vastama 2
=>istuma3 (*klappima 1, passima 2, sobima 1*)
aega kulutama 1, veetma 1, mööda saatma 2
=>istuma2 (*kinni istuma 2*)
olema 8
=>istuma4; *seletus: istuvas asendis olema*

Joonis 31.

sõna+täh	*sagedus*
istuma;1,4	10
istuma;2,4	2
istuma;3,4	2
istuma;4,+1	1

Tabel 31.

TEKsauruses on tähenduserinevused olemas. Puudu on näitelaused;

istuma1: *ta istus heinakotile*;

istuma2: *ta istus alati öösiti üleval;*

istuma3: *see ettevõtmine ei istunud talle*. Tekstis on suurem osa tähendusi eristatavad, vaid mõnes lauses ei ole tähendus selge või on arusaadav eelnevate või järgnevate lausete põhjal, nt: *ema, istume natuke (1); neid paigutati esimesse ritta istuma (1), istu parem (4).*

MÕISTMA

mõistma1
=>**mõistma2** (*jagama 1, taipama 1, aru saama 1, nägema 3, märkama 2*)
võima 1, saama 10
=>**mõistma3** (*valdama 2, oskama 1, asja tundma 1*)
mõtlema 1
=>**mõistma4** (*võtma 6, käsitama 1*)

Joonis 32.

sõna+täh	*sagedus*
mõistma;1,2	8
mõistma;1,3	3
mõistma;2,3	3
mõistma;2,4	1

Tabel 32.

mõistma1 on tipmine mõiste ning *mõistma2*-le hüperonüüm (autohüponüümia suhe). Neid tähendusi on tekstis mõnikord keeruline eristada, nt: *nagu ei mõistaks (1), mida temalt tahetakse; „Ma mõistan", lausus Hardi (1); ja hiljem koduski ei mõistetud teda*

(2) ; ühestatud korpuses esineb alammõiste (*mõistma2*) isegi sagedamini, kuid mitte nii sagedasti, et seda eelistada saaks (vt Lisa 1).

Puudu on TEKsauruses *mõistma3* näide (ühestatud korpusest: *ei mõistnud nii sügavalt kummardada, kui nõuti*).

MÄRKAMA

mõistma 1
=>**märkama2** (*jagama 1, mõistma 2, taipama 1, aru saama 1, nägema 3)*
nägema 1
=>**märkama3** (*tähele panema 2, silmama 1, täheldama 1*)
mõtlema1
=>**märkama4** (*avastama 2, leidma 2*)

Joonis 33.

sõna+täh	***sagedus***
märkama;2,3	6
märkama;3,4	6
märkama;1,3	3

Tabel 33.

märkama1 on eemaldatud TEKsaurusest, kuid erinevuste failides oli see olemas.

märkama2 ja *märkama3* on suhteliselt sarnase tähendusega, kuid siiski on TEKsauruses vahe tajutav. Ühestatud korpuses esineb *märkama3* palju kordi rohkem, seda saaks märgendamisel arvestada. Ka *märkama3* ja *märkama4* puhul on eelistatum *märkama3* (vt Lisa 1).

HOIDMA

hoidma 1; seletus: *millestki v. kellestki kinni pidama, seda haardesse jätma, haardest mitte vabastama*

hoidma 8; seletus: *midagi v. kedagi teat. olukorras, seisundis olla laskma v. olema sundima, teda sellesse jätma*

olema 8
=>**hoidma4** (*hoiduma 4, vältima 2, kõrvale hoidma 1*)
=>**hoidma5** (*halastama 1, säästma 1*)
olema 9
=>**hoidma6** (*säilitama 1, talletama 1*)
=>**hoidma2** *(pidma3)*

jälgima 3, silmas pidama 1
=>**hoidma7** (*karjatama 3, kantseldama 1*)

kogema 1, tundma 3, tunnetama 3
=>**hoidma3** (*armastama 1, lembima 1, väga hoolima 1*)

Joonis 34.

sõna+täh	***sagedus***
hoidma;1,8	2
hoidma;2,4	2
hoidma;2,7	2
hoidma;1,2	1
hoidma;1,4	1
hoidma;1,6	1
hoidma;1,7	1
hoidma;3,4	1
hoidma;4,6	1
hoidma;8,+1	1

Tabel 34.

hoidma1 ja *hoidma8* on tipmised mõisted.

hoidma4 on *hoidma5* hüperonüümiks.

Hoidma6 on *hoidma2*-e hüperonüümiks.

TEKsauruses on kõik tähendused selgelt eristatud ning olulisi tähendusklastreid erinevuste failide põhjal ei teki. Puudu on *hoidma6* näide (ühestatud korpusest: *niiti oli vähe, seda pidi hoidma*).

KARTMA

kogema 1, tundma 3, tunnetama 3
=>**kartma1** (*pelgama 1, arglema 1, hirmu tundma 1*)
oletama 1, uskuma 1, arvama 2, mõtlema 3
=>**kartma2** (*oletama 4, arvama 4*)
tajuma 4, aistima 1, tundma 1
=>**kartma4**
pidama 2, arvama 5, lugema 1
=>**kartma3**

Joonis 35.

sõna+täh	***sagedus***
kartma;1,2	7
kartma;1,3	2
kartma;1,+1	1
kartma;1,4	1
kartma;3,4	1

Tabel 35.

kartma1 ja *kartma 2* tähendusvahe on TEKsauruses selge; ilmselt võib tekkida tekstis probleeme sobiva tähendusnumbri valimisel. Nt lauses *Mats kartis tookord küllap, et poisid pistavad plehku* on sõna *kartma* saanud tähenduse 2, kuid edasisest kontekstist on võimalik välja lugeda, et Mats võis tunda ka hirmu, kartust.

USKUMA

uskuma1 (*oletama 1, arvama 2, mõtlema 3*)

uskuma3 (*kindel olema 1, veendunud olema 1*)
=>**uskuma2**

hindama 2, hinnangut andma 1, otsustama 3
=>**uskuma4** (*usaldama 1*)

Joonis 36.

sõna+täh	***sagedus***
uskuma;1,3	6
uskuma;1,2	2
uskuma;3,4	2
uskuma;1,4	1
uskuma;2,4	1

Tabel 36.

uskuma1 on tipmine mõiste.

uskuma3 on *uskuma2* hüperonüüm.

uskuma3 sünohulgas on fraas, mis on sarnane *uskuma1* seletuses olevaga: *uskuma3* sünohulgas on *veendunud olema* vrd *uskuma1* seletuses on *veendumusel olema*.

TEKsaurusest on puudu ka näited:

uskuma3 näide: *ma ei uskunud, et me koju jõuame*;

uskuma4 näide: *ma ei usu oma silmi; jäin teda uskuma*;

uskuma2-e ei esinenud ühestatud korpuses. *uskuma3* puhul tekkis tähelepanek, et selle tähenduses esineb enamasti antud sõna, kui talle järgneb *et-kõrvallause* (see pole siiski sajaprotsendiline).

VIIMA

liigutama 2
=>**viima1**
=>**viima4** (*sisestama 1, söötma 4*)
liigutama 2, juhatama 1, juhtima 2
=>**viima5**
mõjutama 1, kallutama 1, mõjustama 1, mõju avaldama 1
=>**viima2**
asuma 2, asetsema 1, paiknema 1, olema 7
=>**viima3** *(minema 15)*

Joonis 37.

sõna+täh	***sagedus***
viima;1,2	3
viima;1,+1	2
viima;1,3	2
viima;2,5	2
viiima;1,+1	1
viima;1,5	1
viima;2,3	1
viima;3,5	1

Tabel 37.

viima1 on *viima4*-ja hüperonüüm (autohüponüümia suhe).

viima1 ja *viima5* on osalised ko-hüponüümid.

Sõna *viima* tähendused on TEKsauruses selgelt eristatud; ka erinevuste failide põhjal ei teki olulisi tähendusklastreid. Puudu on *viima4* näide, kuid *viima4* ei esinenud ühestatud korpuses.

LASKMA

laskma 2; seletus: *kuhugi allapoole, madalamale, sügavamale langeda, vajuda võimaldama*

panema 4
=>laskma1
tegutsema 3, toimima 4, tegema 3
=>laskma3 (*tulistama1*)
soostuma 1, nõustuma 1, leppima 1, aktsepteerima 2
=>laskma4 (*luba andma 1, lubama 1, võimaldama 1*)
väljutama 1
=>laskma5

Joonis 38.

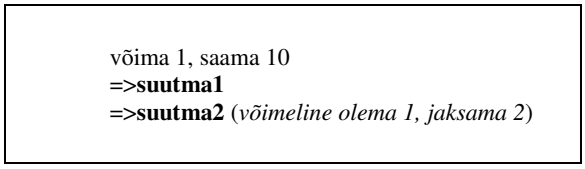
sõna+täh	***sagedus***
laskma;3,4	4
laskma;1,3	2
laskma;1,4	2
laskma;3,+1	2
laskma;4,+1	1

Tabel 38.

laskma2 on tipmine mõiste.

Võib oletada, et *laskma3* ja *laskma4* koosesinemine on tingitud märgendajate näpuvigadest, sest need tähendused on väga selgelt erinevad (nii TEKsauruses kui tekstis).

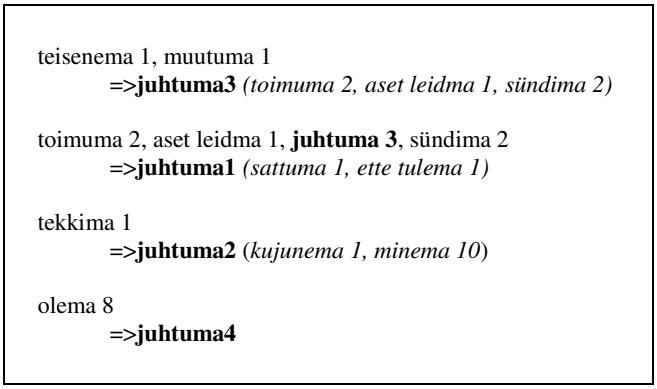
SUUTMA



Joonis 39.

suutma1 ja *suutma2* esinevad erinevuste seas 12 korda. Tähendused on ko-hüponüümid ja seega on neid suhteliselt keeruline eristada. *suutma2* sünohulk on sarnane *suutma1* seletusele: *võimeline olema* vrd *võimeline olema, millegagi toime tulema, hakkama saama*. TEKsauruses on teatav õrn tähendusvahe, kuid tekstis iga kord ilmselt sellest aru pole saada. Nt on ühestatud korpuses saanud mõlemad tähendusnumbrid lauses: *...et ma ei suuda praegu sinna minna*; raske on üheselt määrata tähendusnumbrit lauses: *ma ei suutnudki kellessegi tõsiselt kiinduda (2).*

JUHTUMA



Joonis 40.

sõna+täh	***sagedus***
juhtuma;1,3	7
juhtuma;1,4	2
juhtuma;2,3	2
juhtuma;3,4	2

Tabel 39.

juhtuma3 on *juhtuma1* üks hüperonüüme (autohüponüümia suhe), seega tuleks märgendamisel lähtuda sellest, et hüperonüüm on eelistatud. *juhtuma3* esineb ka ühestatud korpuses mõnevõrra sagedamini (vt Lisa 1).

TEKsaurusest on puudu *juhtuma1* näide (ühestatud korpusest: *mis seal mõisas õieti juhtus; juhtusin rahapaja juurde*).

AJAMA

panema 5, sundima 2
=>**ajama1** (kihutama 7 1, kupatama 2)
sättima 2, paigutama 2, seadma 1
=>**ajama5**
äratama 1, esile kutsuma 4, tekitama 5, süütama 1
=>**ajama6**
viima 2
=>**ajama7**
eraldama 3
=>**ajama8**
otsima 1
=>**ajama9**
rääkima 2, kõnelema 2
=>**ajama10** (puhuma 3, vestma 2)
tegema 4, häält tegema 1, kõlama 2
=>**ajama11**
lendama 4, kihutama 1
=>**ajama12**
toimetama 2, askeldama 1, korraldama 6, õiendama 1
=>**ajama13**
taotlust esitama 1, taotlema 3
=>**ajama14**
püstitama 1, rajama 1, ehitama 1
=>a**jama15**
kinnitama 3, kinni panema 3
=>**ajama16**
tegema 8, valmistama 3
=>**ajama17** (*utma 1, destilleerima 1*)
kasvatama 3
=>**ajama18**
ära võtma 2, eemaldama 2, kõrvaldama 1
=>**ajama19**

Joonis 41.

sõna+täh	***sagedus***
ajama;1,5	2
ajama;1,6	2
ajama;1,12	1
ajama;1,19	1
ajama;1,2	1
ajama;1,7	1
ajama;1,8	1
ajama;6,7	1
ajend;1,2	1

Tabel 40.

TEKsauruses on sõna *ajama* tähendused selgelt eristatud, seda näitab ka asjaolu, et erinevuste failide põhjal ei teki suuremaid tähendusklastreid. Ilmselt viitab täpsetele tähenduserinevustele ka see, et puuduvad ko-hüponüümia ning autohüponüümia suhted.

PANEMA

panema1 (*seadma 4, asetama 3*)

mõjutama 1, kallutama 1, mõjustama 1, mõju avaldama 1
=>**panema2**
sättima 2, paigutama 2, seadma 1
=>**panema3**
esile kutsuma 3, tingima 2
=>**panema4**
taotlust esitama 1, taotlema 3
=>**panema5** (*sundima 2*)
kohale määrama 1, nimetama 2, määrama 2, kinnitama 4
=>**panema6**
pidama 2, arvama 5, lugema 1
=>**panema7**
peksma 2, lööma 1
=>**panema8** (*andma 8, virutama 2, äigama 1*)
minema 1
=>**panema9** (*tuiskama 2, põrutama 1, kihutama 3*)
asuma 4, hakkama 2
=>**panema10** (*pistma 2, kukkuma 3*)
käsutama 1, kohustama 2, käskima 1, ülesandeks tegema 1, kamandama 1
=>**panema11** (*kohustama 1, määrama 4, käsundama 1, nimetama 4*)

Joonis 42.

sõna+täh	***sagedus***
panema;1,3	7
panema;1,4	3
panema;2,3	2
panema;2,5	2
panema;1,2	2
panema;2,6	1
panema;4,5	1
panema;1,8	1
panema;3,4	1
panema;1,7	1
panema;5,6	1
panema;3,8	1
panema;3,9	1
panema;8,+1	1

Tabel 41.

panema1 on tipmine mõiste.

panema1 sünohulgas ja *panema3* hüperonüümides on sama sõna, aga erineva tähendusega (*seadma4* vrd *seadma1*); see võib olla põhjuseks, miks mõningatel juhtudel on raske neid tähendusi eristada.

NÄGEMA

tajuma 4, aistima 1, tundma 1
=>**nägema1**
=>**nägema5**

kohtuma 1, trehvama 1, kokku saama 1
=>**nägema2**

mõistma 1
=>**nägema3** (*jagama 1, mõistma 2, taipama 1, aru saama 1, märkama 2*)

märkama 4, avastama 2, leidma 2
=>**nägema4** (*tunnetama 2, tajuma 3*)

teada saama 1
=>**nägema6**

tahtma 1, soovima 1
=>**nägema7**

Joonis 43.

sõna+täh	***sagedus***
nägema;1,8	12
nägema;1,3	9
nägema;1,4	9
nägema;1,6	7
nägema;3,4	4
nägema;3,8	4
nägema;1,2	3
nägema;4,6	2
nägema;4,8	2
nägema;1,5	1
nägema;1,7	1
nägema;2,3	1
nägema;2,6	1
nägema;2,8	1
nägema;3,+1	1
nägema;3,5	1
nägema;3,6	1
nägema;3,7	1
nägema;4,+1	1
nägema;4,5	1

Tabel 42.

nägema1 ja *nägema5* on ko-hüponüümid.

Enam ei ole TEKsauruses tähendust *nägema8*, kuid erinevuste failides oli see tähendusnumber veel sees.

nägema1 ja *nägema3* omavad TEKsauruses selget tähendusvahet. Märgendamisel tuleks eelistada *nägema1*, sest see esineb ühestatud korpuses sagedamini (vt Lisa 1).

nägema1 ja *nägema4* eristamisel tekkinud probleem võib tuleneda sellest, et *nägema1* üks hüperonüümidest (*tajuma4*) ja *nägema4* üks sünohulga sõna (*tajuma3*) on samad sõnad. Märgendamisel on eelistatud *nägema4*, sest see esineb sagedamini (vt Lisa 1).
nägema1 ja *nägema6* on TEKsauruses tähendusvahega, *nägema6* on spetsiifilisema tähendusega kui *nägema1* ning on seega eristatav ka tekstis.

LIIKUMA

liikuma 3; seletus: *asendit muutma, teise asendisse minema*
=>**liikuma1** (*käima 25*); seletus: *mitte kõndimise kohta*
=>**liikuma2** (*kulgema 2, siirduma 2*); seletus: *oma asukohta muutma, paigast teise siirduma*

Joonis 44.

sõna+täh	***sagedus***
liikuma;1,3	8
liikuma;2,3	5
liikuma;1,2	2
liikuma;1,4	2
liikuma;3,4	1

Tabel 43.

liikuma3 on tipmine mõiste.
liikuma1 ja *liikuma2* on ko-hüponüümid, mille ülemmõisteks on *liikuma3* (autohüponüümia suhe).
TEKsauruses on *liikuma1* ja *liikuma2* ja ka *liikuma3* teatav tähendusvahe ning *liikuma1* ja *liikuma3* on eristatavad ka tekstis (see nõuab teatavat põhjalikku uurimist; mitte-leksikograafile võib nende kahe tähenduse vahe arusaamatu olla).
Antud töö autori subjektiivse arvamuse kohaselt on keerulisem eristada *liikuma2* ja *liikuma3*-e lauseis nagu *...elava liiklusega bulvar, kus liikusid trammid (3)*, milles on võimalik, et trammid muudavad oma asukohta, mitte asendit. Lauses *maja kolmandal korrusel ei olnud õnneks hingelistki liikumas (3)* on samuti võimalik tõlgendada teisiti.

ASTUMA

kulgema 2, liikuma 2, siirduma 2
=>**astuma1;** seletus: *ühe, paari, mõne sammu võrra kuhugipoole liikuma*
minema 1
=>**astuma2** (*sammuma 1, käima 1, kõndima 1*); seletus: *jalgsi (edasi) liikuma*
asuma 4, hakkama 2
=>**astuma3**

Joonis 45.

sõna+täh	***sagedus***
astuma;1,2	12
astuma;1,+1	2
astuma;1,3	1
astuma;2,3	1

Tabel 44.

astuma1 ja *astuma2* on raskelt tajutava tähenduserinevusega – kui vahe on märgatav TEKsauruses, siis tekstis ilmselt mitte. *astuma2* juhud on eristatavad, need ei saa kuuluda *astuma1* alla. Aga *astuma1* mõningad juhud võivad kuuluda ka *astuma2* alla, nt: *kui saali astus kaks inimest; sellepärast astus ta ühe tooli juurde*, kus võib ette kujutada ka võimalust, et paari sammu asemel tegelikult sooritatakse pikem kõnnak (siinkohal peab rõhutama, et tegu on autori subjektiivse keeletajuga). Nende tähenduste eristamiseks peaks proovima *astuma2* puhul asendamist sünohulgas olevate teiste sõnadega, *astuma1* puhul pole selline asendus võimalik.

VÕIMA

võima 1 (saama 10)
võima2 (tohtima 1)

Joonis 46.

võima1 ja *võima2* on tipmised mõisted. *võima1* ja *võima2* esinesid erinevuste failides koos 22 korda. *võima1* esineb ühestatud korpuses palju sagedamini, seda võib eelistada märgendamisel *võima2*-le (vt Lisa 1). Tekstis on neid tähendusi sageli võimatu ja

keeruline eristada, sest antud tähendustel on juba TEKsauruses raske vahet teha; nt lauses: *võisime pliidil vett soojendada* on tõlgendatavad mõlemad variandid.

VISKAMA

viskama 2, heitma 2
=>**viskama1** (*heitma 3, paiskama 1*); seletus: *hooga midagi kuhugi panema v. laskma; midagi hooga kuhugi, mingisse asendisse v. seisundisse heitma, virutama, pilduma v. lennutama*
=>**viskama3** (*ette andma 1, söötma 3*)
edasi liigutama 1
=>**viskama2** (*heitma 2*);seletus: hooga *kuhugi midagi panema või laskma*

Joonis 47.

viskama1 ja *viskama3* on ko-hüponüümid.

Erinevuste failis esines koos tähendusklaster *viskama1* ja *viskama2*. Ilmselt on neid tähendusi eristada raske sellepärast, et seletused kattuvad osaliselt. TEKsauruses on puudu ka *viskama2* näide (*viskas mulle põgusa pilgu; viskasin paberi prügikasti*).

PÖÖRDUMA

liikuma 3
=>**pöörduma1** (*käänduma 1*)
=>**pöörduma3** (*pöörama 6*)
palvega pöörduma 1, nõudlema 1, **pöörduma 5**, taotlema 5
=>**pöörduma2** (*kõnetama 1, apelleerima 1*)
jääma 4, muutuma 3, minema 8, saama 3
=>**pöörduma4** (*konjugeeruma 1*)
paluma 3, küsima 7
=>**pöörduma5** (*palvega pöörduma 1, nõudlema 1, taotlema 5*)

Joonis 48.

sõna+täh	***sagedus***
pöörduma;2,5	5
pöörduma;1,3	4
pöörduma;1,5	2
pöörduma;2,3	1
pöörduma;1,4	1
pöörduma;2,+1	1

Tabel 45.

pöörduma1 ja *pöörduma3* on ko-hüponüümid.

pöörduma5 on üks *pöörduma2* hüperonüümidest, mis tekitab mõnel juhul raskusi nende tähenduste eristamisel.

TEKsaurusest on puudu näitelaused. *pöörduma2*: *ta pöördus abi saamiseks minu poole*; *pöörduma3*: *pöördusin siiapoole, kuigi mul polnud siia asja*; *pöörduma1*: *ei suutnud pöörduda oma kohale*.

JOOKSMA

kulgema 2, liikuma 2, siirduma 2
=>**jooksma1** (*voolama 1*)
minema 1
=>**jooksma2**
laiali lagunema 1
=>**jooksma4** (*hargnema 2, harunema 1*)
käima 25, liikuma 1
=>**jooksma5**
arenema 1
=>**jooksma6** (*edenema 3, sujuma 1, minema 9, laabuma 1*)
kulgema 1, minema 17, käima 9
=>**jooksma7**
paistma 1, näha olema 1, nähtuma 2
=>**jooksma8** *(linastuma 1*)

Joonis 49.

sõna+täh	***sagedus***
jooksma;2,4	4
jooskma;2,3	2
jooksma;1,6	2
jooksma;1,2	1
jooksma;1,4	1
jooksma;5,6	1
jooksma;2,5	1
jooksma;3,4	1

Tabel 46.

Olulisi tähendusklastreid erinevuste faili põhjal ei teki, mis viitab TEKsauruse selgetele tähenduste eristustele. Tähenduste vahedest arusaamise muudab kergemaks ka see, et puuduvad autohüponüümia suhted ja ko-hüponüümid.

jooksma4 näide peab olema *jooksma1* all.

KUULMA

tajuma 4, aistima 1, tundma 1
=>**kuulma1**
tajuma 4, aistima 1, tundma 1 keskenduma 1, kontsentreeruma 1, tähelepanu koondama 1
=>**kuulma4**
teada saama 1
=>**kuulma2**
alluma 1, kuuletuma 2, alistuma 2
=>**kuulma3** (*kuulama 3*)

Joonis 50.

sõna+täh	***sagedus***
kuulma;1,2	5
kuulma;1,+1	2
kuulma;1,3	1
kuulma;1,4	1
kuulma;3,4	1

Tabel 47.

kuulma1 ja *kuulma4* osa hüperonüüme kattuvad; nad on osalised ko-hüponüümid. Tähendusi aitaks eristada TEKsauruses puuduvad näited; *kuulma1*: *ta kuulis, kuidas süda lööb*; *kuulma2*: *kahjuks ma ei kuulnud, mis teie nimi oli*; *kuulma3*: *ta ei teinud kuulmagi; vaat, mida pidin kuulma!*

4.2. Substantiivid

MEES

inimene 1, inimolend 1, hingeline 1, indiviid 1, isik 1, persoon 1, hing 1
=>mees2 *(meesterahvas 1, meessoost inimene 1)*
meesterahvas 1, meessoost inimene 1, **mees 2**
=>mees1
=>mees3 (*abikaasa 2*)

Joonis 51.

sõna+täh	***sagedus***
mees;1,2	27
mees;1,3	3
mees;2,3	1

Tabel 48.

mees2 on *mees1* üks hüperonüümidest. *mees1* on *mees3* hüperonüümiks.

TEKsauruses on tähendusvahed olemas; keeruline, võimatu ja mõnikord ka ebaotstarbekas on üht ja õiget tähendust leida tekstis, nt lauseis: *publikule tegid suurt lõbu kaadrid tohutu paksust tünnakast mehest (2); mitte ilmaski ei kõnni sellel ükski kasimata jalgadega mees (1); mehed uputavad mured viina (1).* Antud töö autor ei taju nende tähenduste vahel erilist vahet. Selle sõna puhul ei kehti reegel, et märgendamisel tuleks eelistada hüperonüümi ilmselt sellepärast, et *mees1* esineb ühestatud korpuses palju sagedamini (vt Lisa 1).

NAINE

inimene 1, inimolend 1, hingeline 1, indiviid 1, isik 1, persoon 1, hing 1
=>**naine2** (*naisterahvas 1, naissoost inimene 1*)

naine 2, naisterahvas 1, naissoost inimene 1
=>**naine1**
=>**naine3** *(abikaasa 1)*

Joonis 52.

sõna+täh	***sagedus***
naine;1,2	10
naine;1,3	9

Tabel 49.

naine2 on *naine1* ja *naine3* üheks hüperonüümiks.

naine1 ja *naine2* on TEKsauruses eristatavad, kuid samuti nagu sõna *mees* puhul, on tekstis õiget tähendusnumbrit raske leida. Märgendamisel on eelistatud *naine1*, sest ta esineb ühestatud korpuses sagedamini (vt Lisa 1).

naine1 ja *naine3* moodustavad erinevuste failis omaette tähendusklastri ilmselt sellepärast, et eesti keeles nimetatakse abielunaist ka lihtsalt naiseks, seega pole tekstis mõnes kontekstis isegi võimalik kindlaks teha, kumb tähendus on õige.

AEG

ajavahemik 1, periood 1
=>**aeg4**
=>**aeg5** (*ajastu 2, ajajärk 2, epohh 1*)

näit 1, näitarv 1, lugem 1
=>**aeg1** (*kellaaeg 1, kell 4*)

abstraktsioon 2, üldmõiste 1
=>**aeg2**

grammatiline kategooria 1
=>**aeg3**

mõõt 1, suurus 1, mõõde 1
=>**aeg6**

Joonis 53.

sõna+täh	*sagedus*
aeg;2,4	21
aeg;1,4	8
aeg;4,6	8
aeg;4,5	6
aeg;1,2	3
aeg;3,4	3
aeg;2,3	2
aeg;4,+1	2
aeg;1,5	1
aeg;1,6	1
aeg;2,6	1

Tabel 50.

aeg4 ja *aeg5* on ko-hüponüümid, mis võib märgendamisel teatavaid arusaamatusi põhjustada, kuid TEKsauruses on tähendusvahe olemas. Puudu on *aeg5* näide (ühestatud korpusest: *sellele ajale iseloomulik trend; läbi aegade*).

Erinevuste faili põhjal tekib kõige suurem tähendusklaster *aeg2* ja *aeg4* vahel. *aeg4* esineb ühestatud korpuses sagedamini kui *aeg2* ning *aeg4* näitelause on TEKsaurusest puudu, mis võib olla erimeelsuse põhjuseks (ühestatud korpusest näide *aeg4*-le: *oli mõnda aega tagaplaanil; ettenähtud aeg; Eesti ja Soome iseseisvusid ühel ajal*).

aeg1 ja *aeg4* vahe on TEKsauruses eristatav.

aeg4 ja *aeg6* eristamise võib keeruliseks teha asjaolu, et kummalgi tähendusel pole näidet; *aeg6*-l puudub ka seletus. Ühestatud korpuses *aeg6* ei esinenud kordagi.

KÄSI

kehaliige 1, liige 1, ihuliige 1, jäse 1
=>**käsi1;** seletus: *inimese v. ahvi ülajäse randmest sõrmeotsteni*
=>**käsi5** (*käsivars 1*)seletus: *ülajäse tervikuna, õlast kuni sõrmeotsteni.*
võime 1
=>**käsi2**
pool 2, külg 2
=>**käsi3**
mängija 1
=>**käsi4;** seletus: *üks vajalikest mängupartnereist kaardimängus; bridzis ka mängija, kelle kätte mäng jääb*

Joonis 54.

sõna+täh	*sagedus*
käsi;1,5	14
käsi;1,4	12
käsi;1,+1	1
käsi;2,4	1
käsi;3,+1	1
käsi;4,+1	1
käsi;4,5	1

Tabel 51.

käsi1 ja *käsi5* on ko-hüponüümid, mis võib olla üheks eristamisprobleemide põhjuseks. TEKsaurusest on puudu ka mõlema tähenduse näited; *käsi1*: *õlised käed; paneb käed rinnale risti*; *käsi5*: *ebaloomulikult pikad käed*. Ühestatud korpuses on sagedustel küll teatav vahe (*käsi1* esineb 89 korda ning *käsi5* esineb 35 korda), kuid sageli on raske määrata just *käsi1* tähendust, nt lauseis: *ma panin käed rinnale risti; politseinik andis käega märku* (ei saa päris kindlalt öelda, et ainult teatud osaga käest saab märku anda). Ilmselt pole see isegi iga kord vajalik.

käsi1 ja *käsi4* eristus on TEKsauruses selge ning ühestatud korpuses ei esine *käsi4* mitte ühtegi korda.

LAPS

järglane 2, järeltulija 2, võsu 2
=>**laps1**; seletus: *kellegi järglane, poeg v. tütar oma vanemate suhtes*

alaealine 1
=>**laps2**; seletus: *inimene sündimisest sugulise küpsemise alguseni*

Joonis 55.

sõna+täh	*sagedus*
laps;1,2	17
laps;1,+1	1

Tabel 52.

Sõna *laps* puhul on raske tekstis aru saada, kumb tähendusnumber sobib sest iga kord ei näita kontekst seda kätte ja tegelikult on ju kõik lapsed kellegi järglased. Kui lauses on

sõna *laps* ees täiendina *minu, meie* vms, siis on suhteliselt selge, et tegemist tähendusnumbriga 1. Samuti, kui lähikontekstis esineb sõna *ema.* TEKsauruse vahe on selge ning ühestatud korpuses esinevad mõlemad tähendused enam-vähem võrdse sagedusega (vt Lisa 1).

ASI

olev 2
=>**asi1** (*objekt 1*)
=>**asi3**
=>**asi4** (*tehisasi 2, artefakt 2*)
tehisasi 2, artefakt 2, **asi 4**
=>**asi12** (*värk 1, tühi-tähi 1, asjad 1, asjandus 2*)
töö 3
=>**asi5** (*toimetus 2, ettevõtmine 1, ülesanne 2*)
ütlus 1, väljend 1
=>**asi6**
juht 4, juhtum 1, sündmus 1
=>**asi7** (*lugu2*)
asi iseeneses 1, idee 2, abstraktsioon 1
=>**asi8** (*nähtus2*)
tegu 2
=>**asi10**
seisund 4, olukord 4, situatsioon 1
=>**asi11**
atribuut 1, omadus 2
=>**asi9**

Joonis 56.

sõna+täh	***sagedus***
asi;7,11	10
asi;3,11	7
asi;3,9	7
asi;3,7	5
asi;1,11	4
asi;3,8	4
asi;3,5	3
asi;7,9	3
asi;1,4	2
asi;1,5	2
asi;1,7	2
asi;10,11	2
asi;5,11	2
asi;1,9	1
asi;10,+1	1
asi;10,5	1
asi;10,7	1
asi;3,6	1
asi;4,8	1
asi;4,9	1
asi;5,8	1
asi;5,9	1
asi;6,7	1
asi;7,+1	1
asi;7,12	1
asi;8,11	1
asi;8,9	1
asi;9,11	1

Tabel 53.

asi1 on *asi3* ja *asi4* üheks hüperonüümiks (autohüponüümia suhe). *asi3* ja *asi4* on ko-hüponüümid.

Kõige suurem tähendusklaster tekib tähenduste 7 ja 11 vahel, mis sõltub sellest, et hüperonüümid on sarnase tähendusega ning TEKsaurusest on puudu ka näited. Võimalikud näitelaused, mis on võetud ühestatud korpusest; *asi11*: *see ei puutu asjasse; nad olevat asjast nii aru saanud, et...*; *asi7*: *asjasse pühendama; kas sellega ongi asi lahendatud*.

asi3 ja *asi11* vahe on TEKsauruses selge; ka tekstis tuleb vahe välja, kuid kuna antud sõnal on suhteliselt palju tähendusi, on inimesel päris raske otsustada, milline on sobivaim.

asi3 ja *asi9* eristus on TEKsauruses olemas, enamjaolt ka tekstis. Ühestatud korpuses esinevad mõlemad võrdselt harva.

AASTA

ajavahemik 1, periood 1
=>aasta2
=>aasta1(*kalendriaasta1*)
=>aasta4
=>aasta3

Joonis 57.

sõna+täh	***sagedus***
aasta;1,2	18
aasta;2,4	6
aasta;2,3	1

Tabel 54.

aasta1 ja *aasta4* hüperonüümiks on *aasta2*. *aasta1* ja *aasta4* on ko-hüponüümid.

aasta1 on spetsiifilisem, *aasta2* hüperonüümina üldisem ning märgendamisel tuleks tavaliselt eelistada üldisema tähendusega hüperonüümi. Ka *aasta2* ja *aasta4* võrdluses tuleks eelistada *aasta2* kui hüperonüümi.

aasta1 puhul saab öelda kindla numbriga ning selle võiks kirjutada ka käsitsiühestaja juhendisse.

PÄEV

ajavahemik 1, periood 1
=>**päev3** (*tööpäev 1*)
=>**päev5** (*kalendripäev 1*)
=>**päev1** (*tähtpäev 1, pidupäev 1*)
=>**päev6** (*valge 2, päevaaeg 1*)
=>**päev7**

aeg 4
=>**päev2**

ajaühik 1
=>**päev4** (*ööpäev 1*)

kiirgus 1
=>**päev8** (*päikesepaiste 2, päikene 1, päike 2*)

Joonis 58.

sõna+täh	***sagedus***
päev;4,6	7
päev;2,4	5
päev;2,5	5
päev;2,6	4
päev;2,3	3
päev;1,2	2
päev;1,4	2
päev;3,4	2
päev;3,5	2
päev;1,6	1
päev;3,6	1
päev;5,6	1

Tabel 55.

päev3, *päev5*, *päev6* ja *päev7* on ko-hüponüümid. *päev5* on *päev*1 hüperonüüm.

Erimeelsused märgendamisel võivad tuleneda sellest, et TEKsauruses on puudu näited;

päev2: *sellest päevast alates; esimesel päeval*;

päev3: *nad tegid pikki päevi*;

päev4: *ühe päevaga ei juhtu midagi; paar päeva tagasi; järgmisel päeval oli kõik jälle normaalne*;

päev5: *viimastel päevadel olid nad rahulikud*;

päev6: *päise päeva ajal; kuidas sinu päev läks*.

TEKsauruses on vahed selged, kuid antud sõna mitmed erinevad tähendused tekitavad raskusi sobivaima valimisel.

PEA

kehaosa 1
=>**pea1;** seletus: *inimese v. looma keha ülemine ajude ja meeleorganitega varustatud ning kerest kaelaga eraldatud osa.*
teadmus 1, teadmine 1, tunnetus 1, kaemus 1
=>**pea2;** seletus: *pea psüühiliste protsesside ja tunnete asupaiga ning võrdkujuna*
intelligents 1, intelligentsus 1, arukus 1, tarkus 1, taibukus 1
=>**pea3** (*nupp 4, mõistus 2, nutt 2, aju 2*)
taimeosa 1
=>**pea4**
omadus 1
=>**pea5**

Joonis 59.

sõna+täh	***sagedus***
pea;1,2	11
pea;2,5	2
pea;1,+1	1
pea;1,1	1
pea;2,3	1
pea;3,5	1

Tabel 56.

Kõige suurema tähendusklastri moodustavad tähendused 1 ja 2. TEKsauruses on vahe selge, vaid mõnes kohas võib tekkida probleeme nende eristamisega. Nt lauses: *pea ja keha muutusid tinaraskeks* ei ole autori keeletaju päris kindel, kas inimese pea saab tõepoolest tinaraskeks muutuda; tekib kahtlus, et siinkohal võib olla ka *pea* tähenduses 2.

Võib oletada, et eristada on neid tähendusi keeruline, sest puuduvad näitelaused;
pea1: *pead vangutama; käed kõrgel üle pea*;
pea2: *korraga tekkis pähe veider mõte*.

KORD

süsteem 3, reeglid 1
=>**kord1** (*korraldus 7*)
seisund 4, olukord 4, situatsioon 1
=>**kord2** (*korrastus 1, korraldus 5, korrapära 1*)
aeg 4
=>**kord3** (puhk2)
koht 5, pind 4
=>**kord4** (*majakorrus 1, majakord 1, korrus 1*)
tehisasi 2, artefakt 2, asi 4
=>**kord5** (*kiht2*)
seisund 2, seisukord 1
=>**kord6** (*korrasolek 1*)
viis 2, stiil 2, laad 3, maneer 1, mood 1
=>**kord7**
määr 1, järk 2, tase 1, aste 2
=>**kord8**

Joonis 60.

sõna+täh	***sagedus***
kord;3,8	5
kord;3,+1	3
kord;2,3	2
kord;3,5	2
kord;1,7	1
kord;2,8	1
kord;3,7	1

Tabel 57.

Antud sõna puhul ei esine autohüponüümiat, ko-hüponüüme, mis ei muuda tähenduste eristamist väga keeruliseks. Ainuke olulisem tähendusklaster tekib *kord3* ja *kord8* vahel. TEKsauruses on see vahe eristatav, märgendamisel tuleks arvestada, et *kord3* esineb ühestatud korpuses palju sagedamini (vt Lisa 1).

MÕTE

mentaalne objekt 1, kognitiivne sisu 1
=>**mõte1**
teadmus 1, teadmine 1, tunnetus 1, kaemus 1
=>**mõte2** (*idee 1, juhtmõte 1)*
olemus 1
=>**mõte3** (*sisu 1, tähendus 2*)
tunnus 3, põhijoon 1, tunnusjoon 1, joon 3
=>**mõte4** (*tarvilikkus 1, tähtsus 3, otstarve 1*)
kõrgem tunnetusprotsess 1
=>**mõte5** (*ajutegevus 2, mõtlemine 1, mõttetegevus 1*)

Joonis 61.

sõna+täh	***sagedus***
mõte;1,2	11
mõte;1,5	10
mõte;3,4	3
mõte;1,3	2
mõte;1,4	2
mõte;2,3	2
mõte;2,5	2
mõte;4,5	2
mõte;3,5	1

Tabel 58.

Kõige olulisemad tähendusklastrid tekivad *mõte1* ja *mõte2* vahel ning *mõte1* ja *mõte5* vahel. Tähendused 1 ja 2 on TEKsauruses eristatavad; *mõte1* esineb ühestatud korpuses ka sagedamini. Mõnes lauses võib tekkida segadus, nt: *võõraste mõtete ümberkohandamine*; antud töö autor ei välista võimalust, et sõna *mõte* võiks saada ka tähendusnumbri 2.

mõte1 ja *mõte5* eristamisel võib tekkida raskusi, sest TEKsauruses puudub *mõte5* näide; ühestatud korpusest: *mõtted rabelesid sihitult; kõik ta meeled ja mõtted olid keskendunud ühele asjale*.

JUTT

suhtlus 2, suhtlemine 1, lävimine 1, kommunikatsioon 2
=>**jutt1** (*kõne 3, keel 5*); seletus: *verbaalne suhtlemine, see, mida keegi räägib, ütleb, jutustab*
jutustus 1
=>**jutt2** (*lugu1*); seletus: *üks eepika väikevorme*
arutelu 1, mõttevahetus 1, diskussioon 1
=>**jutt3** (*kõnelus 1, vestlus 1, jutuajamine 1*); seletus: *inimeste omavaheline rääkimine*

Joonis 62.

sõna+täh	***sagedus***
jutt;1,3	10
jutt;1,2	4
jutt;1,+1	2
jutt;2,3	2

Tabel 59.

TEKsauruses on selgelt eristatav *jutt2* (vt seletust). *Jutt1* ja *jutt3* on töö autorile kohati tunnetuslikult rasked eristada. Segadusse ajab *jutt3* seletus, mille puhul tekib tunne, et juttu peavad ajama kaks või rohkem inimest. Ühestatud korpusest pärit näitelaused ei eelda aga kindlalt mitme inimesevahelist jutuajamist; *kuulab sõbratari juttu; tegi oma jutus pausi; jutt on ikkagi tunnetest*. Näiteks on ühestatud korpuses lause: *ülempreestri juttu kuulates*, kus sõna *jutt* on saanud tähendusnumbri 1 (vrd *jutt3*: *kuulab sõbratari juttu*).

Kõik antud sõna tähendusnumbrid esinevad ühestatud korpuses enam-vähem võrdselt, seega ei saa öelda, millist tähendust eelistada.

MAJA

ehitus 1, ehitatu 1, ehitis 2
=>**maja2** (*hoone1, ehitis1*); seletus: *igasugune katuse ja seintega ehitatu*
=>**maja1** (*elumaja 1, eluhoone 1, elamu 1, eluase 1*); seletus: *ehitis, mis on mõeldud elamiseks ühele v. enamale leibkonnale*

Joonis 63.

sõna+täh	*sagedus*
maja;1,2	15
maja;1,+1	1

Tabel 60.

maja2 on *maja1* suhtes hüperonüüm.

TEKsauruses on tähendusvahe olemas, kuid puudub näide tähendusel *maja2*; ühestatud korpusest: *19.sajandi majad; madalad majad. maja1* esineb ühestatud korpuses poole rohkem kui *maja2* (vt Lisa 1). Mõnede lausete puhul on raske tähendusnumbrit määrata, nt lauses: *teeme maja ees suitsu* on ühestatud korpuses märgendatud kui *maja2* ning lauses: *istusime maja ees lauahunnikul ja jõime õlut* on sõnal *maja* tähendusnumber 1.

UKS

tõke 1, takistus 1, tõkestus 1
=>**uks1**; seletus: *hingedel v. rullidel liikuv takistus, mis sulgeb pääsu mingisse ruumi*
auk 3, ava 2, avaus 1
=>**uks2** (*ukseava2*); seletus: *avaus seinas, mille kaudu saab mingisse ruumi siseneda ja sealt väljuda; avaus, mida suletakse uksega (1. täh)*

Joonis 64.

sõna+täh	*sagedus*
uks;1,2	14
uks;2,+1	1

Tabel 61.

TEKsauruses on tähendusvahed olemas, puudu on näited, mis võivad muuta sobiva tähenduse valimise keeruliseks; ühestatud korpusest *uks1*: *vannitoa uks; ootas ukse taga; lükkasin ukse lahti*. Ilmselt on mõnikord raske tähendustel tekstis vahet teha, nt lauses: *ta liikus ukse poole*, kus on antud töö autori arvates peaaegu võimatu selgeks teha, kas liigutakse konkreetselt ukse poole või ukse avause juurde.

Tekstis on mõnel juhul raske valida sobivat tähendust, näiteks on ühestatud korpuses lause: *uks oli lahti* tähendusnumbriga 1, võrdluseks lause: *Juku teeb ukse lahti*, mis on

tähendusnumbriga 2. Keeletaju ütleb, et sõnal *uks* on tegelikult üks ja seesama tähendus (antud juhul *uks1*). Lauses *jättis ukse sulgemise Heinrichi hooleks* on sõna *uks* saanud tähendusnumbri 2, kuigi peaks olema tähendusnumbriga 1 (ei sulgeta ukseava, vaid konkreetselt ust).

JALG

kehaliige 1, liige 1, ihuliige 1, jäse 1
=>**jalg1**; seletus: *inimese v. ahvi alajäse; linnu tagajäse; paljudel loomadel, putukatel jm. üldse jäse v. kulgemiselund*
toetus 1, tugi 1
=>**jalg2**; seletus: alumine *kandev v. toetav osa (hrl. esemetel)*
pikkusühik1
=>**jalg3**

Joonis 65.

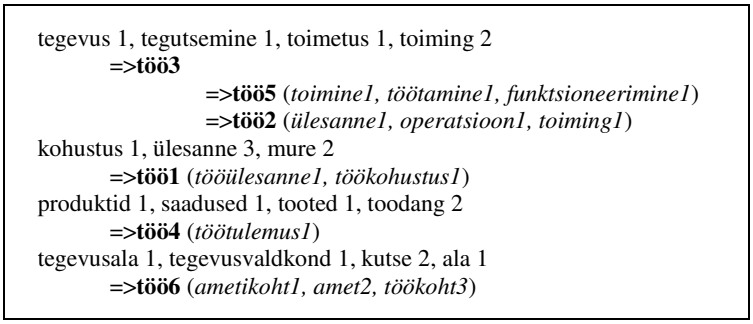
sõna+täh	***sagedus***
jalg;1,2	7
jalg;1,3	2
jalg;1,+1	1

Tabel 62.

TEKsauruses on tähendused selgelt eristatud, ka tekstis tulevad tähenduserinevused välja ning võib oletada, et märgendamisel on tehtud näpuvigu.

Tundub, et puudu on veel üks tähendus; nt lauseis: *jalust maha rabama; jalust maha niitma* jt. Samamoodi on sõnadel *pea* ja *käsi* olemas abstraktsem tähendus.

TÖÖ



Joonis 66.

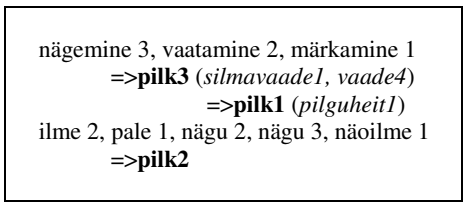
sõna+täh	***sagedus***
töö;1,3	10
töö;1,6	5
töö;1,4	2
töö;1,2	1
töö;2,3	1
töö;3,6	1

Tabel 63.

töö3 on *töö2* ja *töö5* hüperonüümiks; *töö2* ja *töö5* on ko-hüponüümid.

Kõige suurema tähendusklastri moodustavad *töö1* ja *töö3*. TEKsauruses on tähenduste vahe mõningal määral tajutav, kuid tekstis võib mõningate lausete puhul olla keeruline sobiva tähenduse valimine. Ühestatud korpuses on tähendusnumbri 1 saanud lause *mil Lutrin oma töö katkestas* vrd tähendusnumber 3 on sõnal *töö* lauses *jättis töö pooleli*. Puudu on TEKsaurusest *töö5* näide; ühestatud korpusest: *nagu oleks tööle hakanud mingi tohutu õlitamata masinavärk.*

PILK



Joonis 67.

sõna+täh	*sagedus*
pilk;1,3	7
pilk;2,3	7
pilk;2,+1	2
pilk;1,+1	1
pilk;1,2	1

Tabel 64.

pilk3 on *pilk1* üheks hüperonüümiks (autohüponüümia suhe).

pilk1 ja *pilk3* tähendused on TEKsauruses selged; ka tekstis tulevad vahed välja. Antud juhul tuleks eelistada *pilk3*-e, sest see on hüperonüüm.

pilk2 ja *pilk3* tähendus on TEKsauruses tajutav, kuid mõningais lauseid on raske vahet määrata; nt: *...kelle vahetult pärivatesse silmadesse ei tohtinud vaadata süüdlasse eksiva pilguga* (antud juhul on see ühestatud korpuses märgendatud tähendusnumbriga 2, kuid on segamini aetav *pilk3*-ga); *..püüdsin mu silmad kinni neiu sädeleva pilgu, mis mind uudistades vaatas* (siin on tegemist *pilk3*-ga, kuid võib olla tõlgendatav ka *pilk*2-na, kui märgendaja ei süvene edasi *mis*-kõrvallause tähendusse).

KEEL

kommunikatsioon 1, suhtlus 1
=>**keel3**
=>**keel2** (*inimkeel1*)
=>**keel1** (*kõne2, kõnekeel1*)
organ 2, elund 1
=>**keel4**
suhtlus 2, suhtlemine 1, lävimine 1, kommunikatsioon 2
=>**keel5** (*jutt1, kõne3*)
rihm 1
=>**keel6** (*piug1, piitsarihm1*)
asjandus1, vahend2
=>**keel7** (*pillikeel1*)

Joonis 68.

sõna+täh	*sagedus*
keel;2,3	8
keel;1,2	4
keel;2,4	4
keel;2,6	2
keel;1,3	1
keel;1,5	1
keel;2,5	1
keel;4,+1	1
keel;4,6	1
keel;6,+1	1

Tabel 65.

keel1 ja *keel3* on ko-hüponüümid. *keel3* on *keel2* hüperonüüm. *keel2* ja *keel3* eristamisel ei kehti see, et märgendamisel tuleks eelistada hüperonüümi kui üldisemat tähendust, sest *keel3* esineb ühestatud korpuses vaid 2 korda (vrd *keel2* esineb 45 korda). *keel2* tähendust omavad sellised fraasid nagu *soome keel, võru keel* jt. Ühel juhul tundub, et sõna *keel* on lauses saanud tähendusnumbri 3, kuigi peaks olema tähendusnumbriga 2 (*nad kõnelesid saksa keeles*).

TEKsaurusest on puudu *keel4* näide; ühestatud korpusest: *limpsas keelega üle huulte.*

KOHT

koht1; seletus: *keha, eseme, hoone jne. kitsam piirkond*
osa 3, ala 3, piirkond 2, regioon 2
=>**koht3** (*punkt2*)
=>**koht2**
=>**koht6** (*asukoht2, asupaik2*)
koht4 (*asukoht1, paik1, asupaik1*)
ehitus 1, ehitatu 1, ehitis 2
=>**koht5** (*pind4*)
hoone 1, ehitis 1, maja 2
=>**koht7**
positsioon 2, staatus 2
=>**koht8**

Joonis 69.

sõna+täh	*sagedus*
koht;3,4	3
koht;1,4	2
koht;2,+1	2
koht;3,+1	2
koht;5,+1	2
koht;5,6	2
koht;1,7	1
koht;1,8	1
koht;2,6	1
koht;2,7	1
koht;2,8	1
koht;3,6	1
koht;4,5	1
koht;4,8	1
koht;6,8	1
koht;8,+1	1

Tabel 66.

koht1, *koht4* on tipmised mõisted.

koht1 on *koht3* hüperonüüm ning *koht3* on omakorda *koht2* ja *koht6* hüperonüümiks. Sellised keerulised suhted ühe sõna sees on tekitanud märgendamisel palju erimeelsusi. Palju esineb ka puuduvaid tähendusi, +1 – ühestatud korpuses 18 korral, mis annab märku sellest, et sõna *koht* osa tähendusi on veel puudu või ei tule need tähendused keerulistes suheterägastikus välja.

TEE

joon 2, piir 4
=>**tee1**(*siht2, orientatsioon2, kurss1, suund2*)
=>**tee6** (*liikumistee1*)
=>**tee7**
jook1
=>**tee5**
tegu2
=>**tee2** (*viis 1, vahend 1, abinõu 1, kanal 1, meede 1, mood 3*)
tehisasi 2, artefakt 2, asi 4
=>**tee4** (*rada2, teerada1*); näide:

Joonis 70.

sõna+täh	*sagedus*
tee;3,4	6
tee;1,2	2
tee;1,4	2
tee;1,3	1
tee;1,6	1
tee;1,7	1
tee;2,+1	1
tee;3,+1	1
tee;3,6	1
tee;4,6	1
tee;4,7	1
tee;5,7	1

Tabel 67.

tee1 ja *tee6* on ko-hüponüümid ning *tee6* on *tee7* hüperonüüm.

TEKsauruses polnud *tee3* ja *tee4* antud töö autorile väga tajutava erinevusega, kuid ühestatud korpuses ilmnes tähenduserinevuste vahe. Üheks põhjuseks võib olla see, et *tee4* näide on pisut segane (ühestatud korpusest veel võimalikke näitelauseid: *tee läks pärastpoole halvemaks, tee hargneb kolmeks, tagasi läksime teist teed*.

HÄÄL

füüsikaline nähtus 1, füüsiline fenomen 1
=>**hääl5** (*heli1*)
=>**hääl2** (*vokaliseerimine1, inimhääl1*)
=>**hääl3** (*häälekõla2, tämber2*)
=>**hääl4** (häälitsus1)
nägemus 2, seisukoht 2, arvamus 1
=>**hääl1**

Joonis 71.

sõna+täh	*sagedus*
hääl;2,3	6
hääl;2,5	4
hääl;1,2	1
hääl;1,5	1
hääl;2,4	1
hääl;4,+1	1

Tabel 68.

hääl5 on *hääl2* ja *hääl4* üheks hüperonüümiks ning *hääl2* ja *hääl4* on ko-hüponüümid. *hääl2* on *hääl3* üheks hüperonüümiks ning seepärast võibki tekkida raskusi nende eristamisega. Tuleks eelistada *hääl2*, mis on hüperonüüm ja seega üldisem mõiste; samuti esineb *hääl2* ühestatud korpuses palju sagedamini kui *hääl3*. Üks põhjusi, miks neid on raske eristada, on see, et TEKsauruses on puudu *hääl3* näide: *Leenale ei meeldinud võõra mehe hääl* (kuigi siin võib see ka tähenduses 2 olla).
Puudu on ka *hääl4* näide: *padrikust kostis linnu mahe kurblik hääl.*

KOOL

õppeasutus 1
=>kool1; seletus: *õppeasutus, kus õpilased õpetajate juhtimisel omandavad teadmisi, oskusi ja vilumusi (ka koolikollektiiv)*
näpunäide 1, õpetus 2, juhatus 2, juhtnöör 1
=>kool2 (*õppetund1*); selgitus: *mingisuguste oskuste omandamine, koolitus; kogemused*
hoone 1, ehitis 1, maja 2
=>kool3 (*koolimaja1, õppehoone1);* selgitus: *hoone, kus asub mingi õppe- või kasvatusasutus*

Joonis 72.

sõna+täh	***sagedus***
kool;1,3	8
kool;1,2	2

Tabel 69.

Olulisim tähendusklaster tekib *kool1* ja *kool3* vahel. Ühestatud korpuses esineb *kool1* sagedamini kui *kool3*, seega peaks eelistatuim olema *kool1* (vt Lisa 1). Mulle tundub, et kool1 käitub mõnes mõttes kool3 hüperonüümina, sest alati on tegemist kooli kui õppeasutusega ning ainult mõnel juhul saab selgelt öelda, et tegemist on hoonega, kus kool asub: *sõitsime kooli ja koju talumeeste regede päradel; kooli trepp; kooli lähedal oli ta näinud oma kunagist klassivenda.*

TUNNE

psüühiline nähtus 1
=>**tunne2** (*tundmus2*)
=>**tunne3** (*tundeelamus 1, emotsioon 1, tundmus 1*)
kõrgem tunnetusprotsess 1
=>**tunne1** (*tunnetamine 1, taju 3*)

Joonis 73.

sõna+täh	***sagedus***
tunne;2,3	10
tunne;1,2	6
tunne;1,3	1

Tabel 70.

tunne2 on *tunne3* üheks hüperonüümiks (autohüponüümia) ning seepärast võib olla teatavatel juhtudel ka nende eristamine. Eelistada saab siinkohal hüperonüümi (esineb mõneti rohkem ka ühestatud korpuses). Arvatavasti on mitte-leksikograafi taustaga inimesel neid erinevusi raske tajuda, nii TEKsauruses kui ka tekstis. Selleks, et valida tähendusnumbrit 3, peab olema veendunud, et see on spetsiifilisem ja kitsama tähendusega kui tähendus 2.

LÕPP

joon 2, piir 4
=>**lõpp1** (*äär3*)
aeg2
=>**lõpp2**; seletus: *piir, mil miski saab läbi, kauem ei kesta; lõpphetk, lõpposa*
piir 1, piirjoon 2
=>**lõpp3** (*piir2, äär2*)
lõik 2, osa 6, jaotus 2
=>**lõpp4** (*lõpuosa1, lõpposa1*)
juht 4, juhtum 1, sündmus 1
=>**lõpp5** (*lõpetus1*)
otspunkt 1
=>**lõpp6** (*ots4*)
seisukord 2, seisund 3, olek 1, seis 2
=>**lõpp7** (*lõppseisund1*)

Joonis 74.

sõna+täh	*sagedus*
lõpp;2,4	3
lõpp;2,7	3
lõpp;2,5	2
lõpp;5,7	2
lõpp;1,2	1
lõpp;1,4	1
lõpp;1,5	1
lõpp;2,6	1
lõpp;4,5	1
lõpp;4,6	1
lõpp;4,7	1
lõpp;7,8	1

Tabel 71.

lõpp2 ja *lõpp4* eristamine võib olla mõneti keeruline, sest *lõpp4* sünohulga sõna (*lõpposa1*) ja *lõpp2* seletuses olev langevad kokku. TEKsaurusest on puudu ka mõlema näited. Antud töö autori keeletaju ei erista *lõpp2* ja *lõpp4* erinevust.

Näited ühestatud korpusest:

lõpp1 – sellega maailma lõppu ei jõua;

lõpp2 – lõpp orjatööle! ...kuhu teda oli 1939. aasta lõpul tööle saadetud;

lõpp4 – millel pole algust, võib ka lõpp puududa;... jätsid Tallinna maha augusti lõpul;

lõpp5 – kodusõda on võiduka lõpuni viidud; jõudis ettevalmistustega lõpuni;

lõpp7 – teeb lõpu nende jutule.

5. Kokkuvõte

Töö eesmärgiks oli uurida käsitsiühestajate arvamuste erinevust TEKsauruse tähendusnumbrite põhjal. Arvamuste erinevuste uurimine annab võimaluse näha, milliste sõnade tähenduste osas on erimeelsused suured ja millised tähendusklastrid tekivad ning see on oluline tulemus TEKsauruse jaoks. Samuti on püütud leida uusi reegleid ja juhiseid käsitsiühestajatele.

Kokku analüüsiti 74 sõna, neist 50 verbi ning 24 substantiivi.

Osa töös analüüsitud sõnade tähendusi moodustasid olulisi tähendusklastreid, osa mitte. Viimatimainitud sõnade tähendused on selgelt eristatavad nii TEKsauruses kui ka käsitsiühestajale tekstis. Analüüsi tulemusena võib öelda, et erimeelsused on väiksed sõnatähenduste hulgas, mis:

- ei sisalda autohüponüümia suhteid ning ko-hüponüüme;
- on TEKsauruses täielikult ja vigadeta esitatud, s.t et kõik vajalikud esitusväljad on korrektselt täidetud;
- on madalama polüseemiatasemega (see pole küll alati reegel, sest ka väga polüseemsete sõnade hulgas, näiteks: *käima, ajama* on tähendused selgesti eristatud).

Kui sõnatähendustel tekivad olulised tähendusklastrid kindlate tähenduste vahel, on põhjust kahtlustada, et:

- TEKsauruses on tähendused eristatavad, kuid tekstis mitte;
- TEKsauruses on teatavad puudused: näidete puudumine, ähmane seletus, näidete ja/või seletuste ja/või sünohulkade kokkulangemine.

Erimeelsuste tekkimise tavalisimaks põhjuseks on raskus leida sobiv tähendusnumber tekstisõnale, põhjuseks eelmainitud autohüponüümia suhted, ko-hüponüümid ning sõnatähenduste liigne detailsus TEKsauruses. Kõrge polüseemiatasemega sõnade puhul on keerulisem üht ja ainsat sobivat tähendust määrata. Sageli tekib olukordi, kus tekstis polegi võimalik ainult üht ja kindlat tähendust määrata, kontekst lubab mitmeid võimalikke tähendusi või polegi otstarbekas teatud sõnatähendusi eristada.

Tähendusklastrite uurimine annab võimaluse viidata mingi konkreetse sõnatähenduse ebaselgusele, mõni sõnatähendus kombineerub kõikide erinevuste seas või paljude teiste tähendustega. Juhul kui ühe sõna kõik tähendused saavad üksteisega koos esineda (peaaegu kõik tähendused kombineeruvad peaaegu kõigiga), võib oletada, et terve sõna tähendusjaotus on puudulik, ebaselge ning vajab täiendamist. See tendents ilmneb abstraktse ja/või laia tähendusampluaaga sõnatähenduste korral, nt sõnad *asi, aeg, jääma*.

Käsitsiühestajate erimeelsuste uurimise tulemusena ilmnesid ka TEKsauruse teatud puudused. Antud töös toodi välja eelnevalt mainitud näidete, seletuste ja sünohulkade kattumise juhud, mis raskendavad käsitsiühestaja tööd. Samuti on ühestatud korpusest leiti TEKsauruses puuduvad sobilikud näited, mille hulgast on TEKsauruse tegijatel võimalik välja valida sobivaimad. Sõnatähenduste analüüs näitas, et sageli kaheldakse just nende tähenduste vahel, millel TEKsauruses puuduvad näited ning seega on näidete lisamine käsitsimärgendajale kindlasti suureks abiks.

Mõned sõnatähenduste uurijad on viidanud probleemile, et ebaselged ja raskesti märgendatavad on tipmistesse sünohulkadesse kuuluvad sõnad. Ka antud töö sõnade analüüs näitas, et sageli on erimeelsused just tipmiste sünohulkade osas, samuti on tipmisesse sünohulka kuuluvat sõnatähendust raskem teistest eristada.

Praegune käsitsiühestaja juhend abistab vaid mõne üksiku sõna puhul, eksisteerivad ka kirjutamata reeglid. Üheks kirjutamata reegliks on tähelepanek, et märgendamisel tuleks eelistada hüperonüümi ning töö tulemusena selgus, et see peab enamasti ka paika. Muidugi leidub tähendusi, kus hüperonüüm esineb ühestatud korpuses sagedamini ning TEKsauruse haldajatel on mõistlik sellised juhud üle vaadata. Esineb olukordi, kus mõned kirjutamata reeglid peaks lisama märgendamisjuhendisse. Ilmekas näide on sõna *ütlema*, mille puhul ei ole ei TEKsauruses ega juhendis kirjas, milline tähendusnumber vastab otsese kõne saatelauses esinevale sõnale. Ka sõna esinemissagedust ühestatud korpuses saab märgendamisel abiks võtta – mõne sõna kaheldavate tähenduste (kindlate tähendusklastrite) puhul on võimalik eelistatuim (ehk sagedamini esinev) tähendusnumber juba juhendis ära mainida.

Käsitsiühestajate eriarvamuste alamkorpuse moodustamisel oli arvesse võetud fenomeni "üks tähendus ühe teksti kohta" ning antud töö autori arvates võiks käsitsiühestaja juhendis fenomeni paikapidavust mainida.

Sõnatähenduste analüüsi tulemusena peab nentima, et sageli on tähenduserinevused olemas nii TEKsauruses kui ka reaalses tekstis, kuid mitte-leksikograafi taustaga inimesel (ning pealiskaudsel süvenemisel) on tähendustel raske vahet teha. Tähendusklastrid võivad olla ka oluliseks abiks automaatse sõnatähenduste ühestaja töös. Nagu selgus peatükis 2.1.6, on automaatsel ühestamissüsteemil kergem valida sobivat ja õiget tähendusnumbrit, kui tähendused pole liialt üle-eristatud, s.t süsteemi töö kiirus ja täpsus suurenevad. Kui ühe sõna tähendused moodustavad mingeid kindlaid ja olulisi tähendusklastreid, on mõttekas need teatud keeletehnoloogiliste rakenduste jaoks liita üheks tervikuks ning anda neile üks tähendusnumber. Rakenduse leiaks see näiteks masintõlkes – kui inimene ei taju kõiki võimalikke tähenduserinevusi, siis ei ole ilmselt mõttekas ka masintõlkesüsteemil kõiki võimalikke tähendusi eristada.

Kirjandus

Chklovski, Tim; Mihalcea, Rada 2003. Exploiting Agreement and Disagreement of Human Annotators for Word Sense Disambiguation. – Proceedings of the Conference on Recent Advances in Natural Language Processing. Borovetz, Bulgaria, pp 4–12.

Edmonds, Philip; Cotton, Scott 2001. SENSEVAL-2: Overview. – Proceedings of SENSEVAL-2, Tolouse, France, pp. 1–7.

ETF grant 4467, "Eesti keele semantilise ühestaja loomine". 2000–2002.

Gale, William; Church, Kenneth; Yarowsky, David 1992. One Sense Per Discourse. – DARPA Workshop on Speech and Natural Language, New York, pp. 233–237.

Ide, Nancy; Véronis, Jean 1998. Introduction to the special issue on word sense disambiguation: the state of the art. – Computational Linguistics, No 24, pp. 2–40.

Kahusk, Neeme; Kaljurand, Kaarel 2002. *Semyhe* tulemusi: kas tasub *naise* pärast WordNet ümber teha? – Tähendusepüüdja. Toim. R.Pajusalu, T.Hennoste. Tartu: Tartu Ülikooli üldkeeleteaduse õppetooli toimetised 3, lk 185–195.

Karlsson, Fred 2002. Üldkeeleteadus. Tõlkinud ja kohandanud R.Pajusalu, J.Valge, I.Tragel. Tallinn: Eesti Keele Sihtasutus.

Kilgarriff, Adam 1997. "I don't believe in word senses". – Computers and the Humanities, No 32 (2), pp. 91–113.

Muischnek, Kadri; Orav, Heili; Kaalep, Heiki-Jaan; Õim, Haldur 2003. Toim. Talvik, Urve. Eesti keele tehnoloogilised ressursid ja vahendid. Arvutikorpused, arvutisõnastikud, keeletehnoloogiline tarkvara. Tartu: Haridus- ja Teadusministeerium, Eesti keelenõukogu.

Orav, Heili; Vider, Kadri 2002. Kas tesaurus ja tekstid lähevad kasutuses kokku? – Tähendusepüüdja. Toim. R.Pajusalu, T.Hennoste. Tartu: Tartu Ülikooli üldkeeleteaduse õppetooli toimetised 3, lk 297–303.

Peters, Wim; Peters, Ivonne 1998. The Restructuring of the ILI. EuroWordNet (LE-8328) Deliverable 2D004.

Peters, Wim; Peters, Ivonne; Vossen, Piek 1998. Automatic sense clustering in EuroWordnet. – Proceedings of the First Interational Conference on Language Resources and Evaluation, Granada.

Ravin, Yael; Leacock, Claudia (Eds) 2002. Polysemy : theoretical and computational approaches.

Tomuro, Noriko 2001. Tree-cut and a lexicon based on systematic polysemy. In Proceedings of the North American Chapter of the Association for Computational Linguistics.

Vider, Kadri; Kahusk, Neeme; Orav, Heili; Õim, Haldur; Paldre, Leho 2000. Eesti keele tesaurus. – Arvutuslingvistikalt inimesele. Toim. Tiit Hennoste. Tartu: Tartu Ülikooli üldkeeleteaduse õppetooli toimetised 1, lk 127–152.

Vider, Kadri 2001. Eesti keele tesaurus - teooria ja tegelikkus. Leksikograafiaseminar "Sõna tänapäeva maailmas"/ Leksikografinen seminaari "Sanat nykymaailmassa". Ettekannete kogumik. Toim. M. Langemets. Eesti Keele Instituudi toimetised 9. Tallinn, lk 134–156.

Wam, Stephen 1999.
http://www.ics.mq.edu.au/~swan/readingroom/word_sense_disambiguation/timeline.htm#ai

Abstract

Word Senses in Texts and in Thesauri based on Human Annotators Disagreement.

Word sense disambiguation is considered to be one of the most important problems in natural language processing.

This B.A thesis gives a brief overview of the history and methods of word sense disambiguation in chapter two. Chapter three deals with the structure of Estonian thesaurus called TEKsaurus. The TEKsaurus, also known as the Estonian WordNet, is created by the research group of computer linguistics (University of Tartu) since 1997. Word sense disambiguation is based on the sense numbers in TEKsaurus. The word sense disambiguation task is done half-manually and therefore chapter three concentrates on the sense tagging process.

In this thesis, the research focuses on exploiting agreement and disagreement of human annotators: are there any remarkable and important sense clusters? Manual sense tagging refers to problems in TEKsaurus like missing examples, overlaping synsets of explanations and over-grained senses. Sense clusters are made by processing the disagreement files and the most frequent words – 50 verbs and 24 nouns – are analyzed by describing lexical relations like autohyponymy and sisters (co-hyponyms).

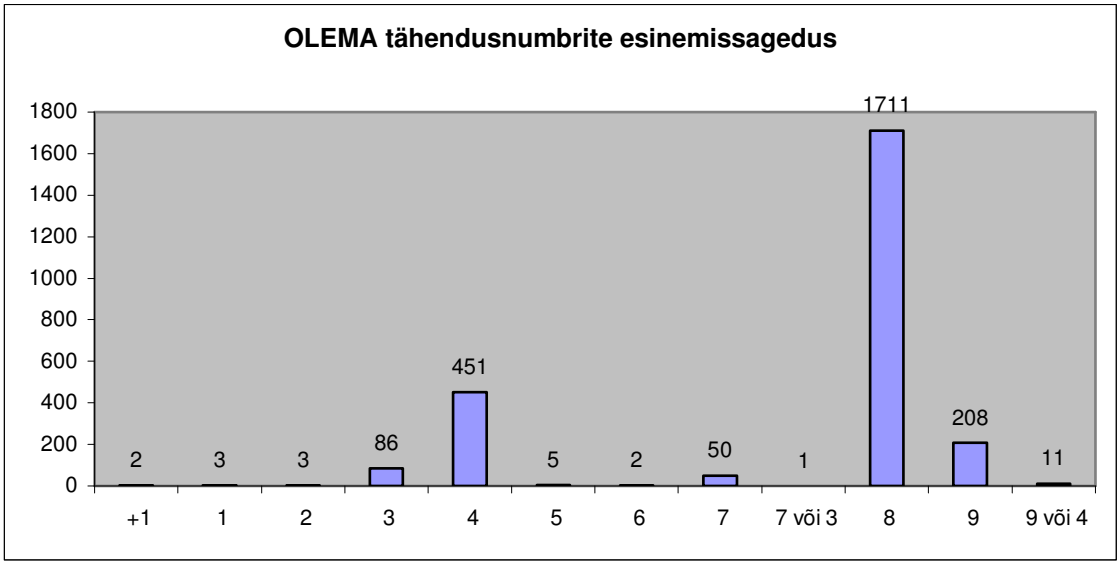
The results are meant to be helpful for the creators of TEKsaurus.

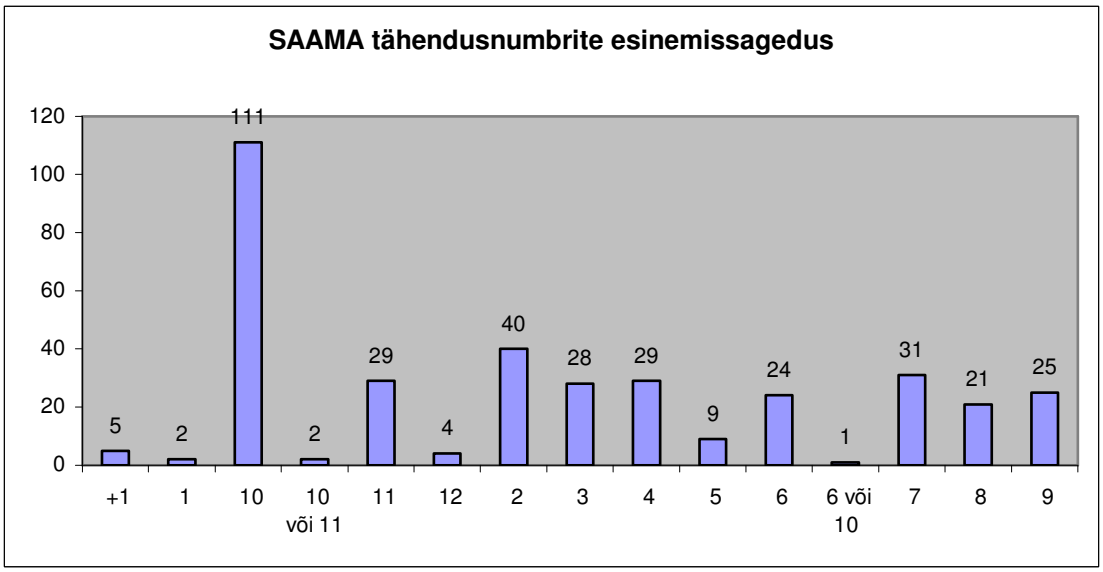
Lisa 1.

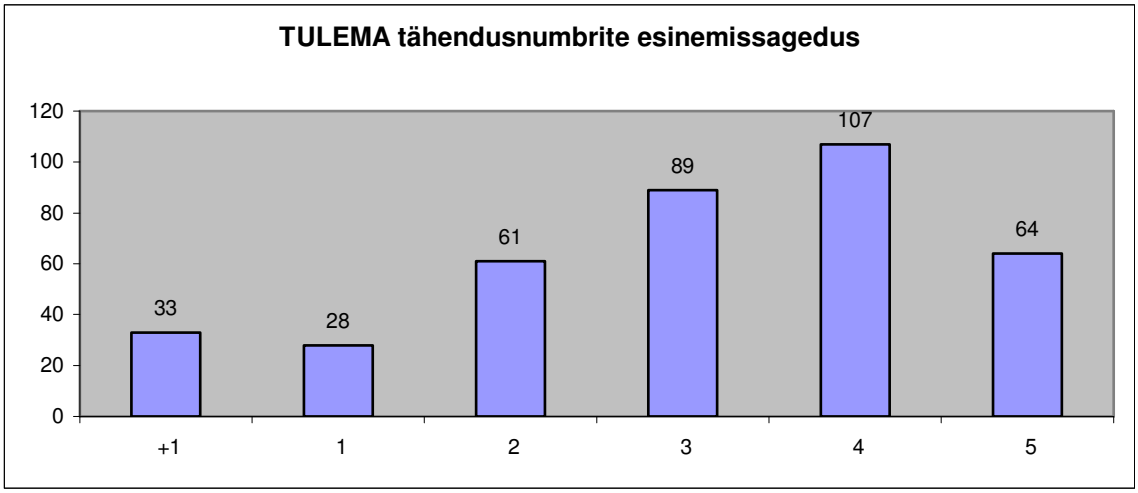
Kõikide analüüsitud sõnade tähendusnumbrite esinemissagedus ühestatud korpuses. Reastatud analüüsi järjekorras.

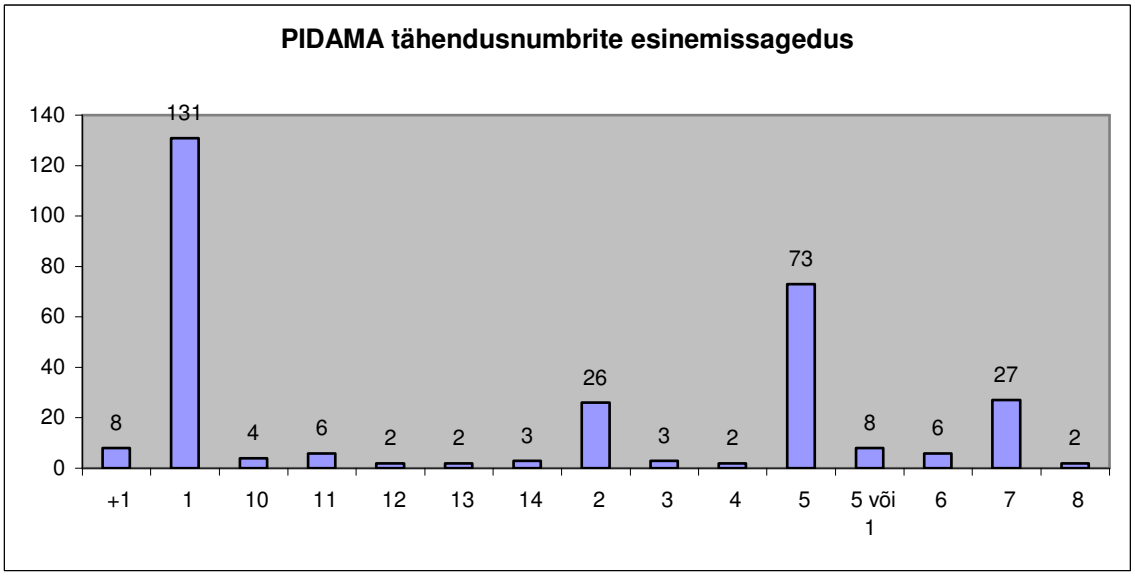
Osa tähendusi on ühestatud korpuses saanud kaks või enam tähendusnumbrit.

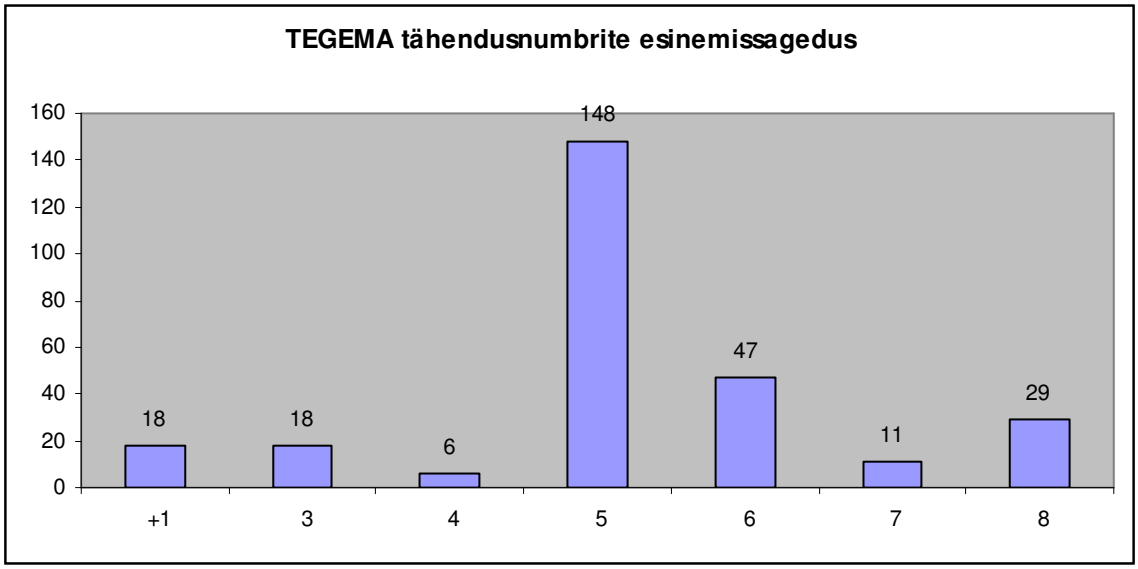
Diagrammi vasakul äärel on esinemissagedus; täpne esinemissagedus on märgitud diagrammil tulba peale. Alumisel, x-teljel, on sõna tähendusnumbrid.

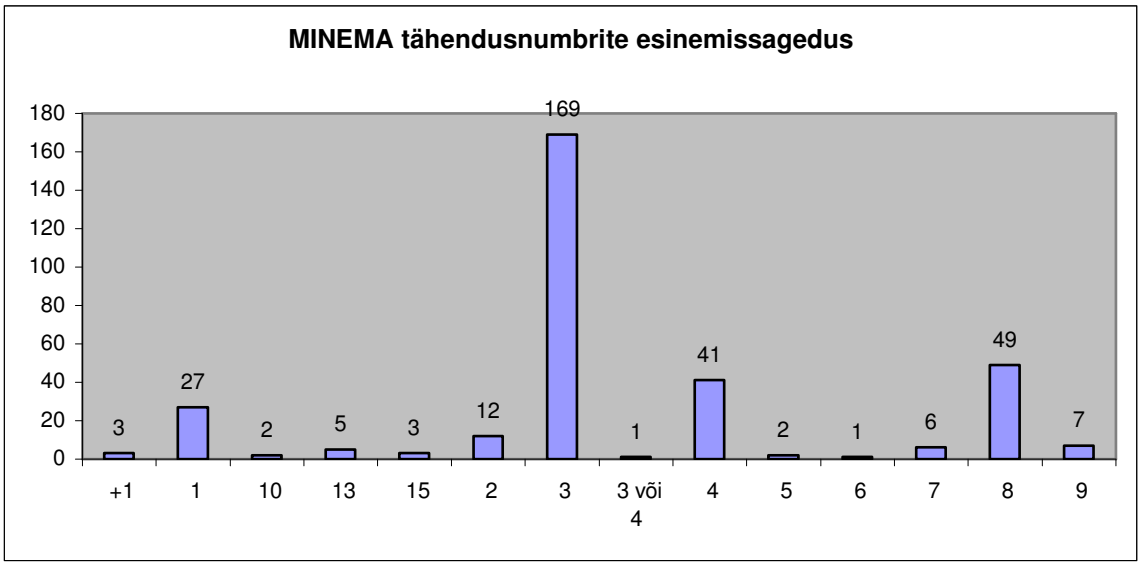


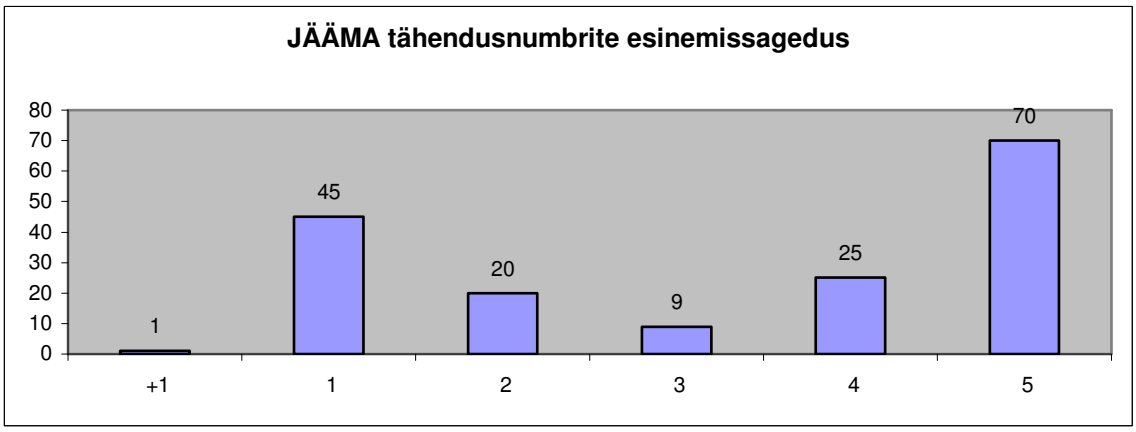


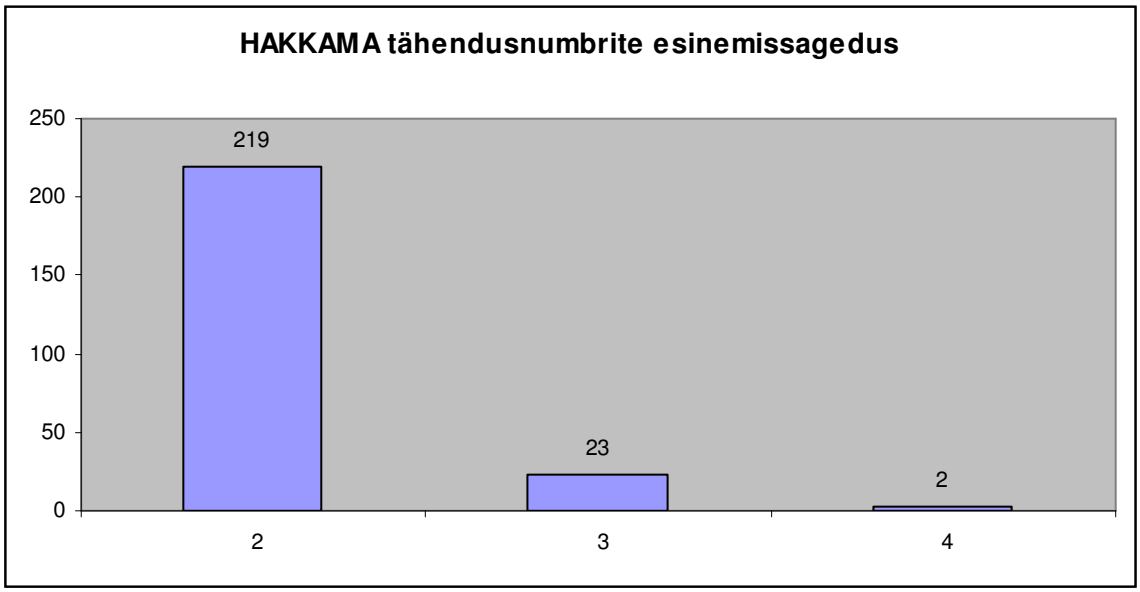

TULEMA tähendusnumbrite esinemissagedus
120
100
80
60
40
20
0
33
28
61
89
107
64
+1
1
2
3
4
5

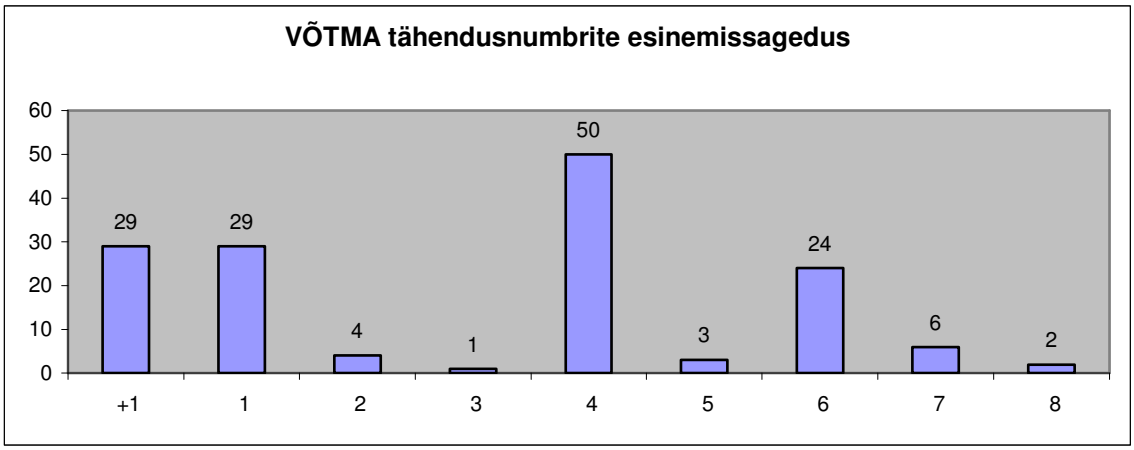

PIDAMA tähendusnumbrite esinemissagedus
140
120
100
80
60
40
20
0
8
131
4
6
2
2
3
26
3
2
73
8
6
27
2
+1
1
10
11
12
13
14
2
3
4
5
5 või 1
6
7
8

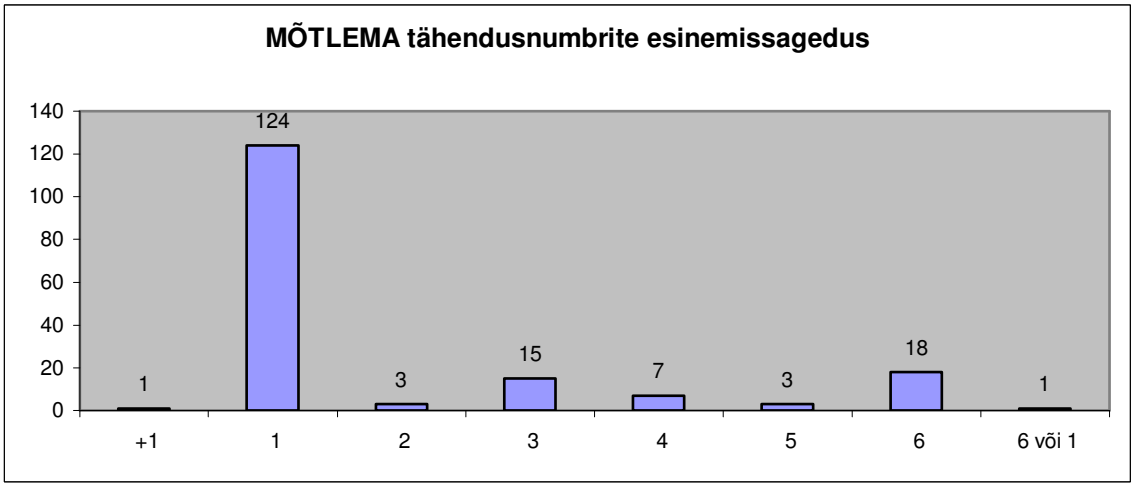

TEGEMA tähendusnumbrite esinemissagedus
160
140
120
100
80
60
40
20
0
18
18
6
148
47
11
29
+1
3
4
5
6
7
8

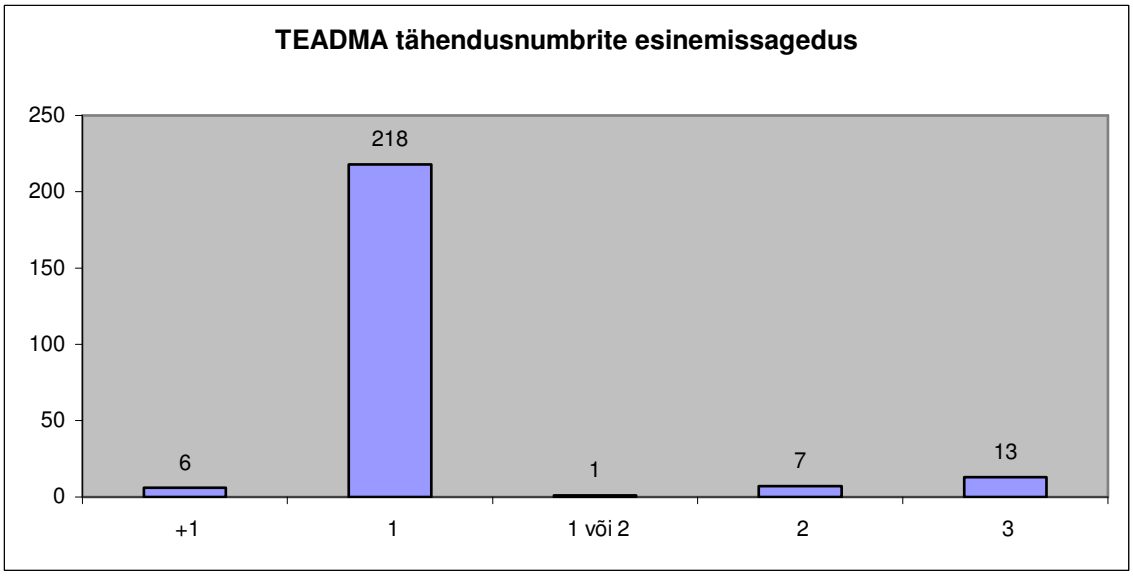

MINEMA tähendusnumbrite esinemissagedus
180
160
140
120
100
80
60
40
20
0
3
27
2
5
3
12
169
1
41
2
1
6
49
7
+1
1
10
13
15
2
3
3 või 4
4
5
6
7
8
9

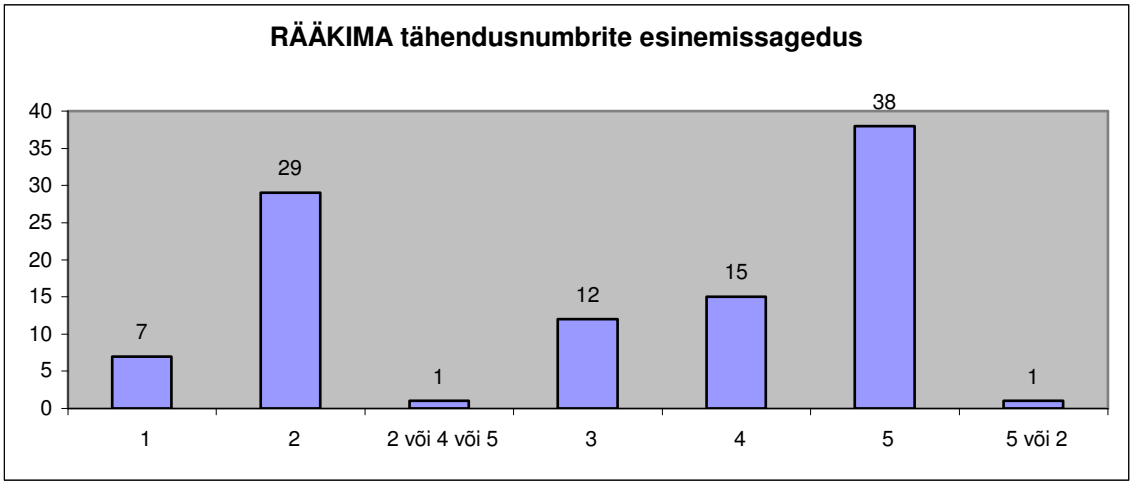

JÄÄMA tähendusnumbrite esinemissagedus
80
70
60
50
40
30
20
10
0
1
45
20
9
25
70
+1
1
2
3
4
5

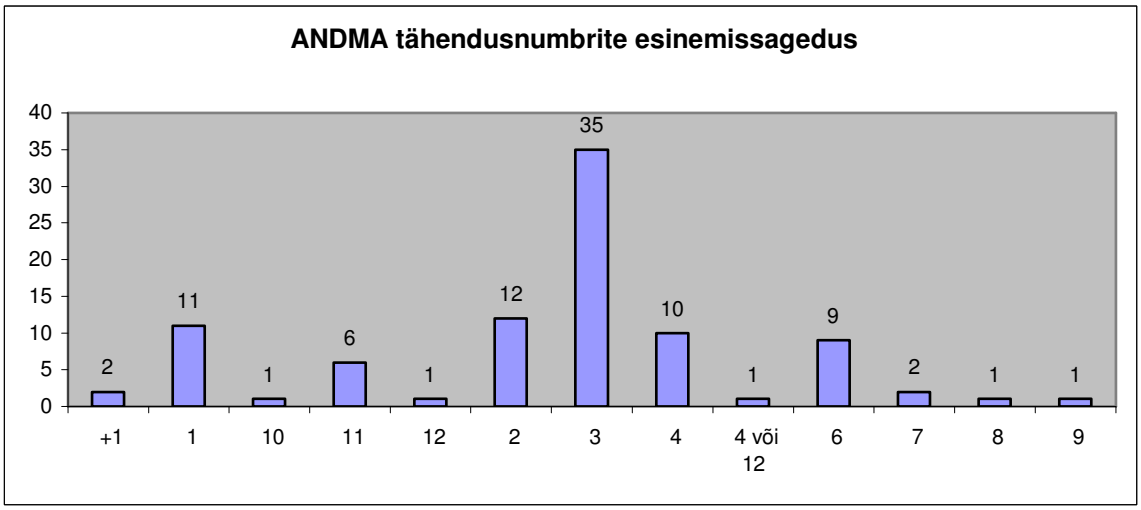

HAKKAMA tähendusnumbrite esinemissagedus
250
200
150
100
50
0
219
23
2
2
3
4

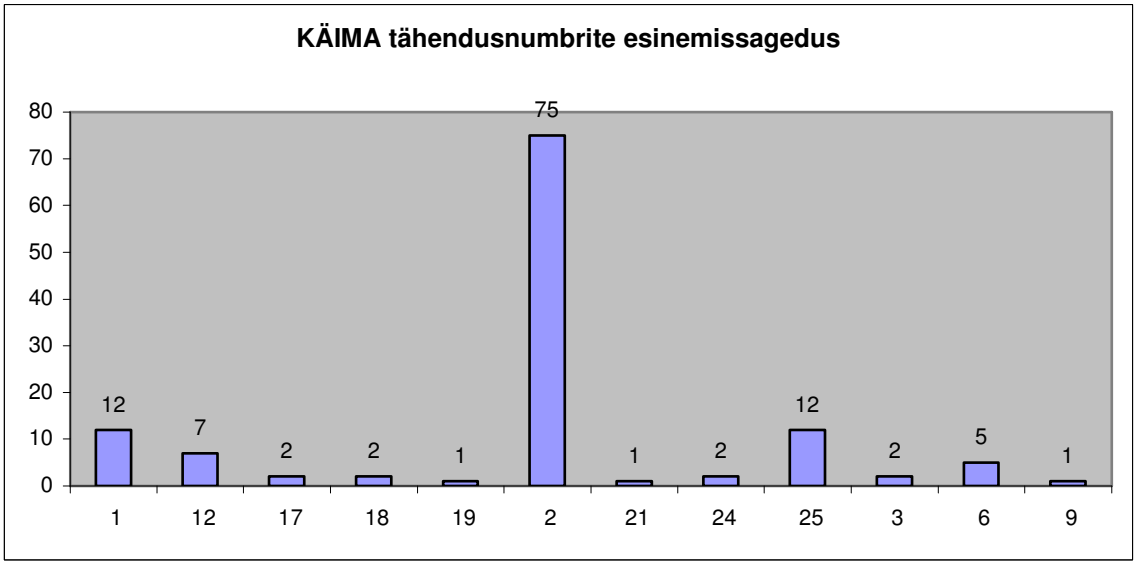

VÕTMA tähendusnumbrite esinemissagedus
60
50
40
30
20
10
0
29
29
4
1
50
3
24
6
2
+1
1
2
3
4
5
6
7
8

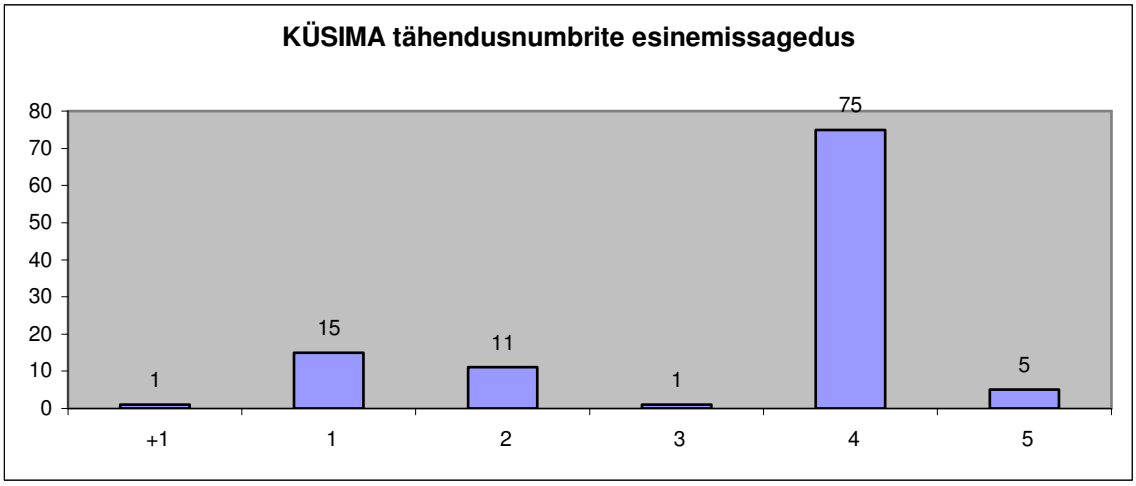

MÕTLEMA tähendusnumbrite esinemissagedus
140
120
100
80
60
40
20
0
1
124
3
15
7
3
18
1
+1
1
2
3
4
5
6
6 või 1

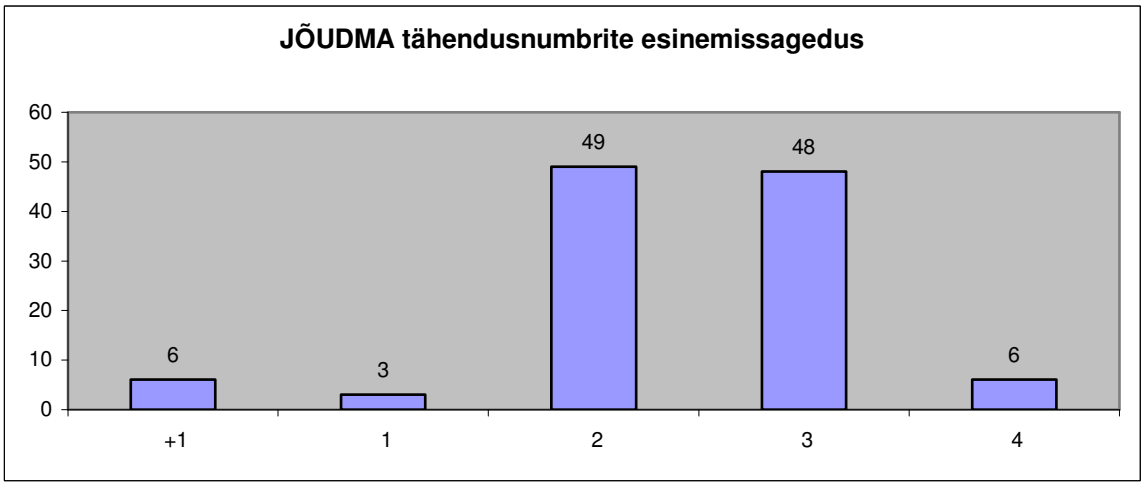

TEADMA tähendusnumbrite esinemissagedus
250
200
150
100
50
0
6
218
1
7
13
+1
1
1 või 2
2
3

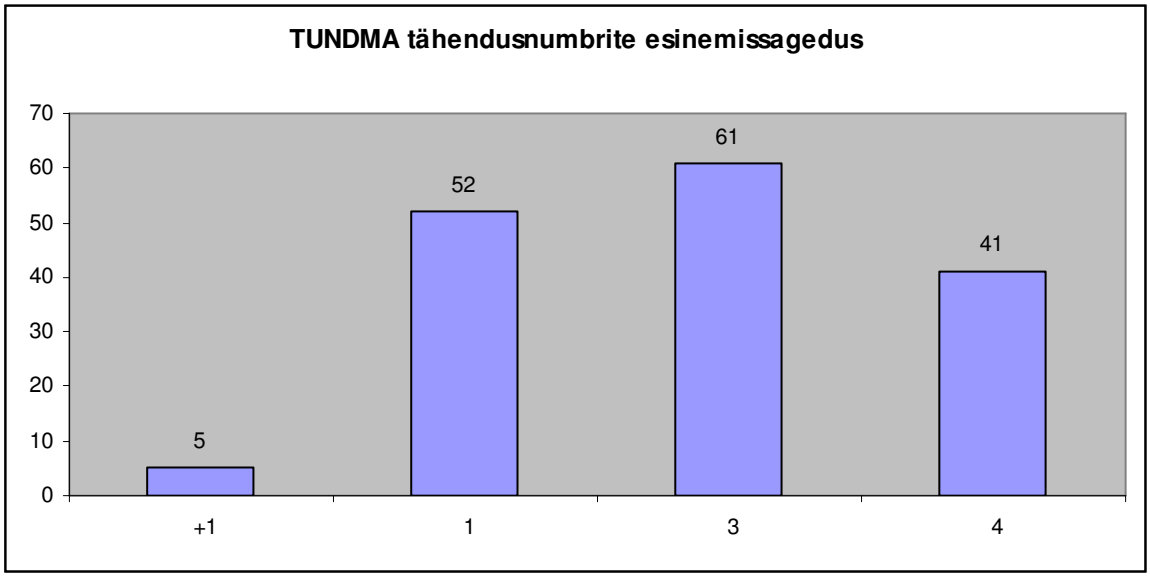

RÄÄKIMA tähendusnumbrite esinemissagedus
40
35
30
25
20
15
10
5
0
7
29
1
12
15
38
1
1
2
2 või 4 või 5
3
4
5
5 või 2

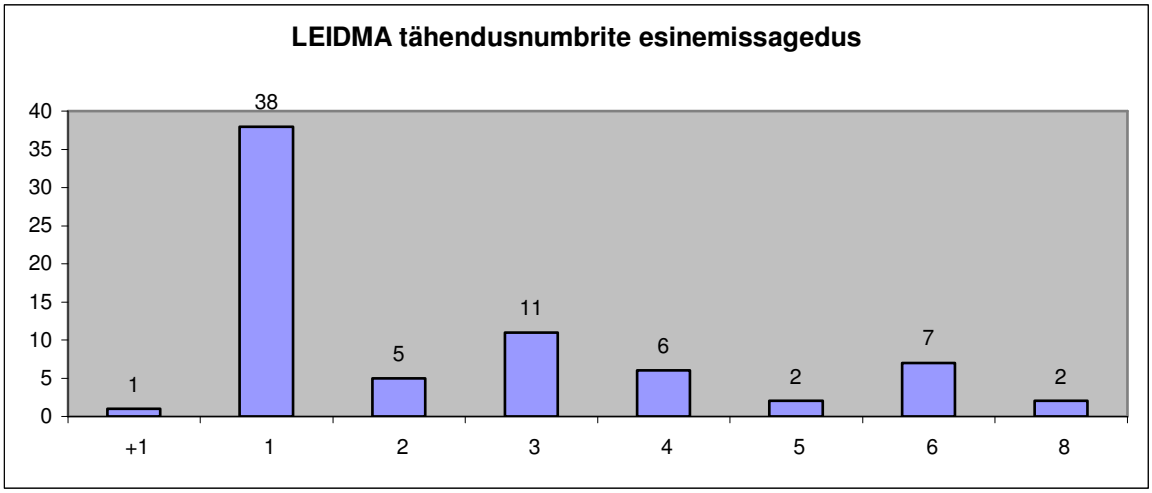

ANDMA tähendusnumbrite esinemissagedus
40
35
30
25
20
15
10
5
0
2
11
1
6
1
12
35
10
1
9
2
1
1
+1
1
10
11
12
2
3
4
4 või 12
6
7
8
9

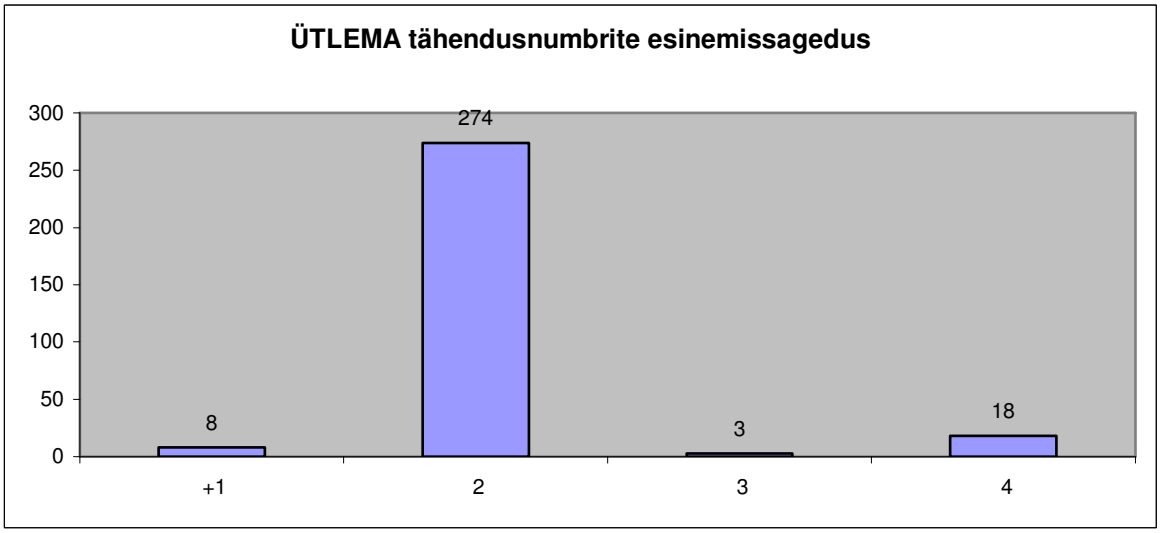

KÄIMA tähendusnumbrite esinemissagedus
80
70
60
50
40
30
20
10
0
12
7
2
2
1
75
1
2
12
2
5
1
1
12
17
18
19
2
21
24
25
3
6
9

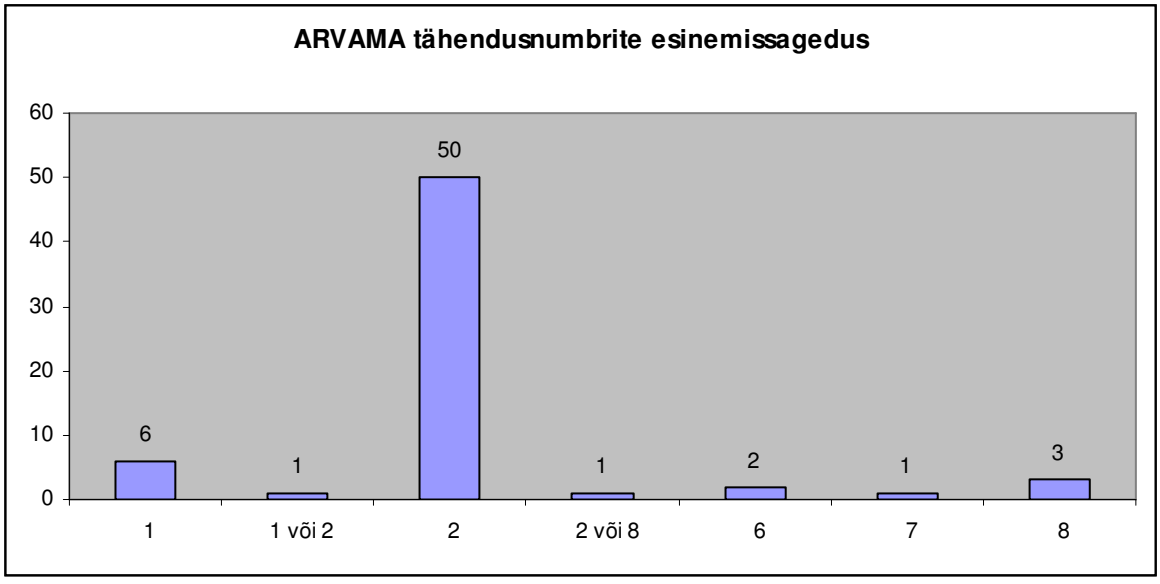

KÜSIMA tähendusnumbrite esinemissagedus
80
70
60
50
40
30
20
10
0
1
15
11
1
75
5
+1
1
2
3
4
5

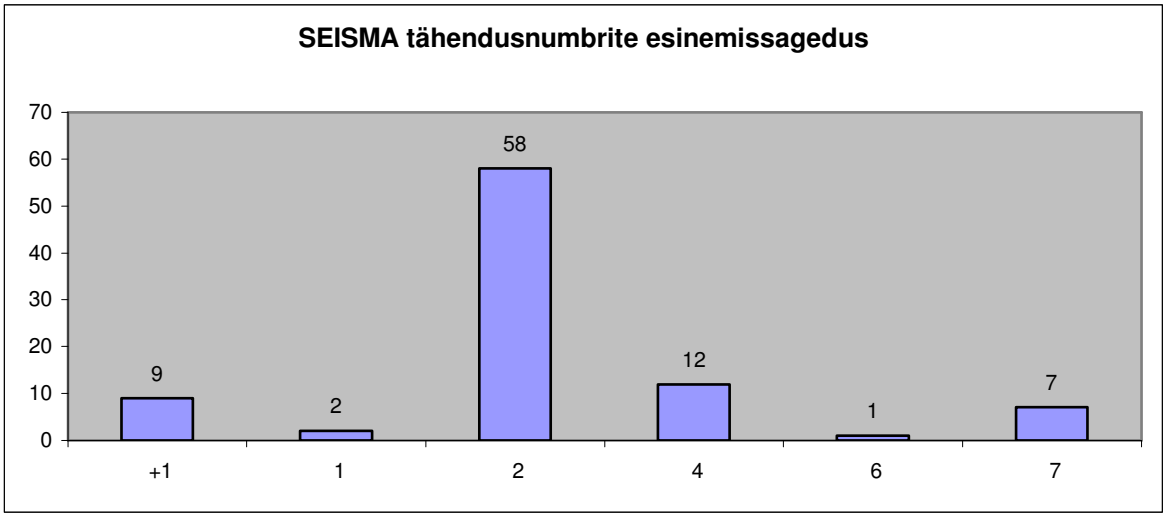

JÕUDMA tähendusnumbrite esinemissagedus
60
50
40
30
20
10
0
6
3
49
48
6
+1
1
2
3
4

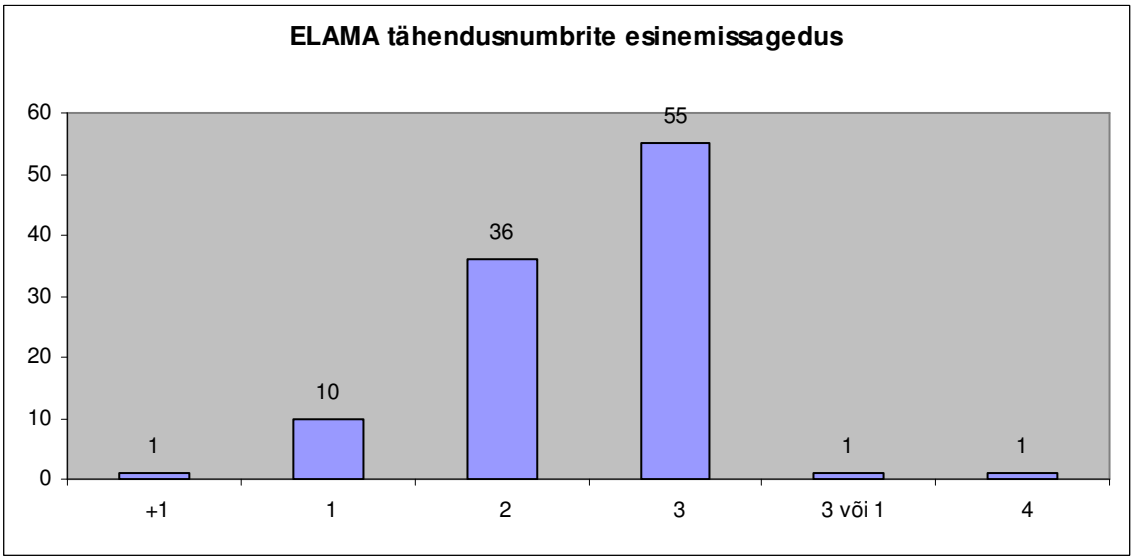

TUNDMA tähendusnumbrite esinemissagedus
70
60
50
40
30
20
10
0
5
52
61
41
+1
1
3
4

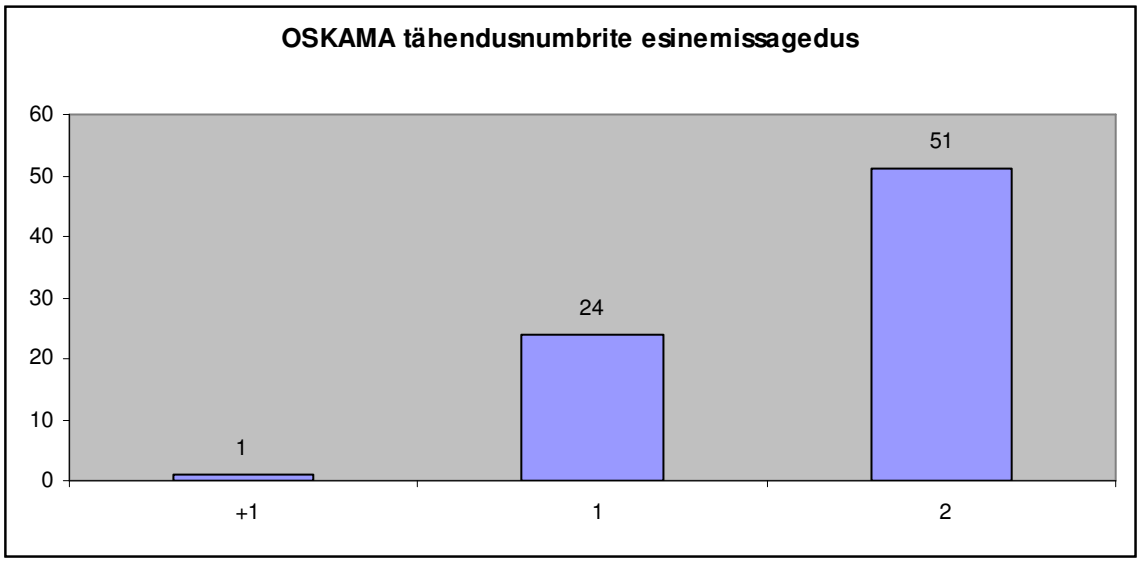

LEIDMA tähendusnumbrite esinemissagedus
40
35
30
25
20
15
10
5
0
38
11
6
7
5
2
2
1
+1
1
2
3
4
5
6
8

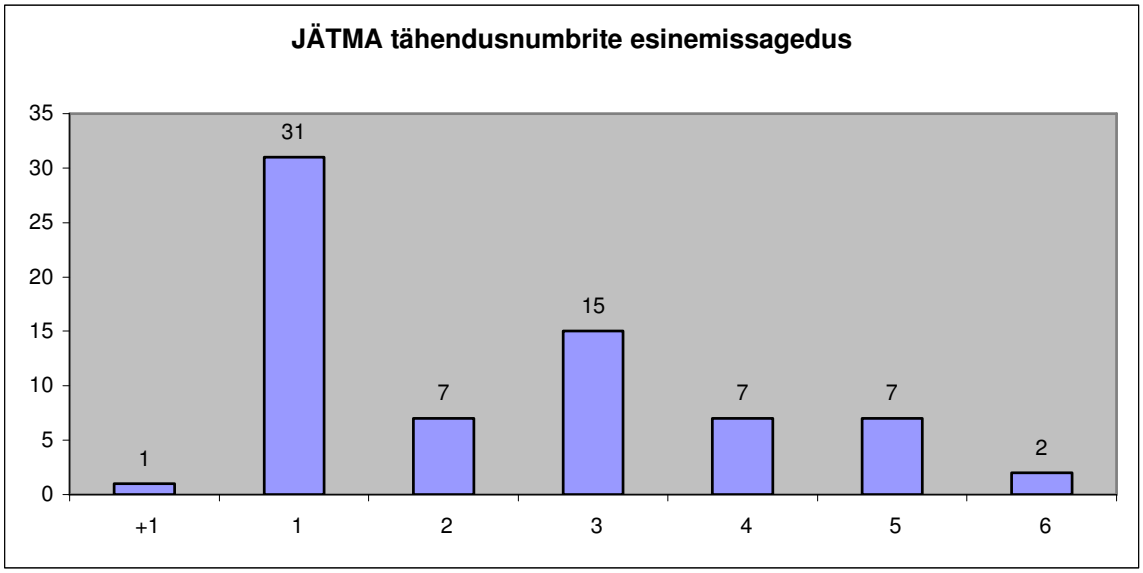

ÜTLEMA tähendusnumbrite esinemissagedus
300
250
200
150
100
50
0
274
8
3
18
+1
2
3
4

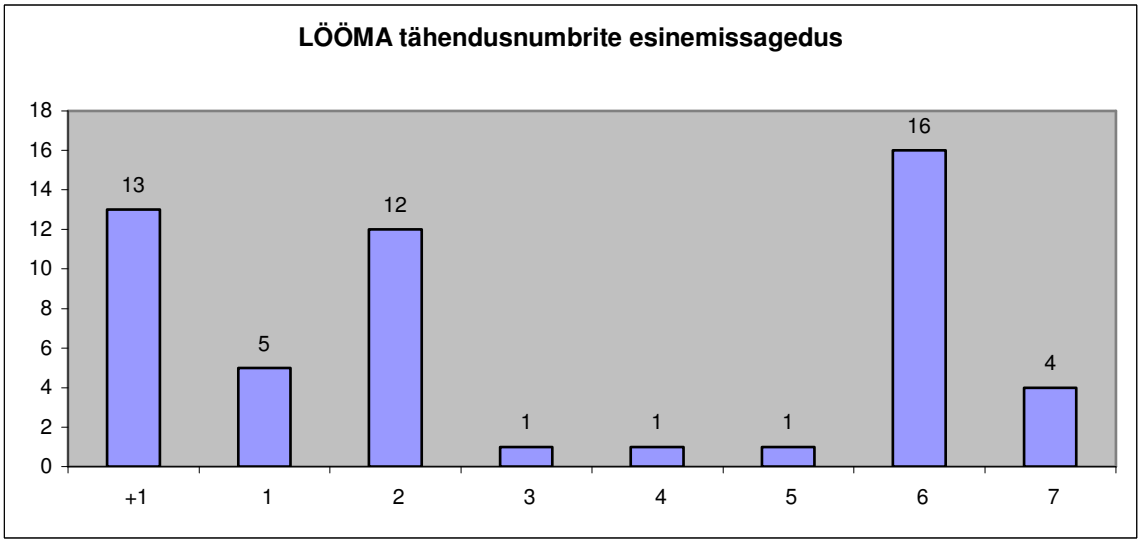

ARVAMA tähendusnumbrite esinemissagedus
60
50
40
30
20
10
0
50
6
1
1
2
1
3
1
1 või 2
2
2 või 8
6
7
8

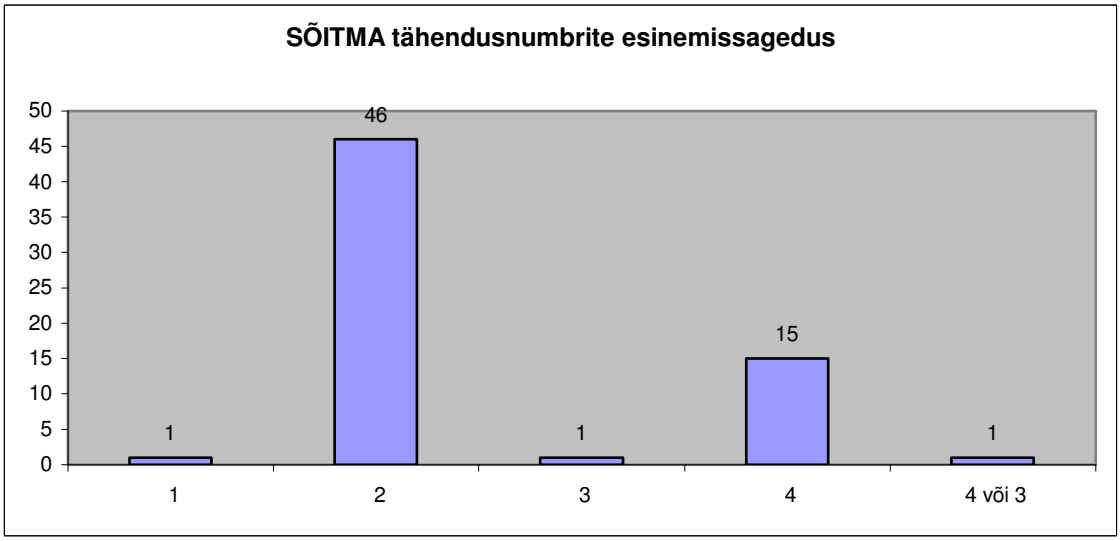

SEISMA tähendusnumbrite esinemissagedus
70
60
50
40
30
20
10
0
9
2
58
12
1
7
+1
1
2
4
6
7

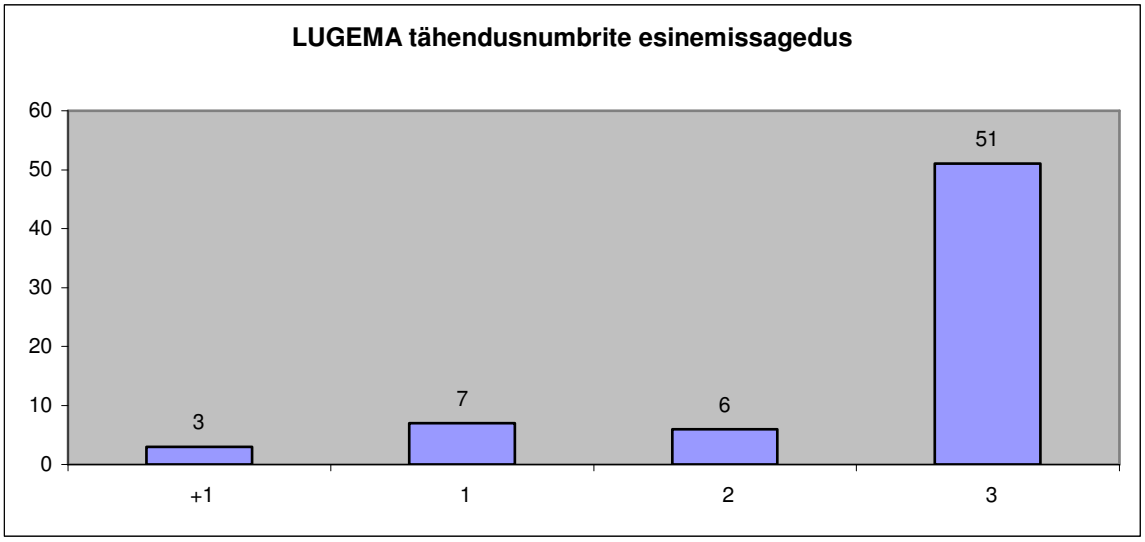

ELAMA tähendusnumbrite esinemissagedus
60
50
40
30
20
10
0
1
10
36
55
1
1
+1
1
2
3
3 või 1
4

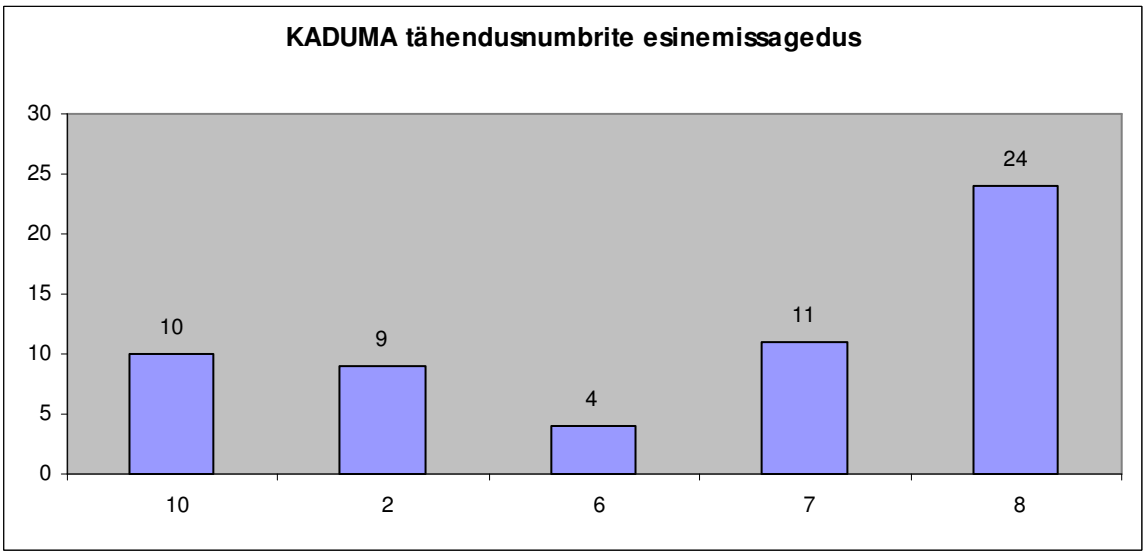

OSKAMA tähendusnumbrite esinemissagedus
60
50
40
30
20
10
0
1
24
51
+1
1
2

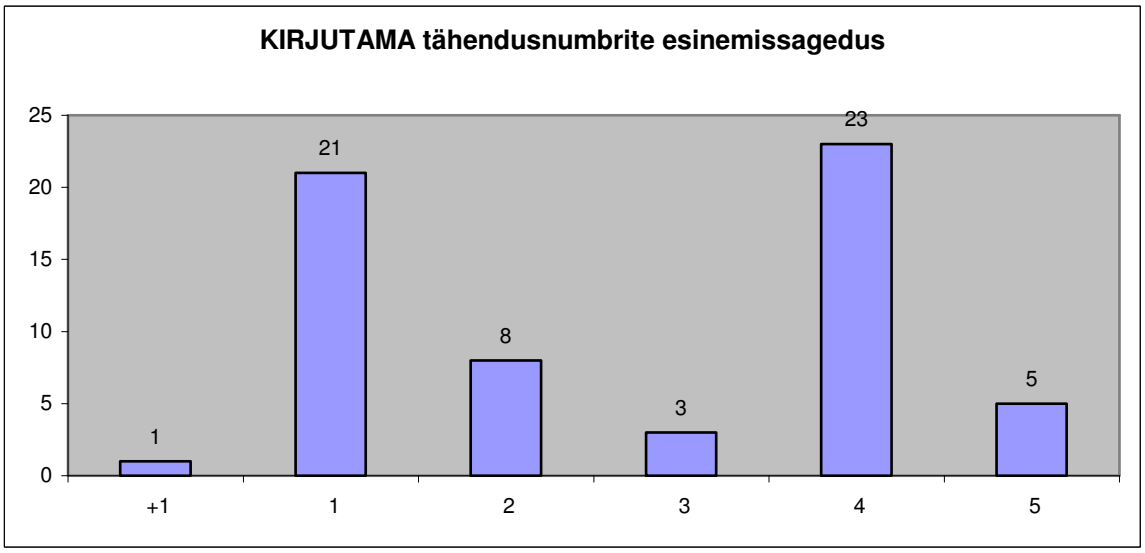

JÄTMA tähendusnumbrite esinemissagedus
35
30
25
20
15
10
5
0
1
31
7
15
7
7
2
+1
1
2
3
4
5
6

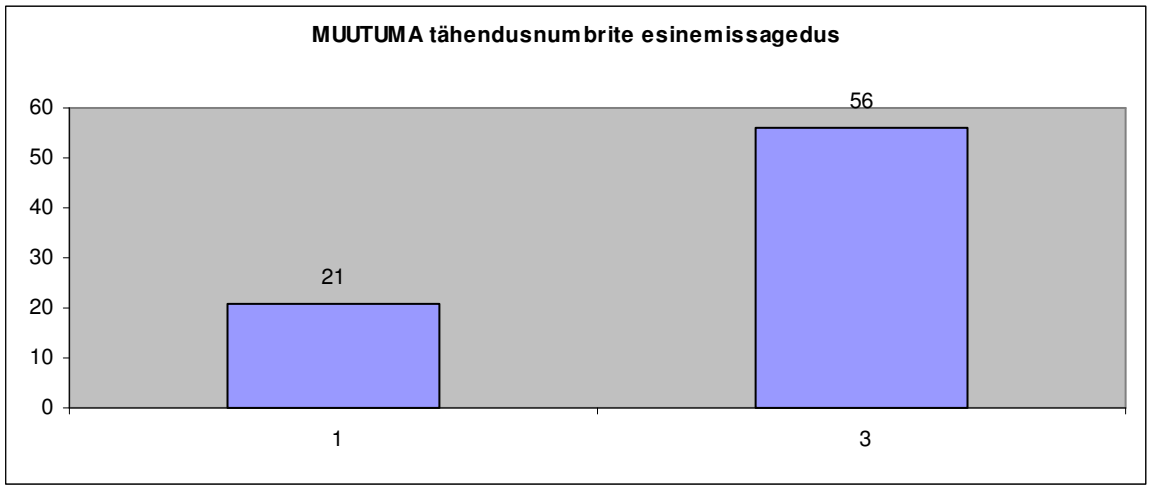

LÖÖMA tähendusnumbrite esinemissagedus
18
16
14
12
10
8
6
4
2
0
13
5
12
1
1
1
16
4
+1
1
2
3
4
5
6
7

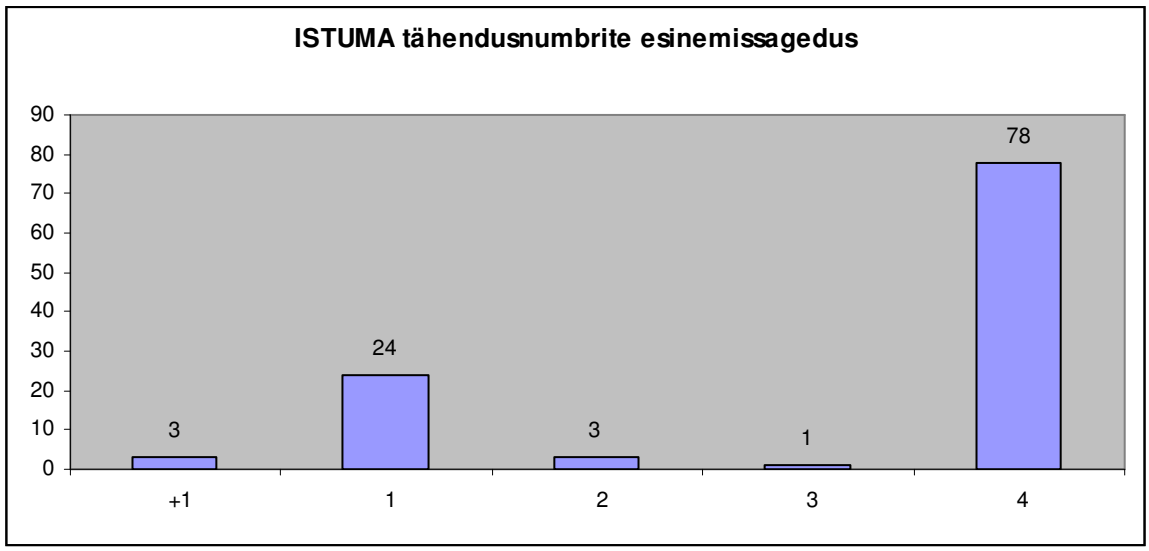

SÕITMA tähendusnumbrite esinemissagedus
50
45
40
35
30
25
20
15
10
5
0
1
46
1
15
1
1
2
3
4
4 või 3

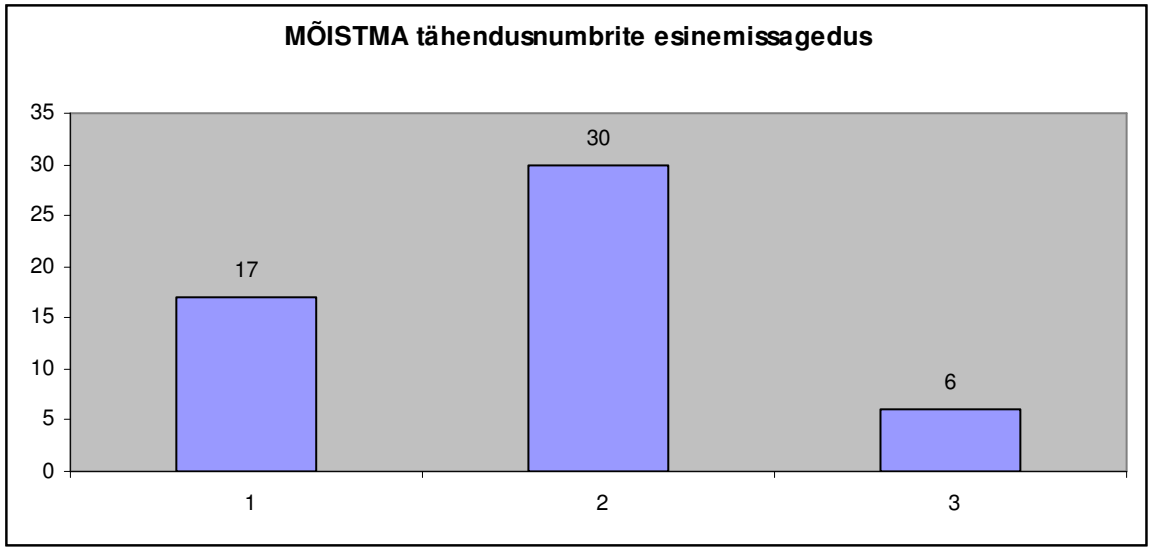

LUGEMA tähendusnumbrite esinemissagedus
60
50
40
30
20
10
0
3
7
6
51
+1
1
2
3

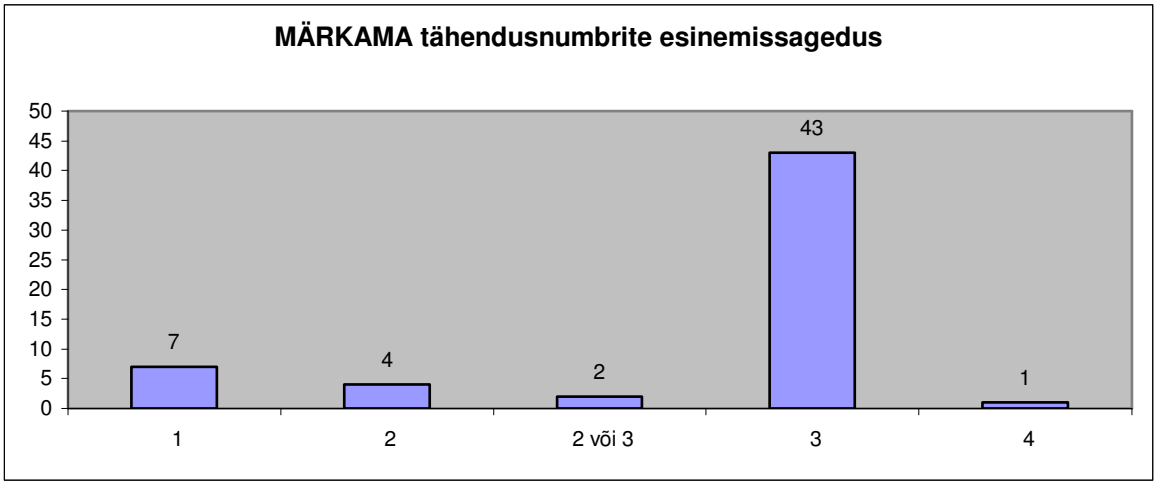

KADUMA tähendusnumbrite esinemissagedus
30
25
20
15
10
5
0
10
9
4
11
24
10
2
6
7
8

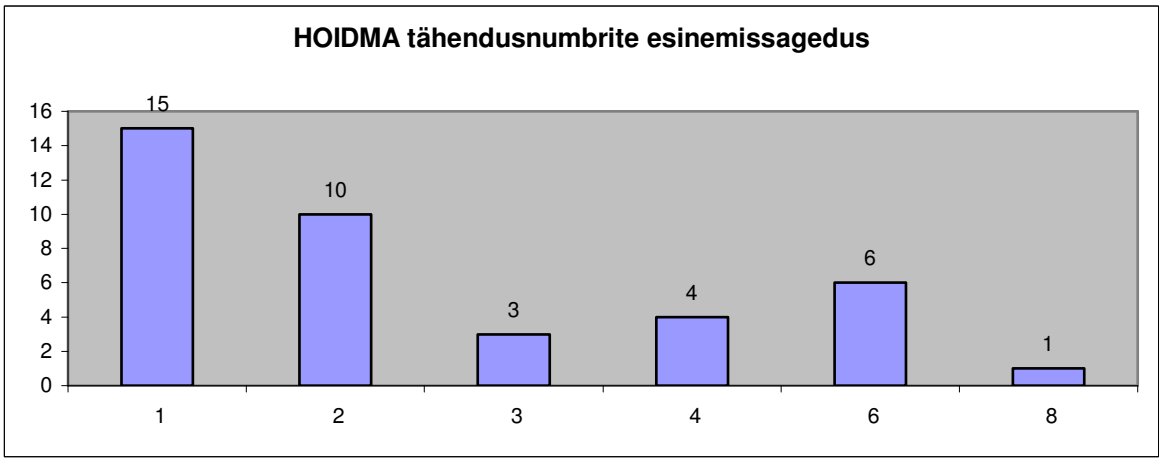

KIRJUTAMA tähendusnumbrite esinemissagedus
25
20
15
10
5
0
1
21
8
3
23
5
+1
1
2
3
4
5

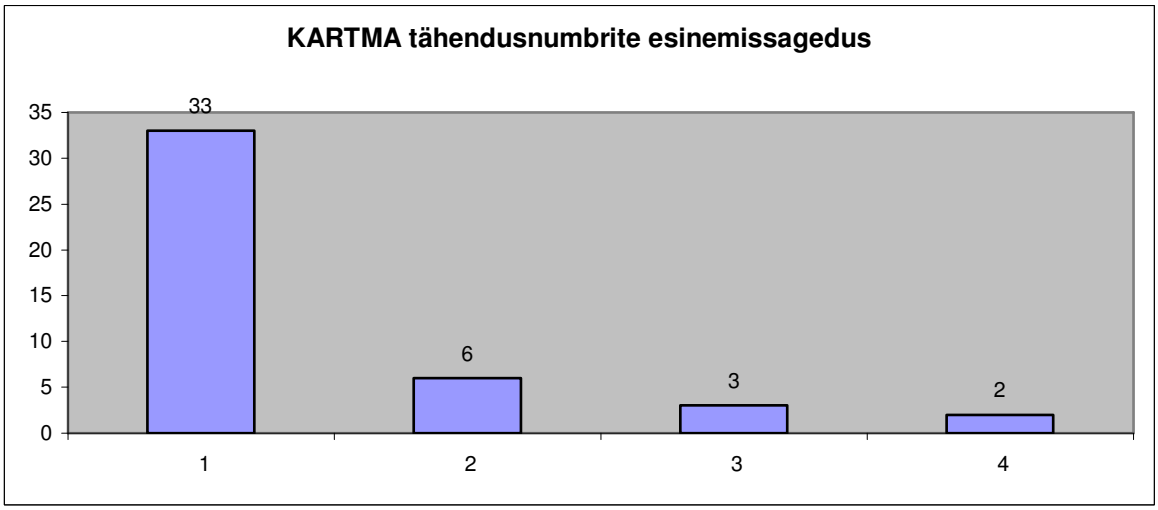

MUUTUMA tähendusnumbrite esinemissagedus
60
50
40
30
20
10
0
21
56
1
3

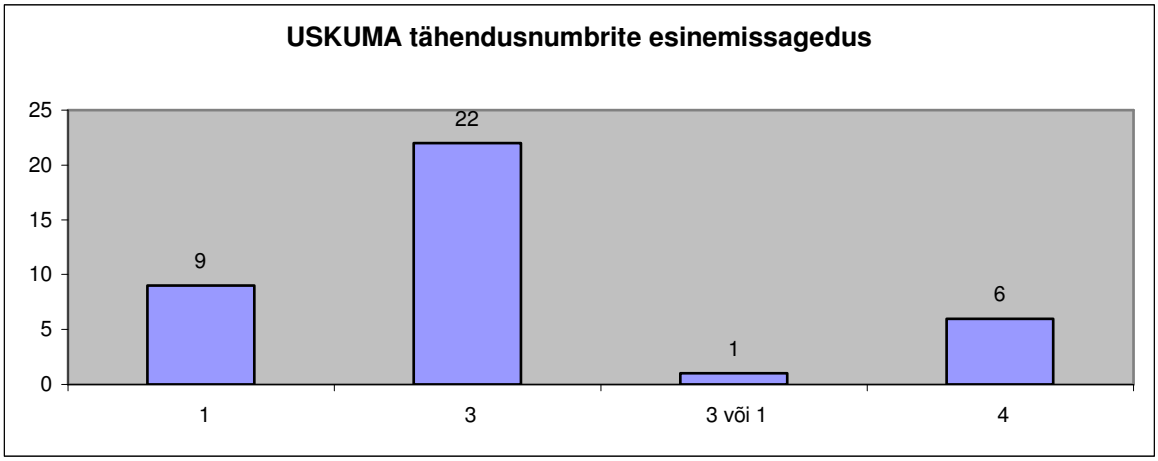

ISTUMA tähendusnumbrite esinemissagedus
90
80
70
60
50
40
30
20
10
0
3
24
3
1
78
+1
1
2
3
4

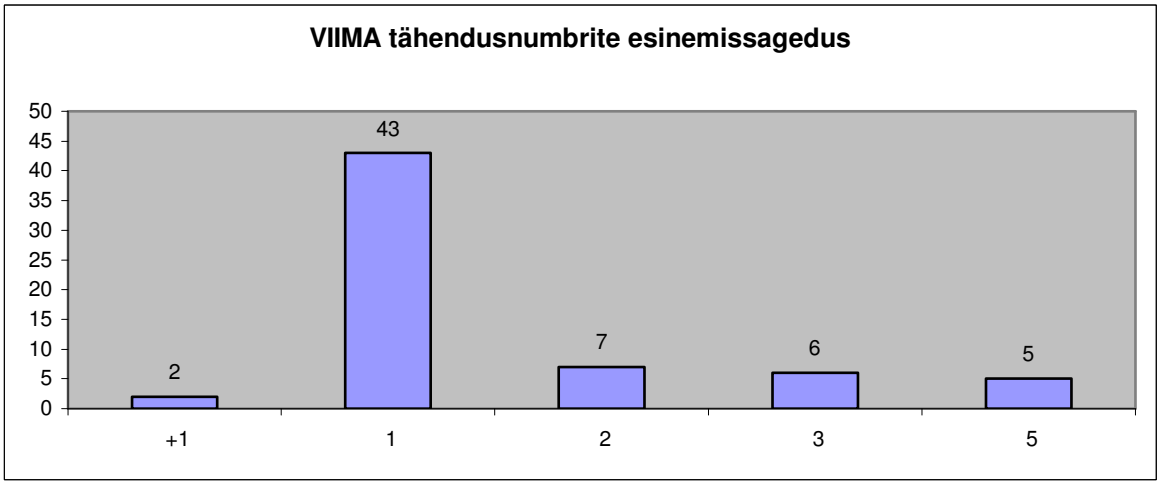

MÕISTMA tähendusnumbrite esinemissagedus
35
30
25
20
15
10
5
0
17
30
6
1
2
3

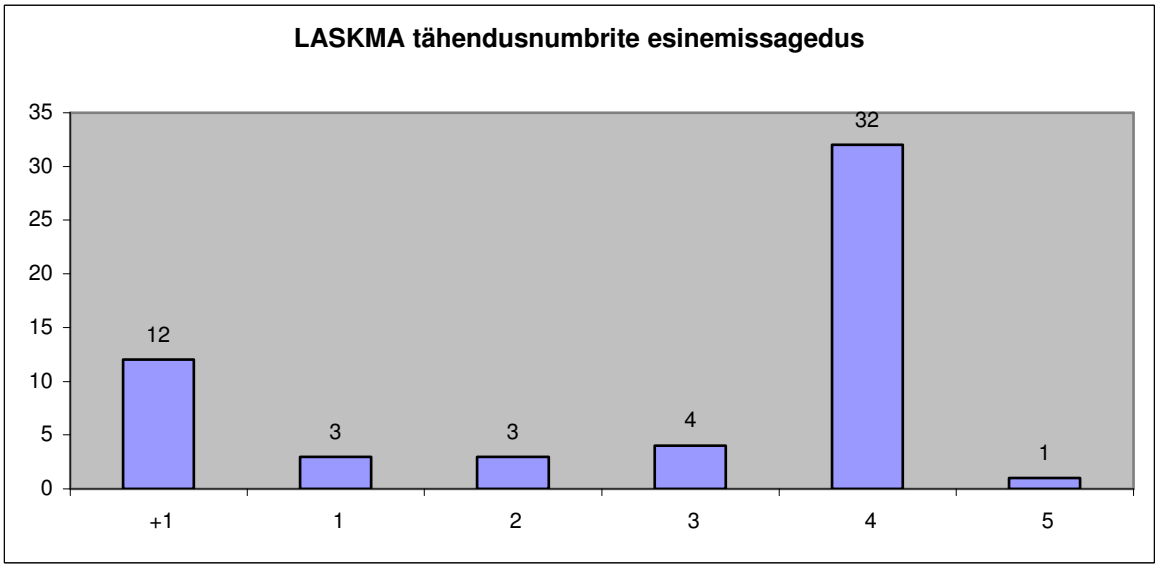

MÄRKAMA tähendusnumbrite esinemissagedus
50
45
40
35
30
25
20
15
10
5
0
7
4
2
43
1
1
2
2 või 3
3
4

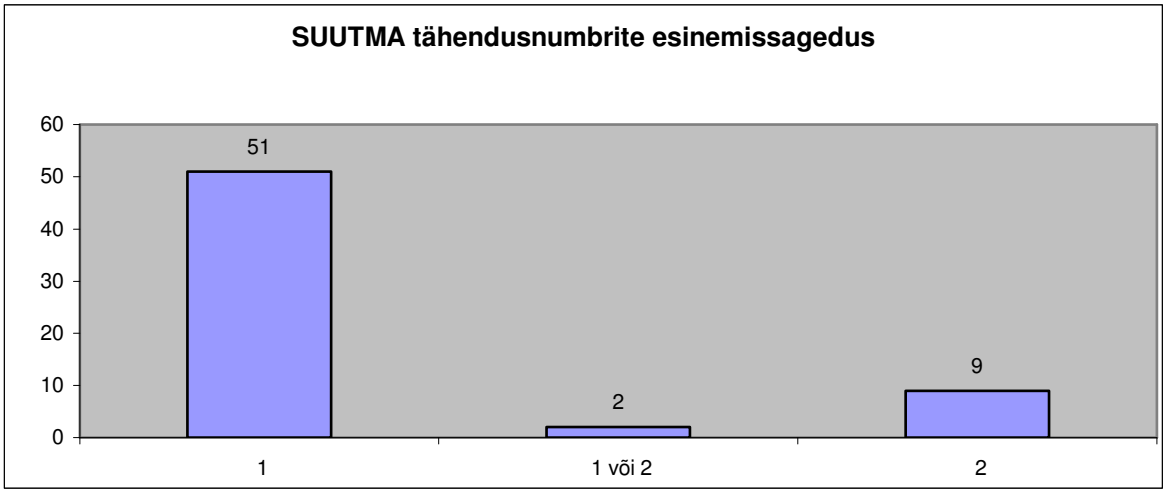

HOIDMA tähendusnumbrite esinemissagedus
16
14
12
10
8
6
4
2
0
15
10
3
4
6
1
1
2
3
4
6
8

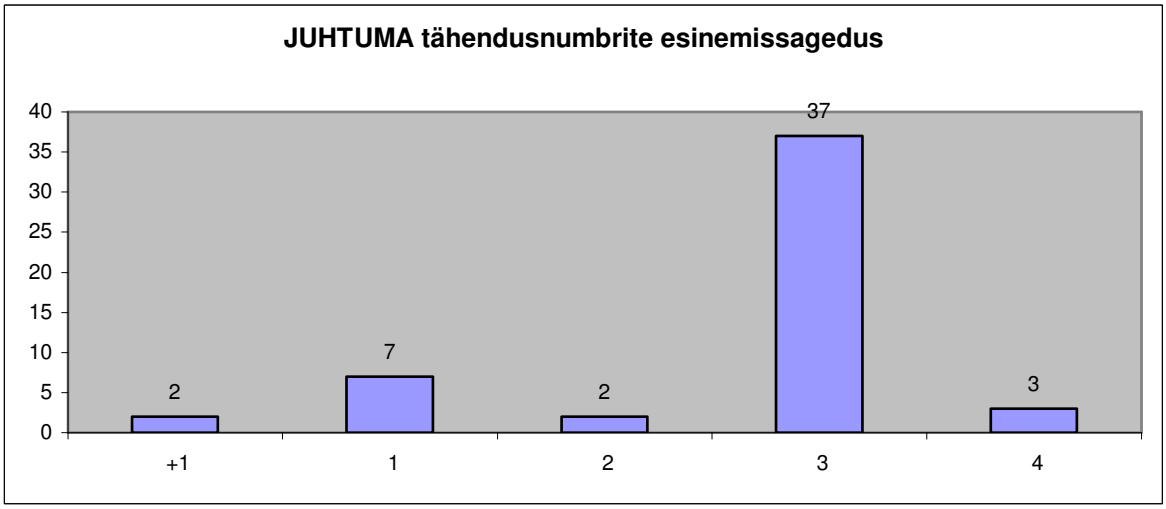

KARTMA tähendusnumbrite esinemissagedus
35
30
25
20
15
10
5
0
33
6
3
2
1
2
3
4

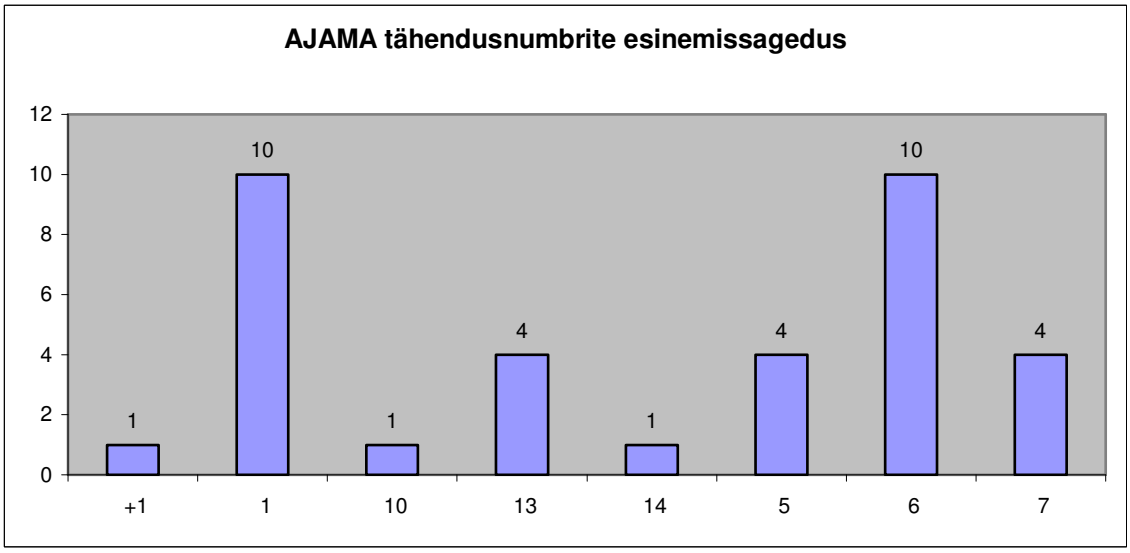

USKUMA tähendusnumbrite esinemissagedus
25
20
15
10
5
0
9
22
1
6
1
3
3 või 1
4

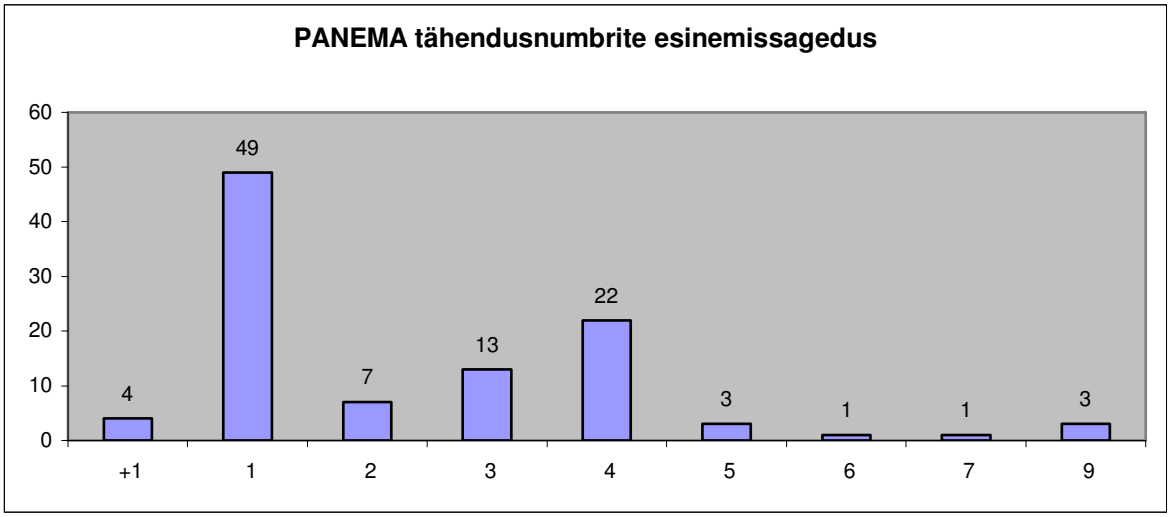

VIIMA tähendusnumbrite esinemissagedus
50
45
40
35
30
25
20
15
10
5
0
2
43
7
6
5
+1
1
2
3
5

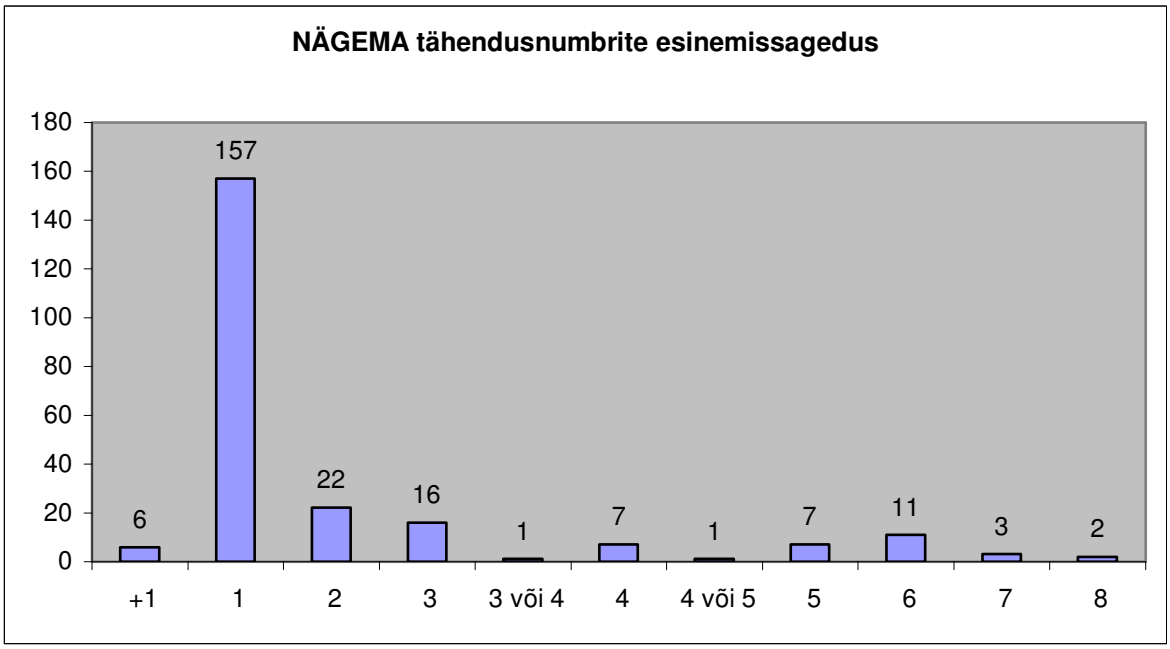

LASKMA tähendusnumbrite esinemissagedus
35
30
25
20
15
10
5
0
12
3
3
4
32
1
+1
1
2
3
4
5

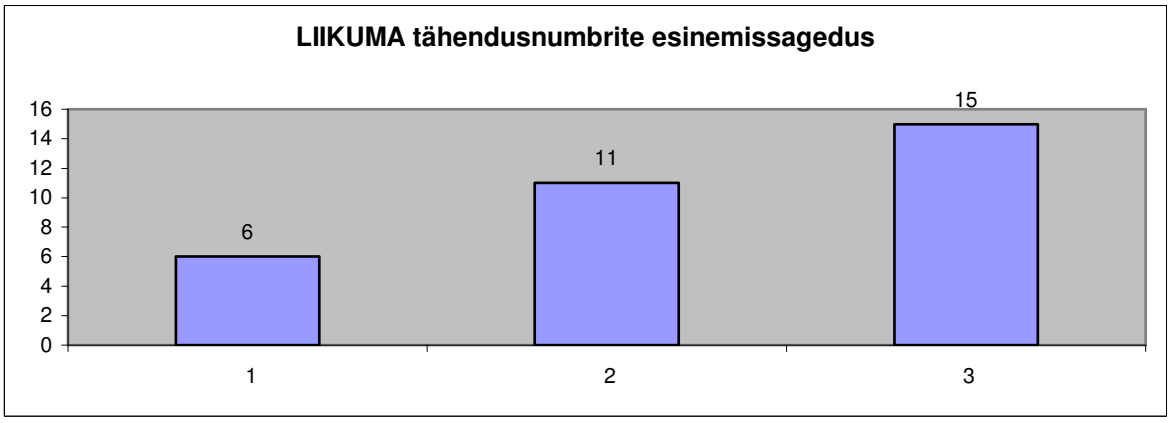

SUUTMA tähendusnumbrite esinemissagedus
60
50
40
30
20
10
0
51
2
9
1
1 või 2
2

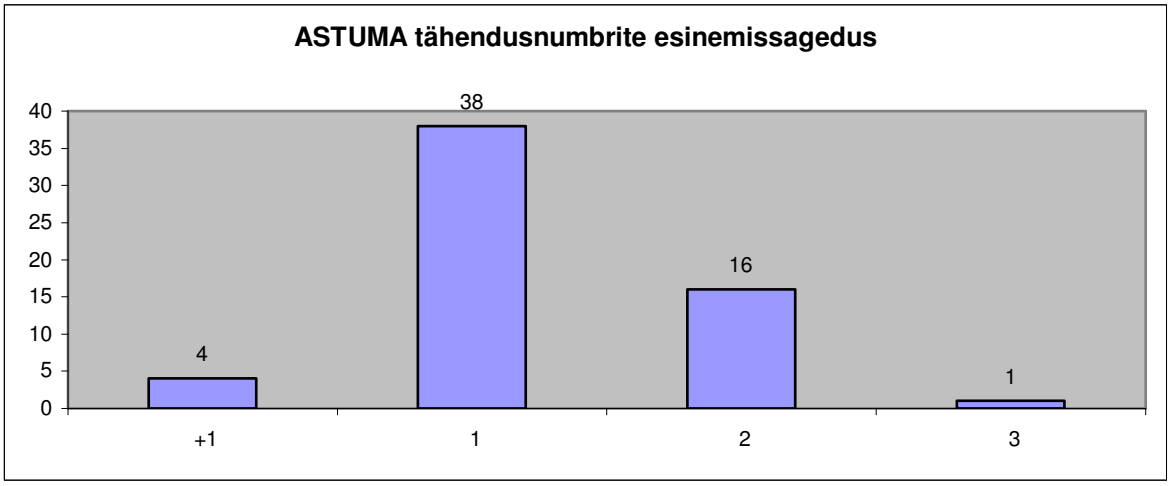

JUHTUMA tähendusnumbrite esinemissagedus
40
35
30
25
20
15
10
5
0
2
7
2
37
3
+1
1
2
3
4

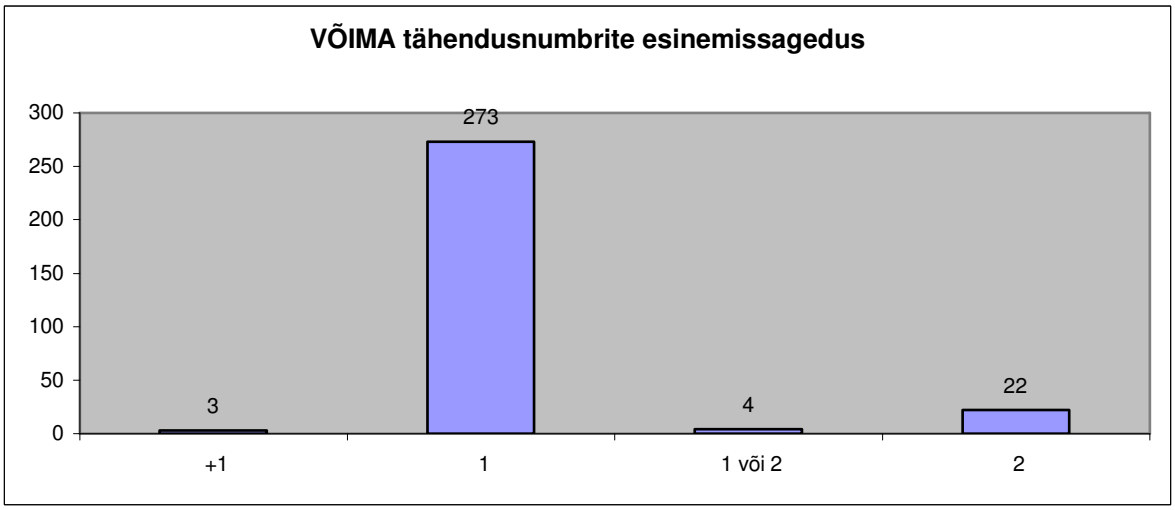

AJAMA tähendusnumbrite esinemissagedus
12
10
8
6
4
2
0
1
10
1
4
1
4
10
4
+1
1
10
13
14
5
6
7

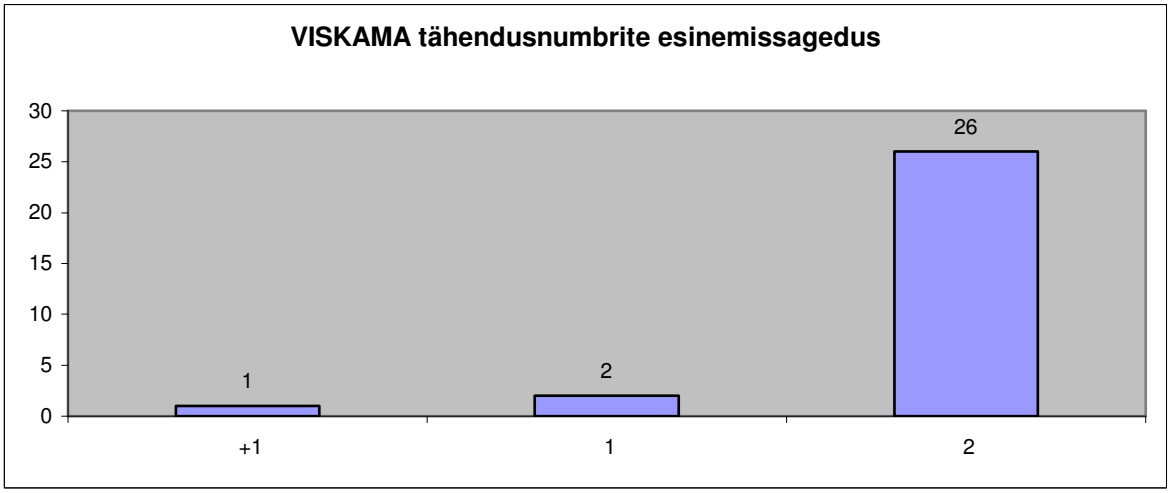

PANEMA tähendusnumbrite esinemissagedus
60
50
40
30
20
10
0
4
49
7
13
22
3
1
1
3
+1
1
2
3
4
5
6
7
9

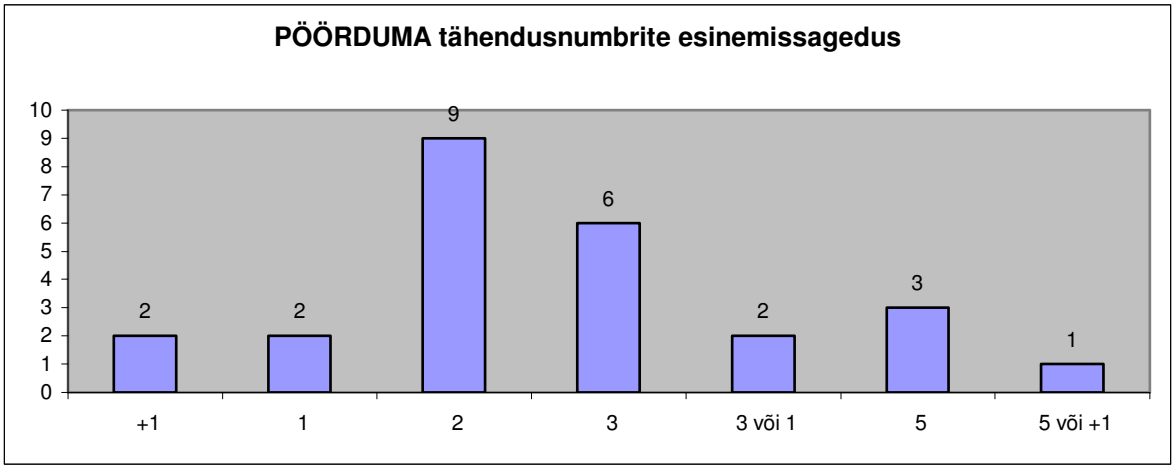

NÄGEMA tähendusnumbrite esinemissagedus
180
160
140
120
100
80
60
40
20
0
6
157
22
16
1
7
1
7
11
3
2
+1
1
2
3
3 või 4
4
4 või 5
5
6
7
8

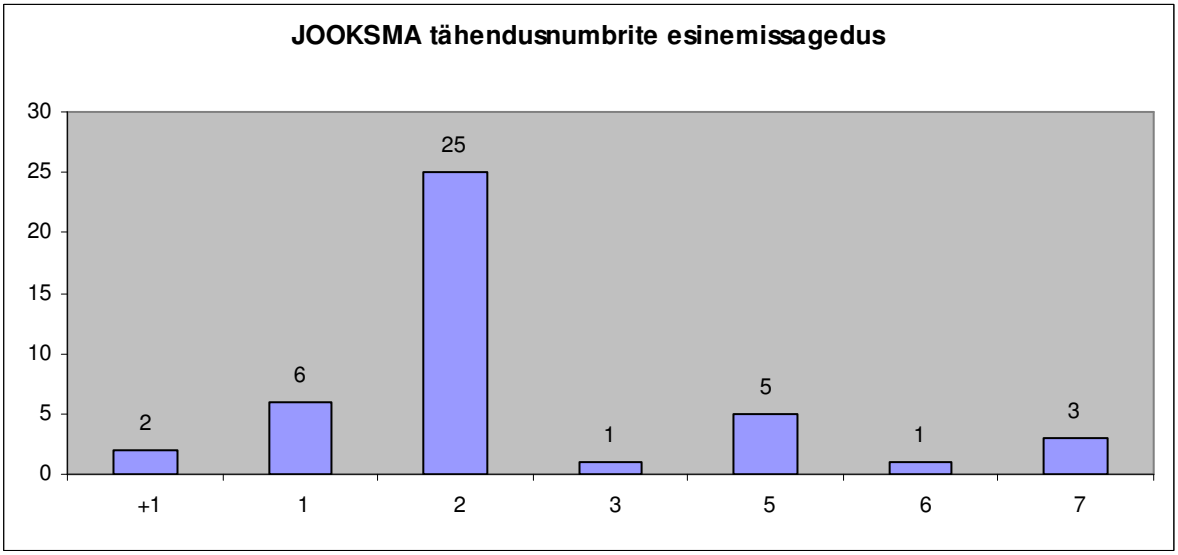

LIIKUMA tähendusnumbrite esinemissagedus
16
14
12
10
8
6
4
2
0
6
11
15
1
2
3

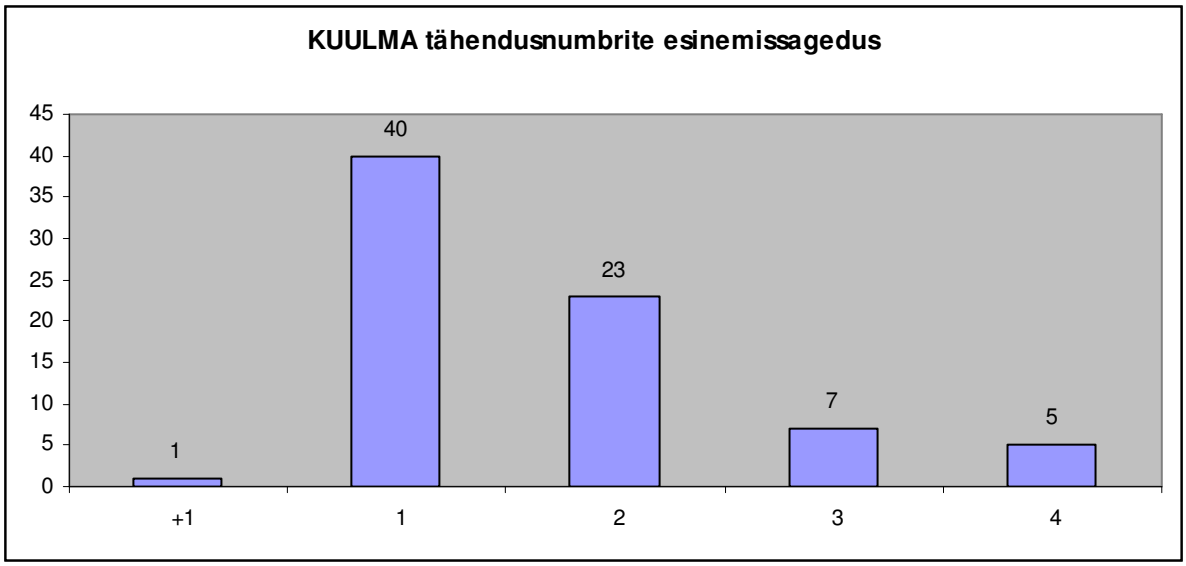

ASTUMA tähendusnumbrite esinemissagedus
40
35
30
25
20
15
10
5
0
4
38
16
1
+1
1
2
3

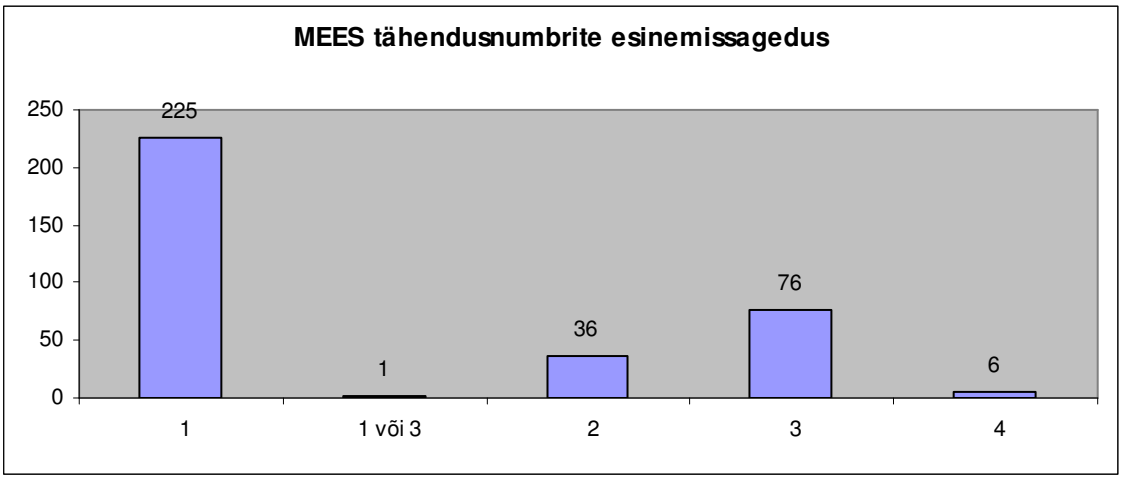

VÕIMA tähendusnumbrite esinemissagedus
300
250
200
150
100
50
0
3
273
4
22
+1
1
1 või 2
2

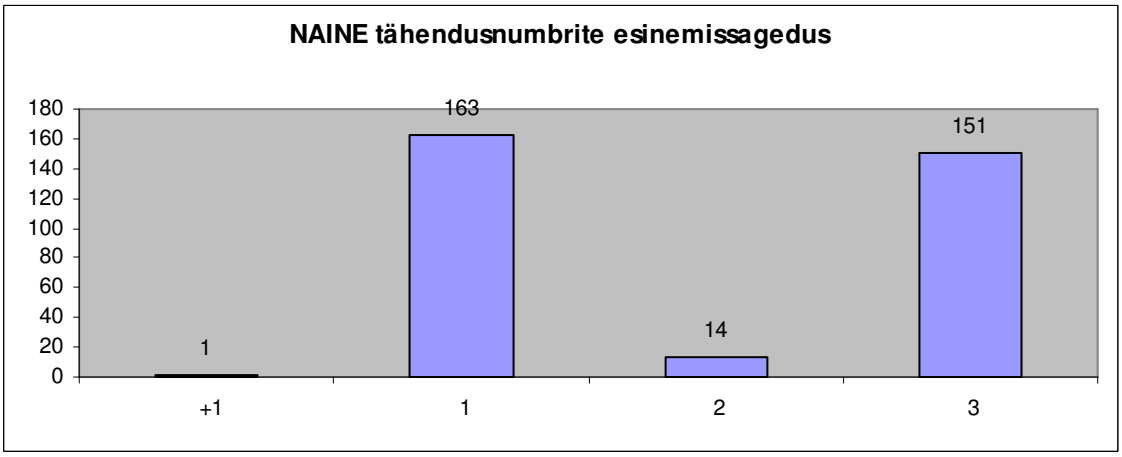

VISKAMA tähendusnumbrite esinemissagedus
30
25
20
15
10
5
0
1
2
26
+1
1
2

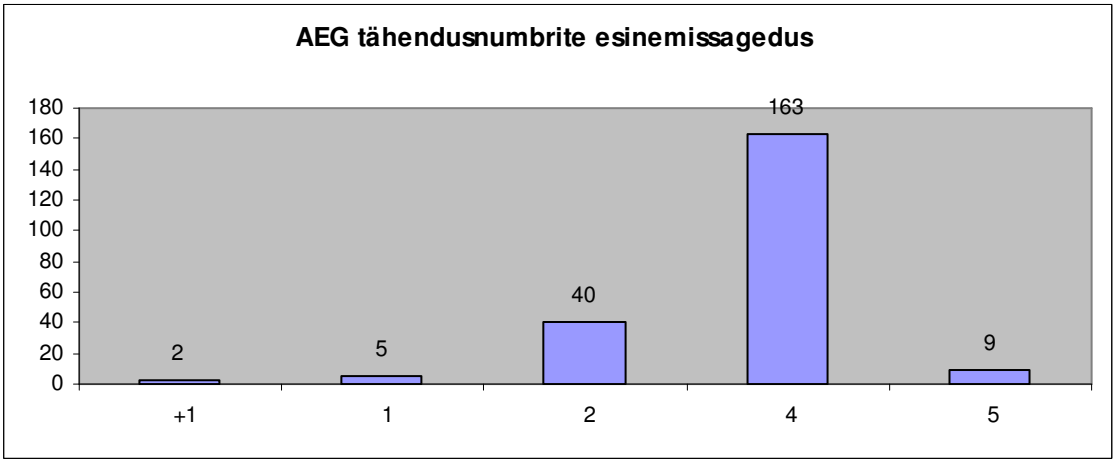

PÖÖRDUMA tähendusnumbrite esinemissagedus
10
9
8
7
6
5
4
3
2
1
0
2
2
9
6
2
3
1
+1
1
2
3
3 või 1
5
5 või +1

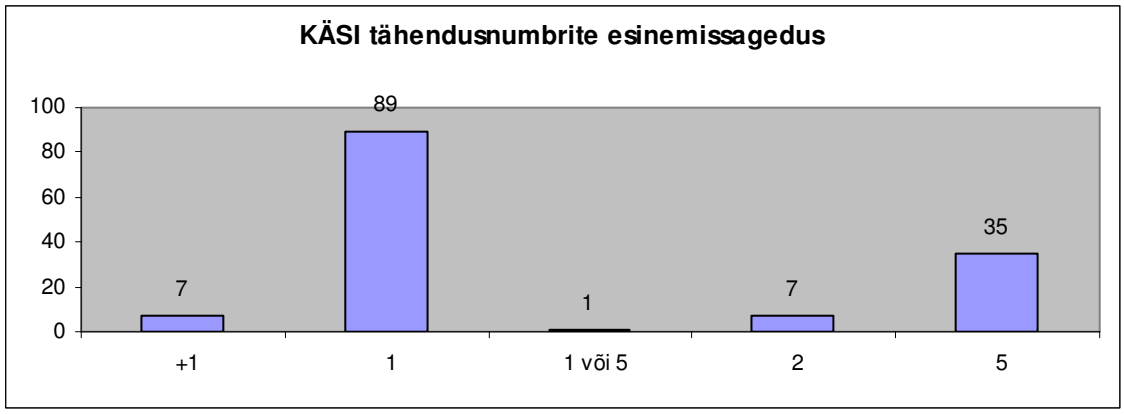

JOOKSMA tähendusnumbrite esinemissagedus
30
25
20
15
10
5
0
2
6
25
1
5
1
3
+1
1
2
3
5
6
7

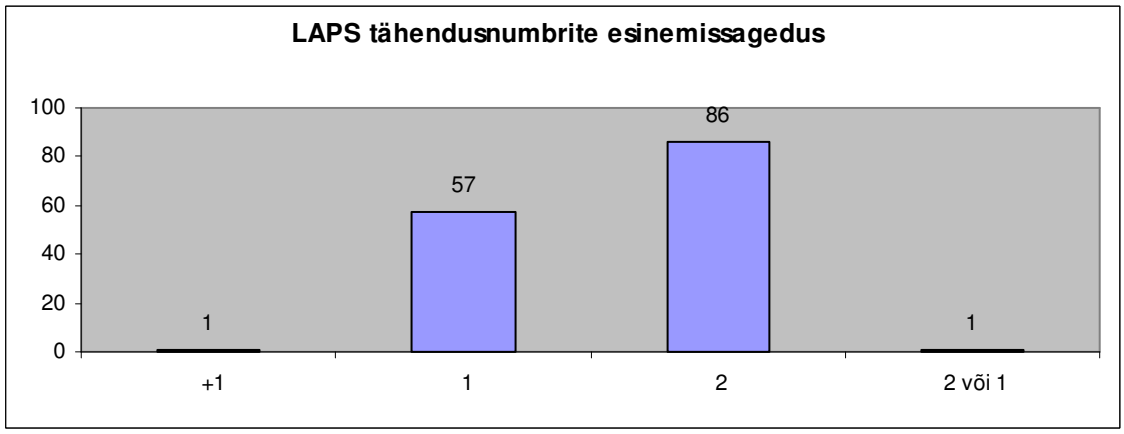

KUULMA tähendusnumbrite esinemissagedus
45
40
35
30
25
20
15
10
5
0
1
40
23
7
5
+1
1
2
3
4

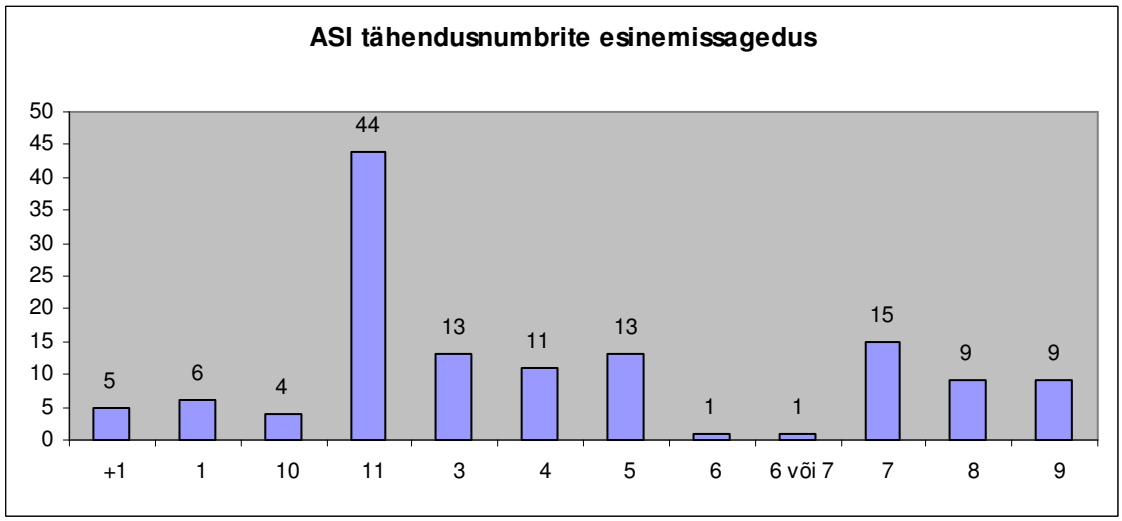

MEES tähendusnumbrite esinemissagedus
250
200
150
100
50
0
225
1
36
76
6
1
1 või 3
2
3
4

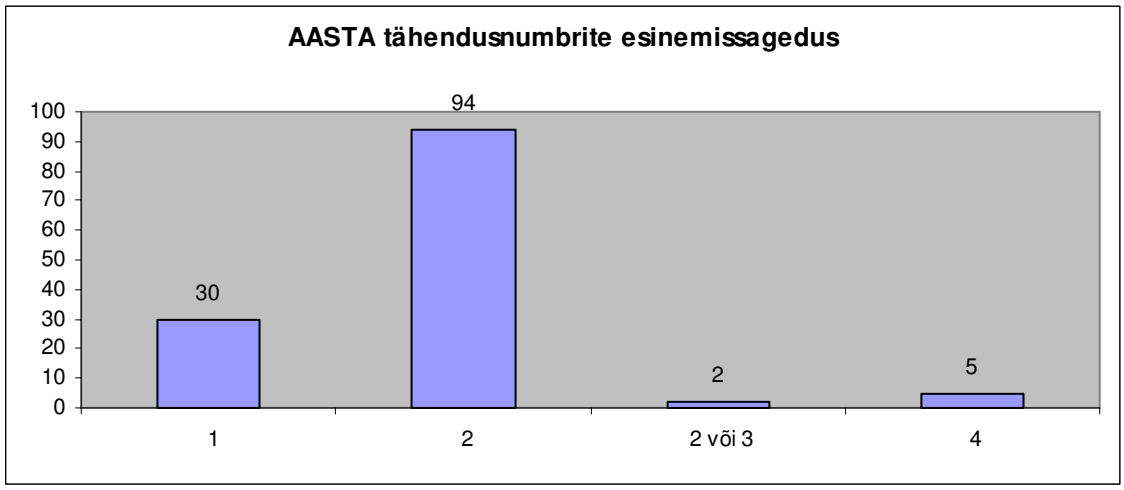

NAINE tähendusnumbrite esinemissagedus
180
160
140
120
100
80
60
40
20
0
1
163
14
151
+1
1
2
3

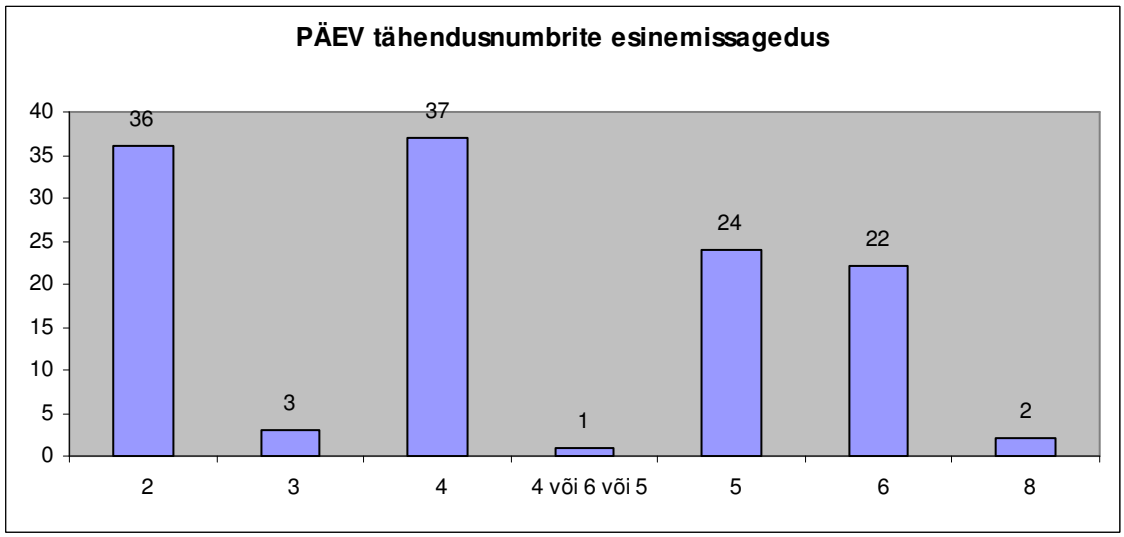

AEG tähendusnumbrite esinemissagedus
180
160
140
120
100
80
60
40
20
0
2
5
40
163
9
+1
1
2
4
5

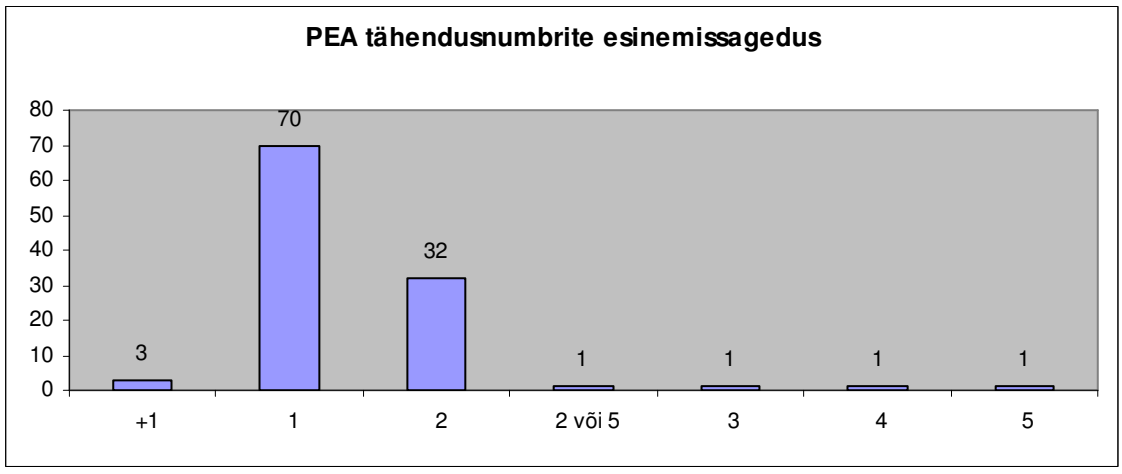

KÄSI tähendusnumbrite esinemissagedus
100
80
60
40
20
0
7
89
1
7
35
+1
1
1 või 5
2
5

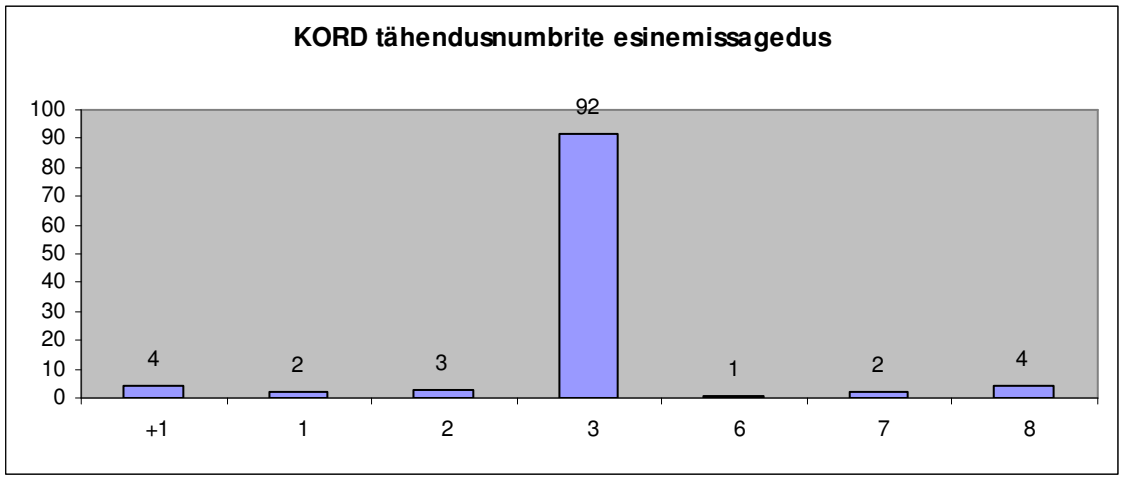

LAPS tähendusnumbrite esinemissagedus
100
80
60
40
20
0
1
57
86
1
+1
1
2
2 või 1

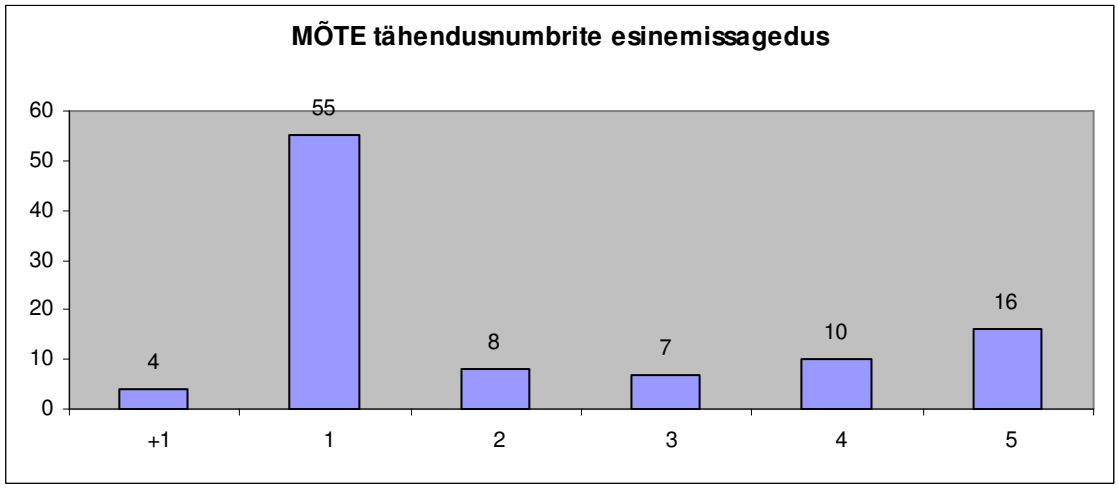

ASI tähendusnumbrite esinemissagedus
50
45
40
35
30
25
20
15
10
5
0
5
6
4
44
13
11
13
1
1
15
9
9
+1
1
10
11
3
4
5
6
6 või 7
7
8
9

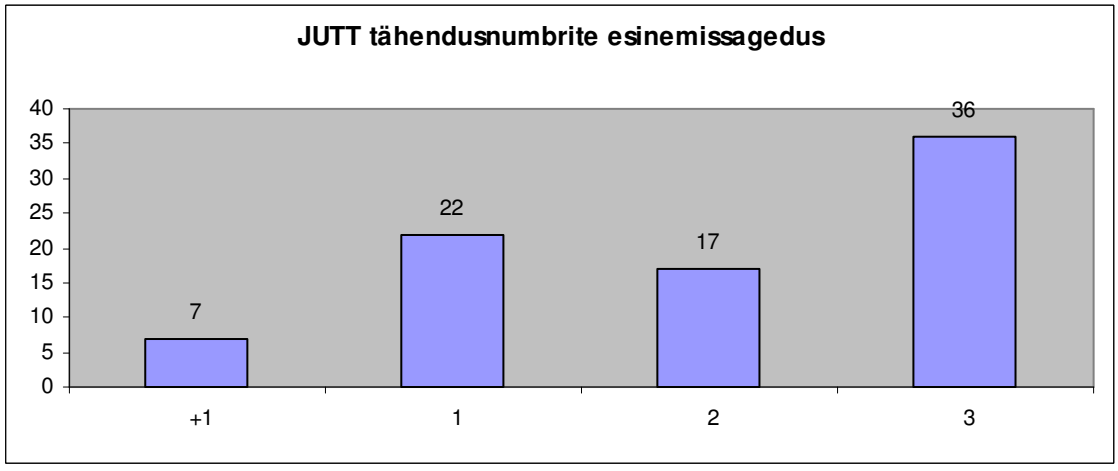

AASTA tähendusnumbrite esinemissagedus
100
90
80
70
60
50
40
30
20
10
0
30
94
2
5
1
2
2 või 3
4

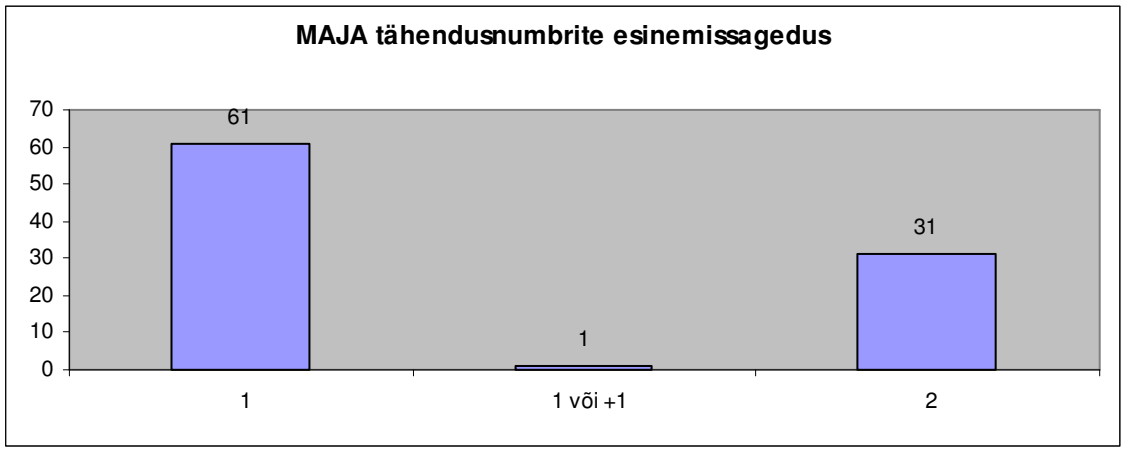

PÄEV tähendusnumbrite esinemissagedus
40
35
30
25
20
15
10
5
0
36
3
37
1
24
22
2
2
3
4
4 või 6 või 5
5
6
8

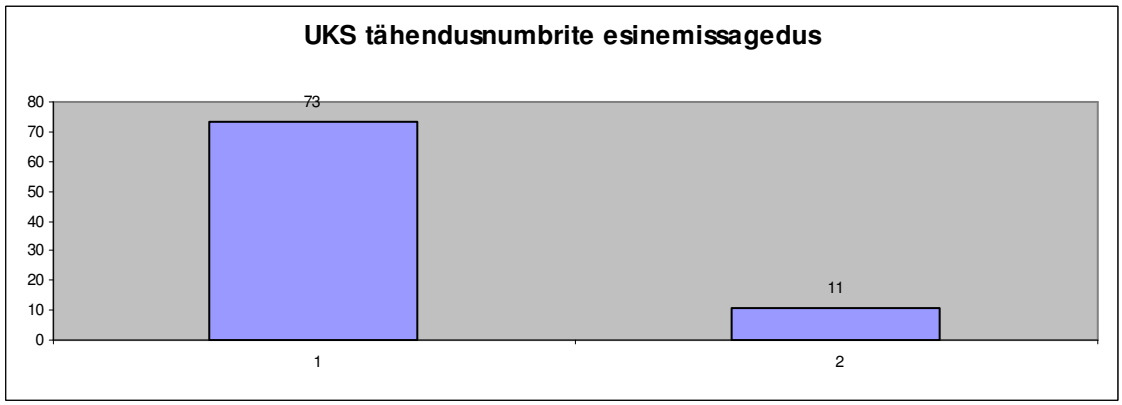

PEA tähendusnumbrite esinemissagedus
80
70
60
50
40
30
20
10
0
3
70
32
1
1
1
1
+1
1
2
2 või 5
3
4
5

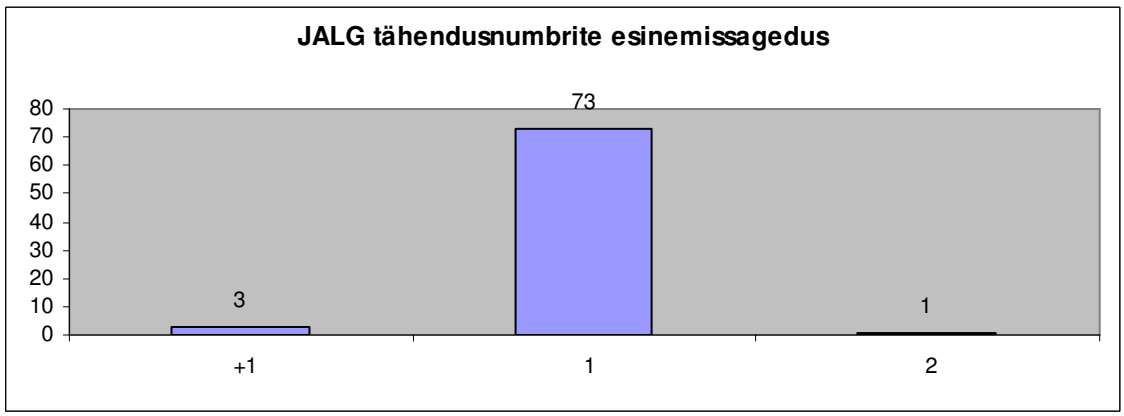

KORD tähendusnumbrite esinemissagedus
100
90
80
70
60
50
40
30
20
10
0
4
2
3
92
1
2
4
+1
1
2
3
6
7
8

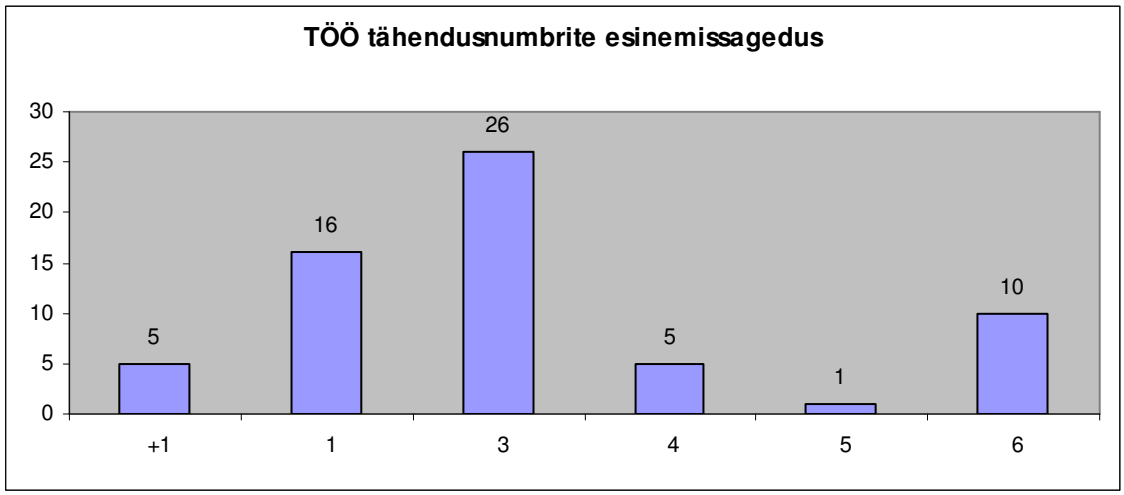

MÕTE tähendusnumbrite esinemissagedus
60
50
40
30
20
10
0
4
55
8
7
10
16
+1
1
2
3
4
5

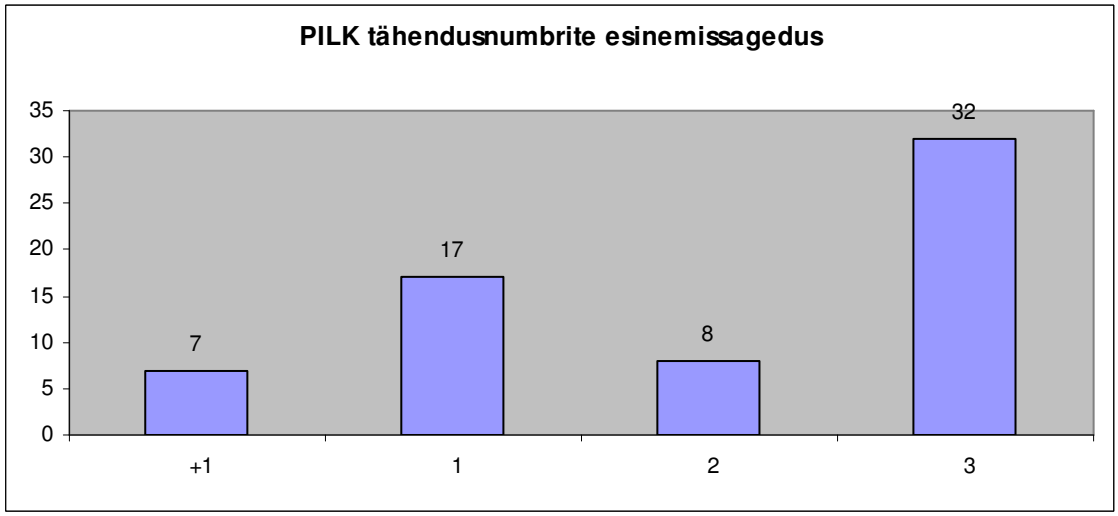

JUTT tähendusnumbrite esinemissagedus
40
35
30
25
20
15
10
5
0
7
22
17
36
+1
1
2
3

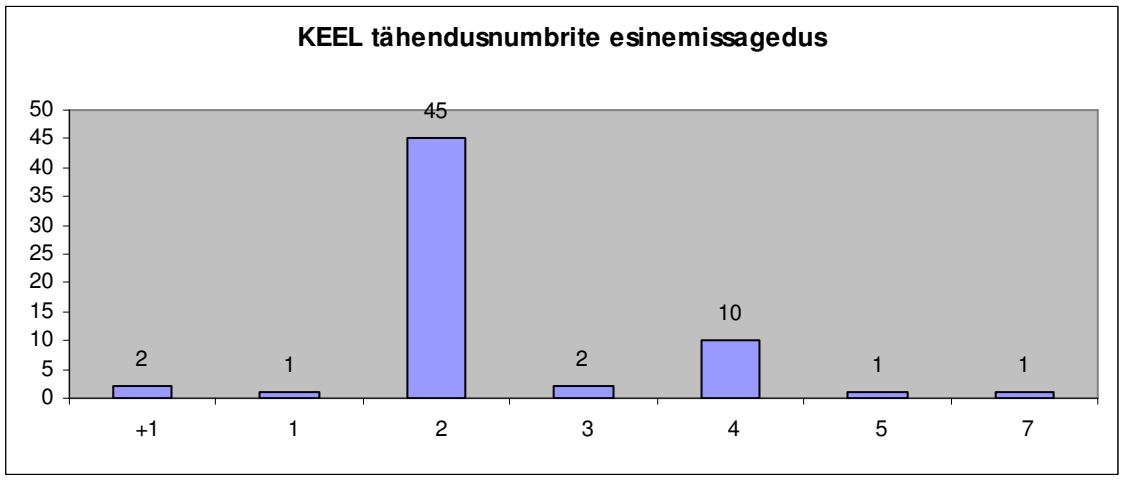

MAJA tähendusnumbrite esinemissagedus
70
60
50
40
30
20
10
0
61
1
31
1
1 või +1
2

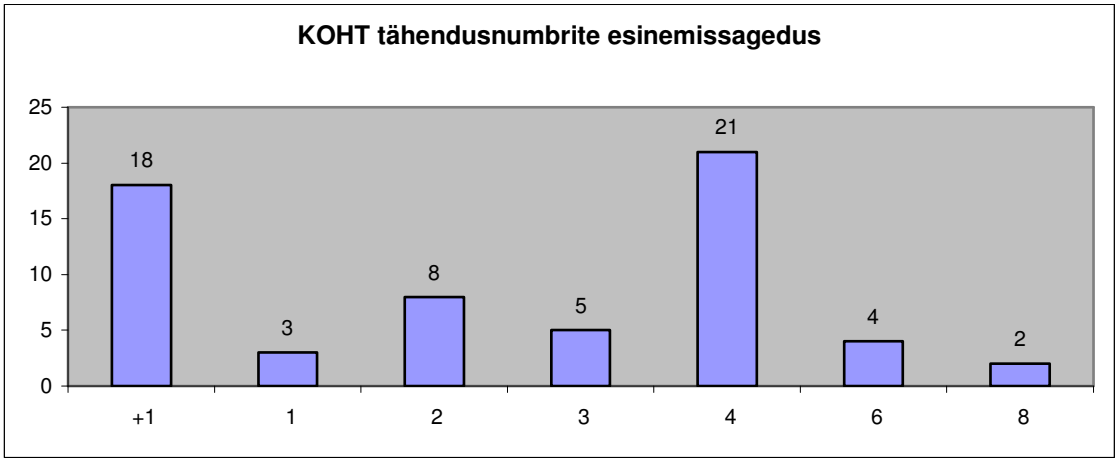

UKS tähendusnumbrite esinemissagedus
80
70
60
50
40
30
20
10
0
73
11
1
2

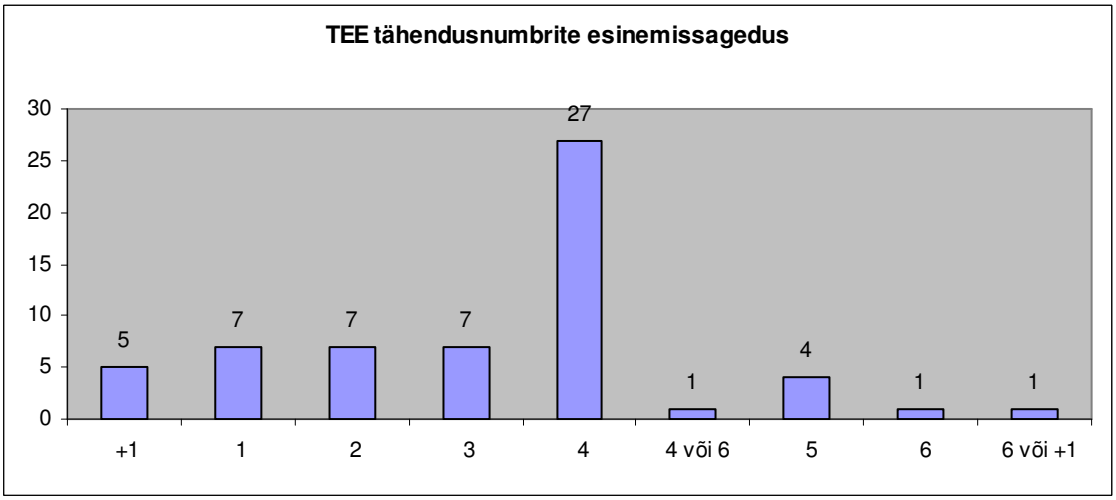

JALG tähendusnumbrite esinemissagedus
80
70
60
50
40
30
20
10
0
3
73
1
+1
1
2

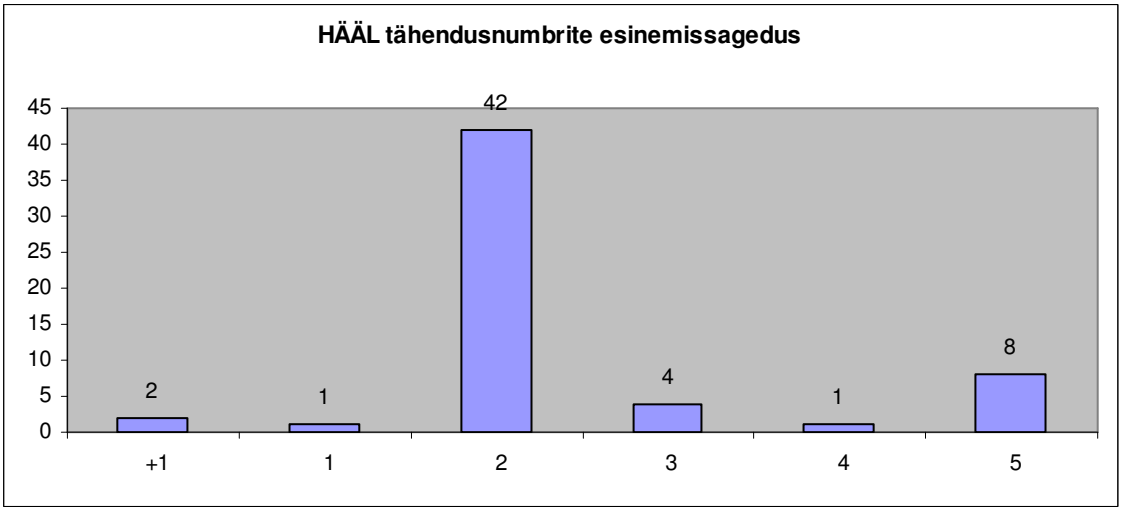

TÖÖ tähendusnumbrite esinemissagedus
30
25
20
15
10
5
0
5
16
26
5
1
10
+1
1
3
4
5
6

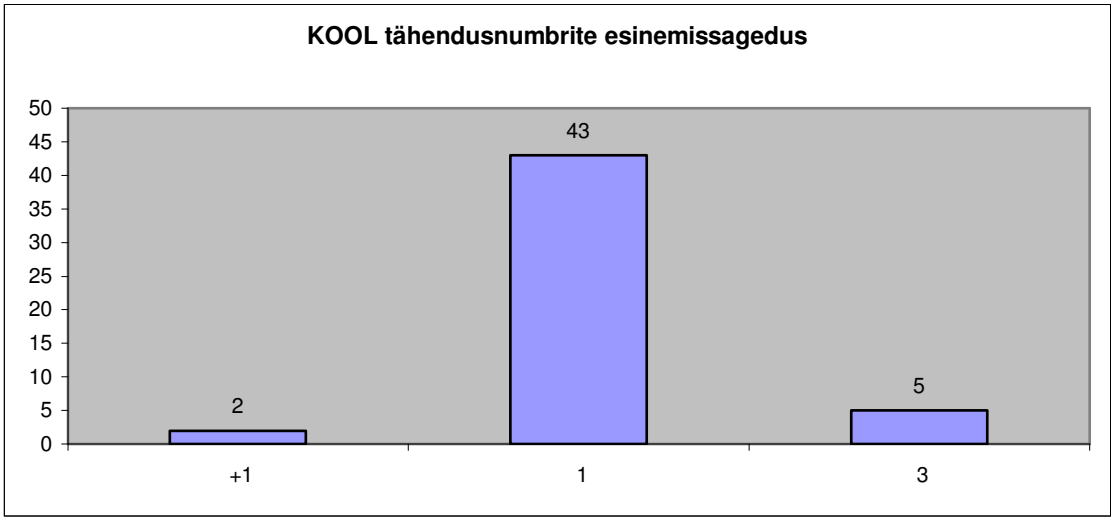

PILK tähendusnumbrite esinemissagedus
35
30
25
20
15
10
5
0
7
17
8
32
+1
1
2
3

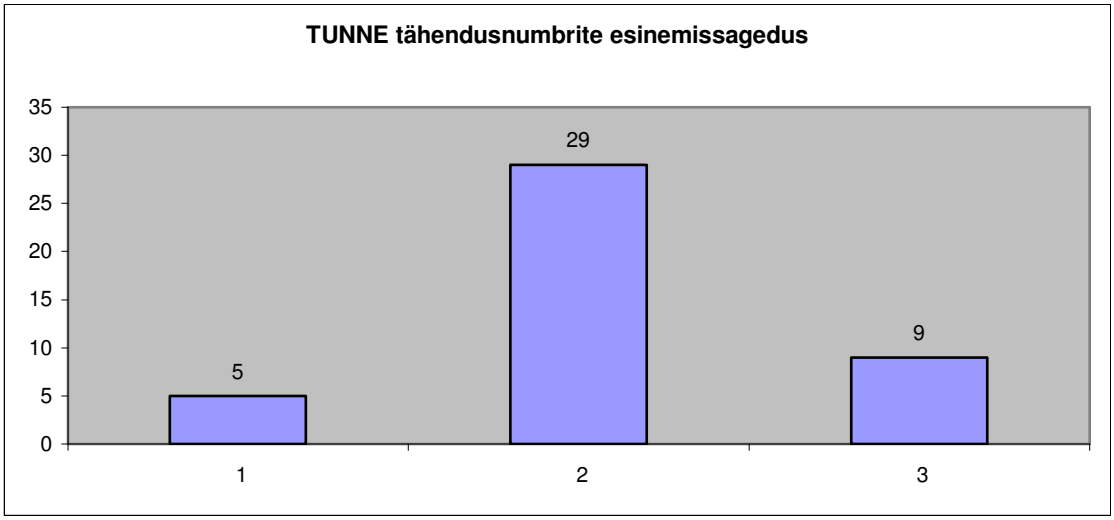

KEEL tähendusnumbrite esinemissagedus
50
45
40
35
30
25
20
15
10
5
0
2
1
45
2
10
1
1
+1
1
2
3
4
5
7

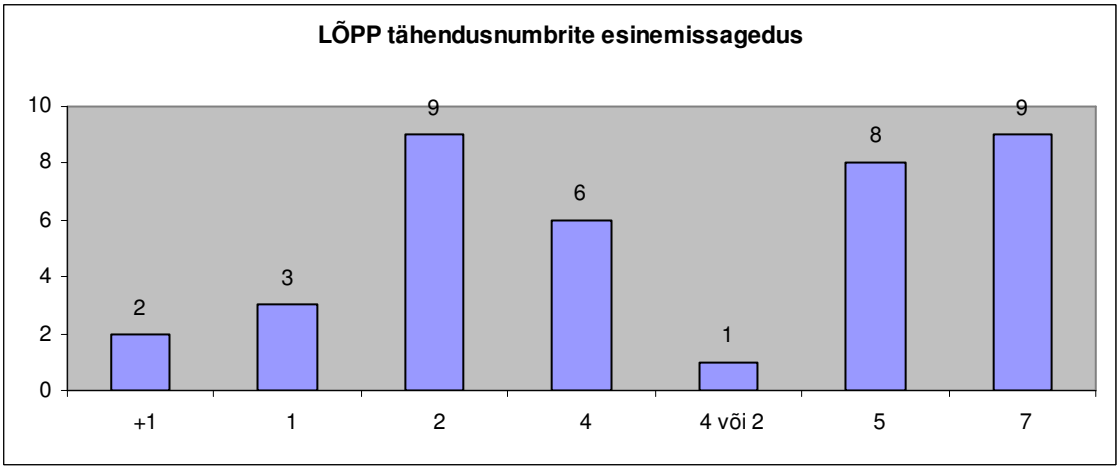

KOHT tähendusnumbrite esinemissagedus
25
20
15
10
5
0
18
3
8
5
21
4
2
+1
1
2
3
4
6
8


TEE tähendusnumbrite esinemissagedus
30
25
20
15
10
5
0
5
7
7
7
27
1
4
1
1
+1
1
2
3
4
4 või 6
5
6
6 või +1


HÄÄL tähendusnumbrite esinemissagedus
45
40
35
30
25
20
15
10
5
0
2
1
42
4
1
8
+1
1
2
3
4
5


KOOL tähendusnumbrite esinemissagedus
50
45
40
35
30
25
20
15
10
5
0
2
43
5
+1
1
3


TUNNE tähendusnumbrite esinemissagedus
35
30
25
20
15
10
5
0
5
29
9
1
2
3


LÕPP tähendusnumbrite esinemissagedus
10
8
6
4
2
0
2
3
9
6
1
8
9
+1
1
2
4
4 või 2
5
7

Lisa 2.

Käsitsimägrendja juhend.

Sõnatähenduste ühestamiseks on vaja kolme asja:

I **Tesaurust** aadressil http://gling.psych.ut.ee:8080/cgi-bin/tesaurusetegija11.py

Selles näed kõigepealt päringuvälja, kuhu võid sisestada sõnu ja sõnaühendeid sõnastikuvormis, näiteks: elu, saama, läbi elama, käsi, ja seejärel vajuta hiirega 'Otsi' või klaviatuuril Enter-klahvi.
Ilmub otsitava sõna kõiki tähendusi esitav tabel, otsisõna ja tema vastav(ad) tähendusnumber(-numbrid) on koos sama sünohulga muude liikmetega II veerus **paksus kirjas**.

II **Tekstifaili**, mis on morfoloogiliselt juba ühestatud ja semantilise ühestaja filtrist läbi lastud. See näeb välja midagi niisugust:

```
<s>
Mitmel
    mitu+l //_P_ indef sg ad //
korral
    kord+l //_S_ com sg ad // @ kord:6523:8
nägime
    näge+ime //_V_ main indic impf ps1 pl ps af // @ nägema:635:7
oma
    oma+0 //_P_ pos sg gen //
silmaga
    silm+ga //_S_ com sg kom // @ silm:1220:1
puude
    puu+de //_S_ com pl gen // @ puu:496:3
langemist
    langemine+t //_S_ com sg part // @ langemine:8153#9761:2
.
    . //_Z_ Fst //
```

Õiget tähendusnumbrit vajavatel ehk semantiliselt ühestatavatel sõnadel on programmi *semyhe* poolt tehtud eelanalüüs, mille tulemused on peale @ märki. Read, millel @ märki pole, seega ühestamise alla ei kuulu.
Käsitsiühestaja ülesandeks on panna `//` ja `@` märkide vahele selle sõna õige tähenduse number, mis saadakse tesaurusest. Tähendusnumber pannakse omakorda tühikute vahele:

```
<s>
Kui
    kui+0 //_J_ sub //
nii
    nii+0 //_D_ //
võtta
    võt+a //_V_ main inf // 6 @ võtma:7021:8
,
```

```
    , //_Z_ Com //
olin
    ole+in //_V_ aux indic impf ps1 sg ps af //
kogu
    kogu+0 //_A_ pos //
orkaani
    orkaan+0 //_S_ com sg gen // 1 @ orkaan:2640:1
kaasa
    kaasa+0 //_D_ // kaasa_tegema 0
teinud
    tege+nud //_V_ main partic past ps // kaasa_tegema 0 @ tegema:41:6
ja
    ja+0 //_J_ crd //
```

Mõnel juhul on *semyhe* vajaliku koha peale juba 0 pannud, mis tähendab, et seda sõna tesaurusest otsida ei tasu.
Tähendusega sõnaühendeid (nagu ühend- ja väljendverbid, fraseologismid ja idiomaatilised ühendid) *semyhe* veel eristada ei suuda. Sel juhul tuleb sõnaühend alakriipsu abil ise juurde märkida mõlemale komponendile (nagu ülemises näites) ja ise ka järgi vaadata, kas sõnaühendil on tesauruses mõni tähendus juba olemas.

NB! Üsna mõttekas on tegeleda ühe sõna ja tema tähendustega tekstis läbivalt, siis ei ole vaja sama sõna järgmise esinemise juures hakata uuesti uurima ja mõtlema, kuidas täpselt tähendused jaotuvad ja mis nende omavahelised erinevused on.

NB! Mõned head reeglid, mida võiks ühtluse mõttes kasutada:
hakkama, millele järgneb ma-infinitiiv (hakkama midagi tegema) = 2
saama 8 = eesti keeles puuduvat tuleviku kategooriat väljendav konstruktsioon
eksistentsiaallauses olema = 4 (või selle hüponüümid)
possessiivlauseis olema = 9
predikatiivlauseis olema = 8
jääma4, muutuma3, saama3, minema8 + Adj-ks või N-ks

Kui sinu meelest ei sobi ühestatavale sõnale ükski tesauruses esitatud tähendustest, kirjutad tekstifaili tähenduse numbriks +1. Ja kirjutad

III **kommentaarifaili** selgituse, milline tähendus on puudu, näiteks:

PUUDU
'taevas' laotuse tähenduses, olemas ainult usklike hingede taevas.
'üle_minema' tähenduses 'mööduma'

Kommentaare võid kirjutada ka tesauruses olevate sõnade ja tähenduste kohta, näiteks:

'võtma' väga metafoorne
võtma5 võiks seletusest rõhuasetuse vedelikule välja jätta
'mahtuma1' on täpselt sama mis 'mahtuma2' !?
'talu2' on minumeelest midagi käsitluses valesti... aga kohe ei oska öelda, mis...
'põdema' tähenduste 1, 2, 3 erinevus?