

Eesti Rakenduslingvistika Ühing

**Eesti Rakenduslingvistika Ühingu
aastaraamat 8**

Estonian Papers in Applied Linguistics 8

Toimetanud
Helle Metslang, Margit Langemets
ja Maria-Maren Sepper

Tallinn 2012

EESTI RAKENDUSLINGVISTIKA ÜHINGU AASTARAAMAT 8 **Estonian Papers in Applied Linguistics 8**

Toimetajad / Editors: Helle Metslang (Tartu), Margit Langemets (Tallinn),
Maria-Maren Sepper (Tallinn)

Kujundaja / Design: Piia Ruber

Toimetuskolleegium / Advisory board:

Martin Ehala (Tallinn), Auli Hakulinen (Helsinki), Birute Klaas (Tartu), Eino Koponen (Helsinki), Irina Külmoja (Tartu), Maisa Martin (Jyväskylä), Jaan Mikk (Tartu), Hille Pajupuu (Tallinn), János Puzstay (Szombathely), Helena Sulkala (Oulu), Urmas Sutrop (Tallinn/Tartu), Eva Toulouze (Tartu), Tiia Tulviste (Tartu), Marilyn Vihman (Bangor, U.K.), Leo Võhandu (Tallinn), Haldur Õim (Tartu)

“Eesti Rakenduslingvistika Ühingu aastaraamat” on registreeritud ja refereeritud järgmistes bibliograafilistes andmebaasides / Estonian Papers in Applied Linguistics is covered by the following abstracting/indexing services:

Arts and Humanities Citation Index (Thomson Reuters)
Central & Eastern European Academic Source (EBSCO)
Central and Eastern European Online Library (CEEOL)
CSA Linguistics and Language Behavior Abstracts (ProQuest)
Directory of Open Access Journals (Lund University)
Eesti Rahvusraamatukogu digitaalne arhiiv (DIGAR)
European Reference Index for the Humanities (European Science Foundation)
Journal Citation Reports / Social Sciences Edition (Thomson Reuters)
Linguistic Abstracts (Wiley-Blackwell Publishing)
Linguistic Bibliography Online (Brill Publishers)
MLA Directory of Periodicals (Modern Language Association)
MLA International Bibliography (Modern Language Association)
SCOPUS (Elsevier)
Social Sciences Citation Index (Thomson Reuters)
Social Scisearch (Thomson Reuters)
Ulrich's Periodicals Directory

Ühingu aadress / Contact information:

Eesti Rakenduslingvistika Ühing / Estonian Association for Applied Linguistics
Eesti Keele Instituut / Institute of the Estonian Language
Roosikrantsi 6
10119 Tallinn
Estonia

ery@eki.ee

www.rakenduslingvistika.ee

“Eesti Rakenduslingvistika Ühingu aastaraamat 8” väljaandmist on toetanud Haridus- ja Teadusministeeriumi riiklik programm “Eesti keel ja kultuurimälu” ja Eesti Keele Instituut.

Autoriõigus / Copyright: autorid ja Eesti Rakenduslingvistika Ühing

ISSN 1736-2563

doi:10.5128/ERYa.1736-2563

SISUKORD

Kausatiivsuse omandamisest eesti keeles <i>Acquisition of causatives in Estonian</i> Reili Argus	5
Lithuanian narrative language at preschool age <i>Leedukeelne narratiiv koolieeliku-eas</i> Ingrida Balčiūnienė	21
Enda algatatud eneseparandus eestikeelsetes MSN-i dialoogides <i>Self-initiated self-repair in Estonian Instant Messaging interaction</i> Tiit Hennoste	37 54
Osalausete tuvastamine eestikeelses tekstis kui iseseisev ülesanne <i>Clause splitting as a separate task (in the analysis of Estonian texts)</i> Heiki-Jaan Kaalep, Kadri Muischnek	55
Lithuanian and Russian child-directed speech: Why do we ask young children so many questions? <i>Leedu- ja venekeelne lapsele suunatud kõne: miks me esitame väikelastele nii palju küsimusi?</i> Victoria V. Kazakovskaya, Ingrida Balčiūnienė	69
Referentsiaalsed noomenifraasid laste narratiivides <i>Referential Noun Phrases in Estonian children's narratives</i> Helen Kirsipuu, Piret Soodla, Renate Pajusalu	91
Konversatsiooniagenti modelleerimine argumenteerimisdialoogis: suhtlus kui infoseisundite värskendamine <i>Modelling of conversational agents in argumentation: Conversation as updating of information states</i> Mare Koit	109
Argumenteerimismetodid riigieksamikirjandites <i>Argumentation methods in the Estonian high school matriculation examination essays</i> Kersti Lepajõe	123

Eesti Wordnet'i struktuuri analüüsisist <i>Structural analysis of the current Estonian Wordnet</i> Ahti Lohk, Leo Vöhandu	139
Automaatne ajaväljendite tuvastamine eestikeelsetes tekstides <i>Automatic recognition and normalization of temporal expressions in Estonian language texts</i> Siim Orasmaa	153
Lexicon-based detection of emotion in different types of texts: Preliminary remarks <i>Eri tüüpi tekstide emotsionaalsuse tuvastamine leksikonipõhisel meetodil: esmased katsetused</i> Hille Pajupuu, Krista Kerge, Rene Altrov	171
Study diaries in translation education at Tallinn University <i>Tõlkepäevikud tõlkijakoolituses</i> Triin Pappel, Tiia Falk	185
Ajasõnad varasel perioodil eesti lapsekeeles <i>Early development of time adverbs in Estonian child language</i> Sirli Parm	195
Direktiivisekventsids isa ja tütre suhtluses: juhtumianalüüs <i>Directive sequences in a conversation of father and daughter: A case study</i> Andriela Rääbis	213
Kas formandid peegeldavad emotsioone? <i>Do formants speak of emotions?</i> Kairi Tamuri	231
Algupäraste ja tõlgitud eestikeelsete tekstide eristatavusest ning neile omistatud omadustest <i>On the identifiability and characteristics of translations and non-translations in Estonian</i> Marju Taukar	245
Kuidas määrata eesti keele sõnavara tundetoone? <i>Detecting emotional valencies for the Estonian vocabulary</i> Ene Vainik	257
Focus on the form versus focus on the message in <i>lingua franca</i> conversation <i>Sõnumile või keelelisele vormile keskendumine võõrkeelses vestluses</i> Natalja Zagura	275

KAUSATIIVSUSE OMANDAMISEST EESTI KEELES

Reili Argus

Ülevaade. Artikli eesmärk on anda ülevaade kausatiivsuse omandamisest eesti keeles. Uurimuse keskmes on küsimus, millal ja milliseid kausatiivsuse väljendusvahendeid lapsed varase keeleomandamise perioodil kasutavad, kas eelistused on seotud kasutatavate kausatiivsete struktuuride läbipaistvusega (analüütilised kausatiivid on läbipaistvamad kui sünteetilised ehk morfoloogilised), produktiivsusega või pigem sisendkeele sagedustega. Eri tüüpi kausatiivsete struktuuride omandamise juures on vaadeldud ka seda, kuidas on võimalik eesti keele kausatiivsuse omandamisel eristada omandamisetappe või õpimudeleid. Kasutatud keeleandmeteks on pikiuuringuga kogutud kahe eesti keelt omandava lapse spontaanse kõne lindistused vanusest 1;8–3;0.*

Võttesõnad: lapsekeel, hoidjakeel, keeleomandamine, kausatiivsus, sõnamoodustus, eesti keel

1. Kausatiivsus leksikaalgrammatilise kategooriana

Kausaalsus ehk põhjuslikkus on nähtus, kus üks sündmus kutsub esile teise sündmuse. Põhjustatud sündmuse puhul võib olla tegemist nii seisundi, tegevuse, soorituse kui ka saavutusega (Van Valin, Lapolla 1997: 93). Kui mittekausatiivses lauses võib olla ka ainult üks argument, nt *Tüdruk sööb*, siis kausatiivsus lisab lausesse veel uue argumendi, mis (või kes) on selle tegevuse või situatsiooni põhjustaja, nt *Poiss sööda tüdrukut*, kusjuures mittekausatiivse situatsiooni agendist saab selle muutusega patsient (eelmises näites *tüdruk*). Tavaliselt seostataksegi kausatiivsusega kaht olulist mõistet – “agentiivsus” ja “(seisundi)muutus”. Prototüüpne kausatiiv näitab, et agent teeb midagi, ning teise kriteeriumina, et eksisteerib ka patsiendi rollis osaline, kellega toimub muutus (Lakoff 1987: 54–55). Jackendoffi (1983) järgi sisaldab kausaalne suhe tegelikult kaht protsessi – üks protsess põhjustab teise ehk tulemprotsessi. Seda, mis juhtub, ehk tulemprotsessi, ei põhjusta mitte alati agendi rollis olev isik otse, vaid tema tegevus või tegevusetus. Kasiku järgi

* Artikli valmimisel on toeks olnud Eesti Teadusfondi grant nr 7492 “3–7-aastaste laste keelelise arengu leksikaalsed ja grammatilised seaduspärad”.

on tüüpilises kausaalses lauses tegemist tahtlikult tegutseva inimese põhjustatud tagajärgedega (Kasik 2001: 77–79). Mitteprototüüpse kausatiivsuse puhul võib agendi asemel olla lauses hoopis selline osaline, kes ei põhjusta sündmust teadlikult ja aktiivselt. Näiteks lauses **Kuum vesi sulatas jää** ei põhjustanud kuum vesi jää sulamist teadlikult ja tahtlikult ning jääd ei saa siin käsitleda agendi, vaid pigem nn mõjutajana¹. Samuti võib tulemsündmuse põhjustajaks ehk siis mõjutajaks olla protsess: **Kulla hinna langemine maailmaturul langetas aktsiate hindu**.

Tuleb aga ka meeles pidada, et kausatiivse propositsiooni ja füüsilise reaalsuse vahel ei ole otsest vastavust, pigem peegeldab see viisi, kuidas me kausatiivset situatsiooni kontseptualiseerime (Paulsen 2011: 81). Ühelt poolt on väidetud, et tegeliku elu kausaalsust mõistetakse varem, kui seda väljendatakse ehk mingi kindla keele abil kontseptualiseerima õpitakse (Tomasello 1999), teisalt sõltub see, kuidas me kausaalsust väljendame, paljuski keelest, mida kõneleme (Clark 2003).

Kausatiivse predikaadi ja selle mittekausatiivse vaste vormilise suhte alusel (vt Comrie 1993: 167) on kausatiivsed struktuurid² jagatud sageli kolmeks:

- 1) analüütilised (ehk süntaktilised), nt *sööma panema*;
- 2) sünteetilised (ehk morfoloogilised), nt *söötma*;
- 3) leksikaalsed, nt *avama*.

Eesti keele puhul võib aga leksikaalsete kausatiivide hulka paigutada peale lihtverbide ka ühendverbid, nagu näiteks *lahti tegema* ja *kinni panema*. Seega saab eesti keeles analüütilised kausatiivsed struktuurid jaotada põhimõtteliselt samuti kaheks: olemuselt leksikaalsed ehk kinnistunud ühendid (ühendverbid) ning produktiivselt ja reeglipäraselt moodustatavad (tugiverb + *ma-* või *da-*infinitiivis leksikaalverb) ahelverbid, nagu *käima panema*, *nutta laskma*.

Leksikaalsete ehk sisemiselt kausatiivsete verbide puhul on sõnamoodustuse seisukohalt tegemist ebaproduktiivse verbirühmaga. Leksikaalsed kausatiivid on transitiivsed verbid, mille tähenduses on samuti nagu morfoloogiliste kausatiivide puhul tegelikult kaks sündmust – agendi põhjustatud tegevus ning patsiendi seisundimuutus³. Selline verb tähistab olukorda, kus tegija läbi viidud tegevus ehk algsündmus põhjustab patsiendi tegutsemise või muutumise ehk tulemsündmuse või seisundimuutuse, nt *poiss lõhub vaasi* > *vaas on katki*.

Kausaalsus saab olla eesti keeles eksplitsiitselt märgistatud, tuletiste puhul afiksiga *-ta* (või afiksi eri variantidega *-t*, *-sta*, *-nda*) ja analüütiliste kausatiivide puhul näiteks leksikaalverbi *ma-*infinitiiviga, või märgistamata (nt sisemiselt kausatiivse tähendusega verbid *avama*, *sulgema* ja ühendverbid, nagu *lahti tegema*). Prototüüpse kausatiivi ja tema tuletusaluse suhe on eesti keeles Kasiku (2001: 82) järgi küll morfoloogiline – kausatiivne vorm moodustatakse mittekausatiivsest verbist või noomenist afiksi teel, kuid kausatiivide tuletamine ei ole eesti keeles siiski täiesti produktiivne, mõnda tähendust saab väljendada ainult morfoloogilise kausatiiviga (*kasvatab last*) ja teist jälle ainult analüütilise kausatiiviga (*laseb laguneda*).

Eelpool esitatud Bernard Comrie kausatiivide kolmikjaotus on seega eesti keele kausatiivsuse väljendusvahendeid silmas pidades pisut tinglik – puhtalt vormi alusel võib tõesti kausatiivsed struktuurid jagada analüütilisteks ja sünteetilisteks, kuid kui analüütilised jagunevad moodustusmehhanismilt ühendverbideks ja tugiverbi

¹ Van Valin ja Lapolla kasutavad tahtmatult ja eesmärgitult tegutseva osalise tähistamiseks terminit *mõjutaja* (ingl *effector*).

² Termin “kausatiivne struktuur” on siinses artiklis kasutusel katmaks nii morfoloogilisi (tuletatud) kausatiive kui ka leksikaalseid kausatiivsuse väljendusvahendeid (analüütilisi konstruktsioone ning kausatiivse tähendusega lihtverbe).

³ Sageli on leksikaalsed kausatiivid keeles kasutusel otse põhjustamise (nt *Jaan lõhkus akna*) ning morfoloogilised ja süntaktilised pigem kaudse põhjustamise (nt *Jaan ajas Malle naerma*) tähistamiseks (vt nt Jackendoff 1990, Matsumoto 1996).

ühendeiks, siis sünteetiliste alla võiks paigutada samuti kaks tüüpi: nii sõnamoostuse abil saadud tuletatud verbid kui ka sisemiselt kausatiivsed lihtverbid. Nii võibki eesti kausatiivsete struktuuride puhul eristada tegelikult kaht eri jaotust – vormil põhinev jaotus analüütilisteks ja sünteetilisteks ning produktiivsusel põhinev jaotus produktiivseteks ehk grammatikareeglite järgi moodustatavateks ja ebaproductiivseteks ehk leksikaalseteks kausatiivideks. Seega on kausatiivsus eesti keeles olemuselt kompleksne leksikaalgrammatiline kategooria, mille väljendusvahendid kuuluvad ühtaegu nii sõnavarasse kui ka grammatikasse.

2. Kausatiivsuse omandamisest

Esimese keele omandamine on protsess, kus nii grammatilised, semantilised kui ka pragmaatilised tegurid on omavahel tihedalt põimunud. Kui noomenite muutemorfoloogia omandamine sõltub peamiselt morfofonoloogilistest teguritest, keelestruktuuri produktiivsusest ning kindlate struktuuride sagedusest lapsele suunatud kõnes (Argus 2008a), siis verbimorfoloogia arengut paistavad sageduste kõrval mõjutavat enam leksikaalsemantilised tegurid (Argus 2008a, 2008b).

Verbide leksikaalgrammatilistest kategooriatest on kausatiivsust peetud eituse kõrval üheks olulisemaks ja sagedasemaks (vt nt Lakoff 1987: 55, Podlesskaya 1993: 165). Kuigi kausatiive on käsitletud laialdaselt näiteks tüpoloogilisest perspektiivist (Comrie 1993), ei ole kausatiivsuse omandamine väga põhjalikult uuritud isegi mitte indo-euroopa keeltes, kus verbi omandamisega seonduv on lingvistide tähelepanu keskmes olnud juba aastakümneid. Üksikasjalikumalt on kirjeldatud jaapani (Okabe 2007), kešua (Courtney 2002) ja inuktituti keele (Allen 1998) kausatiivide omandamist. Eesti keele kohta leidub küll mitu verbi omandamisele keskenduvat kirjutist (vt Vihman, Vija 2006, Argus 2006, Argus, Parm 2010), aga ka mõne grammatilise kategooria, nagu konditsionaali (Pajusalu jt 2011) või tegumoe (Vija jt 2009) omandamise käsitlusi, kuid eesti verbide leksikaalgrammatiliste kategooriate omandamise kohta on esialgu veel väheseid andmeid: kirjeldatud on eesti keele aspektilisuse (Argus 2006) ning *tegema*-verbi (Argus 2010) seotusest verbide leksikaalsemantikaga.

Kuigi kausatiivsus on keeltes oluline ja üsna sage leksikaalsemantiline kategooria, võib selle omandamine olla lapse jaoks ometi keerukas ning eeldada enam kognitiivseid oskusi kui mõni teine kategooria. Kausatiivse verbi puhul on ühes verbis sisuliselt kahe verbi tähendus: algse ehk põhjustava tegevuse kõrval info ka lõppseisundi ehk tulemtegevuse kohta, samuti on kausatiivsusel eri tüüpi väljendusvahendeid, nii leksikaalseid kui ka grammatilisi, nii produktiivseid kui ka ebaproductiivseid, ning ka see võib omandamist mõjutada.

Järgneva eesti keele kausatiivsuse omandamise uuringu taustaks võib seada kolm laiemat teoreetilist lähtekohta: Dan Slobini analüütilise vormi eelistamise operatsiooniprintsiip, funktsionalistlikust lähenemisest tuttava vormi ja tähenduse vastendus (ingl *mapping*) ning kasutuspõhistes lähenemistes kasutatud mudeli-põhise omandamise idee.

Slobini üldkognitiivse **analüütilise vormi printsiibi** järgi eelistavad lapsed siis, kui keerukat tähendussisu saab väljendada nii sünteetilise kui ka analüütilise konstruktsiooniga, viimast kui läbipaistvamat (Slobin 1985: 1166–1229). Eesti keele tugiverbi *tegema*-ühendite omandamisandmete põhjal (Argus 2010)

võib oletada, et sarnaselt analüütiliste transitiivsete verbidega omandatakse ka kausatiivide analüütilised vasted varem kui mitte nii läbipaistvad sünteetilised tuletatud ja leksikaalsed ehk sisemiselt kausatiivsed verbid. Lisaks läbipaistvusele võib üheks mõjutajaks olla ka lapsele suunatud kõne kui suuline ja vabam register. Virve Raag (1998: 132) väidab, et kui keeles on rööpvormid, siis on loomulik, et need erinevad stiililiselt. Helle Metslang (1994: 207) märgib samuti, et kui keeles eksisteerib nii sünteetiline kui ka analüütiline väljendusviis, eelistatakse kirjakeeles sünteetilisi, suulises ja mitteametlikus keeles aga analüütilisi vorme. Seega võib oletada, et lapsele suunatud kõnes on enam analüütilisi kui sünteetilisi kausatiive ja nii mõjutab omandamist peale läbipaistvuse ka sagedus.

Samas on aga Eve Clark väitnud, et see, mis vanuses lapsed kausatiivid omandavad, sõltub väga palju omandatavast keelest: näiteks keeles, kus kausatiivsuse märkimisel on oluline roll muutemorfoloogial (lause patsient on mingis kindlas käändevormis), omandatakse kausatiivsus varem kui sõnajärjega opereerivas keeles. Näiteks türgi keeles, kus kausatiivsust märgitakse verbiliitiga, omandatakse see varem kui keeltes, kus kausatiivsust märgitakse analüütiliselt, sest viimasel juhul on vaja lapsel haarata korraga kogu lause ja see on ilmselt raskem kui ühe vormiga toimetulek. (Clark 2003: 238) Nii on inglise keele kausatiivide omandamise kohta andmeid, et laps võib küll omandamist alustada juba vanuses 1;11, kuid teeb vigu veel väga pikka aega (Bowermann 1974: 168).

Funktsionalistlike lähenemiste keskmes on seisukoht, et omandamine on vormide ja funktsioonide (pragmaatiliste eesmärkide ja tähenduste) **vastendus** (Bates, MacWhinney 1982). Sama seisukohta toetavad ka konstruktsioonigrammatikud (Kay, Fillmore 1999), kes näevadki keelt kui vormide-funktsioonide vastenduste kogumit. Eesti keele kausatiivsuse väljendamisel ei saa aga laps kasutada funktsiooni ja vormi üks-ühest vastendust: ühelt poolt on kausatiivsust võimalik eesti keeles väljendada eri vahenditega, nii sünteetiliselt kui ka analüütiliselt, ning teisalt on kausatiivsel (*s*)*ta*-liitel mitu tähendust, lisaks kausatiivsusele ka üldisemat laadi faktitiivitähendus⁴. Veel on mõned tuletatud kausatiivid leksikaliseerunud ning omandanud teisi tähendusi peale kausatiivsuse ning neid ei saagi kasutada kausatiivse ja mittekausatiivse tegevuse vastendamisel, nt *istuma – istutama*. Suur osa *ta*-liitelistest kausatiivverbidest on labiilsed, s.t võivad esineda nii kausatiivses kui ka mittekausatiivses situatsioonis (nt *ehmatama: Kertu ehmatas – Kertu ehmatas poissi*) (Kehayov, Vihman 2010). Seetõttu võib oletada, et kausatiivsuse sünteetiliste väljendusvahendite omandamine ei ole eesti lapse jaoks kerge ülesanne ja ta omandab need suhteliselt hilja võrreldes selliste verbituletusliidetega, millel on vasteks ühtsem tähendusrühm (nt frekventatiivid või refleksiivid).

Kasutuspõhiste omandamisuringute puhul on vahel seletatud omandamisprotsessi kui toetumist mõnele kindlale **õpimudelile**. Keelesüsteemi mõne kindla osa kõige sagedamini kasutatavad üksuseid on peetud justkui oma valdkonna õpimudeliteks (Ninio 1999: 3), mille abil toimub kogu ülejäänud sarnaste keeleelementide omandamine. Lapsed omandavad esimesena üksuse, mis esineb lapsele suunatud kõnes väga sageli. Mida sagedasemad on mingid keeleüksused lapsele suunatud kõnes, seda vähem auditiivset või ka visuaalset stiimulit on vaja, et see üksus ära tunda ja seda tervikuna tajuda (Lieven 2010: 2546). Sellised esimesena omandatud ja lapsele suunatud kõnes sagedased üksused on oma süsteemiosa n-ö

parimad esindajad – kõige üldisemad, kõige prototüüpsemad ja ka kõige lihtsamad. Kui eesti verbide omandamisel paistab mudeli rollis olema *tegema*-verb (vt Argus 2010), siis võib oletada, et sarnane mudelverb võib aidata omandada ka verbide kitsamaid leksikaalgrammatilisi kategooriaid, nagu kausatiivsus.

Enamasti on leitud, et kausatiivid omandatakse **etapiti**. Kuigi eri keelte puhul on esitatud eri arv etappe, paistab omandamine kulgevat põhijoontes sarnaselt. Esialgu omandavad lapsed väikese hulga leksikaalseid kausatiive (Allen 1998, Courtney 2002). Selline tervikomandamine on üldiseltki iseloomulik varasele keeleomandamisstaadiumile, mil esialgu domineerib mälu töö. Tervikuna omandamise periood võib kausatiivide puhul kesta küllaltki kaua, sest näiteks hakkavad heebrea keelt omandavad lapsed morfoloogilisi kausatiive produktiivselt kasutama alles vanusest 3;0 (Berman 1994). Seejärel hakkavad lapsed kasutama kausatiivseid morfeeme ning seda ainult väikese rühma väga tuttavate verbidega kindlas pragmaatilises kontekstis, nt mõnes keeles esialgu ainult imperatiivsetes lausungites (Allen 1998: 633). Väikese rühma tuttavate verbide põhjal tekib lapsel osaliselt produktiivne reegel, mida ta võib laiendada ka sellele reeglile mitte alluvatele verbidele ja tekitada üleüldistusi. Alles pärast seda etappi hakkab laps morfoloogilisi kausatiive kasutama erinevates pragmaatilistes kontekstides ja täiskasvanupäraste reeglite järgi. (Allen 1998: 633)

Eelkirjeldatu taustal on siinses eesti kausatiivsete struktuuride omandamise uurimuses otsitud vastuseid küsimustele:

- 1) millal hakkavad eesti lapsed kasutama kausatiivseid struktuure;
- 2) millised kausatiivsuse väljendusvahendid omandatakse esimesena: mis vanuses hakkavad lapsed sünteetilisi ja analüütilisi kausatiive kasutama;
- 3) missugused tegurid mängivad kausatiivsuse omandamisel kõige suuremat rolli, kas sisendkeele sagedus, struktuuri läbipaistvus või moodustusmalli produktiivsus;
- 4) kas analüütiliste kausatiivide puhul võib märgata, et lapsed kasutavad mingit leksikaalset malli kui kausatiivsuse väljendamise põhivahendit ehk õpimudelit;
- 5) millal muutub kausatiivide moodustamine produktiivseks nii, et samast tüvest esineb nii kausatiivseid kui ka mittekausatiivseid kasutusi, kuidas kulgeb see protsess võrreldes teiste verbituletusviisidega.

3. Eesti keele kausatiivide omandamine

3.1. Keelematerjali üldkirjeldus ja analüüsipõhimõtted

Uurimuse aluseks on CHILDES-i eesti keele korpuse 1;8–3;0 vanuste laste pikiuuringu materjalid⁵. Kogu keelematerjali puhul on analüüsitud nii lapse kui ka lapsele suunatud kõne. Suulise kõne dialoogidest on registreeritud kõigepealt kõik verbid ning seejärel kõik kausatiivsed struktuurid, mis on omakorda jagatud analüütilisteks ja sünteetilisteks ning produktiivseteks ja ebaproduktiivseteks.

Kausatiivsete struktuuride määramisel on kasutatud nii morfosüntaktilisi (kausatiivset situatsiooni tähistab transitivne verb, morfoloogilisi kausatiive tuletus-

⁵ Analüüsitud keelematerjal on pärit Reili Arguse koostatud Hendriku ja Maigi Vija koostatud Andrease suulise kõne alamkorpusest, viimasest on valitud igast kuust üks tunniajane lindistus.

liide *-ta*) kui ka semantilisi kriteeriume (situatsioon tähistab põhjustatud tegevust või seisundit ning selles on kaks osalist, agent ja patsient⁶). Kuna kausatiivsus on äärmiselt kontekstisidus nähtus (vt Paulsen 2011: 24) ning paljud kausatiivsust väljendavad verbid võivad olla kasutusel ka mittekausatiivse situatsiooni tähistajaina, on siinse materjali puhul analüüsitud alati ka verbide esinemiskonteksti. Kontekstianalüüsi tulemusena jäid materjalist välja need verbid, kus kontekst ei võimaldanud üheselt otsustada, kas verbi oli kasutatud kausatiivse või mittekausatiivse situatsiooni tähistamiseks (näiteks ühesõnalauseaugid, kus patsient ei olnud eksplitseeritud ega mõistetav ei kitsamast (eelmine või järgmine kõnevoor) ega ka laiemast (kogu situatsioon) kontekstist.

Tabel 1 annab ülevaate laste ja lastele suunatud kõne keelematerjalist, kaldkriipsuga on eristatud lekseemide ja esinemisjuhtude arv. Võrreldes kõigepealt tabelis 1 esitatud verbide koguhulka kausatiivsete struktuuride hulgaga, võib väita, et kausatiivseid struktuure ei ole ei lapse ega lapsele suunatud kõnes kuigi arvuvalt, need moodustavad näiteks lapsele suunatud kõnes verbide esinemisjuhtudest enamasti 5–6% ega ületa üheski lindistuses 10%.

Tabel 1. Verbid ja kausatiivsed struktuurid Hendriku ja Andrease andmekogudes

Lapse vanus	Hendrik				Andreas			
	Lapse kõne		Lapsele suunatud kõne		Lapse kõne		Lapsele suunatud kõne	
	Verbide arv	Kausatiivsete struktuuride arv	Verbide arv	Kausatiivsete struktuuride arv	Verbide arv	Kausatiivsete struktuuride arv	Verbide arv	Kausatiivsete struktuuride arv
1;8	0/0	0/0	17/40	4/5	9/25	0/0	67/283	7/9
1;9	0/0	0/0	13/59	2/6	17/33	1/1	58/294	10/19
1;10	0/0	0/0	26/165	4/5	29/135	1/5	94/718	14/46
1;11	3/5	0/0	22/149	2/9	20/50	1/4	32/138	4/9
2;0	4/14	0/0	14/75	4/9	58/287	9/21	89/522	19/50
2;1	6/20	3/5	31/138	5/10	48/219	6/15	72/381	13/30
2;2	11/51	2/3	42/312	10/17	–	–	–	–
2;3	15/59	6/9	42/354	8/20	39/219	5/7	52/157	5/9
2;4	27/94	4/5	44/226	12/20	56/344	9/19	48/203	5/5
2;5	28/112	3/4	53/359	9/24	31/135	7/24	35/151	6/8
2;6	34/223	1/2	53/462	4/12	50/122	6/17	51/259	8/11
2;7	–	–	–	–	55/444	8/19	66/284	7/22
2;8	–	–	–	–	58/313	12/26	61/208	11/18
3;0	–	–	–	–	59/515	8/31	75/486	8/21

Tabelist 1 tuleb muu hulgas selgelt esile, et kahe vaadeldud lapse keeleline areng kulgeb mõnevõrra erineva kiirusega: Hendriku kõnesse ilmuvad esimesed verbid vähemalt kolm kuud hiljem kui Andrease kõnesse, esimesed kausatiivsed struktuurid aga on Hendriku kõnest registreeritud lausa viis kuud hiljem kui Andrease

⁶ Kausatiivsete struktuuride määramisel oli abiks van Valini ja Lapolla (1997: 93) esitatud kriteerium, mille järgi on sündmus kausatiivne, kui seda on võimalik kausatiivselt parafraseerida, nt *poiss avas akna* lausega *poiss põhjustas akna avanemise*.

kõnest. Kui Andrease kõnes leiduvate kausatiivide hulk on talle suunatud kõnega võrreldes peaaegu sama juba alates vanusest 2;3, leidub Hendriku kõnes kausatiive tunduvalt vähem kui talle suunatud kõnes (ka Hendrikule suunatud kõnes on kausatiive ja ka verbe üldiselt vähem kui Andreasele suunatud kõnes) ning kuna tema keelematerjal ei saa kausatiivide regulaarsest kasutusest rääkida, on järgneva analüüsi keskmeks just Andrease keelematerjal ning Hendriku andmeid on seal, kus võimalik, kasutatud võrdluseks.

3.2. Esimestest kausatiivsetest struktuuridest regulaarse kasutuseni

Esimene kausatiivne struktuur leidis Andrease kõnes vanuses 1;9 ning selleks oli tuletatud kausatiiv *näitama*:

- (1) *CHI: **näita**.
*EMA: näitasime jah (arstitädile kätt).

Kuu hiljem, vanuses 1;10 on registreeritud ka esimene leksikaalne kausatiiv *lõikama*:

- (2) *CHI: lõika (varbaküüsi lühemaks⁷).

Vanuses 2;0 leidub Andrease kõnes esimene analüütiline kausatiiv: *pane lampi põlema*⁸. Kui vanuseni 2;0 leidub lapse materjalis igas kuus ainult üks kausatiivne struktuur, siis alates kolmanda eluaasta algusest nende hulk tasahaaval suureneb.

Kausatiivsete struktuuride kasutamine ei paista lapsele raskusi valmistavat, ainus viga leidis Andrease materjalis vanuses 2;0, kus laps kasutas mittekausatiivset verbi *küpsema* kausatiivses kontekstis:

- (3) *CHI: (vaatan), kudas emme <küpsseb> [: küpsetab].

On väidetud, et kausatiivide omandamisel võivad mõne keele puhul teha lapsed ainult ühesuunalisi vigu, kasutades intransitiivset verbi transitiivses situatsioonis (nt Bowermann 1974: 147), mõne teise keele puhul aga kahesuunalisi vigu, nt kasutada kausatiivset morfeemi verbiga, mis on juba olemuslikult kausatiivne, ning intransitiivseid verbe nii, nagu nad oleksid transitiivsed (Aksu-Koç, Slobin 1985). Vaadeldud materjalis oli ainuke leitud viga eelkirjeldatu (3), vastupidiseid vigu, kus laps kasutaks (*s)ta*-liitelist verbi mittekausatiivse situatsiooni tähistamiseks, ei leidunud.

Tuletatud kausatiivverbe on lapse keelematerjalis väga vähe. Kõige rohkem leidis neli eri lekseemi ühes lindistuses ning kogu materjalis oli kokku ainult kümme eri lekseemi, sagedasimaks neist verbid *näitama* ja *vajutama*. Ometi võib vanuses 2;0 lindistatud materjalis leida ka juba esimesi märke kausatiivliite produktiivsest⁹ kasutamisest, ühes ja samas dialoogis leidis samast tüvest nii kausatiivse kui ka mittekausatiivse tähendusega verb:

- (4) *CHI: ei näe, ei **näe** patareisid.
[---]
*CHI: arstile peab **näitama**.

⁷ Andmebaasis tähistab iga kõneleja kõnevooru algust kolmetäheline kood, CHI käib lapse kohta, näidetes on sulgudesse lisatud eelmisest või järgnevast kõnevoorust või kontekstist pärit, kuid lapse lausungis eksplitseerimata jäänud info, mis näite mõistmisele siiski kaasa aitab.

⁸ Analüütiliste kausatiivide puhul on arvestatud ainult neid, kus lapse kõnevoorus esinevad mõlemad konstruktsiooni osad, s.t kui laps vastab vanema kõnevoorule *pane see kinni* lausungiga *ei pane*, ei ole seda analüüsimaterjali hulka arvatud, samuti on materjalist välja jäetud imitatsioonid ehk vanema kõnevooru täpsed kordused.

⁹ Produktiivseks on loetud kausatiiv siis, kui samast tüvest kasutab laps nii kausatiivse kui ka mittekausatiivse tähendusega vormi.

Analüütiliste kausatiivide omandamine paistab Andrease puhul alguses toetuvat *panema*-verbile: vanusest 2;0 kuni vanuseni 2;5 esineb materjalis ainult selle verbiga kausatiivseid konstruktsioone (nt *pani põlema*, *pani kinni*, *panid seisma*, *paneme kokku*, *tuleb tuduma panna*). Samas ei kasuta laps seda verbi kindla kinnistunud mallina, *panema*-verbi kasutused võib jagada leksikaalseteks ja produktiivselt moodustatavateks: *panema*-verb ühendverbi koosseisus (nt *kokku panema*) ja tulemtegevust väljendava *ma*-infinitiiviga konstruktsioonid (nt *panid seisma*). Vanuses 2;5 lisanduvad ka *tegema*-verbiga konstruktsioonid, nagu *teevad ai-ai mulle*, *teen lahti*. Imperatiivsena kasutab laps kõige sagedamini analüütilisi kausatiivseid ühendverbe (nt *pane kinni!*, *tee suu lahti!*). Imperatiivide hulk aga ei ole kogu materjali hulgas suur, need ilmuvad alles vanuses 2;5 ning kausatiivide omandamise algusfaasi kohta ei saa päris kindlasti väita, et kausatiivsete struktuuride kasutus oleks piiratud ainult kitsalt ühe pragmaatilise konteksti ehk imperatiivse situatsiooniga.

Vaadeldes arengut kumulatiivselt ehk uute kausatiivsete struktuuride lisandumist lapse keelekasutusse, on näha, et kuigi selliste struktuuride hulk suureneb lapse vanusega, lisandub päris uusi kausatiive lapse kõnesse ainult 1–2 igas kuus ning üldiselt kasutab laps ikka ühtesid ja samu tuttavaid ja ka lapsele suunatud kõnes sagedasi lekseeme (*näitama*, *vajutama*, *kinni panema*).

Teise vaatlusaluse lapse, Hendriku kõnesse ilmuvad esimesena samuti analüütilised kausatiivid, ühendverb *kinni panema* on registreeritud lapse vanuses 2;1:

(5) *CHI: **kinni pane** kaa.

*EMA: panen kinni või?

*CHI: jaa pane.

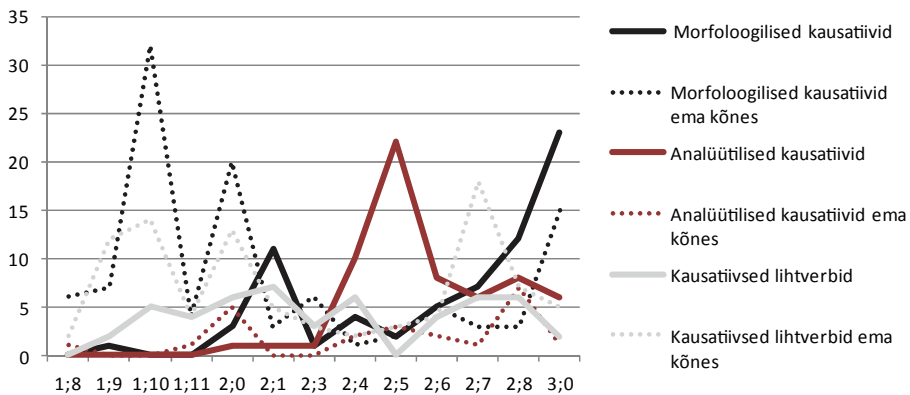
Vanuses 2;1–2;3 kasutabki Hendrik kausatiivsuse väljendamiseks ainult analüütilisi konstruktsioone (nt *katki tegema*, *lukku panema*, *käima panema*). Seejärel, vanuses 2;3 tekivad materjali esimesed kausatiivsed lihtverbid (*vett sinna kalla*, *lükka siia vann*) ning kaks kuud hiljem, vanuses 2;5 esimene ja ainuke tuletatud kausatiiv *liigutama*. Analüütiliste kausatiivide hulk oli kuni vaatlusperioodi lõpuni pea ühesugune: igas lindistuses leidis 2–3 eri kausatiivi 4–5 korral. Sagedasemad olid seisundimuutust tähistavad kausatiivsed ühendverbid, nagu *kinni panema*, *lukku panema*, *lahi tegema*, üksikuid tegevuse põhjustamist tähistavaid kausatiivseid ahelverbe leidis vanusest 2;3 (*käima panema*). Vanuses 2;3 materjalis leiduv ühend *ptrr panema* 'põrisema panema' näitab aga, et laps moodustab, hoolimata sellest, et grammatikavahendid ei ole veel täielikult omandatud ja ahelverbis esineb vaid *ma*-infinitiivi tunnusetat onomatopoeetiline sõna, *panema*-verbiga ka juba ise uusi kausatiivset tähendust väljendavaid struktuure.

Hendriku analüütiliste kausatiivsete konstruktsioonide omandamisel ei paista samuti nagu Andreasegi puhul *panema*-verbil keskset kohta olevat, materjalis esineb nii *panema*- kui ka *tegema*-verbiga ühendeid ning *panema*-verb on kasutusel nii ühendverbides kui ka produktiivselt moodustatavais ahelverbides.

Võrreldes kahe lapse kausatiivsete struktuuride omandamisandmeid võib väita, et erinevalt Andreasest eelistab Hendrik varasel omandamisperioodil analüütilisi kausatiivseid struktuure, neid leidub tema materjalis kõige varem ja kogu materjalis ka kõige rohkem. Andreasega sarnaselt ei ole märgata, et laps kasutaks kausatiive ainult imperatiivses lausungis, ainukesena esines kitsalt imperatiivses lausungis ühendverb *kinni panema*.

3.3. Analüütiliste ja sünteetiliste kausatiivsete struktuuride omandamine

Kausatiivsed struktuurid võib lähtuvalt nende läbipaistvuse astmelt paigutada skaalale analüütilised > morfoloogilised > leksikaalsed, kus ühel pool on struktuurilt kõige läbipaistvamad analüütilised kausatiivid ning teisel pool läbipaistmatud sisemiselt kausatiivsed lihtverbid. Võrreldes eri tüüpi kausatiivsete struktuuride hulka Andrease kõnes (vt joonis 1), võib väita, et esialgu eelistab laps läbipaistmatuid struktuure ehk kausatiivseid lihtverbe ning alles seejärel hakkab ta omandama morfoloogilisi ja analüütilisi kausatiivsuse väljendusvahendeid. Järjekindlalt suureneb materjalis ainult morfoloogiliste kausatiivide hulk, kausatiivseid lihtverbe on lindistustes ühtlaselt alla kümne ning analüütiliste konstruktsioonide hulk sõltub palju lindistusolukorrast. Kuna kausatiivide esinemisjuhte on kokkuvõttes siiski vähe, võib üksiku lekseemi suurem sagedus mõnes lindistuses (nt vanuses 2;5 esines Andrease materjalis analüütiline ühendverb *kinni panema* 12 korral) mõjutada üldpilti olulisel määral (vt joonis 1). Suurema materjalihulga puhul võivad sellised järsemad kõikumised tasanduda. Samuti tuleb arvestada, et nii kausatiivsete lihtverbide kui ka analüütiliste kausatiivide hulka loetud ühendverbide puhul on tegemist loomult ebaproduktiivsete klassidega ning nende hulk ei saagi võrreldes morfoloogiliste kausatiividega kuigi palju suurened.

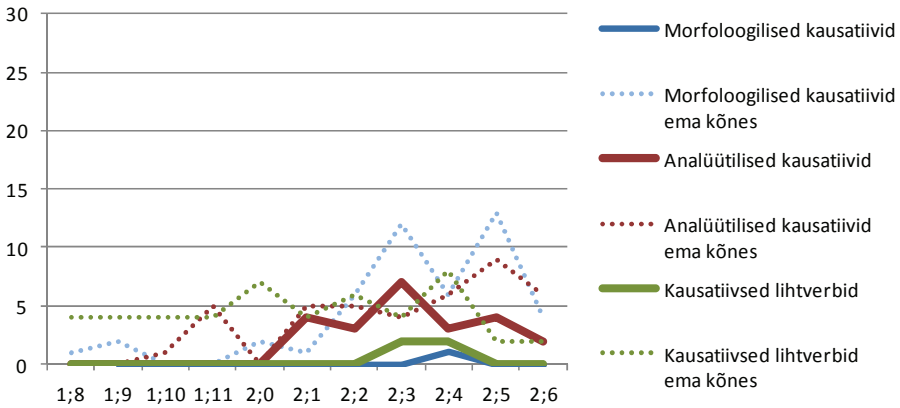


Joonis 1. Kausatiivsete struktuuride hulk ja vormiline jaotus Andrease ja tema ema kõnes

Hendriku kausatiivsete struktuuride hulgas on ülekaalus analüütilised kausatiivid (vt joonis 2). Vaatlusperioodi lõpu poole nende hulk siiski mõnevõrra väheneb, ent kuna kausatiivsete struktuuride koguhulk on tema kõnes üldse väga väike (kõige enam seitse kasutusjuhtu ühes lindistuses), ei saa muutusi siiski pidada märkimisväärseks.

Vaadeldes eri tüüpi struktuuride esinemissagedust sisendkõnes, on näha, et Hendrikule suunatud kõnes on analüütilisi kausatiive mõnevõrra rohkem kui Andreasele suunatud kõnes. Selline erinevus peegeldub ka lapse kõnes: Hendrik alustab omandamist just analüütilistest, Andreas sünteetilistest kausatiividest. Seega võib sisendkõne sagedust pidada eri tüüpi kausatiivsete struktuuride omandamise juures olulisemaks teguriks kui struktuuri läbipaistvust. Kui näiteks mõnes teises keeles omandatakse esmalt leksikaalsed kausatiivid ja alles seejärel analüütilised

(Bowerman 1974), siis eesti keele puhul võib see lasteti erineda ning sõltuda kõige enam sisendkeele sagedustest (vrd kausatiivsete lihtverbide esinemishulka lapsele suunatud kõnes joonisel 1 ja 2).

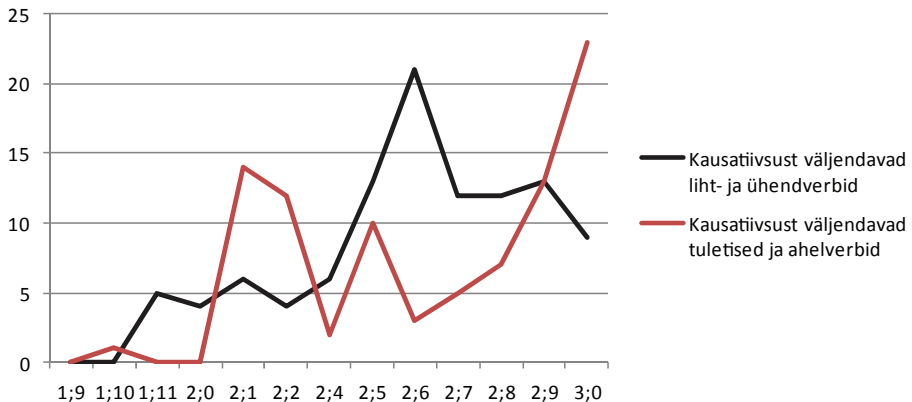


Joonis 2. Kausatiivsete struktuuride hulk ja vormiline jaotus Hendriku ja tema ema kõnes

3.4. Kausatiivsuse produktiivsete ja ebaproduktiivsete väljendusvahendite omandamine

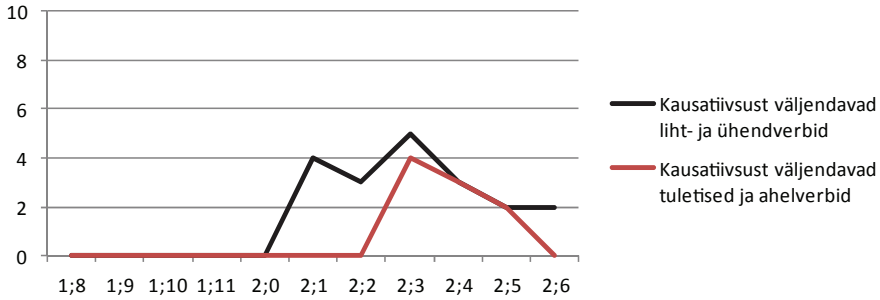
Kui eraldada analüütiliste kausatiivsete konstruktsioonide hulgast kinnistunud ühendid ehk ühendverbid, selgub, et regulaarselt moodustatavaid kausatiivseid ahelverbe on Andrease materjalis väga vähe, kokku ainult neli lekseemi. Esimene kausatiivne ahelverb on Andreasel registreeritud küll juba vanusest 2;0 (*pani lampi põlema*), kuid järgmine näide leidub materjalis alles neli kuud hiljem vanuses 2;4 (*panid seisma*) ning ülejäänud kaks vanusest 2;5 (*tuduma panna*) ja 2;8 (*panen praadima*).

Kui vaatlusperioodi alguses paistab, et laps eelistab leksikaalseid kausatiivsuse väljendusvahendeid (kausatiivsed liitverbid ja ühendverbid), siis vanusest 2;6 hakkab leksikaalsete vahendite hulk langema ja regulaarselt moodustatavate kausatiivsuse väljendusvahendite (tuletised ja ahelverbid) hulk tõusma (vt joonis 3).



Joonis 3. Produktiivsed ja ebaproduktiivsed kausatiivsuse väljendusvahendid Andrease kõnes

Hendriku keelematerjal on produktiivsete ja ebaproductiivsete kausatiivide omandamise kohta järelduste tegemiseks liiga väike. Olemasolevad vähesed andmed näitavad aga siiski, et laps alustab leksikaalsetest vahenditest: vanuses 2;1 esinevad esimesed kausatiivsed ühendverbid ning esimesed ahelverbid on registreeritud kaks kuud hiljem, vanuses 2;3. Kuigi vahe ei ole märkimisväärne, on leksikaalseid kausatiivsuse väljendusvahendeid ka koguhulgalt lapse kõnes siiski rohkem kui tuletisi ja ahelverbe (vt joonis 4).



Joonis 4. Produktiivsed ja ebaproductiivsed kausatiivsuse väljendusvahendid Hendriku kõnes

3.5. Mitmene vastendus ehk (s)ta-liite omandamine

Verbituletuse arengut üldisemalt on võimalik vaadelda mahult suuremas, Andrease andmekogus. Lapse esimesed omandatud verbid on juurverbid, esimese verbi tuletusliite võib Andrease materjalist leida vanusest 1;9 ning see on kausatiivliide (verbis *näita*), seejärel ilmuvad ka *rda-*, *sta-*liitelised faktitiivid (vanuses 1;10 *kummardama*, *pigistama*) ning *ise*-liiteline kontinuatiiv *tilisema*.

Enne kolmanda eluaasta algust on Andrease sõnavaras verbituletiste hulgas ülekaalus kausatiivid, mille kõrval leidub kõige enam *ta*-liitelisi faktitiive (nt vanuses 1;11 *valutama*). Esimesed frekventatiivid ilmuvad lapse keelekasutusse vanuses 2;0 (*toksima*, *kopsima*). Ka kogu kolmanda eluaasta jooksul domineerivad tuletatud verbide hulgas kausatiivid, mille kõrval leidub mõni faktitiiv (nt *lindistama* vanuses 2;2) ja frekventatiiv (nt *limpsima* vanuses 2;1). Tuletatud verbide hulgas muidu üsna tavalisi refleksiive vaadeldud lapse kõnes ei esinenud. Ka teiste keelte kohta on saadud sarnaseid andmeid, näiteks heebrea keeleski omandatakse kausatiivid samuti varem kui refleksiivid (Slobin 1985: 333). Lisaks võib eesti keele puhul oletada, et refleksiivsuse väljendamiseks kasutab Andreas esialgu teisi (leksikaalseid) vahendeid, näiteks pronoomenit *ise*.

Anaüüsiandmetest on näha, et võrreldes teiste verbi tuletuskategooriatega on kausatiivid lapse kõnes varased ning see, et lapsel ei ole nende omandamise puhul võimalik toetuda vormi ja tähenduse üks-ühesele vastendusele, ei paista omandamist kuidagi takistavat. Küll aga võib arvata, et kõik tuletatud verbid on siiski üldiselt tervikuna omandatud, sest laps kasutab oma kõnes ainult neid tuletisi, mis esinevad ka talle suunatud kõnes (enamasti isegi samas lindistuses) ega proovi liidetege *ise* produktiivselt uusi tuletisi moodustada.

4. Kokkuvõtteks

Kahe eesti lapse omandamisandmete põhjal saab öelda, et lapsed võivad alustada kausatiivsete struktuuride kasutamist mõnevõrra eri ajal, kuid üldiselt teise eluaasta teises pooles või kolmanda eluaasta alguses. Võrreldes inglise keele kausatiivide omandamisandmetega (Bowermann 1974, Clark 2003) paistab, et eesti lapsed omandavad kausatiivse tähenduse väljendamise kiiremini ja veatumalt. Esialgu kasutavad lapsed kausatiivsuse väljendamiseks leksikaalseid vahendeid (analüütilisi kausatiivseid ühendverbe või kausatiivseid lihtverbe), kuid kui tuletatud kausatiivid on lapse kõnesse juba ilmunud, hakkab nende hulk tasapisi ja pidevalt suurenema. Hoopis enam kui analüütilise vormi printsiip, mõjutab eri väljendusvahendite omandamise algust mingi kindla struktuuri sagedus sisendkeeles. Vormilt läbipaistev struktuur omandatakse varem ainult sellisel juhul, kui see on sisendkeeles sagedasem kui mitte nii läbipaistev sünteetiline struktuur. Nii võivad sisendkeele sagedustest olenevalt olla tulemused ka ühe keele omandamise puhul lasteti erinevad.

Morfoloogiliste kausatiivide omandamine ei paista hoolimata üks-ühese vastanduse puudumisest lapsele raskusi valmistavat, ühtsema tähendusrühma vasteks olevad refleksiivliited omandatakse võrreldes kausatiividega tunduvalt hiljem. Ka siin mängib kõige suuremat rolli lapsele suunatud kõnes esinevate eri tüüpi tuletiste sagedus: kausatiivliitega verbid on lapsele suunatud kõnes kõige sagedasemad, sageduselt järgmise tähendusrühma, samuti *ta*-liittega moodustatud faktitiivid omandavad lapsed samuti varakult.

Kausatiivide omandamises seni eristatud nii selgeid omandamisetappe ei ole eesti materjali põhjal võimalik tuvastada. Kuigi omandamine tundub olevat isegi kuni vaatlusperioodi lõpuni suhteliselt aeglane ja materjalis esinevad ikka ühed ja samad kausatiivsed struktuurid koos ainult mõne üksiku uue lekseemiga, võib siiski märgata, et omandamise alguses domineerinud moodustuselt ebaproduktiivsed kausatiivsed struktuurid, nagu kausatiivsed lihtverbid ja ühendverbid, hakkavad vaatlusperioodi lõpu poole (alates vanusest 2;8) tagaplaanile jääma ja järjest enam kasutab laps produktiivseid kausatiivsuse väljendusvahendeid, tuletisi ja ahelverbe. Samuti ei ole märgata, et kausatiivide kasutamine piirduks eesti lastel mingi kindla pragmaatilise kontekstiga: kuigi mõnevõrra sagedamini esinesid kausatiivsed ühendverbid imperatiivses kontekstis, ei saa seda teiste kausatiivsete struktuuride kohta väita.

Analüütiliste kausatiivide omandamine ei põhine eesti lastel kitsalt ühel nn mudelverbil: *panema*-verbiga väljendatud analüütilisi kausatiivseid konstruktsioone materjalis küll leidub, kuid neid on kasutatud eri tüüpi konstruktsioonides, nii ühend- kui ka ahelverbides, ega ole need omandamise algetapil ka ainukesed analüütilised kausatiivsuse väljendusvahendid.

Analüüsitud teguritest paistab kausatiivsuse omandamise alguses laste valikuid mõjutavat kõige enam sisendkeele sagedus, struktuuride läbipaistvusel ja produktiivsusel ei ole esialgu kuigi suurt rolli. Alles enne neljanda eluaasta algust võib täheldada produktiivsuse rolli kasvu ning seetõttu võibki kokkuvõtvalt väita, et eesti kausatiivsuse omandamine paistab varases keelelise arengu staadiumis olema pigem leksika ja alles hiljem ka grammatika omandamine.

Viidatud kirjandus

- Aksu-Koç, Ayhan; Slobin, Dan I. 1985. The Acquisition of Turkish. – Dan I. Slobin (Ed.). Crosslinguistic Study of Language Acquisition, 1, The Data. Hillsdale, NJ: Lawrence Erlbaum Associates, 839–878.
- Allen, Shanley E. M. 1998. Categories within the verb categorie: Learning the causative in Inuktitut. – Linguistics, 36 (4), 633–677. <http://dx.doi.org/10.1515/ling.1998.36.4.633>
- Argus, Reili 2006. Eesti keele verbi ajamorfoloogia ja aspektilisuse omandamisest. – Emakeele Seltsi aastaraamat, 52, 7–32.
- Argus, Reili 2008a. Eesti keele muutemorfoloogia omandamine. Tallinna ülikooli humanitaarteaduste dissertatsioonid, 19. Tallinn: TLÜ Kirjastus.
- Argus, Reili 2008b. Pühholingvistiline katse eesti keele objekti käändevahelduse omandamise uurimise meetodina. – Emakeele Seltsi aastaraamat, 54, 22–43.
- Argus, Reili 2010. Mida teeb *tegema*-verb hoidjakeeles. – Eesti ja soome-ugri keeleteaduse ajakiri, 2, 17–34.
- Argus, Reili; Parm, Sirli 2010. Eesti keele ajakategooria omandamisest – ajavormid ja ajasõnad. – Eesti Rakenduslingvistika Ühingu aastaraamat, 6, 25–41. <http://dx.doi.org/10.5128/ERYa6.02>
- Bates, Elizabeth; MacWhinney, Brian 1982. Functionalist approaches to grammar. – E. Wanner, L. R. Gleitman (Eds.). Language Acquisition: The State of the Art. Cambridge: Cambridge University Press, 173–218.
- Berman, Ruth 1994. Developmental perspectives on transitivity: A confluence of cues. – Y. Levy (Ed.). Other Children, Other Languages: Issues in the Theory of Acquisition. Hillsdale, NJ: Lawrence Erlbaum, 189–241.
- Bowerman, Melissa 1974. Learning the structure of causative verbs: A study in the relationship of cognitive, semantic and syntactic development. – Papers and Reports on Child Language Development, 8, 142–178.
- Clark, Eve 2003. First Language Acquisition. Cambridge: Cambridge University Press.
- Comrie, Bernard 1993. Language Universals and Linguistic Typology. Syntax and Morphology. Chicago: University of Chicago Press.
- Courtney, Ellen H. 2002. Child acquisition of Quechua causatives and change-of-state verbs. – First Language, 22 (1), 29–71. <http://dx.doi.org/10.1177/014272370202206402>
- Erelt, Mati; Reet Kasik; Helle Metslang; Henno Rajandi; Kristiina Ross; Henn Saari; Kaja Tael; Silvi Vare 1995. Eesti keele grammatika, I. Tallinn: Eesti Teaduste Akadeemia Eesti Keele Instituut.
- Jackendoff, Ray 1983. Semantics and Cognition. Current Studies in Linguistics Series, 8. Cambridge, Mass.: MIT Press.
- Jackendoff, Ray 1990. Semantic Structures. Current Studies in Linguistics Series, 18. Cambridge, Mass.: MIT Press.
- Kasik, Reet 2001. Analytic causatives in Estonian. – Mati Erelt (Ed.). Estonian: Typological Studies, V. Publications of the Department of Estonian of the University of Tartu, 18. Tartu: University of Tartu, 77–122.
- Kay, Paul; Fillmore, Charles J. 1999. Grammatical constructions and linguistic generalizations: The What's X doing Y? construction. – Language, 75 (1), 1–33. <http://dx.doi.org/10.2307/417472>
- Kehayov, Peter; Vihman, Virve 2010. Labiilsete verbide teke keelekontaktide valguses. – Ettekanne Emakeele Seltsi kõnekoosolekul Tartus 12.10.2010.
- Lakoff, George 1987. Women, Fire, and Dangerous Things. What Categories Reveal about the Mind. Chicago: University of Chicago Press.
- Lieven, Elena 2010. Input and first language acquisition: Evaluating the role of frequency. – Lingua, 120 (11), 2546–2556. <http://dx.doi.org/10.1016/j.lingua.2010.06.005>
- Matsumoto, Yoo 1996. Complex predicates in Japanese: A syntactic and semantic study of the notion 'word'. Stanford: CSLI Publications.

- Metslang, Helle 1994. Kielet ja kontrastit. – Virittäjä, 2, 203–226.
- Ninio, Anat 1999. Model learning in syntactic development: Intransitive verbs. – International Journal of Bilingualism, 3 (2–3), 111–131. <http://dx.doi.org/10.1177/13670069990030020301>
- Okabe, Reiko 2007. Children's acquisition of causatives and bi-clausality in Japanese: An experimental study. – Alyona Belikova, Luisa Meroni, Mari Umeda (Eds.). Proceedings of the 2nd Conference on Generative Approaches to Language Acquisition North America (GALANA). Somerville, MA: Cascadia Proceedings Project, 309–319.
- Pajusalu, Renate; Tõugu, Pirko; Vija, Maigi; Tulviste, Tiia 2011. Konditsionaali omandamisest eesti lapsekeeles. – Eesti Rakenduslingvistika Ühingu aastaraamat, 7, 141–155. <http://dx.doi.org/10.5128/ERYa7.09>
- Paulsen, Geda 2011. Causation and Dominance. A study of Finnish Causative Verbs Expressing Social Dominance. Åbo: Åbo Akademi University Press.
- Podlesskaja, Vera 1993. Causatives and causality: Towards a semantic typology of causal relations. – Bernard Comrie, Maria Polinsky (Eds.). Causatives and Transitivity. Studies in Language Companion Series, 23. Amsterdam: J. Benjamins, 165–176.
- Raag, Virve 1998. The effects of planned change on Estonian morphology. Acta Universitatis Upsaliensis. Studia Uralica Upsaliensia, 29. Uppsala.
- Slobin, Dan Isaac 1985. Crosslinguistic evidence for the language-making capacity. – D. I. Slobin (Ed.). The Crosslinguistic Study of Language Acquisition, 2. Theoretical Issues. Hillsdale, NJ: Lawrence Erlbaum Associates, 1157–1256.
- Tomasello, Michael 1999. The Cultural Origins of Human Cognition. Cambridge, London: Harvard University Press.
- van Valin, Robert D.; Lapolla, Randy J. 1997. Syntax. Structure, Meaning and Function. Cambridge: Cambridge University Press.
- Vihman, Marylin; Vija, Maigi 2006. The acquisition of verbal inflection in Estonian. – Natalia Gagarina, Indza Gülzow (Eds.). The Acquisition of Verbs and their Grammar: The Effect of Particular Languages. Studies in Theoretical Psycholinguistics, 33. Dordrecht: Springer Publishing Company, 263–295.
- Vija, Maigi; Torn-Leesik, Reeli; Pajusalu, Renate 2009. Tegumood eesti lapsekeeles. – Eesti Rakenduslingvistika Ühingu aastaraamat, 5, 329–344. <http://dx.doi.org/10.5128/ERYa5.22>

Reili Argus (Tallinna Ülikool) uurimisvaldkonnaks on esimese keele omandamine, huvikeskmes on siiani olnud muutemorfoloogia omandamisega seonduv.
Reili.Argus@tlu.ee

ACQUISITION OF CAUSATIVES IN ESTONIAN

Reili Argus

Tallinn University

The present paper deals with issues relating to the acquisition of causative structures in Estonian. Longitudinal data of two Estonian-speaking boys aged 1;8–3;0 was used to analyse how the acquisition of causatives take place. The main aim was to specify which kind of structures children prefer when they start to use causatives and what kind of factors play the most important role in the acquisition process – transparency, productivity or input frequency of certain causative structures.

Causativity can be expressed with different means in Estonian: with lexical means like verbs with causative meaning (*avama* ‘open’), with morphological means like suffixes (*liigutama* ‘move’) and with syntactic means like special phrasal verbs (*teeb lahti* ‘open’) and constructions with light verbs like *panema*, *ajama* and *ma*-infinitives (*paneb nutma* ‘makes cry’).

The analysis of spontaneous speech of two Estonian-speaking children reveals that children start the acquisition of causatives from lexical means (root verbs and analytic constructions like phrasal verbs with causative meaning). Derivations with causative suffix appear somewhat later, but the number of these verbs with causative suffix increases steadily. At the beginning of the observation period most causatives in children’s speech were unproductive structures (lexical causatives and phrasal verbs) while from the age 2;8 productive structures (morphological and syntactic causatives) were more frequent than unproductive ones.

The most important factor influencing the acquisition of causative structures in Estonian seems to be the frequency of some kind of causative structure in the input. The transparent structure is acquired later than the opaque one only if this structure is frequent in the input. Therefore, the early acquisition of causative structures in Estonian seems to be more the acquisition of lexicon than the acquisition of grammar.

Keywords: child language, child directed speech, input, language acquisition, causativity, word formation, Estonian

LITHUANIAN NARRATIVE LANGUAGE AT PRESCHOOL AGE

Ingrida Balčiūnienė

Abstract. The paper deals with the main linguistic indications of Lithuanian preschoolers' narratives. The analysis is based on experimental data of 24 typically developing monolingual Lithuanian children (6–7 years of age) from middle-class families, attending a state kindergarten in Kaunas (Lithuania). During the experiment, the children were asked to tell a story according to the *Cat Story* (Hickmann 1993) picture sequence. The stories were recorded, transcribed and annotated for an automatic analysis using CHILDES software. During the analysis, the syntactic complexity, lexical diversity, and general productivity (MLUw and type/token ratio) of the narratives were investigated. The results indicated the main microstructural tendencies of Lithuanian narrative language at preschool age.*

Keywords: narrative analysis, child language, language acquisition, Lithuanian

1. Introduction

During the last few decades, the literacy and general language development of school-age children seems to have become a problematic area. In Lithuania, as well as in other countries (Gardner et al. 2006, Topaj, Gagarina 2009), speech therapists, psychologists and teachers observe an increasing number of children with language disorders (SLI, dyslexia, etc.) or delay who need a speech therapy and/or the help of special needs teachers. Although we still need comprehensive statistical data about language impairments in Lithuanian children, one can observe that impaired phonology, grammar and narrative skills tend to complicate the whole process of learning (not languages only, but also other subjects), lead to low academic results and cause demotivation for learning. Thus language and early literacy development has to be investigated in order to a) indicate the typical development of Lithuanian

* This study was carried out in the framework of the national scientific project "Lietuvių vaikų kalba: įtakos ir tendencijos" ('Lithuanian Child Language: Influences and Tendencies') and supported by grant No. LIT-1-18 from the Research Council of Lithuania. The study was based on a methodology developed during author's post-doctoral studies supported by a grant No. SF-PD-2010-08-10-0199 from the Research Council of Lithuania. I am grateful to the two anonymous reviewers for their comments.

spoken and written language, and b) to identify children, whose language may be impaired.

Various models and methods for analysing linguistic competency have been developed. One of them, analysis of narrative skills, has been applied in many studies in psycholinguistics (Caramelli et al. 1998, Batoréo, Costa 2000, Atance, O'Neill 2005), neurolinguistics (Davenport et al. 1986, Baltaxe, D'Angiola 1992, 1996, Juncos-Rabadán et al. 2005, Boudreau 2008), and sociolinguistics (Benson 1997, Bokus, Wales-Shugar 1998, Fiestas, Pena 2004), for a number of reasons. Following Hayward and Schneider (2000), narratives "constitute instances of language in use rather than in isolated components out of context; they are an integral part of everyday social interactions and the school curriculum". The narratives are "typically monologues that have a recognizable beginning and end, thus are relatively easy units to identify; they are also familiar to people of all ages, excepting only infants and toddlers" (Peterson, McCabe 1991). Narratives "play a critical role in the development of discourse, literacy, and socialization abilities" (McCabe 1996). Children's oral narrative skills "are considered a key resource when children begin the important transition from oral to written communication" (Hayward, Schneider 2000) and can be early predictors of risk of reading difficulties. Following Bliss et al. (1998), narrative discourse should be a major component of assessment and intervention programs for school-age children with language disorders.

Researchers have developed various models for analysing narrative competency. One such model is the analysis of narrative microstructure characterized by a set of linguistic indications, such as productivity, lexical diversity, syntactic complexity (Loban 1976, Hughes et al. 1997, Mäkinen, Kunnari 2009, Soodla 2011), and internal linguistic structures used in narrative construction (Justice et al. 2008). Apart from that, other models for analysing narrative competency have been developed, such as the analysis of macrostructure and story grammar (Stein, Glenn 1979, Berman, Slobin 1994). In this study, the microstructure was selected for the narrative, because linguistic indications can be measured automatically by using special software (such as CHILDES, SALT, etc.), and this leads to more objective and reliable interpretation of the results.

The majority of narrative studies are still based on data from English speaking subjects (Labov, Waletzky 1967, Loban 1976, Shapiro, Hudson 1989, Shapiro 1990, Reilly et al. 1998, Ulatowska et al. 2004, Markowiak 2005), but during the past couple of decades, more languages have been involved in narrative studies, e.g.; Polish (Wales-Shugar 1998, Kiellar-Turska 1999), Swedish (Nordqvist 1998), Georgian (Imedadze, Shartava 1999), Portuguese (Jakubowitz-Batoréo 1999), Turkish (Kyuchukov 2000), Russian (Ovchinnikova 2005, 2007), French (Veneziano 2009, Veneziano, Hudelot 2009), Estonian (Soodla, Kikas 2010, Soodla et al. 2010, Soodla 2011) and Lithuanian (Lemanaitė-Deprati 1999, Balčiūnienė 2011). Cross-linguistic investigations have been carried out also (Berman, Slobin 1994, Bamberg 1997, Nicolopoulou et al. 2011). These studies have identified general tendencies of narrative acquisition and revealed many contextual factors which may influence narrative comprehension and production.

Although the importance of narrative comprehension and production tasks is highlighted in Lithuanian law (*Curriculum Framework for Primary and Basic (Lower Secondary) Education*, 2008) and didactic literature (Gaulienė et al.

2000), narrative studies in Lithuania are still at the earliest stages. Taking this fact into account, investigations of any sample (both children and adults; monolingual and bilingual; typically developing and impaired subjects) are necessary in order to identify general tendencies of Lithuanian narrative production. However, the sample of monolingual typically developing Lithuanian preschoolers was selected for a number of reasons. During the preschool and early school years, “children become members of their culture as they acquire the patterns of acting, speaking, and thinking used by their parents and peers. Following Vygotsky, we can view this acquisition process as involving internalization of the culture’s modes of thought, action, and consciousnesses” (Dickinson 1991). Preschool age is considered critical for the transition from oral to written communication, which appears to be crucial for the later development of literacy and academic attainment. There is evidence that children “who have not mastered the ability to produce adequate narratives when entering Grade 1 have difficulty in making the transition to written texts” (Bliss et al. 1998, citing Peterson 1993) and consequently have more difficulties in learning both languages and other subjects. Thus narrative and general language skills at preschool age should be investigated in order to indicate the standards of this age group and to identify children who need language therapy or a help in learning the written language.

2. Data and research method

The subjects of the study were 24 children (mean age 82 months) from middle-class families, attending state kindergarten in Kaunas (the second largest city in Lithuania). An equal number of boys and girls were selected from those children whose parents provided written permission for them to participate in the national scientific project *Lietuvių vaikų kalba: įtakos ir tendencijos*. A picture sequence, the *Cat Story* (developed by Hickmann 1993), was selected for eliciting children’s narratives. The sequence consists of six black-and-white pictures (10 x 13 cm), without a text. There are four protagonists in the sequence: a mother bird, chicks, a cat, and a dog.

An experimenter tested each child individually, in a quiet room in their kindergarten. First, for warming-up, each child was asked, whether he/she likes fairy-tales and stories, who tells stories to him/her, and then the experimenter said: “Today I would like you to tell me a story.” The experimenter took the pictures and continued: “This is a story in these pictures. First I’ll show you all the pictures, and then you look at each picture carefully and tell me the story you see.” Then the experimenter placed the pictures in the correct sequence in a single, horizontal row in front of the child, without saying anything except, “The story starts like this...”. The child was allowed to look at the pictures for a minute or two to get the gist of the story. Then the experimenter said: “Now I want you to tell the story. This is the beginning of the story. Look at the pictures and try to tell the best story you can.” No questions such as “What is he/she doing here?”; “What is this?”; “Who is coming?”, etc., were used in order not to disrupt or influence the child’s narration. Allowable prompts, if the child was hesitant to continue, were, “Tell me a story about what happens in this picture” or “Tell me what happened”.

All the stories were recorded, transcribed and coded according to CLAN (*Child Language Data Exchange System* – CHILDES, MacWhinney 2010) tools for automatic linguistic analysis, and the main linguistic indications, such as general productivity, lexical diversity, and syntactic complexity, were analysed. General productivity was indicated by MLUw index¹ and type/token ratio². For MLUw index, a mean length of utterance in words of each story was calculated. For type/token ratio, was calculated: a) general type/token ratio, b) noun type/token ratio, c) verb type/token ratio, and d) adjective type/token ratio. Lexical diversity was indicated by a reference to the main protagonists. As mentioned above, there were four protagonists in the sequence: a mother bird, chicks, a cat, and a dog. Each of the noun phrase (NP) references to these protagonists was coded as neutral (e.g., *a bird, a cat, a dog*), generalized (e.g., *an animal, bird's children*), or specified (e.g., *a crow, a pigeon, a collie*), and proportions of these NP references were calculated. Syntactic complexity was indicated by the use different syntactic structures. Each sentence of the stories was coded as a simple (e.g., *The cat wanted to catch the chicks.*) or composite sentence. According the *Practical Grammar of Lithuanian* (Ramonienė, Pribušauskaitė 2008), each composite sentence was coded additionally as complex (e.g., *The bird had to catch some worms because he wanted to feed his children.*), compound (e.g., *The cat climbed up the tree, but the dog pulled her down.*), conjunctionless (also termed asyndetic, e.g., *The bird flew away, the cat came to the tree.*) or mixed composite (e.g., *The dog pulled the cat's tail, the cat fell down, and then the bird came back.*) sentence. The proportions of these sentence types were then calculated.

2. Discussion of the results

2.1. General productivity

The findings indicate that the narrative MLUw index at preschool age is quite high (see Figure 1).

The MLUw index, with a few exceptions, varied between 5.3 and 10.0 words per utterance, and the mean MLUw index of all narratives was 9.1. These results correlate with the findings of previous studies of Lithuanian children and MLUw development, where personal and fictional narratives of preschoolers were analysed (Miklovytė 2009). Naturally, the mean MLUw index of the analysed narratives was higher than in a dialogic speech of the same target subjects (~5.0 words per utterance) (Balčiūnienė, Miklovytė 2010).

¹ Although MLU (Brown 1973) is considered a measure of grammaticality (Parker, Brorson 2005, Tilstra, McMaster 2007) and is still a controversial category (because utterance has no formal characteristics in oral speech), it is analysed in many studies as a measurement of a child's gross language development and a risk of language impairments (Eisenberg et al. 2001).

² Type/token ratio (TTR) is a measure of vocabulary and grammar variation within a written or a spoken text. A type (also termed *form*) can be described as a particular inflected form of lemma (e.g., a lemma *cat* in English has four forms: *cat, cat's, cats, and cats'*). Naturally, there are many more types of lemmas in strongly-inflected languages, such as Russian, Polish and Lithuanian (e.g., in Lithuanian a lemma *kat-ė* 'cat' has twelve types (forms) including seven case forms which are inflected additionally for singular vs. plural). *Token* can be described simply as a number of words within a text.

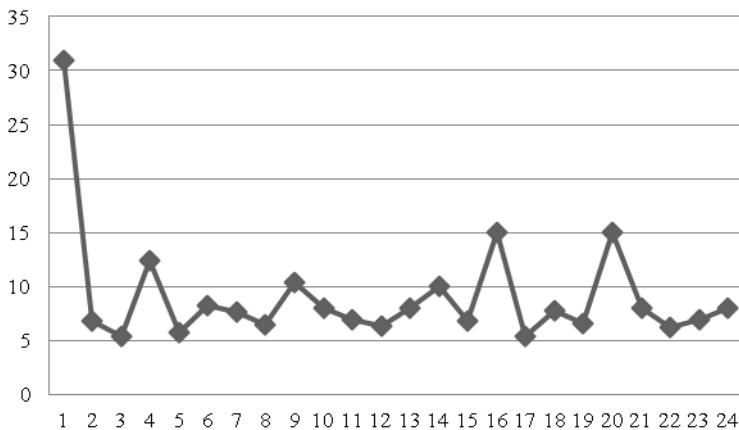


Figure 1. MLUw index in the preschoolers' stories

The MLUw index varied between 10.0 and 15.0 words per utterance in three stories, and one can suppose that these children also demonstrated better skills from the perspective of lexical diversity and syntactic complexity. However, this statement should be examined more comprehensively. The MLUw index was extremely high (31.0 words per utterance) in one story (see Figure 1), but this does not indicate good narration skills; on the contrary, narrative language there generally seems to be impaired (1).

- (1) *Pamatė jq # šuo, kad # katė lipa jau į medį, patempė jai, įkando į uodegą ir tempė, o tuo laiku atskrido paukštis # ir su kirminu, pas vaikus, o paskui eee # katė pasileido nuo sakos [: šakos] ir pradėjo šuo jį gaudyt, o tuomet šuo [//] paukštis [: paukštis] eee va [//] maitino paukščius³.*

‘The dog saw # her that # the cat climbs up the tree already, pulled her, bit her tail and pulled, and then the bird came back, # and with a worm to the children, and then # the cat left the branch and the dog started chasing him, and then the dog [//] the bird ate [//] # fed the birds.’

Despite the fact that word order in Lithuanian is free, the text seems to be “messy”; moreover, one can observe grammatical errors (e.g., the cat is referred as “he”, although “a cat” is a feminine gender noun in Lithuanian), lexical limitations (e.g., the mother bird and her chicks are referred by the same noun “a bird”); inappropriate anaphoric references (e.g. *The dog saw her, that the cat climbs up the tree...* [= The dog saw the cat that she climbs up the tree...]). These findings lead to the presumption that an extremely high MLUw index may indicate narrative and general language disorders/delays, such as lexical limitations, grammar and/or coherence disorders.

These findings indicate that the narrative mean TTR index at preschool age is quite high (see Table 1).

³ According to the tradition of *The Corpora of Spoken Lithuanian* (see Dabašinskienė, Kamandulytė 2009), the segmentation of the transcribed stories was based on intonation and pause, but not on syntactic devices. False starts, fillers, and self-repairs were excluded from calculations.

Table 1. Mean TTR index in the preschoolers' stories

	Types	Tokens	Type/token ratio
All words	66.5	91.45	0.729
Nouns	9.75	13.5	0.736
Verbs	11.8	13.3	0.875
Adjectives	0.33	0.33	1

The general TTR index varied between 0.6 and 0.8, and the mean index of all narratives was 0.7. An extremely low (≤ 0.6) TTR index indicates limited vocabulary and/or grammar (2).

- (2) *Paukštis nematė, kad katinas atėjo. Tada pa [//] katinas pamatė, kad paukštis skrenda. O katinas vėliau pamatė, kad jau nuskrido. Tada skrenda paukštis, o šuo gaudo katę.*

‘The bird did not see that # the cat came. Then saw [//] the cat saw that the bird is flying. And the cat saw then that [the bird] flew away already. Then the bird is flying, and the dog is chasing the cat.’

The noun TTR index varied between 0.5 and 1.0, and the mean index of all stories was 0.7. Similarly to the general TTR, an extremely low noun TTR indicates limited vocabulary and/or grammar, but an extremely high noun TTR index may also indicate poor general skills in narration (3).

- (3) *Katė ėlipo į medį. Atskrido balandis. Ir jis nuo uodegos ir išvijo katę.*
‘**The cat** climbed up **the tree**. The **pigeon** came back. And he [grabbed] **the tail** and chased **the cat** away.’

In this example (3), the story contains only three sentences, and each of the nouns is produced only once within the story. Although the noun TTR is extremely high (1.0) here, the story seems to be poor and limited from the perspective of macrostructure.

The verb TTR index varied between 0.7 and 1.0, and the mean index of all stories was 0.9, i.e., verbs were the most productive content words in the stories. However, the adjective TTR was completely different: adjectives were used in only four stories, and the TTR index was equal to 1.0 in each story. This means that the children used a few adjectives, and their lexical and/or grammatical forms were different. Taking into account that adjectives generally are not numerous in child language (Ceitlin 2000, Clark 2003, Kamandulytė 2010), the productivity of adjectives at preschool age could be a possible indication of accelerated language acquisition.

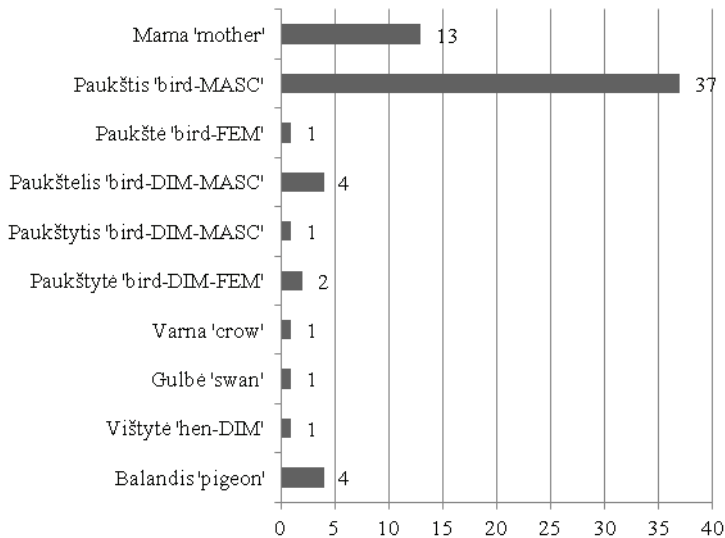
2.2. Lexical diversity

After the analysis, it can be stated that the children tended to use semantically neutral NP references for all protagonists, except chicks, while generalized and specified NP references seem to be less numerous (see Table 2).

Table 2. NP references (number of tokens) to the main protagonists

	Neutral references	Generalized references	Specified references
Mother bird	45	13	7
Chicks	22	23	4
Cat	84	–	–
Dog	50	1	–

The mother bird is the first protagonist to appear on picture 1, accompanied by her chicks. In picture 2, the mother bird leaves the nest, to bring some food for her chicks. She comes back in picture 5 with a worm for the chicks. In the stories, the mother bird was variously referred as; *paukštis* ‘bird’, *mama* ‘mother’, *paukštelis/paukštytis/paukštytė* ‘bird-DIM’, *balandis* ‘pigeon’, *varna* ‘crow’, *gulbė* ‘swan’, and *vištytė* ‘hen-DIM’ (see Figure 2).

**Figure 2.** NP references (number of tokens) to the mother bird

The majority of these NP references belong to the semantically neutral group. The children used various forms of a lexeme *paukšt-* ‘bird’ (45 tokens in total):

- the main form, a root with different endings which mark feminine vs. masculine form, e.g., *paukšt-ė* ‘bird-FEM’ (1 token) vs. *paukšt-is* ‘bird-MASC’ (37 tokens);
- several diminutive forms (the root with different diminutive suffixes, e.g. *paukšt-el-is* (4 tokens), *paukšt-yt-is* ‘bird-DIM-MASC’) which still can be gender-marked, e.g. *paukšt-yt-ė* ‘bird-DIM-FEM’ (2 tokens) vs. *paukšt-yt-is* ‘bird-DIM-MASC’ (1 token). Generalized NP references are not very numerous or various: 13 tokens (and the only grammatical form) of a lexeme *mama* ‘mother’ were observed within the stories. Only a few specified NP

references were produced in the stories: there were four tokens of a lexeme *balandis* ‘pigeon’ and one token of each of the following lexemes: *varna* ‘crow’, *gulbė* ‘swan’, and *višta* ‘hen’.

The chicks are present throughout the story. They appear mostly as an object which is being taken care of, being left alone, being attacked by the cat, being saved by the dog, and finally being fed by the mother. Naturally, when referring to the chicks the children produced many diminutive forms (6/10 lexemes, and 32/49 tokens) (e.g., *paukšė-iuk-ai* ‘chick-DIM-PL’ (16 tokens), *vaik-el-iai* ‘child-DIM-PL’ (2 tokens), *maž-yl-iai* ‘baby-DIM-PL’ (3 tokens), *balandž-iuk-ai* ‘pigeon-DIM-PL’ (2 tokens)) (see Figure 3).

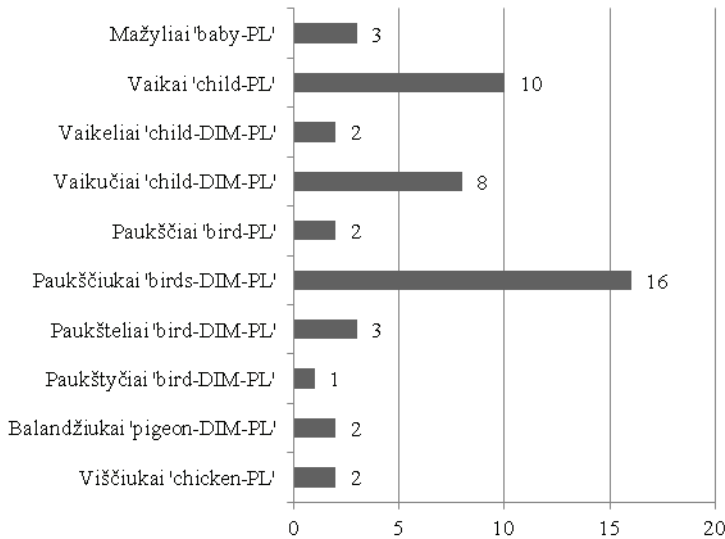


Figure 3. NP references (number of tokens) to the chicks

However, it should be mentioned that in Lithuanian language diminutive suffixes are used also for expressing kinship terms, thus a word *balandž-iu-kai* ‘pigeon-DIM-PL’ can be interpreted as either ‘small pigeons’ or ‘children of a pigeon’; similarly, *paukšė-iuk-ai* ‘bird-DIM-PL’ can be interpreted as either ‘small birds’ or ‘children of a bird’. Concerning this interpretation, all the NP references to the chicks used in the stories contain a concept of ‘children’, but this concept can be semantically neutral (e.g., *paukščiukai* ‘child of bird-PL’), generalized (e.g., *vaikai* ‘child-PL’, *mažyliai* ‘baby-PL’) or specified (e.g., *balandžiukai* ‘child of a pigeon-PL’, *viščiukai* ‘child of a hen-PL’).

The cat is the main and the most active protagonist in the story. It appears in picture 2, and then is trying to catch the chicks and being attacked by the dog. Talking about the cat, the children produced only one lexeme *kat-* ‘cat’ (see Figure 4), but in three different forms: the main form which was marked in several cases by a derivational suffix for masculine gender (*kat-ė* ‘cat-FEM’ (58 tokens) vs. *kat-in-as* ‘cat-MASC’ (24 tokens)), and a diminutive form *kat-yt-ė* ‘cat-DIM-FEM’ (2 tokens).

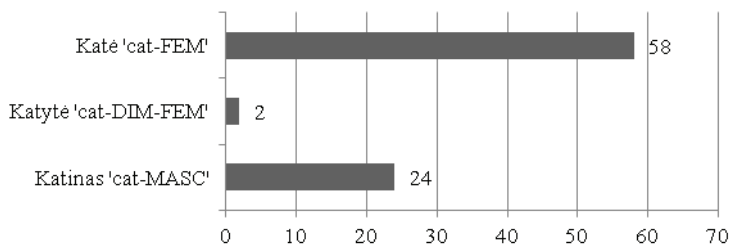


Figure 4. NP references (number of tokens) to the cat

However, only the neutral NP reference was used in the stories.

The dog appears in picture 4 as a positive protagonist who attacks the cat and saves the chicks. The children, referring to the dog, produced mainly neutral NP references: the dog was referred to most often as *šuo* 'dog' (48 tokens) or *šuniukas* 'dog-DIM' (2 tokens) (see Figure 5), and only one token of a generalized NP reference, *žvėris* 'wild animal', was observed.

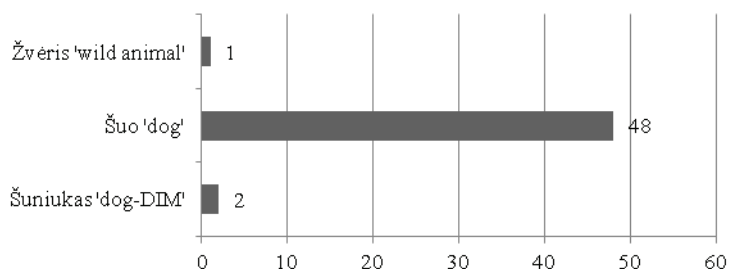


Figure 5. NP references (number of tokens) to the dog

A difference between the general number of lexemes referring to birds (both the mother and her children) and quadrupeds (both the cat and the dog) can be explained by the influence of context (style of the pictures, and linguistic input). First, the species of bird is not easily identifiable from the pictures (besides the lexeme *paukštis* 'bird', birds could be referred as pigeons, crows, magpies, pelicans, cormorants, etc.), thus a child has a wide range of choice, and the preference depends on his/her own experience, input, and general cognitive and language development. The breed of the cat and the dog is also hardly identifiable from the pictures, but it is more usual in spoken Lithuanian (especially in child-directed speech) to use various forms of the lexemes *katė* 'cat' and *šuo* 'dog' than the terms for particular breed. The results indicate the same strategy in the children's stories: the preschoolers produced semantically neutral lexemes *katė* 'cat-FEM' / *katinas* 'cat-MASC' / *katytė* 'cat-DIM-FEM' and *šuo* 'dog' / *šuniukas* 'dog-DIM', and only one child, who probably was not able to identify the dog, preferred the generalized concept *žvėris* 'wild animal'.

2.3. Syntactic productivity

According to these results, the productivity of simple and composite sentences seem to be almost equal: children produced 67 simple sentences, and 63 composite sentences. However, productivity of different types of composite sentences is significantly different. Compound and mixed composites are the most numerous (42 and 21 respectively) among all composite sentences, while other types are rare (only 12 complex sentences, and 6 conjunctionless (asyndetic) sentences were observed within the stories).

Compound sentences (especially those with the conjunction *and*) are considered the easiest type of composite sentences to acquire (Ceitlin 2000: 220). The results indicate that copulative and juxtapositive sentences were preferred (in comparison with other subtypes of compound sentences) by the preschool children, presumably as the simplest and easiest way of expressing temporal and/or causal relations (4a–c):

- (4a) *Tupėjo balandis lizdelyje su balandžiukais, ir atėjo katė.*
'A pigeon was sitting in his nest with his children, and a cat came.'
- (4b) *Po to katinas atėjo, # atsitūpė, o paukštis nuskrido.*
'Then a cat came, sat down, # and the bird left.'
- (4c) *Paskui išskrido savo vaikeliam parnešti valgyt, ir katė norėjo pavogt # vaikelius ir suvalgyt.*
'Then [the bird] left to bring some food to her children, and the cat decided to steal the children and to eat them.'

Other subtypes of compound sentences were not observed in the stories.

Conjunctionless sentences were the rarest type of composite sentence produced in the stories, despite the fact that these sentences express temporal and/or causal relations, similarly to the compound sentences (5a–c):

- (5a) *Lipo katė # į lizdą, tada šuo patempė ir tada vijosi ją.*
'The cat was climbing to the nest, then the dog pulled her and chased her.'
- (5b) *Katinas jau buvo labai arti, # šuo pagriebė katino uodegą.*
'The cat was very close already, # the dog grabbed the cat's tail.'
- (5c) *Katinas lipo į medį, šuo atėjo.*
'The cat climbed up the tree, a dog came.'

Presumably the children were trying to tell a coherent and cohesive story, thus they were trying to use more conjunctions, even in the constructions where the conjunctions are not necessary.

Complex sentences (especially causal and temporal clauses) were still difficult to produce, but even these skills seem to be partially acquired. The children tried to express the goal of the cat, and the goal of the dog, i.e., the goals which are easily identifiable from the pictures (6a–b):

- (6a) *Kai paukštis atskrido, # šuo nusivijo katę.*
'When the bird came back, # the dog was chasing the cat.'

- (6b) *Paskui pagalvojo, kad pati galėtų užsikarti ir [/] ir suvalgyti.*
‘Then [the cat] decided, that she could climb up and [/] and eat [the chicks].’

Other goals are expressed rather by compound or even simple sentences.

According to the *Grammar of Modern Lithuanian* (Ambrazas et al. 1996), mixed composite sentences contain at least two different relationships of the clauses (7a–b):

- (7a) *Paskiau šuo vijosi katiną, kad sudraskytų, o paukščiai kad gyventų gražiai.*
‘Then the dog chased the cat in order [*causal coordination*] to kill her, and [*copulative coordination*] the birds could live happily.’
- (7b) *Paskui katinas įlipo į medį ir [/] ir šuo žiūrėjo, kaip jisai lipo.*
‘Then the cat climbed up the tree, and [/] and [*copulative coordination*] the dog observed, how [*direct coordination*] he was climbing.’

Despite their extremely complicated structure, mixed composite sentences seem to be among the most frequent types of composite sentences within the stories.

3. Conclusion

The results indicated the main microstructural tendencies of Lithuanian narrative language at preschool age:

1. The majority of the subjects (6–7 year old Lithuanian preschoolers) demonstrated quite high general productivity indexes: mean MLUw and type/token ratio indexes were equal to 9.1 and 0.7 respectively. High deviation from the mean MLUw index refer either to accelerated or to weak narrative skills.
2. Semantic analysis indicated a wide-ranging lexical diversity of NP references to the main protagonists; this stands particularly for the *mother bird* that is referred by various semantically neutral, generalized and specified NPs.
3. Complex structures (especially causal and temporal clauses) are still difficult to produce, but even these skills seem to be acquired at preschool age.

In this study, microstructural indications were analysed automatically (by using CHILDES tools), thus only linguistic skills of the narratives can be described. The next steps for future investigations should be macrostructural narrative analysis of the same target group and a comprehensive narrative analysis of other samples, including adults, bilinguals and SLI subjects.

Abbreviations

DIM	diminutive	#	pause
FEM	feminine	[//]	self-correction
MASC	masculine	[:]	incorrect pronunciation
PL	plural		

References

- Ambrasas, Vytautas; Garšva, Kazimieras; Girdenis, Aleksas; Jakaitienė, Evalda; Kniūkšta, Pranas; Krinickaitė, Stasė; Labutis, Vytautas; Laigonaitė, Adelė; Oginskienė, Elena; Pikčilingis, Juozas; Ružė, Albertas; Sližienė, Nijolė; Ulvydas, Kazys; Urbutis, Vincas; Valeckienė, Adelė; Valiulytė, Elena 1996. Dabartinės lietuvių kalbos gramatika [Grammar of Modern Lithuanian]. Vilnius: Mokslo ir enciklopedijų leidykla.
- Atance, Cristina M.; O'Neill, Daniela K. 2005. Preschoolers' talk about future situations. – *First Language*, 25 (1), 5–18. <http://dx.doi.org/10.1177/0142723705045678>
- Balčiūnienė, Ingrida 2011. Narrative skills among Lithuanian preschoolers. – *Онтолингвистика – наука XXI века*. Санкт-Петербург: Златоуст, 279–282.
- Balčiūnienė, Ingrida; Miklovytė, Inga 2010. Priešmokyklinio amžiaus vaikų pokalbiai: ar mergaitės ir berniukai kalba skirtingo ilgio sakiniais? [Conversations between Lithuanian preschoolers: MLU of boys and girls]. – *Psichologiniai tyrimai: menas ar amatas?* [Studies in Psychology: Art or Trade?]. Vilnius: Vilnius University Press, 123–129.
- Baltaxe, Christiane A. M.; D'Angiola, Nora 1992. Cohesion in the discourse interaction of autistic, specifically language-impaired, and normal children. – *Journal of Autism and Developmental Disorders*, 22 (1), 1–21. <http://dx.doi.org/10.1007/BF01046399>
- Baltaxe, Christiane A. M.; D'Angiola, Nora 1996. Referencing skills in children with autism and specific language impairment. – *European Journal of Disorders of Communication*, 31 (3), 245–258. <http://dx.doi.org/10.3109/13682829609033156>
- Batoréo, H. J.; Costa, A. 2000. Reference mechanism in children's oral and written narratives at the age of ten. – *Psycholinguistics on the Treshold of the Year*, 265–270.
- Benson, Margaret S. 1997. Psychological causation and goal-based episodes: Low-income children's emerging narrative skills. – *Early Childhood Research Quarterly*, 12 (4), 439–457. [http://dx.doi.org/10.1016/S0885-2006\(97\)90021-1](http://dx.doi.org/10.1016/S0885-2006(97)90021-1)
- Berman, Ruth A.; Slobin, Dan I. 1994. *Relating Events in Narrative: A crosslinguistic developmental study*. Lawrence Erlbaum Associates, Inc., Publishers.
- Bliss, Lynn S.; MacCabe, Allyssa; Miranda, Elisabeth A. 1998. Narrative assessment profile: Discourse analysis for school-age children. – *Journal of Communication Disorders*, 31 (4), 347–363. [http://dx.doi.org/10.1016/S0021-9924\(98\)00009-4](http://dx.doi.org/10.1016/S0021-9924(98)00009-4)
- Bokus, Barbara; Wales-Shugar, Grace 1998. Social structures of children's narrational activity. – *Psychology of Language and Communication*, 2 (1), 75–81.
- Boudreau, Donna 2008. Narrative abilities: Advances in Research and Implications for Clinical Practice. – *Topics in Language Disorders*, 28 (2), 99–114. <http://dx.doi.org/10.1097/01.TLD.0000318932.08807.da>
- Brown, Roger 1973. *At first language: The early stages*. MA thesis. Harvard University Press.
- Caramelli, Nicoletta; Borghi, Anna M.; Tison, Carla 1998. The textual pre-conditions for action narration in children's story construction. – *Psychology of Language and Communication*, 2 (2), 47–56.
- Ceitlin, Stella N. 2000 = Цейтлин, Стелла Н. 2000. *Язык и ребенок: лингвистика детской речи* [Language and Child: Linguistics of Child Language]. Москва: ВЛАДОС.
- Clark, Eva V. 2003. *The Lexicon in Acquisition*. Cambridge: Cambridge University Press.
- Curriculum Framework for Primary and Basic (Lower Secondary) Education 2008. Approved by the Minister of Education and Science of the Republic of Lithuania. 26 August

2008. Resolution No. ISAK-970. http://www.smm.lt/en/ge/docs/18%20overtimas%20SAC%202008%20Bendrosios%20programos%2008%2009%2022_Anglu%20k.pdf (23.09.2011).
- Dabašinskiėnė, Ineta; Kamandulytė, Laura 2009. Corpora of Spoken Lithuanian. – *Estonian Papers in Applied Linguistics*, 5, 67–77. <http://dx.doi.org/10.5128/ERYa5.05>
- Davenport, Linda; Yingling, Charles D.; Fein, George; Galin, David; Johnstone, Jack 1986. Narrative speech deficits in Dyslexics. – *Journal of Clinical and Experimental Neuropsychology*, 8 (4), 347–361. <http://dx.doi.org/10.1080/01688638608401326>
- Dickinson, David D. 1991. Teacher agenda and setting: constraints of conversation in pre-schools. – Allysya McCabe, Carole Peterson (Eds.). *Developing Narrative Structure*. Lawrence Erlbaum Associates, Inc., Publishers, 255–310.
- Eisenberg, Sarita L.; McGovern-Fersko, Tara; Lundgren, Sheryl 2001. The use of MLU for identifying language impairment in preschool children: A review. – *American Journal of Speech-Language Pathology*, 10 (4), 333–342.
- Fiestas, Christine E.; Pena, Elisabeth D. 2004. Narrative discourse in bilingual children: Language and task effect. – *Language, Speech, and Hearing Services in Schools*, 35, 155–168. [http://dx.doi.org/10.1044/0161-1461\(2004/016\)](http://dx.doi.org/10.1044/0161-1461(2004/016))
- Gardner, Hilary; Froud, Karen; McClelland, Alastair; van der Lely, Heather K. J. 2006. Development of the Grammar and Phonology Screening (GAPS) test to assess key markers of specific language and literacy difficulties in young children. – *International Journal of Language and Communication Disorders*, 41 (5), 513–540. <http://dx.doi.org/10.1080/13682820500442644>
- Gaulienė, Veronika G.; Giedrienė, Regina; Grikainienė, Loretta 2000. Kalbos tyrimas: metodinė priemonė [A Tool for Language Assessment]. Vilnius: Žuvėdra.
- Hayward, Denyse, W.; Schneider, Phyllis 2000. Effectiveness of teaching story grammar knowledge to pre-school children with language impairment: An exploratory study. – *Child Language Teaching and Therapy*, 16 (3), 255–284. <http://dx.doi.org/10.1177/026565900001600303>
- Hickmann, Maya 2003. *Children's Discourse: Person, Space and Time Across Languages*. Cambridge: Cambridge University Press.
- Hughes, Diana; McGillivray, LaRae; Schmidek, Mark 1997. *Guide to Narrative Language: Procedures for Assessment*. PRO-ED.
- Imedadze, Natela; Shartava, Lia 1999. Motion verbs in Georgian children's narratives. – *Psychology of Language and Communication*, 3 (1), 61–74.
- Jakubowitz-Batoré, Hanna 1999. Children's discourse in European Portuguese narratives. – *Psychology of Language and Communication*, 3 (1), 19–27.
- Juncos-Rabadán, Onésimo; Pereiro, Arturo X.; Soledad-Rodríguez, María 2005. Narrative speech in aging: Quantity, information content, and cohesion. – *Brain and Language*, 95 (3), 423–434. <http://dx.doi.org/10.1016/j.bandl.2005.04.001>
- Justice, Laura M.; Bowles, Ryan P.; Kaderavek, Joan N.; Ukrainetz, Teresa A.; Eisenberg, Sarita L.; Gillam, Ronald B. 2006. The index of narrative microstructure: A clinical tool for analyzing school-age children's narrative performances. – *American Journal of Speech-Language Pathology*, 15, 177–191. [http://dx.doi.org/10.1044/1058-0360\(2006/017\)](http://dx.doi.org/10.1044/1058-0360(2006/017))
- Kamandulytė, Laura 2010. *Lietuvių kalbos būdvardžio įsisavinimas: leksinės ir morfosintaksinės ypatybės* [Acquisition of Lithuanian Adjective: Lexical and Morphosyntactic Features]. PhD thesis. Kaunas: Vytautas Magnus University.
- Kielar-Turska, Maria 1999. The inner landscape of characters in stories told by children. – *Psychology of Language and Communication*, 3 (2), 49–56.
- Kyuchukov, Hristo 2000. Introducing referents in Turkish children's narratives. – *Psychology of Language and Communication*, 4 (1), 65–74.

- Labov, William; Waletzky, Joshua 1967. Narrative analysis: Oral versions of personal experience. – June Helm (Ed.). *Essays on the Verbal and Visual Arts: Proceedings of the 1966 Annual Spring Meeting of the American Ethnological Society*. Seattle: University of Washington Press, 12–44.
- Lemanaitė-Deprati, Greta 1999. Perception and evaluation in Lithuanian children's narratives. – *Psychology of Language and Communication*, 3 (2), 57–64.
- Loban, Walter 1976. *Language Development: Kindergarten through grade twelve*. National Council of Teachers of English, Urbana, Ill.
- MacWhinney, Brian 2010. *The CHILDES Project: Tools for Analyzing Talk*. Electronic Edition. <http://childes.psy.cmu.edu/manuals/CLAN.pdf> (23.09.2011).
- Markowiak, Anthea N. 2005. *Narrative Comprehension in Kindergarten: An Analysis of Talk About Narratives by Children Differing in Early Literacy Development*. MA Thesis. Sydney: University of Sydney.
- Mäkinen, Leena; Kunnari, Sari 2009. Narrative as an assessment method in children with SLI and AS/HFA. – COST ISO804 WG2 Narratives Meeting. Amsterdam: Amsterdam University. 16–18 November, 2009.
- McCabe, Allyssa 1991. Preface: Structure as a way of understanding. – Allysa McCabe, Carole Peterson (Eds.). *Developing Narrative Structure*. Lawrence Erlbaum Associates, Inc., Publishers, ix–xvii.
- Miklovytė, Inga 2009. *Priešmokyklinio amžiaus vaikų kalbos morfologijos ypatybės [Morphological Characteristics of Lithuanian Preschoolers' Speech]*. BA thesis. Kaunas: Vytautas Magnus University.
- Nicolopoulou, Ageliki; Aksu-Koc, Ayhan; Küntay, Aylin; Andersen, Christina; Chang, Chien-ii; Ignacio, Matt; Slancová, Daniela 2011. A cross-linguistic study of character reference in young children's narratives. – 12th International Congress for the Study of Child Language. Montreal, Université du Québec à Montréal, 19–23 July, 2011.
- Nordqvist, Asa 1998. Projecting speech to protagonists in oral and written narratives: A developmental study. – *Psychology of Language and Communication*, 2 (2), 37–46.
- Ovchinnikova, Irina 2005. Variety of children's narratives as the reflection of individual differences in mental development. – *Psychology of Language and Communication*, 11 (1), 29–53.
- Ovchinnikova, Irina 2007. Slips of the tongue in children's narratives: Connectionist interpretation. – *Psychology of Language and Communication*, 11 (1), 23–41.
- Parker, Matthew D.; Brorson, Kent 2005. A comparative study between mean length of utterance in morphemes (MLUm) and mean length of utterance in words (MLUw). – *First Language*, 25 (3), 365–376. <http://dx.doi.org/10.1177/0142723705059114>
- Peterson, Carole; McCabe, Allyssa 1991. Linking children's connective use and narrative macrostructure. – Allysa McCabe, Carole Peterson (Eds.). *Developing Narrative Structure*. Lawrence Erlbaum Associates, Inc., Publishers, 29–53.
- Ramonienė, Meilutė; Pribušauskaitė, Joana 2008. *Practical Grammar of Lithuanian*. Vilnius: Baltos lankos.
- Reilly, Judy S.; Bates, Elisabeth A.; Marchman, Virginia A. 1998. Narrative discourse in children with early focal brain injury. – *Brain and Language*, 61 (3), 335–375. <http://dx.doi.org/10.1006/brln.1997.1882>
- Schneider, Phyllis; Vis Dubé, Rita V. 2005. Story presentation effects on children's retell content. – *American Journal of Speech-Language Pathology*, 14, 52–60. [http://dx.doi.org/10.1044/1058-0360\(2005/007\)](http://dx.doi.org/10.1044/1058-0360(2005/007))
- Shapiro, Lauren R.; Hudson, Judith A. 1989. Cohesion and coherence in preschool children's picture-elicited narrative. – National Biennial Meeting of the Society for Research in Child Development. Kansas City, MO, April 27–30, 1989.

- Shapiro, Lauren R. 1990. Developmental changes in young children's ability to produce cohesive and coherent stories. – Conference on Human Development. Richmond, VA, March 29–31, 1990.
- Soodla, Piret 2011. Picture-Elicited Narratives of Estonian Children at the Kindergarten-School Transition as a Measure of Language Competence. *Dissertationes Pedagogicae Universitatis Tartuensis*. Tartu: Tartu University Press.
- Soodla, Piret; Kikas, Eve 2010. Macrostructure in the narratives of Estonian children with typical development and language impairment. – *Journal of Speech, Language, and Hearing Research*, 53, 1321–1333. [http://dx.doi.org/10.1044/1092-4388\(2010/08-0113\)](http://dx.doi.org/10.1044/1092-4388(2010/08-0113))
- Soodla, Piret; Kikas, Eve; Pajusalu, Renate; Adamka, Aive; Parm, Sirli 2010. Vahendamata ja vahendatud narratiiv laste kõnearengu hindamisel [Self-generated and retold narratives as a tool of language assessment]. – *Estonian Papers in Applied Linguistics*, 6, 277–296. <http://dx.doi.org/10.5128/ERYa6.17>
- Tilstra, Janet; McMaster, Kristen 2007. Productivity, fluency, and grammaticality measures from narratives: Potential indicators of language proficiency? – *Communication Disorders Quarterly*, 29 (1), 43–53. <http://dx.doi.org/10.1177/1525740108314866>
- Topaj, Nathalie; Gagarina, Natalia 2009. Language diagnostic and support programs for bilingual children in Germany (Berlin). The start-up meeting of the FREPY (Friendly Resources for Playful Speech Therapy) project. Kaunas, Vytautas Magnus University, December 3–6, 2009.
- Ulatowska, H. K.; Streit-Olness, G.; Samson, A. M.; Keebler, M. W.; Goins, K. E. 2004. On the nature of personal narratives of high quality. – *Advances in Speech-Language Pathology*, 6 (1), 3–14. <http://dx.doi.org/10.1080/14417040410001669453>
- Wales-Shugar, Grace 1998. Entry into narration: uses of operations of reference. – *Psychology of Language and Communication*, 2 (2), 7–16.
- Veneziano, Edy 2009. The effects of two intervention procedures on children's narrative skills. – ICERI2009 Proceedings. Madrid, IATED.
- Veneziano, Edy; Hudelot, Christian 2009. Explaining events in narratives: The impact of scaffolding in 4 to 12 year old children. – *Psychology of Language and Communication*, 13 (1), 3–20. <http://dx.doi.org/10.2478/v10057-009-0001-X>

Ingrida Balčiūnienė (Centre of International and Multilingual Communication at Vytautas Magnus University). Her research interests include psycholinguistics and first language acquisition (longitudinal studies, conversation analysis, language disorders).
i.balciuniene@hmf.vdu.lt

LEEDUKEELNE NARRATIIV KOOLIEELIKU-EAS

Ingrida Balčiūnienė

Vytautas Magnuse Ülikool

Artikkel käsitleb leedu eelkooliealiste laste narratiivide põhilisi lingvistilisi indikaatoreid, mis näivad olevat üldise keelearengu olulisemate kriteeriumide hulgas. Analüüs põhineb kahekümne neljal keskklassi kuuluvate perekondade ükskeelsete normaalse keelelise arenguga leedu laste (6–7-aastased) andmetel. Kõik lapsed käivad riiklikus lasteaias Kaunases (Leedus). Eksperimendi vältel testiti lapsi individuaalselt: neid paluti jutustada “Kassi lugu” (Hickmann 1982) pildisarja järgi. Kogutud jutustused lindistati, litereeriti ning kodeeriti automaatanalüüsi jaoks, mis toimus CHILDES-i vahenditega. Analüüsi eesmärk oli vaadelda jutustuste süntaktilist komplekssust, leksikaalset varieeruvust ning üldist produktiivsust (VKP (väljendite keskmine pikkus sõnades, ingl *MLUw*), lekseemi/esinemisjuhu suhtarv (ingl *type/token ratio*)). Tulemused näitasid leedukeelse narratiivi põhilisi mikrostruktuurilisi omadusi eelkooliealiste astmel.

Enamuse 6–7-aastaste leedu eelkooliealiste laste puhul täheldati päris kõrgeid produktiivsuse suhtarve: VKP ja lekseemi/esinemisjuhu suhtarvude keskmised olid 9,1 ja 0,7. Suured kõrvalekalded keskmisest VKP-st osutavad kas kiirema keelelise arenguga lastele või mõne lapse nõrgale jutustamisoskusele.

- 1) Semantilise analüüsi tulemused näitasid olulist leksikaalset erinevust loo peategelastele viitavates nimisõnafraasides; seda oli eriti märgata emalinnu tegelase puhul, millele viidati erinevate semantiliselt neutraalsete, üldiste ning täpsustavate nimisõnafraasidega.
- 2) Komplekseid struktuure (eriti kausaalseid ning temporaalseid osalauseid) oli lastel ikka veel raske koostada, kuid tundub, et isegi neid oskusi hakatakse omandama juba eelkoolieas.

Selles uurimuses analüüsiti mikrostruktuurilisi indikaatoreid automaatselt CHILDES-i vahendeid kasutades, seetõttu saab kirjeldada vaid narratiivi lingvistilisi oskusi. Järgmiseks peaksid uurimused käsitlema narratiivi makrostruktuuri samas sihtgrupis, ühtlasi ka ulatuslikku narratiivi analüüsi teistsugustes valimites, mis koosnevad täiskasvanutest, kakskeelsetest ning alakõnega lastest.

Võtmesõnad: narratiivid, lastekeel, keeleomandamine, leedu keel

ENDA ALGATATUD ENESEPARANDUS EESTIKEELSETES MSN-I DIALOOGIDES

Tiit Hennoste

Ülevaade. Käesolevas artiklis antakse esmakordselt ülevaade eesti MSN-i suhtluses tehtud eneseparandustest Tartu Ülikooli netikeele korpuse materjali põhjal, kasutades vestlusanalüütilist lähenemist. Esmalt tuuakse välja leitud eneseparanduse tehnoloogiad ja vormivõtted vastavalt nende positsioonile probleemallika suhtes, kasutatud operatsioonile ja lahendatud probleemidele. Seejärel analüüsitakse paranduste suhtluslikke põhjuseid ja eesmärgi ning näidatakse, et paranduste roll on seotud mh sellega, millise aja järel need on tehtud. Lõpus tuuakse välja MSN-i paranduste erinevused ja sarnasused suulise dialoogi ning jututubade parandustega.*

Võtmesõnad: vestlusanalüüs, dialoog, internetikeel, eneseparanduste tehnoloogiad ja rollid, eesti keel

1. Sissejuhatus

Käesolevas artiklis analüüsin probleemi tekitaja enda algatatud eneseparandusi eestikeelses MSN-i suhtluses.

MSN-i suhtlus (*Instant Messaging*) on arvuti vahendusel reaajas (sünkroonselt) toimuv kirjalik dialoog.¹ Ühelt poolt on tegu reaajas toimuva dialoogiga, milles vahetatakse suhteliselt spontaanselt lühikesi kõnevoore nagu suulises suhtluses, teiselt poolt on see kirjalik suhtlus, mis kasutab lisaks netisuhtluse erivahendeid (emotikonid jms) ja on seotud arvutiprogrammi erivõimalustega. Olulisim piirang on selles, et MSN-is saadetakse korraka pikemaid löike ehk saadetisi (Baron 2010: *transmission unit*, Markman 2006: *post*), mitte üksikhäälikuid. Saadetus võib olla lõpetatud kõnevoor või moodustada osa voorust (nt kahesaadetieline voor: *kas kõik need sõnad sealt lehelt // tuleb koos tähendustega pähe õppida?*).

Selline suhtlustüüp puudus väljaarenenud kujul enne vastavate arvutiprogrammide tulekut üldse, seetõttu on suhtlejalatel tulnud arendada välja sobivad suhtlusmallid ja -tehnikad. Selleks on kohandatud mitmeid suulise dialoogi võtteid

* Artikkel on valminud Euroopa Regionaalarengufondi Eesti Arvutiteaduse Tippkeskuse (projekt SF0180078s08), riikliku programmi "Eesti keel ja kultuurimälu" (2009–2013; projekt 09104) ning Eesti Teadusfondi grantide nr 8558 toel.

¹ Arvutisuhtluse tekstitüpoloogiate kohta vt Herring (2007), Hennoste (2009).

(vt Garcia, Jacobs 1999, Herring 1999, Quan-Haase 2009, Anderson jt 2010, Baron 2010). Seetõttu analüüsitakse seda suhtlust tavaliselt suulise dialoogi mudelite abil.

Enda algatatud eneseeparandus on tegevus, mille käigus kõneleja või kirjutaja lahendab teksti tegemisel tekkinud probleeme enda algatusel. Eneseeparandus esineb nii kõnes kui kirjas, aga selle põhimudelid on loodud suulise dialoogi põhjal, kus parandamise protsess on eksplitsiitne. Enim on selle teemaga tegelenud vestlusanalüüs (Schegloff jt 1977, Schegloff 1992, 1996), mille lähenemist on kasutatud ka näiteks jututubade eneseeparanduste analüüsiks (Schönfeldt, Golato 2003, Markman 2006). Ka selles uurimuses järgin vestlusanalüüsi lähenemist ja kasutan selle terminoloogiat.

Eneseeparandus jagatakse kaheks: järje otsimine (takerdumine) ja olevas tekstis millegi ümbertegemine. Käesolevas tegelen ainult ümbertegemisega ja kasutan sõna (enese)parandus edaspidi selle kohta.

Kõnelejal on esimene võimalus ennast parandada samas voores, kus paikneb ka probleemlikas. Põhiliselt tehakse parandus samas lauses (*aga se=on **Tuga-Tuglase** kolm`teist aadressil*), harva valmis voores lõpus ehk vooresvahetusallas (*no se=oli see, kus need **veetorud** ühesõnaga olid läbi lõigatud **või=e üpip** **plastiktorud** eksole.*).

Järgmine võimalus on lahendada probleem mõnes enda hilisemas voores. Need parandused jagunevad kaheks. Üks tüüp on kolmanda positsiooni parandus, milles paranduse põhjustab partneripoolne vahepealne väärtõlgendus (*K: a=kas=te` seda ka oskate öelda kas (.)`välismaalasel ka **on võimalik saada**: = // Ö: =üks aasta. (0.5) üks`aasta **saate sellega sõita** välismaal. (.) // K: **ei, ma mõtlengi=et** vä-`välis`maalane. minu mees on välismaalane, /.../).* Teine tüüp on selline, milles väärtõlgendust ei ole (kolmanda voores parandus) (*H: /.../ aga ma=i=te mis se vibraatoriga **see** maksta võib. // (0.5) // V: vast issi`teab. (.) peab talle: `rääkima ma õhtu [`räägin.] // H: [vibraatori]ga **lihvi**).*).

Probleemlikale järgnev voores on esimene koht, kus partner saab algatada parandust eelkõneleja lausutu kohta (teise voores parandus).

Ümbertegemiseks on võimalik kasutada nelja operatsiooni. Suulises dialoogis valitsevad üksuse (põhiliselt sõna) asendamine teisega ja lisamine tagantjärele; kustutamine ja olevate üksuste ümberjärjestamine on haruldased (vt Hennoste 2008). Probleemidest lähtudes jagatakse ümbertegemine kaheks: veeparandus (hääldeviga, vale sõna vms) ning modifitseerimine (alternatiivi pakkumine), mille alla kuuluvad kõik muud muudatused.

Järgnevas analüüsin esmalt leitud eneseeparanduste tehnoloogiaid ja paranduslõikude vormi, seejärel parandamise suhtluslikke põhjusi ja eesmärgi. Materjal pärineb Tartu Ülikooli arvutikeele korpusest. See sisaldab 18–25-aastaste üliõpilaste ja õpilaste vestlusi, milles suhtlevad sõbrad, tuttavad ja sugulased. Analüüsiks valisin juhuslikult 50 vestlust, kokku 4575 saadetist. Selles materjalis leidsin 60 enda algatatud eneseeparandust.

2. Eneseperanduse tehnoloogiad ja vorm

Selles osas annan ülevaate leitud parandustehnoloogiatest ja vormivõtetest. Süstematiseerin tehnoloogiad selle järgi, a) millises positsioonis vea suhtes parandus tehakse, b) millist operatsiooni kasutatakse ja c) milliseid probleeme lahendatakse.

2.1. Sama voo parandus

Esimene võimalus oma teksti parandada on probleemlikuga samas voo. See jaguneb MSN-i suhtluses kaheks: samas saadetises ja järgmises saadetises tehtud parandus.

2.1.1. Parandus samas saadetises

Kirjutaja saab oma teksti parandada kirjutamise ajal. Partner näeb selliseid parandusi ainult siis, kui kirjutaja teeb oma parandusprotsessi eksplitsiitseks, nagu suulises kõnes. Antud kolleksioonis oli üks selline näide.

Näites (1) pakub Martin Jaanale oma abi kolimisel (saadetised 1–3).² Neljandas saadetises algatab Jaana uue alateema, esitades küsimuse Martini sõidukursuste kohta. See on vormistatud vastust pakkuva küsimusena, milles küsija osutab, et tema teada on õige vastus jaatav (*eks?*).³ Martin vastabki jaatavalt ning seletab, et pakutud abiaeg on seotud kursuste ajaga. Sellele reageerib Jaana saadetisega *taipasin jahhh*, mis osutab, et tema esitatud küsimus ja selle vorm olid seotud varasema teadmisega kursusest.

(1)

01. 20:47:24 Martin ma saaksin tulla 12-st kuueni
02. 20:47:29 Martin eks sa mõtle ja anna teada
03. 20:48:08 Jaana okei, teeme nii, et annan siis teada, kui vaja on.
04. 20:48:35 Jaana aga sul ikka **need** kursused käivad **selle sõidu teemalised** eks?
05. 20:48:46 Martin jep
06. 20:49:05 Martin sellepärast ma saangi ainult kuueni olla
07. 20:49:13 Jaana taipasin jahhh

Neljandas saadetises teeb Jaana eneseperanduse. Ta kasutab alguses täiendit *need sõna kursused* ees. Peale verbi *käivad* on küsimus ja kõnevoor potentsiaalselt valmis, aga Jaana jätkab atribuudiga *sõiduteemalised*, pakkudes sõnale *need* alternatiivi. Atribuudi ees on asesõna *selle* ainsuse genitiivis. See näitab, et *selle* on kohatäitjaks sõnale *sõidu*. Kohatäitja on ase- või määr sõna, mis osaleb ehitatavas lauses samas morfoloogilises vormis ja samas süntaktilises rollis kui täistähenduslik leksikaalne üksus, mille kohta ta täidab (Hayashi, Yoon 2006: 490). Teda kasutatakse suulises kõnes mh kui sõnaotsimise probleemi esmast lahendust (Hennoste 2001: 192–193, vt ka Keevallik 2010).

² Näidetes on esitatud saadetise aeg, saatja ja saadetud tekst. Nimed ja muud identifitseerimist võimaldavad andmed on muudetud. Teksti kirjutusviis on säilitatud. Parandused on paksus kirjas.

³ Küsimuste tüpoloogiatega kohta vt Hennoste jt (2009a: 341–346).

Kuidas interpreteerida sõna *need*? Üks võimalus on tõlgendada seda kui määratust osutavat asesõna, mis asendatakse täistähendusliku sõnaga parema arusaamise loomise huvides (*need* > *sõiduteemalised*). Sellist asendust tõlgendatakse täpsustusjätkuna. Aga *selle* uue täiendi ees osutab, et Jaanal on probleem sõna täiendosa leidmisega, mis viitab võimalusele, et *need* võib olla ka kohatäitja sõnale *sõiduteemalised*. Siis oleks tegemist kahekordse kohatäitmisega, milles alguses kasutatakse kohatäitjat kogu täiendi, hiljem teist kohatäitjat täiendosa tarvis. Sellisel juhul lahendab kirjutaja paranduse abil oma sõnaotsimise probleeme.

2.1.2. Parandus järgmises saadetises

Teine sama voozu paranduse võimalus esineb siis, kui kõneleja on jaganud voozu mitmeks saadetiseks ja parandab järgmises saadetises eelmise viga keset poolleli-olevat voozu. Selliseid näiteid oli kaks (vt ka näide 6).

Näites (2) kõneldakse kavandatavast matkast. Eve küsib, kes poistest tulevad. Rein vastab ja vahetab seejärel teemat (saadetis 1: *ni*).

(2)

- | | | | |
|-----|----------|------|-----------------------------|
| 01. | 22:58:15 | Rein | nii |
| 02. | 22:58:21 | Rein | ma tulen arvatavasti tartu |
| 03. | 22:58:23 | Rein | nüüd |
| 04. | 22:58:24 | Rein | homme |
| 05. | 22:58:27 | Rein | 15 rongiga |
| 06. | 22:58:51 | Eve | sa tuled ka homme? riina ka |
| 07. | 22:58:58 | Rein | mhm |
| 08. | 22:59:00 | Rein | kõik tuleme |
| 09. | 22:59:34 | Eve | kes kõik? |
| 10. | 22:59:40 | Rein | mina ja mart |
| 11. | 22:59:43 | Rein | ei |
| 12. | 22:59:45 | Rein | mina ja see |
| 13. | 22:59:46 | Rein | fakk |
| 14. | 22:59:48 | Rein | riina |
| 15. | 22:59:50 | Rein | ja jüri lubas ka |
| 16. | 22:59:55 | Rein | et vb tuleb |

Reinu info peale, et ta tuleb homme Tartusse (saadetised 2–5), küsib Eve saadud info üle (partneriparanduste kohta vt Liddicoat 2011: 226–233). Rein vastab jaa-tavalt ja lisab: *kõik tuleme* (saadetis 8). Sõna *kõik* osutab rohkemale kui ühele või kahele inimesele. Selle peale algatab Eve üheksandas saadetises täpsustava paranduse (*kes kõik*). Rein vastab *mina ja mart*. See ei ole lõpetatud voor, sest siin on nimetatud kaks inimest. Järgneb veaparanduse algusmarker *ei* (Hennoste 2006). Seejärel saadab Rein *mina ja see*, kus kohatäitja *see* osutab, et viga oli nimes *mart*. Sellele järgneb nime otsimise faas (*fakk*) ja nimi *riina*, kelle tulek oli juba teada. Sellega on eneseparandus lõpetatud. Järgmises saadetises lisatakse kolmas nimi *jüri* ning sellega on ka partneri algatatud parandus valmis. Järgneb veel sihitlause-line järellaiendus (ingl *increment*, vt Ford jt 2002: 16–25) *et vb tuleb*, mille abil vähendatakse *jüri* osas info kindlust.

2.2. Enda järgmise vooru parandus

Järgmine eneseperanduse koht on probleemi tekitaja enda järgmine voor. Need parandused jagunevad kolmanda positsiooni paranduseks, mis reageerib partneri väärtõlgendusele ja kolmanda vooru paranduseks, kus valetõlgendust ei ole.

2.2.1. Kolmanda positsiooni parandus

Materjali hulgas oli üks näide, milles kaks kolmanda positsiooni eneseperandust (3).

(3)

- | | | | |
|-----|----------|-------|---|
| 01. | 17:18:03 | Saima | kuule kas külls läheb itta ka :D see nädalavahetus :D |
| 02. | 17:18:08 | Leenu | jaa homme |
| 03. | 17:18:25 | Saima | :) |
| 04. | 17:18:36 | Saima | ma litsalt küsisin :D |
| 05. | 17:18:42 | Leenu | misasja:D |
| 06. | 17:19:00 | Saima | oi:D |
| 07. | 17:19:05 | Saima | lihtsalt :D |
| 08. | 17:19:12 | Leenu | miks sa seda ütled |
| 09. | 17:20:08 | Saima | ei mul oli kirjaviga seal ja siis ma mõtlesin et kirjutan vigase sõna õigesti |
| 10. | 17:20:16 | Leenu | ei enne seda |
| 11. | 17:20:35 | Saima | mm, mõtlesin et tulen siis koos teiega äkki tartu pühapäeval |

Esimeses saadetises on Saima esitanud küsimuse ja saanud vastuse, millele ta reageerib kolmandas saadetises emotikoniga :). Sellega on voor antud Leenule. 11 sekundit hiljem tuleb aga Saimalt uus saadetus, mis kommenteerib tema eelnevat küsimust. Seda on raske tõlgendada. Üks võimalus on, et Saima proovib vaikivat partnerit rääkima saada. Teine võimalik põhjus on partnerite ühistes taustateadmistes, mis põhjendavad sellist vabandavat seletust. Kolmas võimalus on, et Saima ootas lisainfot, mille soovi ta nüüd tagasi võtab.

Leenu reageerib küsimusega *misasja* (saadetus 5). See on partneriparanduse algatus, mida võib tõlgendada mitmeti (nt osutus kogu saadetise probleemsele või puuduvale lauseliikmele: *misasja {sa küsisid}*).⁴ Saima parandab kirjavea (*litsalt > lihtsalt*), näidates, et ta tõlgendas parandusalgatus kui viitamist sellele. Leenult aga järgneb uus parandusalgatus *miks sa seda ütled*. See on tõlgendatav kahel viisil. Üks võimalus on, et probleem on Saima eelnevas saadetises. Saima nii ka tõlgendab ja seletab oma eelnevat tegevust, viies sellega enda jaoks taas paranduse läbi (saadetus 9).

Järgneb Leenu kümnes saadetus *ei enne seda*. See osutab esiteks, et Saima üheksandas saadetises tehtud tõlgendus Leenu kaheksanda saadetise kohta oli väär. Seega on kümnes saadetus ise kolmanda positsiooni eneseperandus. Kus on siis probleemlikas? On usutav, et see pole sõnas *oi*, vaid neljandas voorus. Seega osutab kümnes saadetus ka seda, et *miks sa seda ütled* oli suunatud neljandale saadetisele ning oli selliselt ise kolmanda positsiooni parandus, mis lähtus sellest, et Saima tõlgendas oma kuuendas ja seitsmendas saadetises vääralt tema viendat

⁴ Looksulgudes on võimalikud lisandused ja parandused konstrueeritud näidetes.

saadetist. Järgneb lause, mis ei paista uurija pilgule paranduse läbiviimisena, vaid pigem Saima seletusena oma esimesele küsimusele, kuid sellega antud järjend (sekvents) lõpeb.

2.2.2. Kolmanda voo parandus

Kolmanda voo paranduse kohta oli kolm sõna- ja üks kirjavahemärgi asenduse näide. Kolm neist on veeparandused ja üks modifitseerimine (*puhata* > *lõõgastuda on parem sõna*). Näites (4) on juttu tüdrukute kahlusest klubis.

- (4)
- | | | | |
|-----|----------|-------|--|
| 01. | 20:11:43 | Jüri | Sa ei ole klubis käsi käiku lasnud +:D |
| 02. | 20:11:57 | Jaana | Nojaaa, kuidas kellelegi, sulle ma usun jah, mul ikka arvamus Tarjast muutus_D |
| 03. | 20:12:09 | Jaana | Muidugi mitte, pole nagu põhjust olnud. |
| 04. | 20:12:16 | Jüri | ?* |
| 05. | 20:12:29 | Jüri | Ok suht kahtlane küsimus tgl. |
| 06. | 20:12:29 | Jüri | Sa ei pea vastama :D |
| 07. | 20:13:03 | Jaana | Ma vastasin juba, sest mis seal varjata, ma ei ole seda tüüpi, et kohe kallale, ma närvi lähen, siis lähen lihtsalt minema:D:D |

Jüri esimene saadetis on väitlause vormis küsimus, mille lõpus on plussmärk, mis on arvutiklaviatuurile küsimärgiga sama klahvi all. Teine saadetis on Jaana reaktsioon Jüri eelnevale jutule, mis näitest puudub, ning kolmandas saadetises vastab ta küsimusele. Järgneb Jüri parandusena märgistatud küsimärk (?*). Parandus järgneb partneri vahepealsele vooale, mis aga ei ole analüüsiv väärtõlgenduseks. Vastupidi, see on korrektne vastus esitatud küsimusele. Jaana ei reageeri küsimärgile (ta on vastanud) ja Jüri kommenteerib oma küsimust ning lisab info, et teine ei pea vastama (saadetised 5–6). Küsimus on naabruspaari esiliige, mis ootab järelliiget-vastust või signaali, et seda ei tule.⁵ Jüri kuuenda saadetisega on järelliikme puudumise probleem lahendatud. Selle järel ütleb Jaana, et ta juba vastas. Jüri viies ja kuues saadetis näitavad, et ta kasutas lõpumärgi asendust mitte paranduseks, vaid tema arvates puuduva järelliikme meeldetuletamiseks.

Probleemi põhjustajaks on ilmselt lõhutud vooaerandus (ingl *disrupted turn adjacency*) ehk olukord, milles sisult seotud vooaerandused (saadetised 1 ja 3) ei paikne füüsiliselt kõrvuti (Herring 1999). Selle põhjuseks on asjaolu, et arvutisüsteem paigutab saadetised saamise aja järgi, arvestamata nende sisulisi seoseid. Kui mõni suhtleja (või arvuti) on aeglane või reageerib järjest mitmele eelnevale vooale, võib see viia naabruspaari liikmete eraldamiseni.

2.3. Teise voo parandus

Nagu nägime, tehti ainult vähesed parandused MSN-is positsioonides, mida kasutab suuline dialoog. Põhiosa parandustest on probleemlikat sisaldavale vooale järjend teise voo parandused. Siin kasutatakse nii asendusi kui lisamisi.

⁵ Naabruspaar (ingl *adjacency pair*) koosneb kahest eri esitaja vooalust (suhtlustegevusest), milles esiliige (ingl *first pair part*) nõuab järelliiget (ingl *second pair part*) (Liddicoat 2011: 138–158).

2.3.1. Asendused

Suurem osa asendustest (32 näidet) on sõnade trükkimisvigade parandused: kirjutaja on löönud klahve vales järjekorras, kahte klahvi korraga, kaks korda sama klahvi, õige tähe kõrval asuvat klahvi või jätnud klahvi vahele.⁶ Näites (5) kirjutab Kartul vigaselt voozu viimase sõna ja parandab üsna arusaamatu tulemuse järgmises voozus (*asuale* > *ausalt*).

- (5)
01. 15:06:37 Kartul teglt ma ilastasin arvuti ekraani peale, aga sealt tagasi ei pörkand, lohises mööda ekraani alla krt
 02. 15:07:45 Raamat iuu, saoled röve **asuale**
 03. 15:07:47 Raamat **ausalt**.
 04. 15:08:11 Kartul sina alustasin
 05. 15:08:15 Kartul **d***
 06. 15:08:25 Raamat no okkk, minu viga. aga mis tegelt tegid ä

Grammatiliste vigade parandamise kohta on neli näidet. Näites (5) kasutab Kartul teise isiku pronoomeni järel esimese isiku verbivormi ja asendab selle järgmises saadetises korrektse lõpuga (*alustasin* > *d*). *Sina alustasin* on tõlgendatav kahel viisil (*{mina} alustasin / sina {alustasid}*), mis on antud järjendis erinevad suhtlustegevused.⁷ *Sina alustasid* on õigustav põhjendus oma eelmise saadetise sisule, aga ühtlasi süüdistus, mis ootab reageerimist. *Mina alustasin* on pigem vabandus. Seega on paranduse eesmärgiks likvideerida voozu ambivalentsus ja näidata, millist reaktsiooni oodatakse partnerilt. Raamat vabandab, osutades selle abil ka oma arusaamist parandusest.

Lisaks oli materjalis kaks referentsiprobleemide lahendamist (teises referentsiparanduse näites muudeti ajamääratlust: *teisipäeviti* > *esmaspäeviti*). Näide (6) on vestlus kingituste tegemisest ja veidi lõõpiva tooniga, eriti Antsu poolelt.

- (6)
01. 12:58:18 Rita aga ma pole ka rahaboss kahjuks...
 02. 12:58:34 Ants no siis pead midagi mitte rahalist tegema :D
 03. 12:59:30 Rita jah, sul ju sünnipäev tulemas tegelt
 04. 12:59:43 Ants **ja naistepäev ju kah**
 05. 12:59:59 Ants oot fraseerin **edasi**
 06. 13:00:02 Ants **uuesti** st :D
 07. 13:00:06 Rita :D
 08. 13:00:06 Ants **sul on tulemas naistepäev**

Neljanda saadetisega annab Ants kõnevooru Ritale, aga võtab selle seejärel tagasi. Tema viies saadetis algab partikliga *oot*, mille abil suulises suhtluses peatatakse käsilolev tegevus ja tuuakse enamasti ajutiselt sisse uus tegevus. See võib olla teema mingi aspekti süvendamise märk, paranduse või seletuse algus (Keevallik 2003: 135–138). Sõna *edasi* osutab, et Ants kavatses süvendada neljandas saadetises esitatud mõtet. Järgneb aga eneseparandus *edasi* > *uuesti*, mis osutab, et viies saadetis oli planeeritud (või vahepeal ümber planeeritud) neljanda saadetise ümbertegemise algatusena (*oot fraseerin {uuesti}*). Seega on tegemist kahekordse

⁶ Trükkimisvea piir pole päris are. Puuduv täht võib olla ka sõna lühendamise (*tsu*) ja mitmekordne täht teadlik kordamine emotsioonide väljendamiseks (*tsauu*). Kõik kahtlased juhtumid on analüüsitud mittevigadeks.

⁷ Suhtlustegevus (dialoogiakt) on tegu, mida kõneleja või kirjutaja keele abil suheldes teeb (küsimine, tervitamine jms).

eneseparandusega: veaparandus *edasi* > *uuesti* ja modifitseerimine *ja naistepäev ju kah* > *sul on tulemas naistepäev*. Esimene on voorujärgses positsioonis, teine aga voo ru sees.

Näidete hulgas oli ka üks voo ru lõpus oleva emotikoni asendamine (*D* > *:D*) ja üks kirjavahemärgi parandus (+ > ?).

(7)

- | | | | |
|-----|----------|-------|---|
| 01. | 20:38:19 | Jüri | Õnneks ma suitsu ei tee, ok too õhtu tegin, erand kinnitab ju reeglit :P. |
| 02. | 20:39:08 | Jaana | No kinnitab jah, väljaarvatud minu puhul:D |
| 03. | 20:39:32 | Jüri | Sa vist ei teinud suitsu eksole :). Et minu arust suht raske elada ni et keegi teeb kõrval suitsu + |
| 04. | 20:39:33 | Jüri | ? |
| 05. | 20:39:51 | Jüri | Vesikat ka mitte ? |

Näites (7) esitab Jüri kolmandas saadetises vastust pakkuva küsimuse (*ekssole*) ning selle järel lause, mille ta lõpetab plussmärgiga, asendades selle järgmises saadetises küsimärgiga. Tegu on näpuvea parandamisega, sest pluss on küsimärgiga samal klahvil.

Suhtlustegevuse seisukohast on see lause raskelt interpreteeritav. Igatahes muudab küsimärk selle ühemõtteliselt küsimuseks, mis ootab partnerilt reaktsiooni. Kui seda pikka aega ei tule, saadab Jüri uue, täpsustava küsimuse, mis seostub tema eelneva küsimusega. See jätab mulje, et küsimärk võib olla mõeldud ka kogu voo ru jaoks.

Haruharva tegi kirjutaja parandust läbi viies uue vea, mille ta mõnikord ka parandas (näide 9). Aga oli üks näide, milles korrektne vorm parandati valeks. Näites (8) oleva suhtluse algatas Jüri kell seitse. Jaana ei soovi eriti suhelda, sest õpib samal ajal järgmise päeva eksamiks. Vestlus kestab siiski pikalt. Üks teema lõpeb kell 20:42:36. Järgneb minut ja 19 sekundit vaikust ja siis küsib Jüri, kuidas Jaanal õppimisega läheb. Jaana vastab, et üldse ei lähe ning ta just mõtles ennast välja logida. Näide algab selle järel oleva infoga, et Jaana saatis teistele eemaleminemise teate (saadetised 1–2). Nii on suheldasoovimatuse info mõeldud kaudselt ka Jürile.

(8)

- | | | | |
|-----|----------|-------|--|
| 01. | 20:47:57 | Jaana | icc, mkmm, mina ei tea, mis täna viga neil on, et muli-
sema kukuvad.
just siis kui ma õppima peaks. |
| 02. | 20:48:12 | Jaana | no nüüd nad enam vast ei räägi, sest ma ütlesin,
et lähen eemaleicc |
| 03. | 20:48:58 | Jüri | Hahaha :D |
| 04. | 20:49:25 | Jüri | Oota aga mille peale Sul see homme eksam siis tuleb? |
| 05. | 20:49:35 | Jüri | homme * |
| 06. | 20:50:44 | Jaana | Eesti keele foneetika ja fonoloogia, ehk siis maa-
keeli völdete määramine (mille töö ma keskas ühe
sain), silbitamine, transkribeerimine jne. |

Jaana saadetise järel on 46 sekundit vaikust. Siis saadab Jüri naeru *hahaha*, mis annab kõnejärje Jaanale. Too ei reageeri ja 27 sekundit hiljem ilmub Jüri uus saadetis. See algab partikliga *oota*, mis näitab soovi püsida teemas (Keevallik 2003: 135–138) ja selle järel küsimus. Ka nüüd ei reageeri partner kohe ja 10 sekundi

pärast tuleb Jüri eneseparandus (*homne* > *hommne*). Tegu on korrektse konsonantühendi parandamisega vigaseks, kusjuures varem pole Jüri kirjavigu teinud. Kuna Jüri eelmiste saadetiste eesmärgiks oli partner suhtlema saada, siis võib suure tõenäosusega järeldada, et ka parandusvõte täidab sama eesmärgi.

Paranduste seas oli kolm näidet, milles asendus tuleb ekraanile peale partneri vahepealset saadetist, kuid ei ole kirjutatud selle järel. Näites (6) järgnevad Antsu kuuendale saadetisele Rita emotikon ning Antsu eneseparandus *sul on tulemas naistepäev* (saadetised 7 ja 8), mõlemad 4 sekundit peale Antsu teksti. Siin on formaalselt tegemist kolmanda vooru parandusega, mis aga on tehtud samal ajal partneri reaktsiooniga ning on seega teise vooru parandus.

2.3.2. Lisamised

Ainult kaks näidet oli sõna lisamisest voorulõpu järgses saadetises.

(9)

- | | | | |
|-----|----------|-------|---|
| 01. | 15:11:55 | Jaana | ongii nohh, ma praegu kaalun seda vestlust suulises ära kasuatdaD: |
| 02. | 15:12:00 | Jaana | suulsies knes* |
| 03. | 15:12:02 | Jaana | kõnes* |
| 04. | 15:12:05 | Jaana | suulises* |
| 05. | 15:12:31 | Alma | aa saadan sulle ja enda karli vestluseeeee |

Näites (9) annab Jaana info selle kohta, et ta kavatseb käsil olevat vestlust kasutada suulise kõne kursuses, kasutades selle nimetamiseks sõna *suulises*. Parandust esitavas saadetises lisab ta täpsustava sõna *kõnes* ja lokaliseerib paranduse sõna *suulises* kordamise abil. Sealjuures teeb ta kirjutusvead ja parandab need järgmiste saadetistega ära. Seega teeb Jaana samas fraasis kolm parandust, jättes aga täiesti parandamata nt verbi ja emotikoni *kasuatdaD*:. Miks ta keskendub just selle fraasi parandamisele? Mõlemad on filoloogid ja sõna *suulises* tähendus ei tekita mõistmisraskusi. Oluline on siin see, et Jaana saadetis oli ka partnerile esitatud soov nõusoleku saamiseks ning keskne selle puhul on nimelt koht, kuhu vestlus antakse. Soovile viitab see, et vastav info partnerile üldse antakse ning sõna *kaalun*, mis osutab, et tegu pole lõpliku otsusega. Alma reageerib informeerituse muutumise partikliga *aa* ning ei vaidle vastu. Tema nõusolekut kinnitab lisaks pakutud vestlus.⁸

Neljal juhul lisati eelmise vooru viimasesse sõnasse puuduv täht või sõnaosa. Näites (10) täidab Sirje töölevõtmise CV-d ja küsib sõpradelt abi. Selle näites on kaks eneseparandust.

(10)

- | | | | |
|-----|----------|-------|--|
| 01. | 10:13:10 | Sirje | ma kirjutasin cv koolid ja kursused, ja siis oma tääkohad |
| 02. | 10:13:16 | Sirje | töökohad |
| 03. | 10:13:49 | Sirje | mis ma peale seda peaks kirjutama, kas oma iseloomustus |
| 04. | 10:13:52 | Sirje | e |
| 05. | 10:14:12 | Mare | tee cv keskuses oma cv |

⁸ Taustakirjeldusest selgub, et vestlus ka saadetakse.

Sirje esimene voor on väitlause, mille viimane sõna on trükkimisveaga. 6 sekundit hiljem tuleb parandus (*täähkohad > töökohad*). Konteksti arvestades see viga mõistmist ilmselt ei seganud. Aga Sirje esimene voor ei olnud mõeldud infona, sest seda partnerid ei vaja, vaid kaudse soovina saada infot, mida veel kirjutada. Seda kinnitab 33 sekundit hiljem tulnud otsene küsimus. Nii võib ka paranduse üheks rolliks olla oma soovi meeldetuletamine.

Kolmanda saadetise küsimus võib partneri seisukohast olla poolik lause ja seega pooleli voor (nt *kas oma iseloomustus / {tuleks ka lisada}*), mis paneb ootama järgmist saadetist. Aga tegu võib olla ka lõpetatud lausega/vooruga, mille viimases sõnas on viga. Seega on tegu ambivalentse saadetisega, mis jätab lahtiseks, kuidas reageerida. Järgmise saadetisega lisab Sirje puuduva *e*, mis partneri jaoks osutab lause ja küsimuse lõppu. Analüüs näitas, et kirjutaja poolelt vaadatuna olid kõik seda tüüpi probleemid trükkimisvead, põhjuseks liiga vara löödud ärasaatmisklahv.

Kui eelmiste lisamiste puhul muudeti olevaid üksusi, siis kolmes näites lisati juurde täiesti uus üksus, lõpumärk. Näites (11) kõnelevad kaks neidu alguses oma poisteprobleemidest ja seejärel veidi lõõpivas toonis sellest, et nad kavatsevad jääda vanatüdrukuteks ja tegelda mõnes üksildases kohas kudumisega. Sellega algab toodud näide.

(11)

- | | | | |
|-----|----------|-------|---|
| 01. | 12:49:57 | Alma | samas meil pole nagu muud vaja kui kudumisasju |
| 02. | 12:50:25 | Jaana | ongii nohh homme läheme, siis saan tartust poest materjale hordides osta või ostad ise ära? |
| 03. | 12:50:39 | Alma | ostame koos ehk |
| 04. | 12:50:40 | Alma | ? |
| 05. | 12:51:45 | Jaana | okk see parem idee jah, võtame erinevat värvi lõngad jne siis hea vahetada jne |

Alma teatab esimeses saadetises, et vaja on ainult kudumisasju. Teises saadetises pakub Jaana kaks varianti asjade ostmiseks. Esimene võimalus on, et ta ostab need ise ära ja teine võimalus on esitatud küsimuse vormis ettepanekuna, et Alma ostaks. Kolmandas saadetises saadab Alma omalt poolt ettepaneku *ostame ehk koos*. Sellel pole küsilause tunnuseid, kuid on ebakindlusele ja sellega nõusoleku ootamisele viitav sõna *ehk*. Sekund hiljem järgneb küsimärk, mis muudab ettepaneku vormilt küsimuseks.

Eesti suulises institutsionaalses suhtluses esitatakse soove nii väit- või käsklausetega (*öelge palun Parkeri kabineti number*) kui ka küsilausestega (*kas te saadaks takso Tulika kolm*). Viimaseid kasutatakse siis, kui soov on ebatavaline ja selle täitmine ei ole eeldatavasti garanteeritud (Hennoste jt 2009b: 124–126). Siin järgib Alma sama mustrit. Kuna Jaana on teinud oma ettepanekud, siis saab Alma järeldada, et ta eelistab neid ja seetõttu ei ole tema pakutava uue variandi sobivus garanteeritud. Seega on tegemist parandamisega sobitamaks saadetis vormilt antud järjendisse.

2.4. Paranduste vorm

Paranduse vormivariandid koonduvad kahte rühma.

Probleemiga samas vöorus tehtud vähesed parandused on vormistatud nagu suulises kõnes (näited 1 ja 2). Neis on eksplitsiitne lineaarne parandusprotsess ja kasutatakse ka sellele omaseid vahendeid (algatusmarker, kohatäitja jm). Näidete põhiosa andvad teise ja järgnevate vöorude parandused protsessi ei eksplitseeri. Nende puhul esitatakse peaaegu alati ainult paranduse tulemus, mis moodustab alati eraldi saadetise. Põhiliselt on omaette esitamine ka ainsaks markeriks, mis ütleb, et tegu on parandusega (näide 7). Mõnikord lisatakse paranduse märkijana asterisk (näide 4), harvemini lühend *st* (= see tähendab; näide 6).⁹ Markerid on peaaegu alati sõna järel. Üksikjuhtudel on lisatud mõni eelmarker, nagu *oi* näites (3).

Asendades korratakse peaaegu alati korrektselt kogu sõna (näide 5: *ausalt*), kahel juhul ainult väära sõnaosa (näide 5: *d**). Lisamisel lisatakse puuduv üksus, nt kirjavahemärk (näide 11), aga sõnaosa lisamisel kogu sõna ei korrata (näide 10). Sõna lisamisel kasutatakse harva ka lokaliseerivat raamimist (näide 9: *suulsies knes*). Seega on vea parandamine lisamisel ja asendamisel erinevalt vormistatud. Väga harva kommenteeritakse parandust (näide 3).

Liiga vähesed näited ei luba praegu analüüsida, millistel tingimustel põhitüübigist erinevaid variante kasutatakse.

3. Parandamise põhjused ja eesmärgid

Analüüsitud 50 suhtlusest oli vigu ja parandusi 41 näites. Parandamise põhjuste ja eesmärkide analüüsimiseks jagan probleemid kaheks: trükkimisvead ja muud vead. Trükkimisvea puhul lööb kirjutaja kogemata õige kõrval asuvat klahvi või kaht klahvi korraga, vahetab ära klahvide järjekorra või lööb liiga nõrgalt. Siin saab võrrelda parandamata ja parandatud vigu. Kokku oli 247 löömisviga, neist 42 parandatud (17%).¹⁰ Samas sõnas tehti parandus 38 juhul ja 4 näidet oli tähe või sõnaosa lisamisest järgmises vöorus. Muude paranduste puhul saab uurija analüüsida ainult olevaid muudatusi ning pole võimalik määrata midagi omalt poolt veaks. Siia kuulub 18 parandust: 5 lisamist, 11 asendust ning 2 üheselt määramatut muudatust. Teistpidi saame siin välja tuua 7 viga, 5 modifitseerimist ja 6 üheselt määramatut tüüpi.¹¹

Paranduse keskne roll suhtluses on vastastikuse mõistmise loomine ning parandus seostub seega suhtlusega. Siin saab teha vahet kahel variandil. Üks variant on referentsiga seotud probleemide lahendamine (vale referentsi asendamine, referentsi täpsustamine jms) ning teine suhtlustegevuse probleemide lahendamine (suhtlustegevuse muutmine, ambivalentseuse likvideerimine või suhtlustegevuse vormi kooskõlastamine suhtlusnormiga).

Ühe rühma moodustavad siin kirjutaja poolt leitud referentsivigade või ebatäpsuste parandused, nagu näites (9), kus täpsustatakse sõna referentsi modifitseerimise abil (*suulises* > *suulsies knes*). Teiseks muudetakse või täpsustatakse parandamisega suhtlustegevust, mis omakorda mõjutab partneri reaktsioone. Näites (11)

⁹ Vrd asterisk inglise keeles (Collister 2011).

¹⁰ Analoogiline pilt on ka USA materjalis, kus parandati 9% vigadest (Baron 2010).

¹¹ Grammatikavea piir ei ole äre, kuna MSN-i suhtluses ei kasutata normikeelt. Seega ei saa vigadena tõlgendada nt ebatavalisi käändelõppe (*Talisi*), *olema* ärajätmist, mis on tavaline noorte netikeele joon (*sa loengus või*) jms. Vigadena saaks tõlgendada ainult selliseid parandamata grammatilisi jooi, kus kirjakeeles ja arvutikeeles järgitakse sama normi (nt subjekti ja verbi isiku kooskõla). Selliseid parandamata vigu ei olnud.

muudeti tegevuse väljendusvorm suhtlusnormile sobivaks (*ostame ehk koos > ?*). Näites (5) likvideeriti isikulõpu ja subjekti isiku erinevusest tingitud voo ru tegevuslik ambivalentlus, mis jättis lahtiseks oodatava reaktsiooni (*sina alustasin > d*). Suhtlustegevuse muutmine ja sisuline parandus võivad esineda ka koos. Näites (6) parandatakse referentsi (*edasi > uuesti*), mille abil osutatakse ühtlasi, et järgnev tegevus ei ole eelneva voo ru süvendamine, vaid selle parandamine.

Tabel 1. Trükkimisvead ja nende parandused¹²

Vea tüüp	Vigade hulk	Parandatud viga	Parandatud vigade %
vale voo ru piir	4	4	100
vale täht või märk	46	19	35
vale tähejärg	33	3	15
puuduv täht või märk	66	9	15
üleliigne täht	49	3	8
vale sõnapiir	49	2	4

Teiseks vaatan trükkimisvigu. Tabelist 1 näeme, et erinevat tüüpi trükkimisvigu parandatakse väga erineva sagedusega. Vale voo ru piir (*iseloormustus > e*) parandatakse alati, valet tähte (*p > o*) parandatakse palju, üleliigset tähte (*swiin > siin*) ja vale sõnapiiri (*kil eka > kile*) harva, vale tähejärg ja puuduv täht (*suulsies knes > suulises kõnes*) jäävad kahe vahele. Kuna trükkimisviga võib põhjustada suhtluses häireid, siis võib oletada, et kirjutajate arvates tekitavad eri tüüpi vead probleeme eri määral.

Analüüs näitab siiski muud. Kolmes näites tõi löömisviga kaasa grammatilise muutuse, mis muutis lause sisult teiseks (*pilk ka > pilt ka*), kolm viga seostuvad suhtlustegevuse probleemidega (näited 7 ja 10). Ülejäänud juhtudel ei ole uurija pilgu jaoks võimalike suhtlushäirete ja paranduste vahel üheseid seoseid. Nt tähejärjeveaga sõna *suulsies* (näide 9) on partnerile arusaadav, kuna see oli eelmises voo rus õigesti kirjutatud, aga *asuale* (näide 5) võib küll arusaamatu olla.

Ka on parandaja suhtumine trükkimisvigadesse ja muudesse probleemidesse erinev. Osa inimesi parandas referentsi- ja suhtlustegevuse viga, kuid ei parandanud trükkimisvigu. Nt viies tekstis olid kõik trükkimisvead parandamata ja samas esinesid seal muude vigade parandused. Kolmes tekstis oli ainult üks ja sealjuures parandatud viga, mille taga oli suhtlusprobleem. Seega jääb osa trükkimisvigade parandusi ja mitteparandusi suhtlusliku seletuseta. Millega võib seda seletada?

Markman (2006: 108) on toonud välja, et ühise probleemi lahendamiseks asutatud jututoas vähenes paranduste hulk aja jooksul. Ta jäeldab sellest, et aja jooksul partnerid harjusid üksteise erisustega ning parandus kui vastastikuse mõistmise loomise võte polnud enam nii oluline. Siinses materjalis suhtlevad ammused tuttavad, mis lubab oletada, et nad on üksteise keelekasutusega harjunud.

Osa parandusi on ilmselt seotud suhtleajate arusaamadega korrektsest või lubatavast keelekasutusest. Analüüs näitas, et üheksas suhtluses polnud ühtegi viga, neljas olid kõik vead parandatud ning 16 suhtluses olid ainult parandamata vead. 21 suhtluses oli mõlemat varianti. Seega ei paranda osa inimesi üldse viga, osa aga ei tee viga või parandavad need enne saatmist alati vaikimisi ära. Väga harva saab

¹² Tabelist puudub kaks sõna, milles tehti korruga kaks viga (*asuale, näutelsk*).

korrekse keele taotlus või ideaal nähtavaks ka eksplitsiitse paranduse kaudu. Ühes näites parandati grammatiline viga, mis mõistmist ilmselt ei häiri ega muuda ka suhtlustegevust (*millega Sa suvel plaanid teha > tegelda*). Selle paranduse tegi kirjutaja, kes järgis tugevalt normikeelt, kasutades mh ka sõbraga suheldes viisakussuurtähte (*Sa*). Teises näites aga kritiseeris kirjutaja ise peale parandust oma veategemist (*issand ma ei oska enam trükkida :D*).

Teiseks ilmnes tugev seos parandamise ja vea asukoha vahel. Nii on voo ru viimase sõna puhul parandatud ja parandamata trükkimisvigade suhe 29 : 26, aga eespool asuvate sõnade puhul 13 : 179. Seega parandatakse voo ru lõpu trükkimisvigadest 53% ja eespoolsetest 7%, teistpidi moodustavad parandatud lõpuvead 67% parandatud vigadest. Analüüsid es suhtlusi, milles oli nii parandatud kui parandamata vigu, tuli välja, et üheksas näites olid parandatud kõik voo ru lõpus olevad ve ad ja parandamata kõik eespool olevad ve ad. Viiel juhul oli parandatud osa sisemisi ja osa lõpuvigu ja parandamata vigade hulgas olid sinult sisemised ve ad. Seitsmes näites oli pilt süsteemitu. Seega on osa inimeste puhul seos vea asukoha ja parandamise vahel väga tugev. Nad parandavad kõik või peaaegu kõik voo ru/saadetise lõpus tehtud ve ad ning ei paranda üldse või parandavad harva eespoolseid vigu.

Selle põhjus võib olla psühholoogiline (nt tähelepanu fookus). Voo ru lõpu vea puhul tehakse parandus vahetult vea järel, varasemate vigade puhul viivitusega. Ka suulises suhtluses algatatakse põhiosa parandustest vahetult probleemi järel ja üle 90% parandustest tehakse samas lauses. Teine võimalik põhjus tuleneb asjaolust, et voo ru lõpus oli selgelt enam ka muid parandusi. Nii parandati teises voo rus 4 varasemat ja 8 eelmise voo ru lõpus olevat viga (67%). See näitab, et üks võimalik seletus on juhuslikus vale klahvi vajutamises. Kui kirjutaja vajutab järgmise tegevuse klahvi (nt viimase tähe asemel sõnavahet) liiga vara voo ru sees, siis saab ta vea parandada ilma, et lugeja seda näeks. Kui aga varem vajutatud klahviks on ärasaatmisklahv, siis jõuab viga ka lugejani. Lisakinnituse annavad sellele seletusele näited, milles saatmisklahvi on vajutatud viimase sõna keskel (näide 10).

4. Parandusmehhanismi kasutus väljaspool parandust

Nii suulise suhtluse kui jututubade analüüsid on näidanud, et parandusvõtteid kasutatakse ka muudel eesmärkidel kui parandamine (Schegloff 1997, Schönfield, Golato 2003). Analüüsisin teise ja kolmanda voo ru eneseparandusi ka sellest seisukohast.

Teise voo ru eneseparandus koosneb peaaegu alati ühest sõnast. Inimese trükkimiskiirus on individuaalne ja oleneb eri teguritest. Inglise keele andmetel lööb keskmine kirjutaja teksti tootes 19 arvestuslikku sõna (= 5 tähte / lööki) minutis.¹³ Seega kestab ühe sõna kirjutamine keskmiselt 3 sekundit. Arvestades seda, et aeg kahe saadetise vahel sisaldab vea märkamise, parandamise, saatmise ja serveriaja, võime suure tõenäosusega oletada, et 0–3 sekundit hiljem tulnud parandus on alustatud vahetult eelmise saadetise lähetamise järel. Pikem vahe võib olla tingitud pikemast serveriajast, aga vähemalt osa hilisematest parandustest võib olla tehtud peale pausi.

¹³ http://en.wikipedia.org/wiki/Typing_speed (25.09.2011).

Tabel 2. Probleemvooru ja eneseparandusvooru ajavahe

Ajavahe sekundites	0	1	2	3	4	5	6	7	8	9	10	Pikem	Teadmata
Teise vooru paranduste hulk (trükkimisvead + muud vead)	1	5	15	11	3	4	1	–	1	1	1	1	7

Tabel 2 näitab, et 59% teada ajavahega parandustest saabub 2–3 sekundi pärast ja 73% kuni 3 sekundit peale probleemvooru. See on tüüpilise trükkimistemperaaga hästi kooskõlas. Kui oletada, et vähemalt osa hilisematest parandustest on tehtud peale pausi, siis tõuseb küsimus, miks tuleb kirjutaja vooru juurde tagasi? Vastuse leidmiseks võrdlesin 0–3 ja 4–10 sekundi pärast tulnud parandusi. Analüüs näitas, et siin on oluline piir nelja ja viie sekundi vahel.

Viie sekundi järel parandatakse teises voorus kolm trükkimisviga, tehakse kolm lisamist ja üks õige vormi valeksparandamine. 2/3 probleemallikatest paikneb vooru lõpus, kolmandik eespool. Kõigis näidetes oli parandatav viga selline, mis konteksti ja ühisteadmiste abiga ei tohiks tekitada mõistmisprobleeme (näide 9: *suulises*, näide 10: *täähohad*) ja ühel juhul parandati korrektne vorm vigaseks (näide 8: *homme*). See tõstab küsimuse, millised tegurid mõjutasid nende paranduste tegemist?

Vaatame uuesti näidet (8). Siin teatab Jaana, et määratles ennast teiste jaoks eemalviibijaks. Selle järel tekib pikk vaikus, mille järel Jüri saadab alguses naeru ja siis küsimuse, et saada partnerit suhtlema. Kui küsimusele vastust ei tule, saadab ta eneseparanduse, parandades õige vormi valeks. Kuna naabruspaari esiliikmele mittereageerimine on suhtlusnormi rikkumine, siis on parandus kõigepealt märgu- anne, et partner ignoreerib suhtlusnormi. Samas võib suure tõenäosusega väita, et sellele lisaks ärgitab parandus huvitatut partnerit suhtlema, kusjuures selle kasutajaks on inimene, kes on suhtlusest huvitatud, erinevalt partnerist, kes peab õppima.¹⁴

Analüüs näitab, et analoogiline muster kordub teistes pausiga teise vooru parandustes, kusjuures seitse parandatavat saadetist on naabruspaari esiliikmed. Näites (9) täpsustab ja parandab Jaana mitu korda kesket märksõna kaudselt vormistatud soovis, mille vastusest ta on huvitatud. Näites (10) parandab Sirje kaudselt esitatud soovi saada infot. Kui peale parandust vastust ei tule, saadab ta otsese küsimuse, kuna ta on infot huvitatud. Mõlemal juhul on tegu kaudselt vormistatud suhtlus- tegevusega, kuid autor ei kasuta parandust selleks, et muuta kaudset soovi otseseks ning sellega võimalikku ambivalentsust likvideerida. Alles siis, kui peale parandust vastust pikka aega ei tule, esitab Sirje näites (10) otsese küsimuse.

Eelnev tekitab küsimuse, mida tehakse kolmanda vooru parandustega, mis on põhimõtteliselt samuti pikema pausi järel saadetud parandused. Ainult kahes neist olid olemas sekundid, mis näitavad eri saadetiste täpset saabumisaega. Analüüs näitas, et ka nendes oli parandus sisuliselt ja tegevuslikult mittevajalik ning kasutusel suhtlema ärgitamise või järelliikme ootuse meenutamise rollis (näide 4).

Seega iseloomustab pausijärgset parandusvõtete kasutust neli joont: 1) parandatakse enamasti naabruspaari esiliiget; 2) saatja ja partneri huvi suhtlemise või info vastu ei ole sümmeetriline: saatja on enam huvitatud kui partner; 3) parandus ei ole vajalik mõistmise tagamiseks; 4) parandusmehhanismi kasutatakse naabruspaari

¹⁴ Partneri suhtlema ärgitamine on MSN-i suhtluses oluline teema, sest MSN-is olemine on enamasti vaid üks suhtleja mitmest paralleelsest tegevusest, mis on tihti nende jaoks tagaplaanil. Ärgitamiseks kasutatakse erinevaid vahendeid, eriti neid, mis dialoogi suhtlusnormide järgi eeldavad kindlamat reageerimist, nagu erinevat tüüpi naabruspaari esiliikmeid.

oodatava järelliikme meenutamiseks või suhtluse hoidmiseks, parandamine kui tegevus on tagaplaanil ja/või mittevajalik.

5. Võrdlusi ja üldistusi

Analüüs näitas, et MSN-i dialoogis kasutatakse samu eneseparanduse tehnoloogiad nagu suulises dialoogis. Samas erineb MSN-i parandus suulisest dialoogist põhimõtteliselt, kuna selles valitseb lõpetatud vooru järel saadetud teise vooru enese-parandus. See positsioon on suulise dialoogi normide järgi reserveeritud partnerile. Tulemus erineb ka teisest reaalarajas toimuvast netivestluse tüübist, jututubadest, milles osaleb korraga palju üksteisele võõraid inimesi. Sealseid parandusi uurinud Schönfield ja Golato (2003: 257–258) leidsid vaid üksikuid teise vooru enese-parandusi (2 näidet 57-st). Markmani (2006: 110) analüüsitud väheste suhtlejatega jututoas oli üle poole parandustest teise vooru enese-parandused. See näitab, et olulisim mõjur paranduse asukohale on suhtlejate hulk. Palju suhtlejaid ei võimalda kasutada järgmist saadetist vabalt, sest pole selge, kuhu see probleemvooruhuhtes satub ja suhtlejad arvestavad seda enese-paranduste tegemisel. Samas ei leidnud kumbki neist sama vooru parandusi. Selle põhjuseks võib olla nii vähene näidete hulk kui teistsugune vooruehituse strateegia.

Tehnoloogialt moodustavad MSN-i parandused kaks sarja. Teise (ja ka kolmanda) vooru paranduste puhul esitatakse ainult paranduse tulemus ehk parandatud sõna/fraas, mis moodustab alati omaette saadetise/vooruhu, kuhu mõnikord lisanduvad markerid (asterisk, *st*). See erineb suulise kõne mudelist, sest parandamise protsess on neis peaaegu alati varjatud. Eksplitseeritud protsessist on vaid üksikud näited. Sel juhul on parandus ehitatud samamoodi nagu suulises kõnes. Huvitav on, et seda tüüpi parandused olid tehtud vooruhu sees. Praegu ei luba liiga vähesed näited selle piiri kohta suuremaid üldistusi teha.

Suurem osa näiteid on trükkimisvigade parandamised asenduse abil. Lisamisi ja muud tüüpi vigade (referentsivead, grammatilised või suhtlustegevuse vead) parandusi esineb harva. See tulemus erineb osalt suulisest kõnest, kus kesksed on nimelt referentsivigade parandused (Hennoste 2008). Ka kasutatakse MSN-is parandusvõtteid, mida suuline kõne ei kasuta ega võimalda (lõpumärgi asendamine või lisamine, puuduvate tähtede lisamine uues saadetises).

Teiseks analüüsisin vigade parandamise põhjusi ja eesmäärke. Analüüs näitas, et parandatakse ainult või eelistatavalt vigu, mis võivad tekitada mõistmis- või suhtlusprobleeme (väär või ebatäpne referents, ambivalentne või väär suhtlustegevus, vormistuse suhtlusnormi mittejärgimine). Osa trükkimisvigade parandusi ja mitteparandusi ei ole aga suhtluslikult seletatavad. Osalt on nende põhjuseks kirjutajate arusaamad lubatavast/korrektsest keelest, osalt aga on need seotud probleem sõna asukohaga. Nimelt on põhiosa parandatud trükkimisvigu vooruhu viimases sõnas. Selle põhjused võivad olla psühholoogilised. Aga kuna ka muudest parandustest puudutab valdav osa voorulõpusõnu, siis võib põhjuseks olla ka enneaegne ärasaatmisklahvi vajutamine.

Samas järgnevad parandused enamasti vahetult probleemüksusele, kuigi järgmises saadetises. See on analoogiline suulisele suhtlusele, kus samuti viiakse parandus läbi võimalikult kiiresti peale probleemkohta.

Kolmandaks näitas analüüs, et eneseparanduste rollid erinevad vastavalt sellele, kas nad on tehtud probleemvooru järel ilma pausita või hiljem. Pausita parandused on analoogilised suulises suhtluses kasutatavate vooruvahetusala parandustega, kuigi paiknevad “füüsiliselt” järgmises voorus. Peale pausi tehtud parandused ei ole aga mõistmise või suhtlustegevuse parandamiseks vajalikud. Võime öelda, et neil juhtudel kasutatakse parandusmehhanismi ainult või eeskätt mitteparanduslikel eesmärkidel, naabruspaari oodatava järellikme meenutamiseks või suhtluse hoidmiseks.

Schönfield ja Golato (2003: 266–268) on näidanud, et ka jututubades algatavad suhtlejad järellikmeid taotlemaid parandusi juhul, kui nende naabruspaari esiliikmeid ignoreeritakse (ingl *sequential repair*). Aga ainult ühes näites korrati selleks kogu vooru ning kahes sõnastati voor ringi. Enamasti topikaliseeriti mittereageerimine. Selleks kasutati üldisi küsimusi, mis kontrollivad partneri kättesaadavust (*mis juhtus?*), nn virtuaalseid žeste (*otsi!*), mitmendal mittereageerimise korral ‘kaebusi’ (ingl *complaints: kas sul pole minu jaoks aega?*). Suulises suhtluses kasutatakse parandusväliste suhtluslike eesmärkide tarvis partneriparanduse algatamise vahendeid, kusjuures neid lausungeid parandusena tõlgendada ei saa (Schegloff 1997: 505).

Samas on MSN-i sellised parandusekasutused positsioonilt ja rollilt sarnased suulise suhtluse laiendustega (ingl *increments*), mida kasutatakse tihti peale pausi, kui partnerilt ei ole tulnud eeldatud reaktsiooni (Ford jt 2002). Aga paranduse vahendeid laienduseks minu teada suulises suhtluses ei kasutata.

Viidatud kirjandus

- Anderson, Jeffrey F.; Beard, Fred K.; Walther, Joseph B. 2010. Turn-taking and the local management of conversation in a highly simultaneous computer-mediated communication system. – *Language@Internet*, 7, article 7. <http://www.languageatinternet.org/articles/2010/2804> (11.12.2011).
- Baron, Naomi S. 2010. Discourse structures in instant messaging: The case of utterance breaks. – *Language@Internet*, 7, article 4. <http://www.languageatinternet.org/articles/2010/2651> (11.12.2011).
- Collister, Lauren Brittany 2011. *-repair in online discourse. – *Journal of Pragmatics*, 43 (3), 918–921. <http://dx.doi.org/10.1016/j.pragma.2010.09.025>
- Ford, Cecilia E.; Fox, Barbara A.; Thompson, Sandra A. 2002. Constituency and the grammar of turn increments. – C. Ford, B. Fox, S. Thompson (Eds.). *The Language of Turn and Sequence*. Oxford: Oxford University Press, 14–38.
- Garcia, Angela Cora; Jacobs, Jennifer Baker 1999. The eyes of the beholder: Understanding the turn-taking system in quasi-synchronous computer-mediated communication. – *Research on Language and Social Interaction*, 32 (4), 337–367. http://dx.doi.org/10.1207/S15327973rls3204_2
- Hayashi, Makoto; Yoon, Kyung-Eun 2006. A cross-linguistic exploration of demonstratives in interaction. – *Studies in Language*, 30 (3), 485–540. <http://dx.doi.org/10.1075/sl.30.3.02hay>
- Hennoste, Tiit 2001. Sissejuhatus suulisesse eesti keelde IX. Lausung suulises kõnes 4. – *Akadeemia*, 1, 179–206.
- Hennoste, Tiit 2006. Self-repair initiators in Estonian conversation compared with Finnish. Paper presented in ICCA-06, Helsinki, May 10–14, 2006. MS.

- Hennoste, Tiit 2008. Sõnakatkestusega algavate lausesisese reformuleerimise tüübid eestikeelses vestluses. – H. Metslang, R. Grünthal (Toim.). Lähivertailuja, 18. Uralica Helsingiensia, 1. Helsinki: Helsingin yliopisto, Suomalais-Ugrilainen Seura, 31–53.
- Hennoste, Tiit 2009. Allkeeled. Revised. – Jaan Lahe, Tiit Pädam (Koost.). Minu Issand ja minu Jumal! Pühendusteos Dr. theol. Toomas Pauli 70. sünnipäevaks. EELK Usuteaduse Instituudi toimetised, XX. Tallinn: EELK Usuteaduse Instituut, 234–243.
- Hennoste, Tiit; Gerassimenko, Olga; Kasterpalu, Riina; Koit, Mare; Rääbis, Andriela; Strandson, Krista 2009a. Küsimused eestikeelses infodialoogis I. Küsimuste vorm. – Keel ja Kirjandus, 5, 341–359.
- Hennoste, Tiit; Gerassimenko, Olga; Kasterpalu, Riina; Koit, Mare; Rääbis, Andriela; Strandson, Krista 2009b. Suulise eesti keele korpus ja inimese suhtlus arvutiga. – Eesti Rakenduslingvistika Ühingu aastaraamat, 5, 111–130. <http://dx.doi.org/10.5128/ERYa5.07>
- Herring, Susan C. 1999. Interactional coherence in CMC. – Journal of Computer-Mediated Communication, 4 (4). <http://jcmc.indiana.edu/vol4/issue4/herring.html> (11.12.2011)
- Herring Susan C. 2007. A faceted classification scheme for computer-mediated discourse. – Language@Internet, 4, article 1. <http://www.languageatinternet.org/articles/2007/761> (11.12.2011).
- Keevallik, Leelo 2003. From Interaction to Grammar. Estonian Finite Verb Forms in Conversation. Studia Uralica Upsaliensia, 34. Uppsala.
- Keevallik, Leelo 2010. The interactional profile of a placeholder: The Estonian demonstrative see. – N. Amiridze, B. H. Davis, M. Maclagan (Eds.). Fillers, Pauses and Placeholders. Amsterdam/Philadelphia: John Benjamins, 139–172.
- Liddicoat, Anthony J. 2011. An Introduction to Conversation Analysis. Second Edition. London, New York: Continuum.
- Markman, Kristine M. 2006. Computer-Mediated Conversation: The Organization of Talk in Chat-Based Virtual Team Meetings. PhD Dissertation. University of Texas at Austin. MS.
- Quan-Haase, Anabel 2009. Text-based conversations over instant messaging. – Ch. Rowe, E. L. Wyss (Eds.). Language and New Media. Linguistic, Cultural, and Technological Evolutions. Cresskill, NJ: Hampton Press, 33–54.
- Schegloff, Emanuel A. 1992. Repair after next turn: The last structurally provided defense of intersubjectivity in conversation. – American Journal of Sociology, 97 (5), 1295–1345. <http://dx.doi.org/10.1086/229903>
- Schegloff, Emanuel A. 1996. Third turn repair. – G. R. Guy, C. Feagin, D. Schiffrin, J. Baugh (Eds.). Toward a Social Science of Language. Papers in Honor of William Labov, Vol. 2. Social Interaction and Discourse Structures. Amsterdam: John Benjamins, 31–40.
- Schegloff, Emanuel A. 1997. Practices and actions: Boundary cases of other-initiated repair. – Discourse Processes, 23 (3), 499–545. <http://dx.doi.org/10.1080/01638539709545001>
- Schegloff, Emanuel A.; Jefferson, Gail; Sacks, Harvey 1977. The preference for self-correction in the organization of repair in conversation. – Language, 53 (2), 361–382. <http://dx.doi.org/10.2307/413107>
- Schönfeldt, Juliane; Golato, Andrea 2003. Repair in chats: A conversation analytic approach. – Research on Language and Social Interaction, 36 (3), 241–284. http://dx.doi.org/10.1207/S15327973RLSI3603_02

Tiit Hennoste (Tartu Ülikool) on eesti filoloog, kelle uurimisalaks on suulise ja internetisuhtluse analüüs.

tiit.hennoste@ut.ee

SELF-INITIATED SELF-REPAIR IN ESTONIAN INSTANT MESSAGING INTERACTION

Tiit Hennoste

University of Tartu

The present article is the first analysis of self-initiated self-repair (SISR) in Estonian Instant Messaging (IM) interaction. The method used is conversation analysis. Data come from the Corpus of Internet Language of the University of Tartu. The collection consists of 58 self-repairs which were found from 50 randomly selected dialogues (4575 transmission units (TU)).

First I describe the SISR technologies used. The analysis shows that speakers use all the technologies which are used in Spoken Estonian. The most frequent type of SISR is made in the next TU after a completed turn. This repair is formally in the position which is reserved for repair-initiation by the recipient in spoken interaction. The result is different from SISR in chat-room interaction where the next turn SISR is much less frequently used.

Two different structural models of SISR are used. Second-turn, third-position and third-turn repairs consist of the repair outcome only, which always form separate TUs. Same turn SISRs are formulated similarly to spoken interaction (repair initiators, explicit process).

Most of the repairs are replacements made to correct typographical errors, but there are also some grammatical and referential problems solved by replacement or by insertion. At the same time most typos are not corrected at all. The situation is similar in IM interaction in other cultures.

In the next section I analyse the reasons and the aims for making repairs. Participants prefer to correct errors which cause referential and/or interactional problems (wrong reference, ambivalent or wrong action etc). Some corrections of typos cannot be explained interactionally. Some of them could be explained by the sender's ideal of using good language, some by the place of the problem source. Errors in the last word of the problem-turn are corrected eight times more often than errors in the previous (intermediate) words. Writers prefer to initiate repair immediately after the problem occurs.

In the final section I show that there is a difference between next-turn repairs which are received 0–4 seconds later (probably made without a gap after the problem-turn), and repairs received 5–10 seconds later. In the latter group the repair mechanism is used not for SISR but to get the second pair part of an adjacency pair (questions, requests) and/or maintaining interaction in a situation where the sender of the SISR is more interested in interaction than the receiver. A similar role of the repair mechanism is found in spoken interaction and in chat-room interaction but the resources used are different. The use of repair here is similar to the use of increments in spoken interaction.

Keywords: conversation analysis, dialogue, computer mediated communication, technologies and role of self-initiated self-repair, Estonian

OSALAUSETE TUVASTAMINE EESTIKEELSE TEKSTIS KUI ISESEISEV ÜLESANNE

Heiki-Jaan Kaalep, Kadri Muischnek

Ülevaade. Artiklis esitatakse üks viis, kuidas eestikeelses tekstis automaatselt tuvastada osalauseid ja nendega võrdsustatud osa infiniit-tarinditest, kusjuures omaette üksusena eristatakse kiilud kui sellised üksused, mis katkestavad endast mõlemal pool asuvat sama osalause.

Kirjeldatav süsteem toetub kirjavahemärkidele, osalause piiril olevatele üksiksõnadele ja verbi finitsetele vormidele, olles põhijoontes kooskõlas EKK ja EKG II osalause-käsitlusega. Sõnade morfoloogiline analüüs ja ühestamine peavad olema tehtud, kuid süntaktilist analüüsi ei eeldata. Ehkki kirjeldatav algoritm on üsna lihtne, võimaldab ta osalausepiiri tuvastada küllalt hästi: kõigist võimalikest osalause ja kiilu piiridest tuvastas süsteem 95% (saak) ja kõigist väljapakutud piiridest olid õiged 96% (täpsus). Programmi abil on märgendatud osalused Tartu Ülikooli koondkorpuse veebiversioonis.*

Võtmesõnad: arvutilingvistika, süntaks, lause, osalause, infiniit-tarind, eesti keel

1. Sissejuhatus

Kirjaliku keelekasutuse pikkade ortograafiliste lausete tükeldamine osalauseteks (ingl *clause splitting*) on oluline samm paljudes loomuliku keele automaattöötuse valdkondades, mh automaatses süntaktilises analüüsis, püsiühendite tuvastamises, anafooride lahendamises, mitmekeelsete tekstide paralleelistamises, masintõlkes jne.

Näiteks eestikeelses tekstis ühendverbide tuvastamisel on info osalausepiiride kohta oluline, sest just osalause on optimaalne kontekst kandidaatpaaride moodustamiseks. Nii ei moodusta lauses *Vaatasin tehtu üle ja jalutasin minema* sõnavormid *üle* ja *minema* ühendverbi, kuigi nende vahel on tekstis ainult kaks sõna, erinevalt lausest *Eesti läheb kevadel samaaegselt oma lähiumbruse riikidega üle suvejale*, kus üksteisest viie tekstisõnaga eraldatud sõnavormid *läheb* ja *üle* moodustavad ühe

* Artikli valmimist on toetanud Euroopa Regionaalarengute Fond Eesti Arvutiteaduse Tippkeskuse kaudu ning Haridus- ja Teadusministeerium (sihtfinantseeritav teema sF0180078s08 "Loomulike keelte arvutitöötuse formalismide ja efektiivsete algoritmide väljatöötamine ning eesti keelele rakendamine"). Täname anonüümseid retsensente asjatundliku kriitika ning soovitude eest.

leksikaalse üksuse. Osalause piiri tuvastamine sellistes lausetes võimaldab kitsendada sõnavormidega *üle* ja *minema* potentsiaalselt kokkukuuluvate sõnade hulka.

Ka Kristel Uiboed (2010) nendib eesti murrete korpusest ühendverbide automaatse tuvastamise katseid kokku võttes, et üsna määrav oli lausestaja töö efektiivsus: mida korrektsemalt olid osalused eraldatud, seda adekvaatsem tuli kandidaatandmestik.

Käesolevas artiklis tutvustatakse selle probleemi ühte võimalikku lahendust: osalausestamise programmi, mille sisendiks on morfoloogiliselt ühestatud tekst ja mis lahendab osalausestamise ülesande ilma täieliku süntaktilise analüüsita.

Selle artiklis kirjeldatava osalausestamise programmiga on osalauseteks jagatud kõik Keeleveebi¹ kaudu kasutatavad tänapäeva eesti keele korpused, mis võimaldab kasutajal korpusest mitme keelendi koosinemisi otsides valida kontekstiks osalause.² Nimetatud korpustest pärinevad ka selles artiklis esitatud näited, mida vahel on küll lihtsustatud ja lühendatud, kontsentreerimaks tähelepanu käsitletavatele nähtustele.

Artikkel on üles ehitatud nii, et selle järgmises osas vaadeldakse osalause mõiste ja piiride käsitlemist eesti keele teaduslikes kirjeldustes, osas 3 määratletakse, mida kirjeldatava osalausestamise süsteemi seisukohalt käsitletakse osalausestamisena, osas 4 esitatakse eestikeelse teksti osalausepiiride leidmise algoritm, 5. osas analüüsitakse selle osalausestaja töö tulemusi ja tehtud vigu ning 6. osas võrreldakse kirjeldatavat programmi teiste samalaadsete süsteemidega.

2. Lause, osalause, kõrvallause, sekundaartarind, klaus

Selles osas vaadeldakse lühidalt lause, osalause, kõrvallause, infiniittarindi ning vähemkäibiva klasi mõiste käsitlemist eesti keele grammatikakirjelduses, loomaks mõistelist tausta artikli järgmistele osadele.

“Eesti keele käsiraamat” kirjeldab lauset kui keelelise suhtluse põhiüksust (EKK: 429), mis tüüpiliselt sisaldab finiiitset verbivormi ja ühte või mitut muud moodustajat, milleks võib olla ka osalause (EKK: 434). Osalause mõiste ühendab endas nii pea- kui kõrvallause: liitlause jaguneb osalauseteks. Tüüpilise osalause öeldisverb on finiiitne, erandina võib *da*-infinitiiv samuti toimida iseseisva öeldisverbina teatud tüüpi lausetes; ka iseseisvas lauses võib teatud tingimustel esineda *da*-infinitiivne öeldisverb, täpsemalt vt EKG II: 244–246.

“Eesti keele grammatika” (EKG II: 276) kirjeldab osalausest, eriti just põimlause osalausest lause ja sekundaartarindi vahepealse nähtusena. Eriti *da*-infinitiivse öeldisega kõrvallused sarnanevad infiniittarinditega mitme tunnuse poolest, millest olulisematena olgu nimetatud finiiitse öeldisverbi ja subjekti esinemise võimaluse puudumine.

Osalause kui ühte sündmust väljendava süntaktilise tervikuga piirneb sekundaartarind – sekundaarne sündmuse väljendamise vahend, millel on lausega teatud ühisjooni (EKG II: 232). Infiniitsete sekundaartarindite kohta kasutatakse eesti keeleteaduses ka nimetust lauselühend, mis “Eesti keele käsiraamatu” (EKK: 436) järgi on verbifraasi ja osalause vahepealne moodustaja, ja mis, hoolimata sellest, et

¹ www.keeveebi.ee (01.03.2012).

² Korpusepäringu vastuses on osalausepiirid nähtavad roheliste kandiliste sulgudena.

tema peasõnaks on käändeline verbivorm ja temas alust esineda ei saa, sarnaneb rohkem lause kui fraasiga.

Kõrvallused jagunevad EKG II järgi komplement-, adverbiaal- ja relatiivlauseteks, viimane laiendab tüüpiliselt nimisõna, erandjuhul ka pealauset kui tervikut. Kuna relatiivlause järgneb enamasti vahetult oma peasõnale, paikneb ta sageli oma pealause sees, tükeldades kirjas selle kaheks üksteisest eraldatud katkeks – nt lauses *See mees, kes meile vastu tuli, oli meie direktor* on üksteisest eraldatud lauseosad *see mees* ja *oli meie direktor*.

Lauses üldlaiendina esinev, ülejäänud lausega grammatiliselt seostamata lause on kiillause, nt *Uus Peetri katlamaja (asub Peetri pargi lähedal) läheb käiku jaanuaris*. Kiiluga lauset eristab põimlausest see, et kiildumisega ei kaasne grammatilisi muutusi ei kiilduvas lauses ega teda hõlmavas lauses, mõlemad säilitavad oma süntaktilise sõltumatus ja terviklikkuse. (EKG II: 102)

Uuemates käsitlustes koondab Mati Erelt kõik sündmust väljendavad struktuurid – nii osalused kui sekundaartarandid – klasi nimetuse alla (Erelt 2004, Erelt 2011). Selles artiklis kirjeldatav programm tuvastab teatud osa klausidest, nimelt osalused ja väikese osa sekundaartaranditest. Nii on klaus nende tuvastatavate üksuste jaoks liiga lai termin ja osalause jälle liiga kitsas. Edaspidi kasutatakse selles artiklis käsitletavate üksuste kohta siiski koondnimetust osalause, mõeldes selle all osalauseid ja allpool (osas 3) täpsemalt kirjeldatud tunnustele vastavaid infiniitseid klause.

3. Mida kirjeldatav süsteem käsitleb osalause ja kiiluna?

Kirjeldatav süsteem mõistab osalause või osalausega võrdsustatud klausina keeleüksust, millel on lihtlause või lausesarnase süntaktiliselt tervikliku üksuse struktuur. Osalausestamise eesmärgiks on, nagu öeldud, kirjalike tekstide pikkade ortograafiliste lausete tükeldamine “mõistlikeks juppideks” – väiksemateks süntaktilisteks terviküksusteks, mille süntaktiliseks keskmeks on tüüpjuhul finiiitne verb, ent võib olla ka infiniitne – seda nii *da*-infiniitvise öeldisverbiga osalauses kui teatud tüüpi infiniittarandites.

Osalausestena märgendab siinkirjeldatav süsteem osa *des-*, *mata-* ja *maks-* klausidest. Esiteks võrdsustatakse osalausega sellised põhilause järel paiknevad komaga eraldatud *des-*, *mata-* ja *maks-* klausid, mille peasõna on tarindi algul, ning teiseks lausealgulised ja komaga eraldatud *des-* ja *maks-* klausid, mille peasõna asub tarindi (ja ka kogu lause) algul. Põhjenduseks esiteks EKG II sedastus (lk 272), et sekundaartarindi eraldamine komadega oleneb tema lauselisuse astmest ja teiseks – komaga eraldatud tarindit on lihtsam automaatselt tuvastada.

Keelenorm suhtub erinevalt *des-* ja *mata-* tarinditesse ühelt poolt ning *maks-* tarindisse teiselt poolt. Kohustuslik on eraldada komaga põhilause järel paiknevad *des-* ja *mata-* tarandid, kui infiniitne verbivorm asub tarindi algul, ning soovitatav on eraldada komaga põhilause algul paiknevad pikemad *des-* ja *mata-* tarandid, mille peasõna on tarindi algul. *maks-* tarindite komakasutus ei ole normeeritud, soovi korral võib järgida *des-* ja *mata-* tarindite puhul kehtivaid reegleid.

Eelnevast järeldub, et eeldusel, et kirjakeeles järgitakse täpselt sekundaartarindite komastamise reegleid, ei sõltu osalausestamise tulemus mitte ainult

sekundaartarindi lauselisuse astmest, vaid ka kirjutaja suvast eraldada *maks*-klaus ning põhilause algul paiknev *des*- ja *mata*-klaus komaga või mitte.

Nagu on näidanud Ellen Uuspõld (2001), on tekstides sagedased grammatiseeruvad *des*- ja *mata*-tarindid, mis on muutumas infiniitset klausist kaassõnafraasiks ja mille komakasutus kõigub. Seejuures on keelekasutaja suhtumine kaassõnastuvatesse *des*- vs. *mata*-tarinditesse erinev: tarindeid, mille peasõnaks on kaassõnastuvad *des*-vormid *alates* ja *võrreldes*, tüüpiliselt komadega ei eraldata, kuid samuti kaassõnafraasiks muutumise teel olevaid tarindeid *mata*-vormidega *hoolimata* ja *vaatamata* kirjavahemärgistatakse suhteliselt tihti (Uuspõld 2001: 320). Sellest tulenevalt ei käsitleta meie osalausestaja väljundis osalausest lausealgulist *mata*-tarindit, mille põhisõna on tarindi algul, sest selles positsioonis kipuvad sageli esinema kaassõnastuvad *mata*-vormid *hoolimata*, *vaatamata* ja *rääkimata* ning komakasutus ei erista kaassõnafraasi infiniittarindist.

Osalausestaja praegune versioon ei erista osalauseid nende süntaktiliste omaduste alusel, s.t ei erista nt relatiiv-, adverbiaal- või komplementlauseid, kuid see võiks olla üks töö edasiarendusi, seda enam, et relatiivlause te äratundmine vormitunnuste põhjal on tegelikult suhteliselt lihtne, erinevalt komplement- ja adverbiaallause eristamisest (vt ka Ereht 2004: 402–404). Samuti ei seata osalauseid omavahel hierarhiasse (pealause, kõrvallause, kõrvallause funktsioon pealause suhtes).

Ent kirjeldatav süsteem märgendab eraldi nn kiilud, mida käesolevas artiklis (erinevalt EKG II ja EKK terminikasutusest) mõistetakse laiemalt kui kiillauset. Kiil on siinses käsitluses tervikuna teise süntaktilise üksuse sees paiknev süntaktiline üksus, s.t. sama süntaktiline struktuur algab enne kiilu ja jätkub pärast kiilu lõppu. Just see, et kiilu taga võib jätkuda eespool alanud osalause, eristab kiilu tavalisest osalausest, mille taga algab tingimata juba uus osalause. Kiiluks võib olla relatiivlause (1) või kiillause (2), aga ka fraas või üksiksõna (3).

- (1) Täna need noored eestlased, kel pole huvi kaasaegse maailma vastu kustunud, saavad tekstilise elamuse pigem ..
- (2) Oli väga ilus tüdruk, aga nüüd on (tasandab häält) vanaeit!!
- (3) See stiil tähendab ülisubjektiivset teksti üldjuhul räpastes (narkootikumid, seks, joomarlus, poliitika) teemadel.

Kiile püütakse eristada muudest osalausest esmajärjekorras seetõttu, et kiil lõhub lause terviklikkuse, kui ta aga ära tunda ja tinglikult eemaldada, muutub alles jääv lause palju kergemini analüüsitavaks. Kiilu alguse või lõpu märkimine pelgalt osalausepiirina jätkaks programmi ilma teadmises, et samas osalausest on kuskil (s.t teisel pool kiilu) veel sõnu, mis vaadeldavatega kokku kuuluvad, ja seega teeks programmi töö raskemaks.

Kiile on kerge eristada kirjavahemärgistusele tuginedes, kuna nad paigutatakse kirjas (üla)komade, mõttekriipsude või sulgude vahele.

Kiilude eristamine tavalisest osalausest, mis võimaldab kiilu ümbritsevat struktuuri kohelda tervikuna, on autorite arvates üks kirjeldatava süsteemi eelseid teiste omasarnaste ees.

4. Osalausestamise algoritmi kirjeldus

Algoritmi luues oletati, et lauseid konstrueerides kasutatakse teatud võtteid – kirjavahemärke, sidesõnu, sõnade järjekorda jms – süsteemi huvitavate klauside üksteisest eraldamiseks.

Liitlause loomisel kasutatakse teatud kindlaid malle või lausekonstruktsioone, mille tuvastamisel võib olla kindel, et lause on liitlauselise struktuuriga, olenemata sellest, kas mõned lausele muidu kohustuslikuks peetavad elemendid on igas lauseosas olemas või mitte. Näiteks lauses *Kui osta külmkapp, siis ikka selleks, et toitu säilitada* ei ole ühtegi finiitset verbi, aga on selgelt kolm iseseisvat osa ja neid signaliseerivad sõnad *kui, siis ja et*.

Oletati ka seda, et kirjavahemärkide panemise tava vastab küll suurel määral eesti keele grammatikates esitatud ortograafiareeglitele, kuid ei pruugi nendega päris täpselt kattuda.

Osalause leidmise puhul seisab programm tavaliselt vastamisi pika kirjaliku keele lausega, milles on rohkem kui üks sidesõna, sidend või osalause piiri tähistav kirjavahemärk ja seejuures pole teada, millised neist märkidest või sõnadest eristavad osalauseid, millised aga rinnastatud sõnu või fraase. Nii ongi algoritmi üldiseks ideeks, et osalauseid ja kiile tuvastatakse järk-järgult, alustades lihtsamatest ja kindlamatest juhtumitest. Loodetavasti muutub lause struktuur pärast iga etappi programmi jaoks lihtsamaks. Iga lause vaadatakse läbi mitu korda.

Sisendiks on morfoloogiliselt ühestatud tekst, s.t iga tekstisõna on varustatud informatsiooniga tema lemma ja grammatiliste kategooriate kohta. Morfoloogiline analüüs (s.h sõnastikust puuduvate sõnade analüüs) ja statistiline ühestamine tehti OÜ Filosoofi programmiga *etmrf*.

Kõigepealt märgitakse kiilud, siis tähistatakse osalause keskmeks sobivad verbid, seejärel märgitakse teatud kirjavahemärgid ja sõnad võimalike osalausepiiridena ja siis püütakse otsustada, millised neist võimalikest on kindlad osalausepiirid ja millised omakorda kindlasti osalauseid ei eralda.

Kõige lõpuks tähistatakse kiiludena veel mõned kindlalt tuvastatud osalused, et hilisemad tekstitötlusprogrammid saaksid paremini töötada; nt (4).

(4) Avastus, <kiilu algus> et asi on halb <kiilu lõpp>, tekitas hirmu.

Järgneb algoritmi detailsem kirjeldus.

1. Sulgudes olev tekst, ka ühesõnaline, märgitakse kui kiil. Sulud on kindel indikaator selle kohta, et nendes olev tekst ei ole süntaktiliselt ümbritseva teksti osa.
2. Tähistatakse osalause keskmeks sobivad verbid: verbi finiitset vormid; *nud*-partitsiibid, kui nad on kindlalt määratletavad kui mineviku (*on teinud*) või eituse (*ei teinud*) vormid ja seega ei esine atribuudi rollis (*kadunud auto*), *tud*-partitsiibid eituse vormidena (*ei tehtud*) ja *des-*, *mata-* ning *maks*-vormid, kui nad on vahetult koma järel ehk esimene sõna lauselühendis, mis on järelasendis. (Ka korrelaatsidendite *hoolimata sellest et ja vaatamata sellele et* puhul, kus ainuke koma on terve sidendi ees, ei tee *hoolimata ja vaatamata* ekslik tähistamine osalause keskmeks sobivana otseselt kahju, sest hiljem kasutatakse potentsiaalsete osalausepiiride säilitamiseks/eemaldamiseks ainult keskmeks sobiva verbi olemasolu

informatsiooni, ja antud juhul on kindel, et selline keskmeks sobiv verb samas osalauses ka leidub.) Ka lausealgulisi *des-* ja *maks-* vorme käsitletakse osalause keskmeks sobivatena.

Kõiki liitöeldise koosseisu kuuluvaid sõnavorme ei püütagi tuvastada, sest tähtis on ainult teada, kas potentsiaalses osalauses leidub tema keskmeks sobiv verbivorm või mitte, verbi liitvormi või perifrastilise verbivormi täpne koosseis pole selle ülesande seisukohalt oluline.

On selge, et vead verbi määratlemisel tingivad hiljem vigu ka osalauses-tamisel. Praeguse algoritmi puuduseks on asjaolu, et liitajavormide puhul õnnestub verbi *olema* vorm ja *nud*-partitsiip ühtekuuluvaks määrata ainult siis, kui nad asuvad lauses kõrvuti. Kui *olema*-vormi ja *nud*-partitsiibi vahel on sõnu, nt *oli Mart uskunud*, siis partitsiibivormi ei märgita öeldise osaks, aga kuna *olema* vorm on nagunii märgitud osalause keskmeks sobiva finiitse verbivormina, siis osalausestamiseks sellest piisab. Osalausestamises tekivad vead juhul, kui lauses on hiljem veel kord kasutatud täismineviku liitaega, kuid selliselt, et *olema*-verb on väljajäteline ning reaalselt esineb ainult *nud*-partitsiibi vorm. Siis on vaja otsustada, kas tegemist on atribuudina kasutatud partitsiibi või väljajäetlase öeldisega, selle otsustamiseks tuleb aga omakorda kontrollida, kas *nud*-vormi on samas lauses juba varem kasutatud koos verbi *olema* vormiga (s.t öeldisena). Lauses (5) määratakse *lõhkunud* võimalikuks öeldiseks, sest talle eelnevas lauseosas on *avanud* märgitud öeldise osana, lauses (6) jätab aga algoritm sõnavormi *avanud* öeldise osa märgendita, seetõttu jääb sellele ka *lõhkunud* ning osalause piir jääbki antud lauses lisamata. Lauses (5) on korrektselt tuvastatud osalausepiir märgitud püstkriipsuga.

(5) Ta oli avanud akna ja | lõhkunud ukse.

(6) Ta oli kiiresti avanud akna ja lõhkunud ukse.

Et selliseid *nud*-partitsiipide öeldisena mitte-äratundmisest tingitud vigu vähendada, on edaspidi kavas uurida, kas väljajäetlase öeldise kasutamine järgib mingeid sagedasi ja kindlaid malle, nt hüpoteetiline reegel “kui *nud*-partitsiip asub vahetult sidendi järel ja ta ei kuulu mõne omadussõnaga samasse loendisse (nt *õnnelik ja väsinud*), siis on ta tõenäoliselt öeldis”.

tud-partitsiibile ei rakendata *nud*-partitsiibi reegleid, põhjuseks asjaolu, et automaatselt ei suudeta ilma süntaksianalüüsita nagunii eristada impersonaali liitmineviku vorme seisundipassiivi vormidest. EKG II (lk 30–31) põhjal on lauses (7) *tud*-partitsiibi vormid impersonaali liitajavormi osad ja järelikult on tegu kahe osalausega, lauses (8) aga on tegu seisundipassiiviga – rinnastatud predikatiividega ühes osalauses. Ka EKK (lk 456–457) mõonab, et passiivi korral on partitsiip pigem adjektiiv ja verbi piirijuhtum kui üks neist. On veel teinegi põhjus *tud*-partitsiibi käsitlemiseks *nud*-partitsiibist erinevana: nimelt tundus esialgse korpusuuringu põhjal, et *tud*-partitsiipi kasutatakse atribuudina *nud*-partitsiibist sagedamini, s.t liiga palju on vigu põhjustavaid lauseid nagu (9).

(7) Siin on magatud ja õpitud.

(8) Praod olid vatti täis topitud ja paberiga kinni kleebitud.

(9) Palju on kollaseid värvitud pindu ja poleeritud alumiiniumi.

3. Märgitakse võimalikud osalauseste piiride kohad, milleks on kirjavahemärgid : , ; – . ? ! ja sidesõnad *ja, ning, ega, või*.
4. Otsese kõne algus (koolon, millele järgneb jutumärk) ja lõpp (punkt, hüüumärk, küsimärk või koma, millele järgneb jutumärk) märgendatakse kui kindlad osalausepiirid.
5. Kui jutumärkide paarsus on kindlaks tehtav ja osalause keskmeks sobiv verb asub nii jutumärkidest seespool kui ka väljas, siis on jutumärkide vahel tõenäoliselt tsitaat (s.t kiil) ja see märgitaksegi sellisena; muud jutumärgid ei tähista ei osalause ega kiilu piiri (nt “*Sõda ja rahu*”).
6. Kirjavahemärgid : ja ; märgitakse kui kindlad osalause piirid.
7. Koma või mõttekriips, millele järgneb vahetult kas *ja, ning, ega, või, et, kui, kus, kuhu, kust, sest, kuid, nagu, ehkki, siis, kuni, otsekui, justkui, kuna, kuidas* või *kas*, märgitakse kui kindel osalause piir.

Koma või mõttekriips, millele järgneb sõna *mis, kes, missugune, mil-line* või *see* mistahes käändes, märgitakse samuti kui kindel osalause piir.

Kui komale või mõttekriipsule järgneb mingi sõna, millele omakorda vahetult järgneb sõna *et (ilma et, nii et, ainult et jms)*, siis ka seal on kindel osalause piir.

Nagu näha, peetakse mõningaid kirjavahemärgi ja sõna kombinatsioone niivõrd usaldusväärseteks osalausepiiride signaalseerijateks, et ei kontrollita isegi osalause keskmesse sobiva verbi olemasolu. Põhjenduseks kaalutlus, et selline kindlal moel tähistatav koht lauses tähendab seda, et järgnevad sõnad selles lauseosas kindlasti ei kuulu eelnevatega sama tihedalt kokku kui need, mis asuvad enne piiri.

Seevastu koma või mõttekriips, millele järgneb vahetult sõna *aga* või *kuigi*, märgitakse kui kindel osalause piir ainult juhul, kui neile sõnadele omakorda enne järgmist võimalikku osalausepiiri järgneb osalause keskmeks sobiv verb. Sellega välditakse liigse piiri panemist näiteks kontekstides *aitab lastel, aga ka täiskasvanutel head tuju säilitada; jooksis vaikselt, aga kiiresti ja sisaldab vett, kuigi väheses koguses*.

Tuleb tunnistada, et asjaolu, et sõnade *kuid* ja *ehkki* puhul osalause keskmesse sobiva verbi olemasolu ei kontrollita, *aga* ja *kuigi* puhul aga kontrollitakse, näitab programmi ebajärjekindlust. Kas ja kuidas seda muuta, nii et osalausestaja muutuks paremaks, vajab edasist uurimist.

8. Vaadeldakse allesjäänud oletuslikke piire. Kui mõlemal pool piiri on osalause keskmeks sobiv verb, siis lihtsalt tähistatakse see piir kindlana.

Agas mida teha juhul, kui oletuslike piiride vahel on sellist verbi mittesisaldav fraas, aga mõlemal pool fraasi on osalause keskmeks sobiv verb, nt fraas *endine üliõpilane* lauses (10)? Kas osalause piiriks tuleks valida sellise fraasi ees või taga olev piir? Kõige üldisem reegel sellistel puhkudel on, et kui *ja*-le, *ning*-ile või komale järgneb vahetult osalause keskmeks sobiv verb, siis see ongi osalause piir, antud näite puhul seega jääb osalauseks *Mari on tegelikult Maiu, endine üliõpilane*; korrektselt tuvastatud osalausepiir on näites (10) märgitud püstkriipsuga.

- (10) Mari on tegelikult Maiu, endine üliõpilane, | elab Tammsaare teel ja | armastab rahvalaule laulda.

9. Eemaldatakse oletuslikud osalause piirid, kui need paistavad olevat rinnastatud elementide vahel. S.t kui mõlemal pool oletuslikku piiri on ühesuguses käändes sõna (nt *rohelistes, punastes ja sinistes pükstes*). Mõnikord saabub sellise eemaldamise tulemusena olukord, kus veel mõnede oletuslike osalausepiiride puhul on näha, et neist mõlemale poole jäävas osalause leidub selle keskmeks sobiv verb ja seega saabki antud piiri kindlana tähistada.

Lause (11) on kahe *ja* ja ühe *ning*-iga jagatud neljaks osaks, millest esimeses ja viimases on osalause keskmeks sobiv verb, kahes keskmises aga mitte. Kus peaks olema osalausepiir? Märganud, et *pikkade ja pingeliste* on samas käändes, võib nendevahelise *ja* oletuslike osalausepiiride hulgast kustutada; samal moel võib käituda *tujukuse ja isepäisuse* vahelise *ning*-iga. Järele jääb ainult üks oletuslik piir, millest mõlemal pool on osalause keskmeks sobiv verb, seega võib selle tähistada kindlana (näites püstkriipsuga).

(11) Ma ei nurisenud pikkade ja pingeliste tööpäevade üle ja | Vabariigi Presidendi paljukirjutud tujukus ning isepäisus ei häirinud mind.

10. Mõned osalused nimetatakse ümber kiiludeks; seda saab teha ainult sellistel juhtudel, kui on kindel, et osalause järel jätkub sama lause, mis algas osalause ees, nt (12). S.t ühel pool oletatavat kiilu peab osalause keskmeks sobiv verb kindlasti puuduma. Kiiluks sobivad kõrvallused, mis algavad sõnaga *kes, mis, missugune, milline* (mistahes käändes), *kus, kuhu, kust, millal* või *et*, s.t relatiivlused.

(12) Mees, <kiilu algus> kes tuli vastu <kiilu lõpp>, kandis musta kaabut.

5. Tulemused

Osalausestaja töö hindamiseks kontrolliti käsitsi u 16 000 sõna mahus osalausestatud tekste, mis jagunesid enam-vähem võrdselt ajakirjanduse, ilukirjanduse ja populaarteaduse (ajakiri Horisont) tekstiklasside vahel. Tulemuste hindamiseks arvutati osalausepiiride märgendamise täpsus (süsteemi poolt õigesti tuvastatud osalausepiiride arv jagatud kõigi süsteemi poolt tuvastatud osalausepiiride arvuga) ja saak (süsteemi poolt õigesti tuvastatud osalausepiiride arv jagatud tegelike osalausepiiride arvuga). Seejuures loeti veaks ka sellised juhtumid, kui kiilu alguse või lõpu asemel oli pandud lihtsalt osalausepiir. Üldiselt võib tulemusi nimetada korralikeks: üldine täpsus oli 94% ja saak 96%. Tulemused tekstiklassiti on esitatud tabelis 1.

Tabel 1. Osalausepiiride (sh kiilude) tuvastamise kvaliteet

Tekstiklass	Sõnu	Lauseid	Süsteemi poolt märgendatud osalausepiire	Neist õigeid	Süsteemi poolt märgendamata jäänud osalausepiire	Saak	Täpsus
ajakirjandus	5205	328	308	294	15	95%	95%
populaarteadus	5944	439	333	318	18	95%	95%
ilukirjandus	4926	286	440	427	21	95%	97%
KOKKU	16075	1053	1081	1039	54	95%	96%

Kui vaadata eraldi tavalise osalausepiiri ja kiilu tuvastamist, siis tavalise osalausepiiri tuvastamine on lihtsam ülesanne (saak 98% ja täpsus 96%) kui kiilu piiride märgendamine (saak 79% ja täpsus 100%). Kõige sagedasem viga osalausepiiride märgendamisel seisneski selles, et üks või mõlemad kiilu piirid märgendati tavaliste osalausepiiridena, näiteks on süsteem märgendanud lauses (13) esimese relatiivlause eraldava koma kindla osalausepiirina (näites märgitud püstkriipsuga), kuid teisele komale pole lisanud mingit piiri märgendit. Lauses (14) on mõlemad kiillauset *kui see kedagi huvitab* eraldavad komad märgendatud lihtsalt osalausepiiridena. Kuna kiilu märgendid on asendunud tavaliste osalausepiiride märgenditega, jääb kiilu poolt jagatud lause *Meie neljaliikmelise meeskonna taust oli selline* tervikuks ühendamata.

(13) Seega oli samm, | mille astus Eesti, palju pikem ja otsustavam.

(14) Meie neljaliikmelise meeskonna taust, | kui see kedagi huvitab, | oli selline ..

Vigu põhjustab ka vale morfoloogiline analüüs, näiteks lauses (15) oli sõnavorm *kaaluvad* analüüsitud ekslikult verbi *kaaluma* oleviku mitmuse kolmanda isiku vormiks, mitte oleviku partitsiibi mitmuse nimetavaliseks vormiks, ja see morfoloogilise analüüsi viga põhjustas vale osalausepiiri lisamise viimase koma juurde. Lauses (16) on liigne osalausepiir lisatud selletõttu, et sõnavorm *minevat* on ekslikult morfoloogiliselt tõlgendatud kui kaudse kõneviisi vorm, mitte kui *vat*-infinitiiv (nad on nimelt alati homonüümsed) ja selle tulemusena on meil lause, kus mõlemal pool sidesõna *ja* on verbi finitiivne vorm, ehk siis kindel tunnus, et tegemist on kahe osalausega.

(15) Seda arvestavad ärimehed, investorid, | krediidiiline kaaluvad pangad.

(16) Tüdruk paistis ikka rohkem ja | rohkem närvi minevat.

Muidugi põhjustavad viga ortograafiaeeglitele mittevastavad kirjavahemärgid, näiteks lauses (17).

(17) .. mida lähiaastatel saab kujundama Eesti, | kui kindla EL-iga ühineja maine.

Märgendaja jaoks on problemaatilised rinnastatud osalused, mille öeldisverbid on liitaegades ja teises osaluses on *olema*-verb välja jäetud (vt ka osa 4). Kuna lause (18) teises osaluses *luule triviaalsüsteem teisenenud* on *olema*-verbi finitiivne vorm välja jäetud ja osaluses on eksplitsiitselt olemas ainult *nud*-partitsiip *teisenenud*, jääb tuvastamata ka osalausepiir.

(18) Aeg oli temast mööda läinud, luule triviaalsüsteem teisenenud ..

Osalausepiiride automaatse määramise süsteemi töö hindamisel loeti veaks ainult sellised süsteemi poolt märkimata jäänud osalausepiirid, mida süsteem oma aluspõhimõtetest lähtuvalt oleks pidanud tuvastama. Osalausestaja väljundis on ka selliseid kohti, kuhu järjekindluse huvides oleks tulnud osalausepiir panna, aga kuna tekstis polnud selgeid pidepunkte, millele toetudes seda teha, siis, et vältida paratamatuid vigu, ei üritatudki neid piire märgendada. Järgnevalt kirjeldatakse näidetena mõningaid selliseid konstruktsioone.

Automaatselt on raske eristada rinnastatud öeldisverbe ja rinnastatud osalauseid (mis võivad ju samuti koosneda ainult öeldisverbist). EKG II (lk 214) järgi pole korduvate verbifraasidega lauseid enamasti võimalik üheselt rindlauseteks ja rinnastatud öeldistega koondlauseteks jagada; öeldiste endi rinnastusega on kindlalt tegemist ainult siis, kui rinnastuse funktsiooniks on eelkõige verbi tähenduse intensiivistamine, nt *Ma aina ootan ja ootan* või *Ta nuttis ja naeris ühekorraga*. Tähenduse intensiivistamist automaatselt tuvastada muidugi ei õnnestu ja nii pannakse osalauseste automaatsel tuvastamisel kahe finiiitse verbivormi vahele alati osalausepiir, nt (19–20).

- (19) Ühe treeninguga kadus paar kilo, | pärast sõi ja | jõi need aga jälle tagasi.
(20) Ainult ootan ja | ootan.

Osa infiniitsete klause, mh ka teatud tingimustele vastavad gerundiivitarindid peaksid olema märgendatud omaette osalausestena (vt osa 3). *des*-vormi minevikulisteks vasteteks on *olema*-verbist ja *nud*- või *tud*-partitsiibist koosnevad analüütilised verbivormid (nt *olles teinud*, *olles tehtud*). Tekstis jäetakse sõnavorm *olles* tavaliselt ära ja gerundiivi minevikuvorm esinebki *nud*- või *tud*-partitsiibi kujul. Mineviku partitsiipidel on aga ka palju teisi kasutusi, mh täiendina, nii et iga *nud*- või *tud*-vormi kuulutamine potentsiaalseks osalausestarnase süntaktilise struktuuri verbiliseks keskmeks põhjustaks palju vigu. (Siinkohal tuleb jällegi rõhutada, et liitajavormide puhul piisab üksi *olema*-verbi finiiitsetest vormist selleks, et osalause oleks finiiitne kese.) Nii jäävad gerundiivi minevikuvormis tarindid, kus sõnavorm *olles* on väljajääteline, omaette osalausestena märgendamata, nt (21).

- (21) Tajunud oma võõrdumist tegelikust elust, asub ta ..

6. Võrdlus eesti ning teiste keelte jaoks loodud sarnaste süsteemidega

Osalausepiiride tuvastamiseks kasutatakse nii reeglipõhiseid (nt Müürisep 2000, Puolakainen 2001), masinõppel põhinevaid (nt Mitkov jt 1999) kui ka hübriidseid (nt Orasan 2000, Puscasu 2004) süsteeme. Osalausesteks jagamise ülesande lahendamise algab alati osalausestena mõistetava üksuse täpse piiritlemisega; selle kohta täpsemalt nt (Orasan 2000: 130–131).

Käesolevas artiklis kirjeldatud süsteemile sarnaneb näiteks Georgiana Puscasu (2004) algselt rumeeniakeelsete tekstide jaoks loodud ja hiljem inglise keele jaoks kohandatud osalausestamise süsteem, mille eesmärgiks on küll ainult finiiitverbi sisaldavate osalauseste tuvastamine. Puscasu süsteem saab samuti sisendiks morfoloogiliselt ühestatud teksti ning kasutab osalausepiiride määramiseks põhiliselt kirjavahemärke, sidesõnu ning infot verbi finiiituse/infiniituse kohta. Puscasu süsteemi *F-mõõt* osalause alguse määramisel oli 95% rumeenia ja 92% inglise keele jaoks.

F-mõõt on täpsuse ja saagi harmooniline keskmine, mida arvutatakse järgmise valemi järgi:

$$F = \frac{2 \cdot \text{saak} \cdot \text{täpsus}}{\text{saak} + \text{täpsus}}$$

Käesolevas artiklis kirjeldatud eesti keele osalausestaja *F-mõõt* on järelikult 95%.

Sarnasest tööst võiks nimetada veel saksa keele korpuse DEREKO (DEutsches REferenzKORpus) märgendamise projekti, mille lõpparuandes (Dipper jt 2002) käsitletakse eraldi osalauseste märgendamise probleemi, kusjuures osalause on defineeritud kui üksus, millel on üks põhiverb, mis võib olla nii finiiitne kui infiniitne. DEREKO lingvistiliste märgendite süsteemis eristatakse kolme tüüpi osalauseid: üldisest osalause klassist eristatakse relatiivlauseid ja eraldi märgendatakse infiniitse verbiga osalauseid (originaalis: *clauses*). Osalauseste omavahelisi suhteid eksplitsiitselt ei esitata (s.t lausepuus pole näidatud, millist sõna/fraasi/osalausest mingi osalause laiendab).

Eesti keele automaattöötamise vallas on osalausestamisega varem tegeldud reegli-põhise morfoloogilise ühestamise (Puolakainen 2001) ning reeglipõhise süntaktilise analüüsi (Müürisep 2000) raames. Sarnaselt siin artiklis kirjeldatud osalausestajale kasutavad mõlemad nimetatud reeglipõhised süsteemid osalausepiiride määramisel pidepunktidena sidesõnu, kirjavahemärke ja finiiitseid verbivorme, kuid määratlevad osalauseid täpselt EKG II järgi, s.t ühtegi infiniititarindit ei käsitleta osalausestajana. Nii morfoloogilise ühestaja kui ka süntaksianalüsaatori koosseisus olevate osalausepiiride leidmise reeglite eesmärgiks ei ole kiilude (kiillauseste, relatiivlauseste, sulgudes lauseste või lauseosade jms) eristamine tavalisest osalausest. Selle tulemusena lõhub kiil alati oma pealause, s.t lauses *Mees, kes seal seisab, ostis minu naabermaja ära* jääb lausealguline moodustaja *mees* ühendamata ülejäänud lauseosaga *ostis minu naabermaja ära*.

Müürisep ega Puolakainen ei ole eraldi hinnanud osalausepiiride määramise tulemuslikkust, mistõttu ei ole võimalik võrrelda siinkirjeldatud süsteemi tulemusi nende omadega.

Müürisepa väljatöötatud eesti keele kitsenduste grammatika süntaksianalüsaatorit (Müürisep 2000) on kohandatud ka suulise keele (Müürisep jt 2006) ja murdetekstide (Lindström, Müürisep 2009) analüüsiks. Osalausepiiride määramisel suulise kõne transkriptsioonides ei saanud lähtuda kirjavahemärkidest, sest neid ei kasutata transkribeeritud tekstides normeeritud kirjakeele reeglite järgi, vaid lähtuda tuli transkriptsiooniks kasutatud intonatsioonimärkidest ja partiklitest. Suulise kõne automaatse süntaktilise analüüsi kvaliteedi hindamisel väidavadki autorid, et osalausepiiride valesti määramine põhjustas enim vigu (Müürisep jt 2006: 79). Ka murdetekstide – mis esindavad samuti suulist keelekasutust – analüüsil on tuvastamata osalausepiir kõige sagedasem veatüüp (Lindström, Müürisep 2009: 27).

7. Kokkuvõte

Paljude lingvistiliste probleemide (nii teoreetiliste kui praktiliste) lahendamiseks oleks hea, kui pikad laused oleks jagatud mõistlikeks üksusteks, mille piiridesse antud probleemi seisukohalt olulised lingvistilised nähtused teadaolevalt jäävad. Üheks selliseks mõistlikuks üksuseks on osalause ja osalausele lähedasem osa infiniitsetest klausidest. Asjaolu, et alati ei ole võimalik üheselt otsustada, kas ja mitmeks mõistlikuks osaks tuleks lause jagada, ei tähenda, et sellist jagamist üldse ei saaks teha või et automaatne jagamine oleks väga vearohke.

Selles artiklis esitati üks viis, kuidas automaatselt tuvastada osalauseid ja nendega samastatud osa infiniitartidest, nimelt teatud osa *des-*, *mata-* ning *maks-*artidest. Tuvastatavate konstruktsioonide hulga piiritlemine on paratamatult seotud ka tuvastamise viisiga: milleks defineerida kategooriat, mille esinemisi ei suudeta teistest eritleda?

Kirjeldatud süsteem toetub kirjavahemärkidele, osalause piiril olevatele üksik-sõnadele ja verbi finitsetele vormidele, olles põhijoontes kooskõlas EKK ja EKG II osalausekäsitlusega. Sõnade morfoloogiline analüüs ja ühestamine peavad olema tehtud, kuid süntaktilist analüüsi ei eeldata. Omaette üksusena eristatakse kiilud kui sellised üksused, mis katkestavad endast mõlemal pool asuvat sama osalause.

Ehkki kirjeldatav algoritm on üsna lihtne, võimaldab ta osalausepiire tuvastada küllalt hästi. Käitsi kontrollimise tulemusel selgus, et kõigist võimalikest osalause ja kiilu piiridest tuvastas süsteem 95% (saak) ja et kõigist väljapakutud piiridest olid õiged 96% (täpsus). Programmi abil on märgendatud osalused TÜ koondkorpuse veebiversioonis.

Viidatud kirjandus

- Dipper, Stefanie; Kermes, Hannah; König-Baumer, Esther; Lezius, Wolfhang; Müller, Frank H.; Tylman, Ule 2002. DEREKO. German Reference Corpus. Final Report (Part I). Internetidokument aadressil www.sfs.uni-tuebingen.de/dereko/DEREKOR-report.pdf (20.09.2011).
- EKG II = Erelt, Mati; Kasik, Reet; Metslang, Helle; Rajandi, Henno; Ross, Kristiina; Saari, Henn; Tael, Kaja; Vare, Silvi 1993. Eesti keele grammatika II. Süntaks. Tallinn: Eesti Teaduste Akadeemia Keele ja Kirjanduse Instituut.
- EKK = Erelt, Mati; Erelt, Tiiu; Ross, Kristiina 2007. Eesti keele käsiraamat. Tallinn: Eesti Keele Sihtasutus.
- Erelt, Mati 2004. Märkmeid eesti keele komplekslause kohta. – Keel ja Kirjandus, 6, 401–413.
- Erelt, Mati 2011. Tänapäeva eesti kirjakeele morfosüntaksi ja süntaksi uurimisest Tartu ülikoolis. – Emakeele Seltsi aastaraamat, 56 (2010), 37–62.
- Lindström, Liina; Müürisep, Kaili 2009. Parsing corpus of Estonian dialects. – E. Bick, K. Hagen, K. Müürisep, T. Trosterud (Eds.). Proceedings of the NODALIDA 2009 workshop Constraint Grammar and robust parsing, Odense, Denmark, 14.05.2009, NEALT Proceedings Series, 8. Tartu: Tartu University Library.
- Mitkov, Ruslan; Orasan, Constantin; Evans, Richard 1999. The importance of annotated corpora for NLP: The cases of anaphora resolution and clause splitting. – Proceeding of “Corpora and NLP: Reflecting on Methodology Workshop”, TALN’99. Veebidokument aadressil http://clg.wlv.ac.uk/papers/show_paper.php?ID=10 (16.09.2011).
- Müürisep, Kaili 2000. Eesti keele arvutigrammatika: süntaks. Dissertationes mathematicae Universitatis Tartuensis, 22. Tartu: TÜ kirjastus.
- Müürisep, Kaili; Nigol, Helen; Uibo, Heli 2006. Eesti suulise keele korpuse automaatne pindsüntaktiline analüüs. – M. Koit, R. Pajusalu, H. Õim (Toim.). Keel ja arvuti. Tartu: TÜ Kirjastus, 72–84.
- Orasan, Constantin 2000. A hybrid method for clause splitting in unrestricted English texts. – Proceedings of ACIDCA 2000, Corpora and Natural Language Processing, March 22–24, Monastir, Tunisia, 129–134.
- Puolakainen, Tiina 2001. Eesti keele arvutigrammatika: morfoloogiline ühestamine. Dissertationes mathematicae Universitatis Tartuensis, 27. Tartu: TÜ kirjastus.
- Puscasu, Georgiana 2004. A Multilingual method for clause splitting. – Proceedings of the 7th Annual Colloquium for the UK Special Interest Group for Computational Linguistics,

Birmingham, UK. Internetidokument aadressil <http://clg.wlv.ac.uk/papers/puscasu-04a.pdf> (04.09.2011).

Uiboaed, Kristel 2010. Statistilised meetodid murdekorpuse ühendverbide tuvastamisel. – Eesti Rakenduslingvistika Ühingu aastaraamat, 6, 307–326. <http://dx.doi.org/10.5128/ERYa6.19>

Uuspõld, Ellen 2001. *des-* ja *mata-*vormide kaassõnastumine ja eesti komareeglid. – Reet Kasik (Toim.). Keele kannul. Pühendusteos Mati Ereli 60. sünnipäevaks. TÜ eesti keele õppetooli toimetised, 17. Tartu: TÜ kirjastus, 306–321.

Heiki-Jaan Kaalepi (Tartu Ülikool) põhilised uurimisvaldkonnad on korpuslingvistika ja eesti keele morfoloogia.

heiki-jaan.kaalep@ut.ee

Kadri Muischneki (Tartu Ülikool) teaduslikeks huvialadeks on korpuslingvistika ning eesti keele (automaatne) morfosüntaktiline ja süntaktiline analüüs.

kadri.muischnek@ut.ee

CLAUSE SPLITTING AS A SEPARATE TASK (IN THE ANALYSIS OF ESTONIAN TEXTS)

Heiki-Jaan Kaalep, Kadri Muischnek

University of Tartu

Clause splitting is usually included in syntactic analysis, but it can be also regarded as a task in itself. For example, for a corpus query system it would be convenient to have some knowledge about clause boundaries in order to enable the user to retrieve co-occurrences of words or grammatical categories not in the usual “window” of three or four words, but in the whole clause. This paper presents a rule-based clause splitting system for written Estonian. The input text has to be morphologically disambiguated, but no parsing is required. So the presented system offers a convenient option for tasks that need information about clause boundaries. Two types of clauses are recognized: parentheses are treated separately from other clauses. By parenthesis we mean a clause that is situated in another clause dividing the latter into two separate strings. Recognizing parentheses enables us to treat the divided clause as a whole. The clause splitting system achieves 96% precision and 95% recall in unrestricted Estonian texts.

Keywords: computational linguistics, syntax, sentence, clause, non-finite clause, Estonian

LITHUANIAN AND RUSSIAN CHILD-DIRECTED SPEECH: WHY DO WE ASK YOUNG CHILDREN SO MANY QUESTIONS?

Victoria V. Kazakovskaya, Ingrida Balčiūnienė

Abstract. The main objective of the study was to identify and compare the variety and distribution of interrogatives in Lithuanian and Russian child-directed speech (CDS) from the perspective of the functional and structural characteristics of questions addressed to young children.

The analysis was based on the longitudinal data of two monolingual typically developing children, a Lithuanian girl (2;0–2;8) and a Russian boy (2;0–2;8). The transcribed corpus of conversations between the children and their parents was annotated for multipurpose automatic linguistic analysis, using tools of the program CHILDES (Child Language Data Exchange System). During the investigation, the functional and structural features of parental interrogatives were analysed.

After the analysis of conversations between the children and their parents, the following can be stated: in both Lithuanian and Russian CDS, interrogatives are more numerous than imperatives, statements and exclamations. A number of parental interrogatives are used (similar to natural adult conversation) as requests for information, clarifications of incomprehensible utterances or demonstrations of disagreement. Despite this, the majority of them appear to be used for a very specific purpose (e.g., negative evidence) and in specific forms (e.g., repetitions, reformulations or corrections), which would be inappropriate and/or redundant in a natural adult conversation.

With the exception of a few differences, interrogatives in both Lithuanian and Russian CDS are generally used for the same purpose, and their forms and structures seem to be similar or even identical. This leads us to the conclusion that a correlation can be identified between parental conversational strategy and the type of language, i.e., adults speaking typologically, culturally and geographically close languages demonstrate the same or similar strategy of conversation with their children.*

Keywords: language acquisition, CDS, conversational strategy, Russian, Lithuanian

* The research was carried out with the financial support of the RF President's Foundation (grant No. 3433.2010.6 "The School of Functional Grammar in St. Petersburg") and the Russian Academy of Sciences (the Programme of Fundamental Historical and Philological Studies "Semantic and Formal Redundancy of Texts in Modern Russian"). We are grateful to the two anonymous reviewers for their comments.

1. Introduction

1.1. Aims of the study

Child-directed speech (CDS, also termed *motherese*, or *baby-talk*) is considered one of the most important factors for native language acquisition (Tomasello 2003, Clark 2009). Despite the cultural and individual features of CDS (for example, see Schieffelin 1979, Ochs 1982, Heath 1983, Ochs, Schieffelin 1984, Rye 1986, Fernald et al. 1989, Choi 1997, de León 1988), one can observe that children growing up in a rich linguistic environment acquire better skills in their native language (e.g., Girolametto et al. 2002); and, in contrast, inadequate or pure input can cause delayed or even impaired language acquisition (e.g., Sachs, Johnson 1976, Sachs et al. 1981, Daukšytė 2005). Consequently, insufficient skills of spoken language (pure vocabulary, limited grammar structures etc.) delay acquisition of reading and writing; this causes learning difficulties, and leads to lower academic attainment. Thus CDS, as a primary basis of child language acquisition, needs comprehensive investigation, especially in those countries where child language generally is still a relatively new topic of (psycho-) linguistic research. The majority of the previous investigations of CDS were based on English data (e.g., Snow, Ferguson 1977, Gallaway, Richards 1994). During the last few decades, several comprehensive cross-linguistic studies have been carried out (Slobin 1997, Tulviste 2002), and a variety of other languages have been investigated (e.g., Rūke-Draviņa 1977, Wójcik 1994, Ceitlin 2001, Kazakovskaya 2004; 2008, 2011; Kamandulytė 2005; Kjellrun 2005; Balčiūnienė 2009, Kazakovskaya, Balčiūnienė (forthcoming)). However, more comprehensive investigations are still needed in order to be able to make an assessment of the possible connection between language type and CDS style.

Previous studies have indicated many similarities in parental conversational strategy between typologically closer languages, i.e. between Austrian-German and French, and between Lithuanian and Russian¹. Despite the fact that parents generally tend to react rather to the content than to the form of a child's previous utterance (see Kilani-Schoch et al. 2008, Kazakovskaya 2010, Kazakovskaya, Balčiūnienė (forthcoming)), one can observe a dominant didactic strategy in both Lithuanian and Russian CDS, and a high index of interrogative production. A number of parental interrogatives are used (similar to natural adult conversation) as requests for information, clarifications of incomprehensible utterances or demonstrations of disagreement, but the majority of the questions appear to be used for a very specific purposes (didactic, supporting language acquisition, e.g., negative evidence, see Hirsh-Pasek et al. (1984), Demetras et al. (1986), Bohannon, Stanowitz (1988), Farrar (1992), Sokolov, Snow (1994), Saxton (1997, 2000), Saxton et al. (1998), Chouinard, Clark (2003), Saxton et al. (2005), Strapp et al. (2008), Markus (2003)) and in specific forms (e.g., repetitions, reformulations or corrections), which would be inappropriate or redundant in a natural adult conversation (Jefferson 1982, Clark, Wong 2002, Clark, Bernicot 2008). Studies in CDS have identified two maternal conversational styles, directive vs. conversation-eliciting, and confirmed that mothers with conversation-eliciting style ask a lot of questions to elicit children's conversational participation and their children have

¹ The studies were carried out in the framework of cross-linguistic project "Pre- and Protomorphology in Language Acquisition". The project is supervised by W. U. Dressler (Austrian Academy of Sciences).

better language abilities (Hoff-Ginsberg 1991, Tulviste, Mizera, De Geer, 2004). Taking these observations into account, this paper concentrates on the *functions* and *structures* of parental questions addressed to young children. The main aim of the study was to identify and compare the variety and distribution of functional and structural types of interrogatives in Russian and Lithuanian CDS. The study is still at the initial stage, thus it is mainly the descriptive analysis of the research issue which is presented here.

1.2. Data and methods

The study was based on longitudinal data² on two monolingual, typically developing children: a Lithuanian-speaking girl, Monika³, and a Russian-speaking boy, Vanja⁴. Since maternal CDS is influenced by the context of the interaction (Sorsby, Martlew 1991, Tulviste 2003), subjects characterized by similar (or even identical) social and linguistic context were selected for the study. During the longitudinal observation period, both children were living in the second largest cities of their countries (Monika was living in Kaunas, Lithuania, and Vanja was living in St. Petersburg, Russia) and in upper-middle-class families. During this period, Monika and Vanja had no siblings and were not attending a kindergarten (the girl was being taken care of by her parents and the boy by his grandmother and parents). Monika's parents use the northern Panevėžiškiai sub-dialect of the eastern Aukshtaitish dialect when they speak with each other, but they speak in Standard Lithuanian with their daughter. Consequently, Monika acquired Standard Lithuanian, although she comprehends dialectal speech as well. Vanja, like his family, speaks Standard Russian. Both children were recorded twice or three times a week by a portable tape-recorder in a familiar environment (parents' or grandparents' homes, garden, etc.). Attempts were made to record as many different situations as possible: games, preparation of food, eating, communication with guests, bathing, getting ready for bed, etc. The recordings were done at different times of the day, which mostly depended on the child's willingness to communicate. Most of the recordings are dialogues between the child and her/his mother or grandmother, and there are also quite a few dialogues with more than two participants (mainly child, mother, and father; or child, mother, and grandmother). The size of Monika's corpus is 128,517 words which covers 27 hours of the child's (1;8–2;8) conversations with her parents and relatives. The size of Vanja's corpus is 191,949 words, which contains 63 hours of the child's (1;5–4;0) conversations with his grandmother and parents. For the study, the period from 1;8 to 2;8 of both the corpora was selected (see Table 1).

² Longitudinal language sampling method was developed in the late 1950s by three independent groups of investigators: Martin Braine (Walter Reed Hospital), Susan Ervin and Wick Miller (University of California), and Roger Brown (Harvard University). The specific of the longitudinal language sampling is that the children are selected specifically because they meet predetermined criteria. Usually more than one child is observed in order to identify the most typically developing child for the particular study. At the beginning of observation, they usually are at the transition stage from single-word to multi-word speech. The children are recorded on a regular schedule, e.g., twice a week, 30 min. per session (Ingram 1989).

³ The corpus was collected by Monika's mother I. Balčiūnienė (the co-author of the paper).

⁴ The corpus was collected under the supervision of N. V. Gagarina.

Table 1. The analysed corpus

Age	Total number of adults' and child's utterances		Number of adults' questions (% of all adults' utterances)	
	LT	RU	LT	RU
1;8	3871	5988	1744 (69.7%)	1567 (45.4%)
1;9	4659	3396	1963 (65.8%)	981 (43.9%)
1;10	3913	3384	1853 (70.0%)	977 (56.0%)
1;11	4878	3565	3238 (91.4%)	1118 (50.4%)
2;0	2512	3582	1573 (62.5%)	931 (46.5%)
2;1	4925	5112	3151 (85.7%)	1485 (59.6%)
2;2	1845	3927	1073 (90.7%)	1058 (55.0%)
2;3	3848	4244	2360 (42.5%)	1187 (55.5%)
2;4	2314	3500	1288 (79.0%)	1020 (58.9%)
2;5	5525	3778	2343 (69.7%)	1115 (53.7%)
2;6	3403	3223	1444 (65.5%)	1057 (58.5%)
2;7	2857	2908	1272 (70.1%)	871 (47.9%)
2;8	2622	3748	1140 (72.0%)	978 (42.5%)
Total	47172	50355	24442	14345

One can observe that, despite an identical biological age, the development of Monika and Vanja's mean length of utterance (MLU), which is considered one of the most relevant criteria of grammar and general language acquisition, is significantly different. Monika develops consistently from 1.285 words per utterance at age 1;8 to 2.468 words per utterance at age 2;8; while Vanja's MLU index does not exceed 1.1 words per utterance until age 2;2, but from 2;3 a rapid spurt is observed, and finally (at age 2;8) Vanja's MLU index equals 2.603 words per utterance (see Figure1).

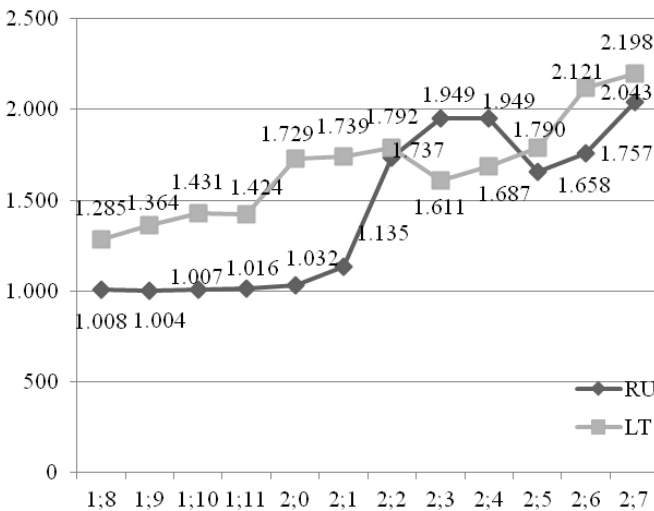


Figure 1. Children's MLU development

However, the development of the MLU index proves that during the observation period both of the children are at the transition stage, i.e., their language develops from a holophrastic (based on single-word utterances) speech to binomial syntactic structures. For the study, a random sample of 300 adult questions⁵ was selected from each month, and these questions were annotated for automatic linguistic analysis using the tools of the CHILDES (*Child Language Data Exchange System*) software (MacWhinney 2010).

2. Classification of interrogatives according to their function and structure

The classification of interrogatives according to their *pragmatic role, position, function, form* and structure was based on the methodology of previous investigations (see Kilani-Schoch et al. 2008, Kazakovskaya 2008, 2010, 2011, Kazakovskaya, Balčiūnienė (forthcoming)), thus here we concentrate mainly on the *functions* and *structure*.

2.1. Functions of interrogatives

Questions can be used to convey the following intentions of a speaker:

- to get some unknown information, e.g., (1a–b):
 - (1a) CHILD: *Aš ieškojau, ku [= kur] peilis.*
‘I [was] looking for a knife.’
ADULT: *Nu kaip? Suradai?*
‘And? [Did you] find?’ (MON 2;7)
 - (1b) ADULT: *A kakie ty xochesh’ konfetki?*
‘So what candies: DIM [do] you want?’ (VAN 2;3)
- to confirm some information, e.g., (2a–b):
 - (2a) ADULT: *Salotų yra, Monika. Pomidorų ir kopūstų. Nori salotų?*
‘There is salad, Monika. Tomatoes and cabbage. [Do you] want salad?’ (MON 1;11)
 - (2b) ADULT: *A jogurt dat’ ili smetanku?*
‘[Should]’ give you yoghurt or cream?’
CHILD: *Net.*
‘No.’
ADULT: *Ne nado ni togo, ni drugogo?*
‘[Don’t you] need any?’ (VAN 2;7)

These could also be termed *proper interrogative sentences*.

⁵ In Monika’s corpus, the majority of the selected questions were produced by her mother, and other questions were produced by her father, grandparents and aunts. In Vanja’s corpus, the majority of the selected questions were produced by his grandmother, and other questions were produced by his mother and father. However, no significant differences concerning structure and forms of interrogatives were observed between the mother’s and grandmother’s or between mother’s and father’s speech. Thus the analysed material could be generally described as a random sample of adult questions addressed to a child.

Another function of questions is:

- to initiate a conversation, e.g., (3a–b):
 - (3a) Situation: Monika enters the room.
ADULT: **O, Monika, kq tu čia atsinešei?**
‘Oh, Monika, what [did] you bring here?’ (MON 1;8)
 - (3b) Situation: Vanja and his grandmother are listening to the record.
ADULT: **Kto tam govorit?**
‘Who [is] speaking there?’ (VAN 2;0)
- to develop (support) a conversation, e.g., (4a–b):
 - (4a) CHILD: *Ladoni* [= raudoni] *meškiukai*.
‘Red bears:DIMINUTIVE.’
ADULT: *Raudoni. Gražūs meškiukai?*
‘Red. Beautiful bears?’ (MON 2;3)
 - (4b) CHILD: *Ja tipej* [= teper] *akulja* [= akula].
‘I [am] a shark now.’
ADULT: *Akula? A čto delaet akula?*
‘A shark? And what [is] a shark doing?’ (VAN 2;8)
- or to repair a conversation, e.g., (5a–b):
 - (5a) CHILD: *Atidengiau*.
‘[I] opened.’
ADULT: **Kq padarei?**
‘What [did you] do?’
CHILD: *Atidengiau*.
‘[I] opened.’
ADULT: **Atidengei?**
‘[Did you] open?’
CHILD: *Atidengiau*.
‘[I] opened.’ (MON 2;2)
 - (5b) CHILD: *Ja kos’ka*.
‘I [am] a cat.’
ADULT: **Kto koshka? Ili v okoshko? Ty čto, pro čto govorish’?**
‘Who [is] a cat? Or “to the window:DIMINUTIVE”? What [are] you talking about?’⁶ (VAN 2;6)

These could be also termed *metainterrogative sentences*. In some cases these intentions (seeking information vs. initiating/developing/repairing a conversation) may overlap, but usually one of them is obviously dominant.

Moreover, all the questions may be divided into two groups according to their pragmatic role: to focus on the content (6a–b) or on the linguistic form (7a–b) of interlocutors’ speech⁷.

⁶ Vanja speaks unclearly, thus the adult cannot identify the real lexeme.

⁷ One should mention here that *initiative* questions are related mainly to the content of conversation, while all other questions can focus either on the content or linguistic form of the interlocutor’s previous utterances.

- (6a) CHILD: *Gélytė.*
 ‘A flower:DIMINUTIVE.’
 ADULT: ***Kur gėlytė?***
 ‘Where [is] a flower:DIMINUTIVE?’ (MON 2;3)
- (6b) ADULT: *Spasibo, mozhno ja ego budu katat’?*
 ‘Thank you, may I drive it?’
 CHILD: *Da.*
 ‘Yes.’
 ADULT: ***A ty chto budesh’ katat’?***
 ‘And what [will] you drive?’ (VAN 1;10)
- (7a) CHILD: *Nėla [= nėra] pakutuko [= plaktuko].*
 ‘Hammer [is] missing.’
 ADULT: ***Ko nėra?***
 ‘What [is] missing?’
 CHILD: *Užkalti.*
 ‘To hammer.’
 ADULT: ***Ai. Pirma sakei, nėra plaktuko?***
 ‘Ah. [You] said before, hammer [is] missing?’ (MON 2;6)
- (7b) ADULT: *Chto ty sobiraesh’?*
 ‘What [are] you collecting?’
 CHILD: *Smeshnoe.*
 ‘Funny.’
 ADULT: ***Veseloe?***
 ‘Merry?’ (VAN 2;7)

2.2. Structure of reaction interrogatives

All reaction interrogatives can be classified into several structural types, such as *repetitions*, *focuses*, *reformulations*, *expansions*, *indirect corrections*, *clarifications*, *topic continuations*, and *topic shifts*. These structural types will be discussed below.

Repetitions can be described as echo-repetitions of the whole preceding phrase of the interlocutor, e.g., (8a–b):

- (8a) ADULT: *Ką piešim?*
 ‘What [will we] draw?’
 CHILD: *Bliūdą.*
 ‘A bowl.’
 ADULT: ***Bliūdą?***
 ‘A bowl?’ (MON 2;2)
- (8b) CHILD: *Xodit akulja [= akula].*
 ‘A shark [is] walking.’
 ADULT: ***Xodit akula?***
 ‘A shark [is] walking.’ (VAN 2;8)

Focuses are partial repetitions of the previous phrase, e.g., (9a–b):

(9a) CHILD: *Kq dabal* [= dabar] *šnekès?*
'What [will she] say now?'

ADULT: ***Dabar?***
'Now?' (MON 2;8)

(9b) CHILD: *Eto budet puska* [= pushka].
'Here [will] be a gun.'

ADULT: ***Pushka?***
'A gun?' (VAN 2;8)

Reformulations are adult rephrasing of what they think the child intended to say, e.g., (10a–b):

(10a) CHILD: *Čia katinukas.*
'Here [is] a cat:DIMINUTIVE.'

ADULT: ***Katinukas čia?***
'A cat [is] here?' (MON 2;2)

(10b) CHILD: *Igus'ki* [= igrushki] *se* [= vse].
'All toys.'

ADULT: ***Vo vse igrushki?***
'With all toys?' (VAN 2;7)

Expansions occur when the child's phrase is lexically and grammatically supplemented by an adult, e.g., (11a–b):

(11a) CHILD: *Skylutė.*
'A hole:DIMINUTIVE.'

ADULT: ***Skylutę radai?***
'[Did you] find a hole:DIMINUTIVE?' (MON 2;0)

(11b) ADULT: *Ja chaj budu pit', s molokom. Vanja, a ty s chem chaj p'esh'?*
'I [will] drink tea with milk. Vanja, what [will] you drink a tea with?'

CHILD: *Maokom* [= s molokom].
'With milk.'

ADULT: ***S molokom tozhe?***
'Also with milk?' (VAN 2;4)

Indirect corrections are an adult's indications that the form (12a) or content (12b) of the child's previous phrase was erroneous, e.g., (12a–b):

(12a) CHILD: *Pat'* [= spat'].
'To sleep.'

ADULT: *Ne spat'. Spit. Spit koza. Spit?*
'Not to sleep. Sleeps. The goat sleeps. Sleeps?' (VAN 2;2)

- (12b) ADULT: *Čhto by mne poest', kak ty думаеш'?*
 'What could I eat, what [do] you think?'
 CHILD: *Tjiku [= čajku].*
 'A tea: DIMINUTIVE.'
 ADULT: ***Čajku? Razve babushka chaek ljubit?***
 'A tea: DIMINUTIVE? [Does] Granny prefer a tea: DIMINUTIVE?'
 CHILD: *Kofe.*
 'Coffee.'
 ADULT: *Kofe, pravilno'.*
 'Coffee, right.' (VAN 2;4)

Clarifications are requests to repeat/clarify the previous phrase, e.g., (13a–b):

- (13a) CHILD: *Vafiukq [= vafliuką] suvalgiau.*
 '[I] ate a waffle: DIMINUTIVE.'
 ADULT: ***Kq suvalgei?***
 'What [did] you eat?'
 CHILD: *Vafiukq [= vafliuką] suvalgiau.*
 '[I] ate a waffle: DIMINUTIVE.'
 ADULT: *Ai, vafliukq.*
 'Ah, a waffle: DIMINUTIVE.' (MON 2;8)
- (13b) ADULT: *Kto ego chinit' teper' budet?*
 'Who [will] repair it?'
 CHILD: *Usjat'ka.*
 Incomprehensible word.
 ADULT: ***Kto?***
 'Who?' (VAN 2;5)

Topic continuations occur when an adult does not respond explicitly to the child's phrase but continues a natural flow of conversation, e.g., (14a–b):

- (14a) CHILD: *Noju [= noriu] tintuko [= trintuko].*
 '[I] want a rubber.'
 ADULT: ***Tai kq dabar daryti?***
 'So what [should I] do now?' (MON 2;2)
- (14b) ADULT: *Vanja, a u vas v sadike mnogo mashinok?*
 'Vanja, [are there] a lot of toy cars in your kindergarten?'
 CHILD: *Da.*
 'Yes.'
 ADULT: ***A bol'shie mashinki, ili raznye?***
 'And [are] the toy cars big or different size?' (VAN 2;5)

Topic shifts occur when an adult does not respond explicitly to the child's phrase but changes the topic, e.g., (15a–b):

- (15a) CHILD: *Mamytè dainuos.*
 'Mummy [will] sing.'
 ADULT: *Gal ir dainuos. Arbatytès nori?*
 'Maybe [she will] sing. [Would you] like a tea?' (MON 2;6)

(15b) CHILD: *Baba [= babushka] katat' gorki [= s gorki].*

'Granny to slide down the hill.'

ADULT: *Babushka budet malen'kie katat' s gorki? Ponjatno. Vanja, ty kushat' xochesh'?*

'Granny [will] slide down the hill? Well. Vanja, [do] you want to eat?' (VAN 2;3)

As mentioned above, the target corpus was annotated for automatic statistical analysis, and the production of different functional and structural types of interrogatives was analysed.

3. Research findings

3.1. General productivity of interrogatives in Lithuanian and Russian CDS

The analysis indicated that interrogatives seem to be the most productive communicative type of utterances in both Russian and Lithuanian “motherese”. They comprise approximately 71.8% of all utterances in Lithuanian and 51.8% in Russian CDS. One can observe (see Figure 2) the tendencies of distribution of parental interrogatives during the ages studied.

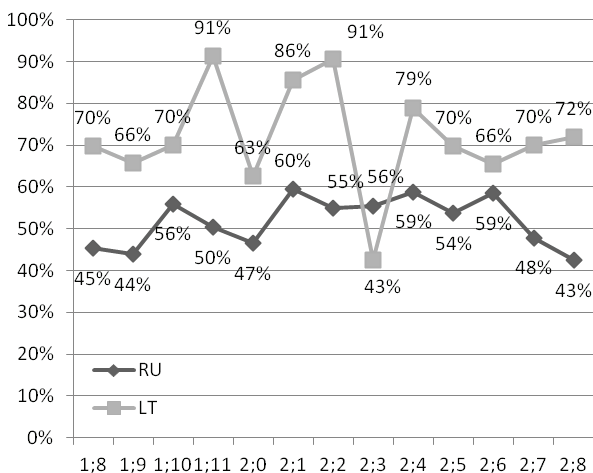


Figure 2. Distribution of interrogatives in Russian and Lithuanian CDS (% of all parental utterances)

Despite the fact that the general number of interrogatives is smaller in Russian CDS than in Lithuanian (it has to be pointed out here that the great number of interrogatives in Lithuanian CDS were indicated in previous investigations as well, see, for example, Kamandulytė 2007, 2008, and also Kazakovskaya 2004, 2011), the development of interrogative production seems to be similar in both corpora. In both cases a considerable increase in the production of interrogatives was registered from the period of 1;8 to 2;1–2;2, after which it decreases (from 2;2

to 2;5–2;6); however, from 2;6 a slight increase was registered in Lithuanian CDS, while a significant decrease was registered in Russian CDS.

Moreover, parental communicative strategy tends to develop similarly, according to the children’s age (see Figure 3).

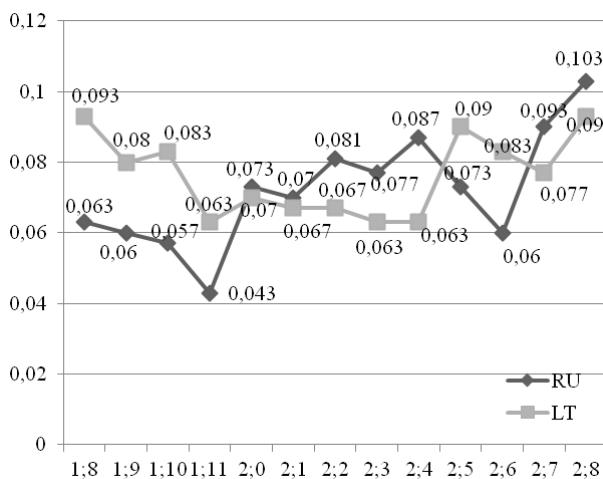


Figure 3. Development of type/token ratio of interrogatives in Russian and Lithuanian CDS

The type/token ratio of questions asked by the parents in both corpora decreases rapidly during the earliest investigated period (1;8–1;11), and then increases consequently as the children grow, i.e., both Russian and Lithuanian speaking parents seem to use more varied interrogative forms in later stages than earlier.

3.2. Functions of interrogatives in Lithuanian and Russian CDS

As mentioned above, questions were divided into two functional types: a) *requests for information* (see examples 1a–b, 2a–b), and b) *development/amendment, correction, repairing conversation* (see examples 3a–b, 4–b, 5a–b). Our analysis indicates that questions which are used in order to develop/repair a conversation are the dominant functional type of interrogatives in both Russian and Lithuanian CDS: they make up 73%–94% of all questions in Lithuanian CDS and 84%–98% in Russian CDS (see Table 2).

These findings prove that parents put in great effort in helping children maintain a conversation: they ask many specific initiative questions (also termed *pre-sequences*, see Levinson 1983) in order to attract child’s attention and to stimulate his interaction; they help to extend a topic of conversation and to add more details to the child’s utterances. Finally, parents improve and reformulate child’s speech, i.e. their questions perform pure didactic functions. These findings correspond to the general tendencies of CDS in many Western cultures (see Ervin Tripp, Strage 1985).

The findings of our analysis correspond to the results of previous studies (see Kilani-Schoch et al. 2008, Balčiūnienė 2009, Kazakovskaya 2010, Kazakovskaya, Balčiūnienė 2011, (forthcoming)) in that the *conversational* questions are more numerous than the *metadiscursive* ones (see Table 3).

Table 2. Distribution of different functional types of reaction interrogatives in Lithuanian and Russian CDS

Age	Corpus of Lithuanian CDS		Corpus of Russian CDS	
	Seeking information %	Developing a conversation %	Seeking information %	Developing a conversation %
1;8	23	77	16	84
1;9	6	94	15	85
1;10	6	94	2	98
1;11	7	93	6	94
2;0	9	91	12	88
2;1	11	89	8	92
2;2	9	91	6	94
2;3	14	86	6	94
2;4	11	89	4	96
2;5	17	83	7	93
2;6	16	84	5	95
2;7	12	88	12	88
2;8	27	73	9	91
Average	13%	87%	8%	92%

Table 3. Distribution of different pragmatic types of reaction interrogatives in Lithuanian and Russian CDS

Age	Corpus of Lithuanian CDS		Corpus of Russian CDS	
	META-interrogatives %	CONV-interrogatives %	META-interrogatives %	CONV-interrogatives %
1;8	5	95	5	95
1;9	22	88	2	98
1;10	15	85	2	98
1;11	5	95	1	99
2;0	12	88	4	96
2;1	11	89	8	92
2;2	10	90	6	94
2;3	10	90	6	94
2;4	13	87	14	86
2;5	8	92	6	94
2;6	11	89	7	93
2;7	8	92	13	87
2;8	12	88	20	80
Average	11%	89%	7%	93%

Conversational interrogatives comprise 85%–95% of all questions in Lithuanian CDS and 80%–99% in Russian CDS. These results confirm that parents prefer to react to the content of conversation or to help a child with a topic maintenance, rather than focus on the linguistic form of the child’s phrases.

3.3. Structure of explicit reaction interrogatives in Lithuanian and Russian CDS

As mentioned above, interrogative reactions can be either explicit (*repetitions* (see examples 8a–b), *focuses* (see examples 9a–b), *reformulations* (see examples 10a–b), *expansions* (see examples 11a–b), *indirect corrections* (see examples 12a–b), and *clarifications* (see examples 13a–b)) or implicit (*topic continuations* (see examples 14a–b), and *topic shifts* (see examples 15a–b)). For the study, only explicit reactions were analysed, since they indicate the main differences between adult conversation and child-directed speech.

Having analysed the *structural* types of explicit reaction interrogatives, we can state that, generally, the distribution of different types in both Russian and Lithuanian CDS is similar (see Table 4a–b).

Expansions and clarifications are the most frequent type of explicit reaction interrogatives in both Russian and Lithuanian CDS, whereas other types are less numerous or were not observed at all in the target corpora.

Table 4a. Distribution of different structural types of explicit reaction interrogatives in Lithuanian CDS

Age	Repetitions %	Focuses %	Reformulations %	Expansions %	Corrections %	Clarifications %
1;8	17	0	7	51	12	13
1;9	12	0	4	23	6	55
1;10	16	2	4	28	2	48
1;11	19	7	11	38	0	25
2;0	16	4	13	29	1	37
2;1	14	7	9	17	4	49
2;2	15	10	13	19	4	39
2;3	15	6	12	22	1	44
2;4	11	4	10	26	6	43
2;5	26	3	8	19	5	39
2;6	9	10	12	19	1	49
2;7	15	13	23	13	5	31
2;8	8	6	11	15	5	55
Average	15%	6%	11%	25%	4%	41%

Table 4b. Distribution of different structural types of explicit reaction interrogatives in Russian CDS

Age	Repetitions %	Focuses %	Reformulations %	Expansions %	Corrections %	Clarifications %
1;8	5	0	37	28	9	21
1;9	11	0	34	45	8	2
1;10	6	0	44	38	4	8
1;11	12	0	31	47	4	6
2;0	14	2	8	48	10	18
2;1	21	0	11	41	7	20
2;2	20	2	14	37	11	16
2;3	19	2	15	42	8	14
2;4	20	16	11	21	9	23
2;5	17	11	15	37	3	17
2;6	27	13	14	24	2	20
2;7	17	24	15	23	4	17
2;8	16	25	19	14	9	17
Average	16%	7%	21%	34%	7%	15%

Expansions make up 25% of all explicit reaction interrogatives in Lithuanian CDS and 34% in Russian CDS. During the holophrastic speech period, expansions usually occur as a rephrasing of the child's single-word utterance into semantically equal syntactic construction and stimulate the child to maintain a topic of conversation, e.g., (16a–b):

(16a) CHILD: *Skylutė.*

'A hole:DIMINUTIVE'

ADULT: *Kur tu ten radai skylutę?*

'Where [did] you find a hole:DIMINUTIVE?'

CHILD: *Matatonas [= magnetofonas].*

'A player.'

ADULT: *Magnetofonas skylutę turi?*

'A player has a hole?' (MON 1;11)

(16b) CHILD: *Mama.*

'Mommy.'

ADULT: *Mama kupila Vane eshche igrushku, da?*

'Mommy bought one more toy to Vanja, right?' (VAN 1;10)

The majority of pure metalinguistic expansions contain a repeated child's utterance with inserted (17a) modal or (17b) auxiliary verb, or a preposition (18a–b):

(17a) CHILD: *Pasitiki.*

'To meet.'

ADULT: *Pasitiki reikia?*

'[Do you] need to meet?' (MON 2;0)

(17b) CHILD: *Kakii* [= *kakie*] *isjo* [= *eshche*]?
'What else?'

ADULT: ***Kakie eshche est***?
'What else are?' (VAN 2;7)

(18a) ADULT: *Saulutė šviečia*.
'The sun: DIMINUTIVE [is] shining.'

CHILD: *Akytė*.
'Eye: DIMINUTIVE.'

ADULT: ***Į akytę šviečia?***
'[Is] to the eye: DIMINUTIVE shining?' (MON 2;0)

(18b) ADULT: *Sobachka kuda poedet?*
'Where [will] the dog: DIMINUTIVE go?'

CHILD: *Pak* [= *park*].
'Park.'

ADULT: ***V park?***
'To the park?' (VAN 2;3)

Reformulations make up 11% of all explicit reaction interrogatives in Lithuanian CDS and 21% in Russian CDS. The first group of morphological reformulations is the correction of premorpheme *a*⁸, e.g., (19a–b):

(19a) CHILD: *A* [= *į*] *Bižus* [= *Biržus*] *atvažiavome*.
'[We] arrived a: PREMORPHEME *Biržai*.'

ADULT: ***Į Biržus atvažiavome?***
'[Did we] arrive in *Biržai*?' (MON 2;0)

(19b) ADULT: *Stishki kakie budem chitat*?
'Which poem [will we] read?'

CHILD: *A* [= *pro*] *bezinju* [= *mashinu*], *a* [= *pro*] *basjuju*
[= *bol'shuju*].
'PREMORPHEME *car*, a: PREMORPHEME *big*.'

ADULT: ***Pro bol'shuju mashinu?***
'About the big car?' (VAN 2;4)

During the earliest period, at a younger age, in Lithuanian CDS one can observe that a great number of child utterances are reformulated by parents from the child's native dialect into Standard Lithuanian, e.g., (20):

(20) CHILD: *Nieka*⁹.
'Nothing.'

ADULT: ***Nieko?***
'Nothing?' (MON 1;8)

⁸ The premorpheme *a* or "filler" is a prosodically appropriate but semantically empty element used in the early period of speech. In the child language these gradually develop into free or bound morphemes (Pepinsky et al. 2001, Savickienė 2002), and are absent from typically developed adult speech. Premorphemes are more conventional and universal than individual or language-specific elements; usually they are produced instead of particular lexical and/or grammatical item and take the position of an unstressed syllable. Following previous studies (Savickienė 2002, Balčiūnienė 2009), Lithuanian children use premorphemes instead of prefixes (*adaryti* [= *uždaryti*] 'to close'), prepositions (*a lova* [= *po lova*] 'under a bed') and inflections (*peija* [= *peilis*] 'knife').

⁹ Dialectal phonological variation.

Naturally, there were no similar examples in the Russian CDS, since Vanja acquired Standard Russian as his first language.

Clarifications represent approximately 41% of all explicit reaction interrogatives in Lithuanian CDS and 15% in Russian CDS. Clarifications usually occur as reactions to an incorrect, inappropriate or incomprehensible child utterance, e.g., (21a– b):

(21a) CHILD: *Sėdi piniginė.*

‘A wallet [is] sitting.’

ADULT: ***Ką sakai?***

‘What [are you] saying?’ (2;0)

(21b) CHILD: *Tipej [= teper’] kupi kasju [= krasnuju].*

‘Now buy the red [one].’

ADULT: ***Chto krasnuju?***

‘What red?’ (VAN 2;8)

Repetitions make up 15% of explicit reaction interrogatives in Lithuanian CDS and 16% in Russian CDS.

Focuses make up 5.6% of all explicit reaction interrogatives in Lithuanian CDS and 10% in Russian CDS. Focuses occur at the later stage of language development, when the children are able already to produce more complex structures, and the parents can focus on the word or words they find to be the most important. Parents presumably repeat the most informative words (nouns and verbs) of the child’s utterance, e.g., (22a–b):

(22a) CHILD: *Dar šitų dažų leikia [= reikia].*

‘[I] still need this colour.’

ADULT: ***Šitų dažų?***

‘This colour?’ (MON 2;8)

(22b) CHILD: *Fati [= smotri], on vez [= vlez].*

‘Look, it fitted.’

ADULT: ***Vlez?***

‘[Did it] fit?’ (VAN 2;8)

However, this prediction should be verified by additional semantic analysis.

Indirect corrections are 4% of all explicit reaction interrogatives in Lithuanian CDS and 7% in Russian CDS; this leads to the conclusion that didactic function is more obvious in the Russian CDS.

The results of the corpus analysis show that, basically, the distribution of structural types of explicit reaction interrogatives is similar in both Russian and Lithuanian CDS. The difference lies in a different production of clarifications – these interrogatives are much more frequent in the Lithuanian corpus. These findings could be explained by individual, cultural or language-specific differences in communicative strategy in Lithuanian and Russian CDS¹⁰. Despite the fact that didactic function is one of the most dominant of functions in both Lithuanian and Russian CDS, it is expressed more explicitly in Russian CDS than in Lithuanian.

¹⁰ Since only one subject from each language community participated in the study, the results still do not indicate to what extent the differences observed between Russian and Lithuanian CDS reflect individual differences between two individuals and to what extent these are language-specific, cultural, age or gender differences.

One can observe many corrections of the child's utterances and requests to repeat a particular word/phrase correctly, while Lithuanian speaking parents ask more clarification questions and stimulate the child to explain what she intended to say. This fact may be explained by age difference between the main target subjects; however, cultural, language-specific and contextual factors should be investigated additionally in order to generalize this statement.

4. Summary and discussion

The statistical analysis indicated that interrogatives seem to be the most productive communicative type of parental utterances in both Lithuanian and Russian CDS. They comprise approximately 71.8% of all utterances in Lithuanian and 51.8% in Russian CDS. Despite the different general number of interrogatives, their development seems to be similar in both target corpora. A considerable increase and decrease in the production of interrogatives was registered during the same periods; moreover, the type/token ratio of parental interrogatives increased consistently as the children grew older.

A number of the interrogatives function (similarly to natural adult conversation) as requests for information, clarifications of incomprehensible utterances or demonstrations of disagreement. Despite this, the majority of them appear to be used for a very specific purposes (didactic, supporting language acquisition, e.g., negative evidence) and in specific forms (e.g., repetitions, reformulations or corrections), which would be inappropriate or redundant in a natural adult conversation. In both Russian and Lithuanian CDS, expansions were the most produced structural type of explicit reaction interrogatives (they comprise 34% of all explicit reaction interrogatives in Russian and 25% in Lithuanian CDS), while other types were less numerous, and their production was different. However, the majority of parental interrogatives (approximately 89% in Lithuanian and 93% in Russian) were related to the content of conversation (i.e., *conversational* questions), while reactions to the linguistic form of the child's previous phrase (i.e., *metadiscursive* questions) were rare.

The study indicated that the caregivers, representing speakers of typologically, culturally and geographically close languages, demonstrated similar communicative behaviour despite some individual differences (the Lithuanian speaking mother seems to use conversation-eliciting style, whereas the Russian speaking mother prefers more directive style; however, this prediction should be verified by more comprehensive studies). This leads to the general idea that a correlation can be identified between parental conversational strategy and the type of language. However, larger corpora should be analysed in order to confirm the hypothesis, thus the research material will be supplemented by longitudinal data of other Lithuanian and Russian children for further investigations. The cultural aspect should not be excluded either and, consequently, other languages (e.g., Estonian which is close culturally and geographically, but far apart typologically, and Austrian-German which is, in contrast, closer typologically, but far apart culturally and geographically) are planned to be included as the additional subject of further studies.

References

- Balčiūnienė, Ingrida 2009. Pokalbio struktūros analizė kalbos įsisavinimo požiūriu. [Analysis of Conversational Structure from the Perspective of Language Acquisition.]. PhD Thesis. Kaunas: Vytautas Magnus University. http://vddb.library.lt/fedora/get/LT-eLABa-0001:E.02~2009~D_20090915_081635-07813/DS.005.1.02.ETD
- Bohannon, John N.; Stanowicz, Laura 1988. The issue of negative evidence: Adult responses to children's language errors. – *Developmental Psychology*, 24 (5), 684–689. <http://dx.doi.org/10.1037/0012-1649.24.5.684>
- Ceitlin, Stella N. 2001. Nekotorye osobennosti dialoga 'vzroslyj – rebenok': Funkciji replik-povtorov'. [Some characteristics of child-adult dialogue: functions of repetitions.] – Stella N. Ceitlin, Victoria V. Kazakovskaya (Eds.). *Rebenok kak partner v dialoge* [A Child as a Partner in a Dialogue]. Saint Petersburg: Sojuz, 9–25.
- Choi, Soonja 1997. Language-specific input and early semantic development: Early from children Korean. – D. I. Slobin (Ed.). *The Cross-Linguistic Study of Language Acquisition*, Vol. 5. Hillsdale, N.J.: Erlbaum, 41–133.
- Chouinard, Michelle M.; Clark, Eve V. 2003. Adult reformulations of child errors as negative evidence. – *Journal of Child Language*, 30 (3), 637–669. <http://dx.doi.org/10.1017/S03050009030005701>
- Clark, Eve V. 2009. *First Language Acquisition*. Cambridge: Cambridge University Press.
- Clark, Eve V.; Wong, Andrew 2002. Pragmatic directions about language use: Words and word meanings. – *Language in Society*, 31 (2), 181–212. <http://dx.doi.org/10.1017/S0047404501020152>
- Daukšytė, Jovita 2005. Skirtingos kalbinės aplinkos nevienodo amžiaus moksleivių žodinės asociacijos gimtąja (lietuvių) kalba [Semantic associations in Lithuanian (as a native) language among various age students from different linguistic environment]. – *Pedagogika* [Pedagogy], 77, 64–71.
- De León, Lourdes 1998. The emergent participant: Interactive patterns in the socialization of Tzotzil (Mayan) infants. – *Journal of Linguistic Anthropology*, 8 (2), 131–161.
- Demetras, M. J.; Post, Kathryn N.; Snow, Catherine E. 1986. Feedback to first language learners: The role of repetitions and clarification questions. – *Journal of Child Language*, 13, 275–292. <http://dx.doi.org/10.1017/S0305000900008059>
- Englund, Kjellrun T. 2005. Voice onset time in infant directed speech over the first six months. – *First Language*, 25 (2), 219–234. <http://dx.doi.org/10.1177/0142723705050286>
- Ervin-Tripp, Susan M.; Strage, Amy 1985. Parent-child discourse. – Teun van Dijk (Ed.). *Handbook of Discourse Analysis*, Vol. 3. London: Academic Press, 67–77.
- Farrar, Michael J. 1992. Negative evidence and grammatical morpheme acquisition. – *Developmental Psychology*, 28 (1), 90–98. <http://dx.doi.org/10.1037/0012-1649.28.1.90>
- Fernald, Anne; Taeschner, Traute; Dunn, Judy; Papousek, Mechthild; De Boysson-Bardies, Benedicte; Fukui, Ikuko 1989. A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. – *Journal of Child Language*, 16 (3), 477–501. <http://dx.doi.org/10.1017/S0305000900010679>
- Girolametto, Luigi; Bonifacio, Serena; Visini, Cristiana; Weitzman, Elaine; Zocconi, Elisabetta; Steig Pearce, Patsy 2002. Mother-child interactions in Canada and Italy: Linguistic responsiveness to late-talking toddlers. – *Journal of Language and Communication Disorders*, 37 (2), 153–171. <http://dx.doi.org/10.1080/13682820110116794>
- Heath, Shirley B. 1983. *Ways with Words*. Cambridge: Cambridge University Press.
- Hirsh-Pasek, Kathy; Treiman, Rebecca; Schneiderman, Maita 1984. Brown & Hanlon revisited: Mothers' sensitivity to ungrammatical forms. – *Journal of Child Language*, 11, 81–88. <http://dx.doi.org/10.1017/S0305000900005596>
- Hoff-Ginsberg, Erika 1991. Mother-child conversation in different social classes and communicative settings. – *Child Development*, 62 (4), 782–796. <http://dx.doi.org/10.2307/1131177>

- Jefferson, Gail 1982. On exposed and embedded correction in conversation. – *Studium Linguisticum*, 14, 58–68.
- Kamandulytė, Laura 2005. Vaikiškosios kalbos modifikacijos [Modifications of Child-directed Speech]. MA Thesis. Kaunas: Vytautas Magnus University.
- Kazakovskaya, Victoria V. 2004. Voprosno-otvetnye edinstva v dialoge vzroslyj–rebenok' [Question-Response Units in the Adult-Child Dialogue]. – *Voprosy yazykoznanija* [Issues in Linguistics], 2, 89–110.
- Kazakovskaya, Victoria V. 2008. Funkcional'naya kategorizaciya voprositel'nykh replik v russkom dialoge: analiz rechi vzroslogo, obrashchennoi k rebenku' [Functional Categorization of Questions in Russian: Analysis of Child Directed Speech]. – A. V. Bondarko, S. A. Shubik (Eds.). *Problemy funkcional'noj grammatiki: Kategorizacija semantiki* [Problems of the Functional Grammar: Semantic Categorization]. Saint Petersburg: Nauka, 331–364.
- Kazakovskaya, Victoria V. 2010. Reaktivnye repliki vzroslogo i usvoenie rebenkom grammatiki rodnogo yazyka' [Adult Reactions and Acquisition of the First Language Grammar]. – *Voprosy yazykoznanija* [Issues in Linguistics], 3, 3–29.
- Kazakovskaya, Victoria V. 2011. Question and answer in adult–child dialogue. Moscow.
- Kazakovskaya, Victoria V.; Balčiūnienė, Ingrida [forthcoming]. Interrogatives in Russian and Lithuanian child-directed speech: Do we communicate in the same way? – *Journal of Baltic Studies*, in press.
- Kazakovskaya, Victoria V.; Balčiūnienė, Ingrida 2011. Osobennosti rechi vzroslogo, obrashchennoj k rebenku: Voprositel'nye repliki v russkom i litovskom yazukakh' [Characteristics of child directed speech: interrogatives in Russian and Lithuanian CDS]. – *Issledovaniya po slavyanskim yazykam* [Studies in Slavic Languages], 16 (1), 259–292.
- Kilani-Schoch, Marianne; Balčiūnienė, Ingrida; Korecky-Kröll, Katharina; Laaha, Sabine; Dressler, Wolfgang U. 2008. On the Role of Pragmatics in Child-Directed Speech for the Acquisition of Verb Morphology. – *Journal of Pragmatics*, 41 (2), 129–159.
- Levinson, Stephen C. 1983. *Pragmatics*. Cambridge: Cambridge University Press.
- MacWhinney, Brian 2010. *The CHILDES Project: Tools for Analyzing Talk*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Ingram, David 1989. *First Language Acquisition: Method, Description, and Explanation*. Cambridge: Cambridge University Press.
- Marcus, Gary F. 1993. Negative evidence in language acquisition. – *Cognition*, 46 (1), 53–85. [http://dx.doi.org/10.1016/0010-0277\(93\)90022-N](http://dx.doi.org/10.1016/0010-0277(93)90022-N)
- Ochs, Elinor 1982. Talking to children in Western Samoa. – *Language in Society*, 11 (1), 77–104. <http://dx.doi.org/10.1017/S0047404500009040>
- Ochs, Elinor; Schieffelin, Bambi B. 1984. Language acquisition and socialization: Three developmental stories. – R. Schweder, R. LeVine (Eds.). *Culture Theory: Essays in Mind, Self and Emotion*. Cambridge: Cambridge University Press, 335–358.
- Pepinsky, Thomas; Demuth, Katherine; Roark, Brian 2001. The status of filler syllables in children's early speech. – Anna H. J. Do et al. (Eds.). *BUCLD 25 Proceedings*. Somerville, MA: Cascadilla Press, 575–586.
- Pye, Clifton 1986. Quiché Mayan speech to children. – *Language*, 59 (3), 583–604. <http://dx.doi.org/10.2307/413905>
- Rūķe-Draviņa, Velta 1977. Modifications of speech addressed to young children in Latvian. – Catherine E. Snow, Charles A. Ferguson (Eds.). *Talking to Children: Language Input and Acquisition*. Cambridge: Cambridge University Press, 237–254.
- Sachs, Jacqueline; Johnson, Marie L. 1976. Language development in a hearing child of deaf parents. – W. von Raffler-Engel, Y. LeBrun (Eds.). *Baby Talk and Infant Speech*. (Special issue), *Neurolinguistics*, 5, 246–252.

- Sachs, Jacqueline; Bard, Barbara; Johnson, Marie L. 1981. Language learning with restricted input: Case studies of two hearing children of deaf parents. – *Applied Psycholinguistics*, 2 (1), 33–54. <http://dx.doi.org/10.1017/S0142716400000643>
- Savickienė, Ineta 2002. The emergence of case distinctions in Lithuanian. – Maria D. Voelikova, Wolfgang U. Dressler (Eds.). *Pre- and Protomorphology: Early Phases of Morphological Development in Nouns and Verbs*. *Lincom Studies in Theoretical Linguistics*, 9. München: Lincom, 105–115.
- Saxton, Matthew 1997. The contrast theory of negative evidence. – *Journal of Child Language*, 24 (1), 139–161. <http://dx.doi.org/10.1017/S030500099600298X>
- Saxton, Matthew 2000. Negative evidence and negative feedback: Immediate effects on the grammaticality of child speech. – *First Language*, 20 (60), 221–252. <http://dx.doi.org/10.1177/014272370002006001>
- Saxton, Matthew; Kulcsar, Bela; Marshall, Greer; Rupra, Mandeep 1998. Longerterm effects of corrective input: An experimental approach. – *Journal of Child Language*, 25 (3), 701–721. <http://dx.doi.org/10.1017/S0305000998003559>
- Saxton, Matthew; Backley, Philipp; Gallaway, Clare 2005. Negative input for grammatical errors: Effects after a lag of 12 weeks. – *Journal of Child Language*, 32 (3), 643–672. <http://dx.doi.org/10.1017/S0305000905006999>
- Schieffelin, Bambi B. 1979. Getting it together: An ethnographic approach to the study of the development of communicative competence. – Elinor Ochs, Bambi B. Schieffelin (Eds.). *Developmental Pragmatics*. New York: Academic Press, 73–108.
- Sokolov, Jeffrey L.; Snow, Catherine E. 1994. The changing role of negative evidence in theories of language development. – Clare Gallaway, Brian J. Richards (Eds.). *Input and Interaction in Language Acquisition*. Cambridge: Cambridge University Press, 38–55.
- Sorsby, Angela J.; Martlew, Margaret 1991. Representational demands in mothers' talk to pre-school children in two contexts: Picture book reading and a modelling task. – *Journal of Child Language*, 18 (2), 373–95. <http://dx.doi.org/10.1017/S0305000900011119>
- Strapp, Chehalis M.; Bleakney, Dana M.; Helmick, Augusta L.; Tonkovich, Hayley M. 2008. Developmental differences in the effects of negative and positive evidence. – *First Language*, 28 (1), 35–53. <http://dx.doi.org/10.1177/0142723707084840>
- Tomasello, Michael 2003. *Constructing a Language: A Usage-Based Theory of Language Acquisition*. London: Harvard University Press.
- Tulviste, Tiia 2002. *Language Socialization Across Socio-Cultural Contexts*. Stockholm: Stockholm University.
- Tulviste, Tiia 2003. Contextual variability in interactions between mothers and 2-year-olds. – *First Language*, 23 (3), 311–325. <http://dx.doi.org/10.1177/01427237030233004>
- Tulviste, Tiia; Mizera, Luule; De Geer, Boel 2004. Expressing communicative intents in Estonian, Finnish, and Swedish mother-adolescent interactions. – *Journal of Child Language*, 31 (4), 801–819. <http://dx.doi.org/10.1017/S0305000904006488>
- Wójcik, Pawel 1994. Some characteristic features of Lithuanian Baby Talk. – *Linguistica Baltica*, 3, 71–86.

Victoria V. Kazakovskaya (Department of Theoretical Grammar, Institute of Linguistic Studies at Russian Academy of Sciences). Her academic research interests include psycholinguistics, language acquisition (longitudinal studies, conversation analysis, and communicative competence).
victory805@mail.ru

Ingrida Balčiūnienė (Centre of International and Multilingual Communication at Vytautas Magnus University). Her research interests include psycholinguistics and first language acquisition (longitudinal studies, conversation analysis, language disorders).
i.balciuniene@hmf.vdu.lt

LEEDU- JA VENEKEELNE LAPSELE SUUNATUD KÕNE: MIKS ME ESITAME VÄIKELASTELE NII PALJU KÜSIMUSI?

Victoria V. Kazakovskaya, Ingrida Balčiūnienė

Vene Teaduste Akadeemia, Vytautas Magnuse Ülikool

Käesoleva uurimuse põhieesmärgiks oli kindlaks määrata ning võrrelda interrogatiivsete struktuuride erinevusi ning esinemist leedu- ning venekeelses lapsele suunatud kõnes. Rõhuasetus oli väikelastele esitatud küsimuste funktsionaalsel ning struktuurilistel omadustel.

Analüüsi aluseks olid pikiuuringuga saadud andmed kahe normaalse keelise arenguga ükskeelse lapse kohta, üks neist leedu tüdruk (2;0–2;8), teine vene poiss (2;0–2;8). Laste ja nende vanemate dialoogide litereeritud korpus annoteeriti mitmeotstarbelise automaatse lingvistilise analüüsi tarvis, kasutades programmi CHILDES vahendeid. Uurimuse käigus analüüsiti lapsevanemate küsilause funktsionaalseid ning struktuurilisi tunnuseid.

Laste ja nende vanemate omavaheliste dialoogide analüüsimise põhjal saame teha järgnevaid järeldusi: nii leedu- kui ka venekeelses lapsele suunatud kõnes on interrogatiive rohkem kui imperatiive, väiteid või hüüatusi. Täiskasvanute omavaheliste vestlustega sarnaselt kasutatakse interrogatiive informatsiooni küsimiseks, arusaamatute ütluste selgitamiseks ning mittenõustumise näitamiseks. Vaatamata sellele esineb enamik neist interrogatiividest väga spetsiifilisel otstarbel (nt negatiivne kinnitus või tagasiside) ning spetsiifilistes vormides (nt kordused, ümbersõnastused, parandused), mis oleksid ebasobivad ja/või üleaarused tavapärasel täiskasvanutevahelises vestluses.

Võtmesõnad: keeleomandamine, lapsele suunatud kõne, vestlusstrateegia, vene keel, leedu keel

REFERENTSIAALSED NOOMENIFRAASID LASTE NARRATIIVIDES

Helen Kirsipuu, Piret Soodla, Renate Pajusalu

Ülevaade. Artikkel otsib vastust küsimusele, millised on laste narratiivide viitamisstrateegiaid, ja võrdleb neid täiskasvanute omadega. Eesmärgiks on leida, kuidas võrd on vaadeldavas eas viitamisstrateegiaid omandatud. Uurimuse andmestik pärineb 6–7-aastaste laste ja täiskasvanute ühe pildiseeria põhjal suuliselt jutustatud lugudest. Vaatleme seda, milliseid strateegiaid lapsed kasutavad erineva infostaatusega referentidele viitamisel ja millised nimisõnafraasid viiteahelates esinevad. Eraldi uurime, kuidas erinevad teineteisest eri lauseliikmeid kodeerivad fraasid ja kuidas kasutatakse 3. isiku personaalpronoomeid ning määratlejaid (eelkõige *see* ja *üks*).*

Võttesõnad: referentsiaalsus, keele omandamine, pronoomenid, määratlejad, eesti keel

1. Sissejuhatus

Kõneldes viitab inimene entiteetidele, millest ta räägib. See tähendab, et kas reaalsuses või diskursuses maailmas olevad referendid paigaldatakse kõneleja ja kuulaja referentsiaalsesse raamistikku. Eduka suhtluse jaoks peavad need raamistikud olema võimalikult sarnased, kuulaja peab suutma identifitseerida kõneleja poolt viidatud referendid, või kui referent on uus, siis looma uue diskursuse referendi. Teoreetiliste infostaatuse käsitluste keskmes on eeldus, et kõneleja annab jutustamise vältel ka keeleliste vahenditega kuulajale märku, kas tegu on uue või vana referendiga ehk kas kuulaja peab oma referentsiaalses raamistikus vastava representatsiooni alles paigaldama või on see juba kõne all olnud ja seega staatuses *antud* või staatuses *saadaval* mingi muu informatsiooni kaudu (vt nt Chafe 1994, Gundel jt 1993). Kõike seda omandab laps võrreldes keele tuumgrammatikaga suhteliselt hilja, ja nagu näeme allpool, ei ole see areng 6–7 aasta vanuses veel lõpule jõudnud.

Talmy Givón (2005: 135) esitab põhjaliku nimekirja grammatilistest vahenditest, millega luuakse referentsiaalset sidusust. Selles on pronoomenid (mille hulgas

* Uurimust on toetanud Eesti Teadusfondi grantid nr 7492 ja 7388.

ka nn nullpronoomenid), aga ka süntaktilised kategooriad nagu käändemarkeering või passiiv. Siinses uurimuses vaadeldakse eesti keele referentsiaalsetest vahenditest substantiivifraase, mis võivad olla laiendatud määratlejaga (eelkõige *see* ja *üks*, nt *üks poiss, see tüdruk*) või olla laiendamata (*poiss, tüdruk*); personaalpronoomeneid ja pronoomenita verbivorme (nn nullpronoomeneid). Need indeksikaalsed vahendid aitavad suhtluspartneritel koos luua referentsiaalse raamistiku (Cornish 1999: 5).

Täiskasvanute narratiivides võetakse pidevalt arvesse referendi infostaatust: esimene mainimine ei kordu enam samasugusel kujul, pronoomenid vahetavad pidevalt referente jne. Jeanette Gundel jt (1993) eristavad oma antuse hierarhias kuut erinevat kognitiivset staatust, mis määravad referentsiaalse fraasi keelise vormi. Ehkki igal staatusel ei pruugi keeles olla üksühel seost mingi vormiga, on nad kõik siiski viitamisraamistiku seisukohast relevantset.

Laps omandab emakeele referentsiaalsed vahendid aegamööda. Eesti keele pronoomenite omandamist on käsitlenud eelkõige Maigi Vija (Vija 2007, Vija, Pajusalu 2009). Need käsitlused puudutavad aga keele arengu varasemaid etappe. Siinse artikli eesmärk on kindlaks teha, milliseks on kujunenud laste viitamisüsteem 6–7 aasta vanuses, s.t vahetult enne kooliminekut. Kuigi esimesed pronoomenid (*mina, sina, tema* ja *see*) ilmuvad lapse keelde juba keele omandamise kõige varasemal etapil (enamasti enne kaheaastaseks saamist), ei tähenda see veel, et kogu referentsiaalne süsteem oleks siis omandatud. Oletame, et just referendi infostaatuse muutumine ja selle muutuse väljendamine võiks lapsele olla liiga keeruline ka siis, kui lihtsamat tüüpi anafooriline viitamine on hästi omandatud.

Siinses käsitluses piirdume narratiividega, mis seab referentsiaalsete vahendite kasutusele selged piirid. Narratiivis ei viidata üldjuhul entiteetidele deiktiliselt, s.t entiteete ei ole olemas suhtlussituatsioonis, vaid nad eksisteerivad kõnelejate jaoks diskursuse referentsiaalse raamistikuna. See, kuidas selles vanuses lapsed viitavad vestluse füüsilises ruumis olevatele entiteetidele, jääb siinkohal käsitlusest täiesti välja.

Lapsed on suutelised piltide järgi narratiive looma juba vähemalt nelja-aastaselt (Foster-Cohen 1999: 177). Esimesed jutustused kasutavad valdavalt eksofoorilisi referentsiaalseid vahendeid, jutus puudub laiem referentsiaalne raamistik. Olulise muutuse ajana lapse referentsiaalsetes vahendites nimetatakse üldiselt kuendat eluaastat: inglise, prantsuse ja saksa keele põhjal tehtud uurimused näitavad, et lapsed hakkavad siis kasutama nn temaatilise subjekti strateegiat, mis seisneb kogu jutustuse ülesehitamises peategelas(t)e ümber ja pronoomenite reserveerimises ainult nendele peategelastele. Senistes uurimustes on tüüpiliselt kasutatud pildiseeriaid, kus tegutsevad poiss ja tüdruk, mis germaani ja romaani keelte puhul tähendab seda, et kasutusel on kaks 3. isiku pronoomenit – üks meessoost, teine naissoost tegelase jaoks. Alles hiljem hakkavad lapsed kasutama täiskasvanupärast anafoorilist strateegiat: lokaalse koreferentsi korral kasutatakse pronoomenit ja referentsi vahetuse korral nominaalfraasi. (Vt ülevaadet Hickmann 1995: 206) Umbes seitsmeaastaselt hakkab laps tajuma ennast kui jutustajat, kes võib ka sündmustele hinnanguid anda (Toolan 2001: 185–186).

Siinses artiklis vaatleme seega viitamisvahendeid vanuses, mil täiskasvanupärane narratiiv peaks juba kujunema hakkama, kuid etteruttavalt võib öelda, et erinevusi 6–7-aastaste laste ja täiskasvanute vahel siiski leidub.

2. Meetod ja materjal

Siinse uurimuse eesmärgiks on vaadelda referentsiaalsete nimisõnafaaside kasutust eesti keeles eelkooliealiste laste (6–7 aastat) suulistes narratiivides. Uurimuses oleme kasutanud 41 lapse narratiivi. Vaatleme, millised seaduspärasused ilmnevad laste jutustustes viitesuhte loomisel ning kas ja kuidas eristavad lapsed oma kõnes uusi (varem mainimata) ning vanu (varem mainitud) referente. Seega on lisaks laiendatud ja laiendamata substantiividele käsitletud ka pronoomenite ja määratlejate kasutust. Laste narratiive võrdleme ka 19 täiskasvanu jutustusega sama pildiseeria põhjal, et näha, kas laste ja täiskasvanute viitavate NP-de kasutuses esineb erinevusi, ja kui esineb, siis mida nende erinevuste põhjal võib öelda referentsiaalsete suhete omandamise kohta.

Analüüs põhineb jutustustel, mille loomiseks kasutati kolme viieosalist pildi-seeriat. Nimetatud pildiseeria alusel koostatud laste narratiive on varasemalt analüüsitud ka jutustuse sisulise ülesehituse (Adamka 2008, Soodla 2011), lausungite mahu ning grammatikavigade aspektist, sealhulgas erineva keelelise arenguga lapsi võrdlevalt (Soodla 2011). Täiskasvanute narratiivid on kogutud spetsiaalselt selle artikli tarbeks eesmärgiga saada võrreldavad andmed samasuguse täiskasvanuliku narratiivi kohta. Käesolevas artiklis analüüsitakse vaid jutustusi ühe pildiseeria alusel, millel oli kujutatud sündmus kelgumäel (lisa 1).¹ Vaadeldud on ainult nende laste jutustusi, kelle keeleoskusele oli lasteaiaõpetaja andnud hinnangu “tugev” või “keskmine”. Lapsi, kelle keeleoskust oli hinnatud “nõrgaks”, sinne uuring ei hõlma, sest artikli eesmärgiks on tavaarenguga laste jutustamisstrateegiatega uurimine. Pildiseerial kujutatud keskkond, tegelased ja tegevused olid eeldatavalt lastele tuttavad ning arusaadavad. Pildiseerial kujutatud sündmus vastas Nancy Steini ja Christine Glenni (1979) jutugrammatika mudelile (ingl *Story Grammar Model*), osutades mudeli sisukomponentidele (taust, käivitav sündmus, sisemine vastus, sisemine plaan, tegevus, tagajärg, reaktsioon, vt täpsemalt Soodla jt 2010).

Testimine viidi läbi individuaalselt, jutustused lindistati ning hiljem transkribeeriti. Testi läbiviija ja katseisik istusid laua taga, näod vastamisi. Testi läbiviija asetaskatseisiku ette lauale pildiseeria esimese pildi ning palus ülejäänud pildid järjestada nii, et neist tekiks jutuke. Kui katseisik järjestas pildid valesti, pani testija need ise õigesse järjekorda. Katseisikule anti piltide vaatamiseks aega ning paluti jutustada nende järgi üks lugu. Testija lisis, et tema pilte hästi ei näe ning palus seetõttu jutustada hästi arusaadavalt. Sellega püüti saavutada olukorda, kus katsealune ei saa kasutada ühiseid füüsilisest kontekstist (pildiseeria nägemisest) tingitud teadmisi ja kasutab seetõttu ainult tekstisiseseid (anafoorilisi) viitamistrateegiaid. Katseisik jutustas loo iseseisvalt, pikkade pauside tekkimisel võis testi läbiviija jutustajat ergutada vaid avatud repliikidega (“Jutusta edasi!”, “Räägi edasi!”). Erinevalt eelkirjeldatud laste jutustuste salvestustest ei palunud testi läbiviija täiskasvanud katseisikutel pilte järjestada.

Analüüsi tarbeks on kodeeritud kõik narratiivides esinevad NP-d: laste narratiivides kokku 519 NP-d, täiskasvanute narratiivides kokku 439 NP-d. NP-dena oleme käsitlenud kõiki sõnu ja fraase, millega saab viidata reaalses või diskursuses maailmas eksisteerivatele entiteetidele. Seega oleme käesolevas analüüsis lähtunud fraasi pragmaatilisest eesmärgist ning kodeerinud lisaks substantiividele tekstis

¹ Pildiseeria koostasid Piret Soodla ja kunstnik Jolana Laidma.

esinevad pronomendid, pärisnimed ja adpositsiooniga koos esinevad substantiivid/pronomendid, kui nende eesmärgiks on viidata konkreetsele asjale, olendile, nähtusele vms, mida käsitletakse terviklikuna ning ajas suhteliselt muutumatuna. (Vt ka nt Abbott 2010, Fraurud 1990.)

Kodeeritud NP-de juures on määratud järgmised tunnused: fraasi referent, määratleja (kui esineb), mainimiskord, fraasi struktuur, referendi semantiline klass, arv, kääne, lauseliige ja aktiveerituse aste. Kõiki tunnuseid on statistiliselt analüüsitud. Sinne artikkel keskendub eelkõige erinevatele fraasitüüpidele ning määratlejatele, kuna need väljendavad kõige otsesemalt referendi infostaatust.

3. Üldised viitamisstrateegiad laste narratiivides

Kasutatud pildiseerias on kaks peategelast, poiss ja tüdruk, kellele jutustamise käigus korduvalt viitama peab. Teistele entiteetidele viitamise sagedus varieerub sõltuvalt jutustaja valikust. Seega oleme keskendunud peamiselt just loo peategelastele viitavatele NP-dele, kuna need esinevad eranditult kõikides narratiivides. Laste jutustuste puhul võib narratiivide kui terviktekstide loomisel täheldada peamiselt kahe täiskasvanutest erineva viitamisstrateegia kasutust. Esimesel juhul kasutatakse narratiivi peaosalistele viitamiseks läbivalt pärisnime või substantiivi. Teise võimalusena esines ka narratiive, kus lapsed kasutasid jutus läbivalt vaid pronomeni või ei viidanud entiteedile eksplitsiitse keelilise vahendiga üldse. Kumbki neist ei vasta täielikult eelpool käsitletud temaatilise subjekti strateegiale, kuid kuna eesti keeles pole soolise eristusega kolmanda isiku pronomeneid, ei saagi mudel olla samasugune. Mõlemale strateegiale on omane see, et lapsed vahetavad referentsiaalset vahendit (nime või substantiivi asemel pronomeni või vastupidi) suhteliselt harvemini kui täiskasvanud. Referentsiaalsete vahendite kinnistumine referendile ja seega harvem vahetumine on omane ka temaatilise subjekti strateegiale (Hickmann 1995), kuid täpsemate paralleelide tõmbamine teistsuguse pronomenisüsteemiga keeltega vajab veel selgitamist. Järgnevalt vaatleme laste viitamisstrateegiaid lähemalt.

Esimene strateegia, pärisnimede või substantiivide kasutamine entiteedile viitamiseks, on siinsete narratiivide kontekstis n-ö kuulajasõbralikum. Jutustaja annab loo tegelastele nimed ning kasutab neid nimesid vastavale referendile viitamiseks kogu loo vältel. Juhul kui kõneleja kasutab samale referendile viitamiseks ka mõnda muud NP-d, siis on selleks pronomeni. Tüüpiliselt algab narratiivi referentsiaalne ahel tegelasele viitava fraasiga, mille puhul kõneleja viitab entiteedile *a* ning kuulaja, kellel enne ei olnud teadmisi ega vahendeid *a* identifitseerimiseks, loob kõnelejalt kuuldu põhjal sama viitesuhte entiteedi *a* jaoks ka oma referentsiaalses raamistikus. Personaalpronomendid, mis tulevad selle strateegia puhul kasutusse vaid loo lõpus, on eelnenud pärisnimede anafoorid.

Narratiiv (1) on hea näide esimesest jutustamisstrateegiast, kus domineerib pärisnimede kasutus. Esimeses kuues lauses viitab jutustaja kõne all olevatele entiteetidele vaid pärisnimedega (neljandas lauses esinev *sul* on siin erandiks, kuna esineb otsekõnes). Viimastes lausetes ei ole kõneleja aga enam läbivalt pärisnimesid kasutanud, vaid tüdrukule viitamiseks kasutatakse personaalpronomeni *ta*. Sealjuures viimases lauses on subjekt (*nad*) lausest hoopis välja jäetud. Erinevus

tüüpilisest täiskasvanu narratiivist on suhteliselt rohkes pärisnimede kasutuses ja kollektiivse referendi (poiss ja tüdruk koos) sisetoomises verbivormiga. Jutustaja jätkab viimases lauses *siis läksid rahulikult ülesse mäele* eelmise lause rinnastussuhet, kuigi referent ei ole enam ainult Jaan.

- (1) Jaan ja Grete läksid kelgutama mäele. Grete lasi alla. siis Grete läks vas-
sõitis vastu puud. Jaan ütles kas sul kõik on korras. Grete ütles ei ole. siis
Jaan lasi ka alla ja tuli Gretele appi. Jaan aitas ta ilusti lumest puhtaks ja
aitas tal mütsi pähe panna ja siis läksid rahulikult ülesse mäele.

Narratiivis (2) on näha, et laps on järjekindlalt kasutanud laiendamata substantiive *poiss* ja *tüdruk* kogu loo ulatuses, nii nagu näites (1) oli kasutatud nimesid. Pronoomenit on jutustamisel kasutatud vaid korra, narratiivi viimases lauses, viitamaks pronomeni *nad* mõlemale narratiivi tegelasele. Nagu ka eelmise narratiivi puhul ei teki kuulajal referentide tuvastamisega probleeme, kuid täiskasvanu narratiividega võrreldes on referentsiaalsete vahendite hulk piiratum, sest puuduvad määratlajaga fraasid ja pronomeniid ei vaheta referenti.

- (2) tüdruk ja poiss läksid kelgutama. tüdruk sõitis vastu puud. siis sõitis poiss
ka mäest alla. aga poiss ei sõitnud vastu puud. siis poiss pani tüdrukule
mütsi pähe ja nad läksid uuesti mäest üles.

Narratiivis (3) võib märgata aga laialdasemat pronomeniite kasutust – vastupidiselt lugudele (1) ja (2), kus domineerivad pärisnimed või substantiivid, on selles loos põhirõhk pronomeniitel. Substantiivi on siin kasutatud vaid uue referendi diskursusesse toomiseks, kusjuures korra on selleks täiendavalt kasutatud ka indefiniitset määratlejat *üks*². Nagu ka kahes eelnevas loos, on siingi referentide infostaatuse tuvastamine kuulajale suures osas võimalik isegi siis, kui ta eelnevalt ei tea, mis sündmusest parasjagu räägitakse. Sellega meenutab narratiiv (3) vägagi täiskasvanute omi ja selles ei saa enam näha lapsepärast nimeandmisstrateegiat. Ainsa mitte täielikult täiskasvanupäraste viitamisenähtena võiks esile tuua fraasi *poiss*: võiks küsida, *kust see poiss sinna sai?* Siin lülituks justkui sisse narratiivides (1) ja (2) analüüsitud mudel, mille järjekindlal rakendamisel oleks ka tüdrukule viidatud läbivalt NP-ga *tüdruk*. Meie arvates on näites (3) tegemist üleminekuga nimeandmisstrateegialt täiskasvanupäraste referendi infostaatuse varieerimise strateegiale.

- (3) üks tüdruk lasi liugu. aga ta laskis lasi niiviisi liugu et ta läks vastu puud.
sis ta hakkas nutma ta sai haiget. sis poiss aitas tal mütsi pähe panna ja siis
nad läksid mäest ülesse tagasi.

Esimese strateegia alusel loodud narratiivides esines ka lauseid, kus referendile ei viidatud nähtava keelelise vahendiga. Selliste nullpronomeni lausete hulk ühe narratiivi kohta oli aga väike. Ainsuse 3. isiku pronomeni väljajätt on üldjuhul lähedane täiskasvanupäraste keelekasutusele. Kui referenti on vahetult enne mainitud, stseen ei muutu või on tegu vahetult üksteisele järgnevate sündmuste jadaga liitlauses, on jutustuse eelnev kontekst piisav, et kuulaja saaks nullpronomeni väljendatud referendi tuvastada (*tüdruk kelgutab mäest alla. sõitis vastu puud ja sai haiget*). Ometigi ei sarnanenud laste nullpronomeni kasutus täielikult

² Siin paistab välja, et lapsel on ilmselt varem kuulnud lugudest sündinud "loo alguse mudel": *üks X tegi Y*. Seda, et just lugude algused on n-õ paremini vormistatud, on märgatud ka teiste keelte põhjal tehtud uurimustes (Hickmann 1995).

täiskasvanute omaga. Nimelt on lapsed referendi sageli nimetamata jätnud narratiivi viimases lauses, kus tuleks kasutada mitmuse 3. isiku pronoomenit *nad* (4), (5). Selline kontekst loo lõpus esines 32 lapse narratiivis, neist 15 juhul kasutati pronoomenit *nad* ning 17 juhul ei kasutatud. Ülejäänud üheksa lapse narratiivis ei esinenud taolist kollektiivsele referendile viitavat fraasi üldse Seega on näha, et olukorras, kus on vaja luua kollektiivreferent (poiss ja tüdruk), jätavad enam kui pooled lapsed selle moodustamata, justkui oletades, et referent püsib sama.

(4) siis poiss andis talle mütsi ja \emptyset läksid ülesse³.

(5) siis Kevin tuli püsti ja ait- pani Mariale ilusti mütsi pähe. siis läksid \emptyset uuesti mäe otsa.

Ainsuse 3. isiku pronoomeni mittetäiskasvanupärast väljajätu esineb seda strateegiat kasutanud lastel märkimisväärselt vähem. Väikese kõhklusega võib selliseks pidada näidet (6), kus esineb kaks subjektita verbi, kuid neid võiks ka käsitleda viimase pronoomeni *ta* rinnastatud predikaatidena.

(6) see tüdruk sõitis alla ta vaatas kuhu sõitis aga ta ei näinud puud. \emptyset kogemata sõitis puule otsa ja nuttis.

Kui valdavalt valivad lapsed jutustamiseks eelkirjeldatud strateegia, siis esines ka narratiive, mille puhul kasutati mittetäiskasvanupäraselt palju kas personaalpronoomeneid või pronoomenita verbi. Selliste narratiivide puhul ei ole kuulajal võimalik referente tuvastada, kui ta kõne all olevate sündmustega kursis ei ole. On võimalik, et nende narratiivide puhul eeldas laps, et ka kuulaja näeb pildiseeriat (ehkki testi läbiviija rõhutas, et tema pilte hästi ei näe) ning seega ei pidanud laps vajalikuks referente täpsemalt tuvastada. Samas on mõeldav, et need lapsed ei ole veel täielikult omandanud sidusat viitamist või ei ole nende jutustamisoskus veel teiste lastega samal tasemel.

Narratiiv (7) on näide juhtumist, kus laps ei kasutanud terve narratiivi vältel kordagi substantiivset NP-d, et subjekti positsioonis esinevale loo peategelastele viidata. Objekti rollis olevale referendile on korra viidatud pronoomeniga (*tal*).

(7) kelguga laseb mäest alla. ja lasi vastu puud ja siis sai aiget. *T: noo jutusta edasi. noh jutusta edasi.* aitas tal mütsi panna pähe. siis tulid koos üles.

Järgmises narratiivis (8) aga on laps jutustamist väga korralikult alustanud, kasutades nii substantiive kui määratlejaid (seega täiskasvanupärast strateegiat). Ent loo teises pooles ei ole enam täit selgust referentidele viitamisel, kuna laps hakkab kasutama pronoomenit *ta* subjekti positsioonis igas lausungis. Sealjuures ei säili sidusad viiteahelad, sest jutustaja tekstist ei selgu enam, millisele referendile pronoomen täpselt viitab. Seetõttu on kuulajal neist narratiividest keeruline aru saada.

(8) ükskord ilusal soojal päeval läksid lapsed kelgutama. siis kui nad siis kui oli üks laps poole mäe peal kuk- libises ta kelgu pealt ära. siis ta libises vastu puud. siis ta sai haiget. ja teine laps ruttu lasi kelguga mäest alla vaatama mis tal juhtus. ja siis ta nuttis. siis ta läks lohutas? siis ta läks rõõmsaks ja läksid jälle uisut- kelgutama.

Täiskasvanute narratiivides oli referentide infostaatuse palju selgemalt väljendatud. Erinevalt lastest, kes alustasid jutustamist enamasti konkreetsest sündmusest (nt *tüdruk laseb mäest alla*), juhatasid kõik täiskasvanud oma narratiivi sisse üldist situatsiooni kirjeldava fraasiga ning tõid kohe diskursusesse mõlemad peategelased (*poiss ja tüdruk kelgutavad, lapsed kelgutavad* vms). Mitte ükski täiskasvanu ei kasutanud jutustamiseks läbivalt pronoomeneid. Lisaks substantiividele ja pronoomenitele kasutati ka määratlejatega NP-sid (*see tüdruk, üks tüdruk, teine laps* jne). Lisaks sellele on täiskasvanud (9), (10) võrreldes lastega oluliselt rohkem rõhku pannud tegelaste ja situatsiooni kirjeldamisele ega piirdu pelgalt pildiseerial toimuvate episoodide mainimisega. Olulise erinevusena laste ja täiskasvanute vahel võib välja tuua ka relatiivlause kasutuse (10). Kui täiskasvanute narratiivides esines relatiivlauseid kuuel korral, siis laste narratiivides ei kasutatud relatiivlauseid korraga. Et relatiivlauseid võib käsitleda ühe koherentse viiteahela loomise vahendina, siis nähtub siitki, et eelkooliealised lapsed ei kasuta veel kõiki täiskasvanupäraseid viitamishendideid.

- (9) ä sis jah poiss ja tüdruk kelgutavad. ääh tundub et tüdruk laseb parasjagu mäest alla ja söidab vastu puud. ja poiss vaatab mäe otsas või nätuke üleval pool, ahhetab. ja siss tüdruk parasjagu nutab kui poiss ka röömsalt akkab alla laskma ja sis kui ta on nagu alla jõudnud sis ta akab seda tüdrukut lohutama. võtab tal mütsi peast ära ja vaatab tüdruk on vist pea ära löönud ja sis nad võtavad käest kinni ja röömsalt lähevad mäest üles tagasi.
- (10) kaks last on väga röömsad. üks laps hoiab kelku käes ja teine laps laseb kelguga alla. oi mis juhtus! **see laps kes lasi kelguga alla**, lendas vastu puud, kelguga. teine laps vahib o ee imeliku molliga. nüüd. teine laps laseb **sellele lapsele kes puu otsa sõitis** kelguga järele, et talle appi minna jaa pead paitada arvatavasti. jaa! tundubki et nad ä üks lohutab teist last. kelgud pillapalla laiaili. nüüd lähevad mõlemad uuesti suusamäe otsa. et sealt tõenäoliselt siis lõbusasti uuesti alla lasta, puule otsa.

4. Fraasitüüpide võrdlus laste ja täiskasvanute narratiivides

Järgnevalt vaatleme põhjalikumalt laste narratiivides esinenud referentsiaalseid fraasitüüpe ning nende kasutust. Et näha, mille poolest 6–7 aasta vanused lapsed täiskasvanutest erinevad, oleme võrdluseks lisanud ka täiskasvanute andmed. Tabelis 1 on esitatud erinevate referentsiaalsete fraasitüüpide sagedus võrdlevalt täiskasvanute ja laste narratiivides. Kokku esines laste narratiivides 11 eri fraasitüüpi. Enim on lapsed ootuspäraselt kasutanud laiendamata substantiive (N; 44%) ja pronoomeneid (PRO; 26%), sageduse poolest tõusevad esile ka pärisnimed (PN), kaassõnaga NP-d (ADP + N; N + ADP) ning määratlejaga NP-d (DET + N). Muude fraaside kasutussagedus on marginaalne. Seega kasutavad lapsed viitamiseks lihtsaid fraasikonstruktsioone, mille eesmärgiks on vaid referendi mainimine. Täiendavat informatsiooni referendi omaduste kohta, näiteks selle omaduste kirjeldamist adjektiividega või infostaatuse täpsustamist määratlejaga, esitavad lapsed harva, mistõttu nii määratlejaid kui adjektiive esineb vähe. Sama tendentsi – tegelaste

iseloomustamata jätmist – on märgatud ka nt ingliskeelsetes laste narratiivides (nn konnalugudes, Berman 2001).

Ka täiskasvanute narratiivides on enim kasutatud fraasitüüpideks laiendamata substantiiv (36%), pronoomen (23%), adpositsiooniga laiendatud NP, pärisnimi ja määratlejaga NP. Samas esines täiskasvanute narratiivides eri fraasitüüpe rohkem (16 tüüpi) ning komplekssemate fraaside sagedus oli suurem kui lastel.

Tabel 1. Fraasitüüpide jagunemine täiskasvanute ja laste narratiivides

Fraasitüüp	Täiskasvanud (%)	Lapsed (%)
N	36,0	43,5
PRO	22,7	25,6
N + ADP	12,3	8,7
PN	7,4	8,5
ADP + N	4,9	4,8
DET + N	4,9	5,4
ADJ + N	3,4	0,8
NUM + N	2,0	0
ADJ + N + ADP	1,7	0
ADJ + NP	1,5	0
PRO + ADP	1,0	1,9
DET + ADJ + N	0,7	0
DET + N + ADP	0,5	0
PN + ADP	0,5	0
ADJ + ADJ + N	0,2	0,2
NUM + ADJ + N	0,2	0
N + N	0	0,4
ADV + N + ADP	0	0,2
Fraase kokku	429 (100%)	481 (100%)

Kõige enam esinevad referentsiaalsed fraasid lauses subjektina ning objektina (sarnased tulemused on saadud ka inglise, saksa, prantsuse ja mandariini hiina laste narratiivide põhjal, vt Hickmann ja Hendriks 1999). Kuna siinses materjalis oli tegemist ainult ühe sündmustikuga ja seega on ka semantiliste rollide hulk piiratud (näiteks inimreferent saab objektina esineda õieti ainult lohutamisolukorras), ei saa teha väga kaugeleulatuvaid järeldusi eri lauseliikmete kodeerimise kohta. Esitame siiski laste narratiivide NP-de esinemise lauseliikmeti tabelis 2, kus on näha eri positsioonides esinevad fraasitüübid ning nende sagedus protsentides laste narratiivides. Ilmnevad olulised erinevused subjekti ja objekti rollis olevale referendile viitamisel – kui pronoomeneid kasutatakse mõlemas positsioonis enam-vähem võrdselt (subjekti puhul 29%, objekti puhul 26%), siis laiendamata substantiivide kasutuses on erinevus märgatav (subjekti puhul 48%, objekti puhul 71%). Ka määratlejaga fraase esineb nii subjekti kui objekti positsioonis, mõnevõrra rohkem siiski subjekti puhul (vastavalt 7% ja 4%). Pärisnimed aga objekti positsioonis ei esine, küll aga subjekti positsioonis (15%). Laste narratiivides viidatakse objekti rollis olevale referendile vaid kolme erineva fraasitüübiga, subjektidele viie eri fraasitüübiga, adverbiaali positsioonis esinevad aga kõik 11 narratiivides esinenud fraasitüüpi.

Tabel 2. Fraasitüüpide ja grammatiliste rollide suhe laste narratiivides (%)

Fraasitüüp	Lauseliige (esinemiskorrad)		
	ADV (206)	OBJ (51)	SUBJ (245)
N	36,9	70,6	48,2
ADJ + ADJ + N	0,5	0	0
ADJ + N	1,0	0	0,8
ADP + N	11,2	0	0
ADV + N + ADP	0,5	0	0
DET + N	3,4	3,9	6,9
N + ADP	20,4	0	0
N+ N	1,0	0	0
PN	2,4	0	14,7
PRO	18,4	25,5	29,4
PRO + ADP	4,4	0	0

5. Personaalpronoomenid laste narratiivides

Pronoomenitel on jutustamisel oluline roll, kuna need võimaldavad lühidalt ja selgelt viidata referendile, keda jutustamise käigus korduvalt nimetada tuleb. Samas on pronoomenite kasutus viitamise aspektist lapsele keerukam kui substantiivide kasutus (vt nt Chien, Wexler 1990). Kuna pronoomeni tähendus selgub alles kontekstis, peab kõneleja pronoomenit kasutades olema kindel, et ka kuulajal on vajalik info pronoomeniga viidatu tuvastamiseks (Vija, Pajusalu 2009). Vaatluse all olevates narratiivides ei olnud jutustajatel tarvis kasutada 1. ja 2. isiku pronoomeneid (v.a mõned üksikud korrad, kui jutustaja kasutas otsekõnet), küll aga 3. isiku pronoomenit loo tegelastele viitamiseks.

Narratiivides kasutavad lapsed pronoomenite lühikesi vorme *ta* ja *nad*. Pikka vormi kasutati väga vähe ning see esines kas rõhulises positsioonis, näiteks koos rõhumäärsõnaga *ka* (*sõitis tema ka otsa*), või koos adpositsiooniga (*tema juurde*, *tema poole*). Seega lauseid kontekstist lahutatuna vaadeldes kasutavad lapsed 3. isiku personaalpronoomeneid grammatiliselt korrektselt (11)–(12), mõningane erinevus täiskasvanute ja laste pronoomenikasutuses ilmneb aga narratiivi kui terviku puhul (4)–(8).

- (11) et üks tüdruk hakkas liugu laskma. siis **ta** läks vastu puud siis **ta** hakkas nutma.
- (12) ee: tüdruk sõitis kelguga. aga **ta** sõitis vastu puud sest ta ei näinud et puu oli ees. siis poiss sõitis järgi **talle**.

Eelpool kirjeldasime pragmaatiliselt ebaselget pronoomenite (ja pronoomenita verbivormide) kasutust (7), (8). Sel puhul ei saa kuulaja olla täiesti kindel, millisele referendile pronoomen parajasti viitab, seda eriti juhul, kui kuulaja ei ole loo kontekstiga varasemalt kursis. Huvitaval kombel esineb selline pronoomeni kasutus aga pigem narratiivi teises osas, jutustuste alguses on viitesuhted selged. Näide (13) pärineb narratiivi teisest poolest, kusjuures narratiivi esimene pool on

jutustatud selgelt ja arusaadavalt ning referentidele on viidatud substantiivi või pärisnimega. Seevastu loo teises pooles hakkab laps mõlemale referendile viitama pronoomeniga *ta*. Sarnane muster ilmneb ka näites (8). Võimalik, et pronoomenite kasutus muutub ebapädevaks just narratiivi teises pooles, kuna jutustatav sündmus muutub siin ise intensiivsemaks ning seega keskenduvad lapsed pigem sündmuste käigule kui tegelastele endale.

- (13) siis vend lasi alla. siis õdel kukkus müts peast ära. aga **ta** sõitis **tast** koge-mata natuke mööda aga jooksis **tema** poole. siis pani **talle** mütsi pähe.

Üldiselt on täheldatud, et lastel on tendents oma lausungitest argumente – nii subjekti kui objekti – välja jätta, seda nii keeltes, mis nõuavad argumenti, kui ka keeltes, mis lubavad argumendi väljajätu (Huang 2010). Eesti keeles on tavaline subjekti positsioonis oleva pronoomeni väljajätt, objekt üldjuhul nullpronoomenina ei esine. Meie laste andmestikus esines nullpronoomenit subjekti positsioonis sageli (14), (15), vt ka näide (4).

- (14) siis sõitis kelguga alla ja ütles pole viga läeme tagasi.

- (15) ja aitab teda ja siis lähevad jälle kelgutama.

Täiskasvanute narratiivides oli pronoomenite kasutus märgatavalt sidusam kui laste narratiivides. Ei esinenud ulatuslikku pronoomeni väljajätu terve loo vältel ega ka pronoomeni liigkasutust. Pronoomeneid ei kasutata referendi esimesel mainimisel, vaid ainult siis, kui kõneleja on veendunud, et kuulaja on pronoomeniga viidatava referendi juba oma referentide raamistikku lisanud. Erinevalt lastest võib täiskasvanute narratiivides täheldada ka 1. isiku pronoomeni kasutust (kolm täiskasvanut 19-st) – sellega soovitakse narratiivi alguses rõhutada, et tegu on jutustaja isikliku tõlgendusega sündmusest (16).

- (16) **Ma** arvan et kaks sõpra kelgutavad mäe peal.

6. Määratlejad laste ja täiskasvanute narratiivides

Ehkki määratlejaid esines lastel narratiivide jutustamisel protsentuaalselt vähe (vt tabel 1), on neil viitamise ja entiteetide infostaatuse seisukohast oluline roll, mistõttu ei saa jätta neid siinses analüüsis käsitlemata. Varasemate uuringute põhjal (Laak 2006, Vija, Pajusalu 2009) on järeldatud, et lapsed vanuses 4–5 aastat kasutavad vestlustes suhteliselt regulaarselt indefiniitseid määratlejaid *üks* ja *mingi* uue referendi ning definiitset määratlejat *see* tuntud referendi mainimiseks. Ka narratiivides tõusid selgelt esile määratlejad *see* ja *üks*, määratlejatena esinesid veel ka *sealt*, *see*, *sinna* ja possessiivpronoomen (*oma*, *tema*, *ta*). Narratiivides ei kasutanud lapsed aga kordagi indefiniitset määratlejat *mingi*. Kui tavaliselt kasutatakse määratlejat *üks* uue referendi mainimiseks ning demonstratiivpronoomen *see* markeerib tuntud referenti (vt Vija, Pajusalu 2009, Pajusalu 1997), siis narratiivide puhul on huvitav, et valdavalt olid määratlejad kasutusel vaid uuele referendile viitava NP koosseisus, antud ja saadaval olevale referendile viidati määratleja kasutusega oluliselt harvem (tabel 3). Seega on määratlejad laste narratiivides kasutusel valdavalt referendi

esimesel mainimisel, kuna ka infostaatus *saadaval* tähendab üldjuhul seda, et selge keelelise vahendiga viidatakse referendile tekstis esimest korda. Märkimist väärrib tõsiasi, et ka määratlejat *see* kasutati pigem uuele kui vanale referendile viitamiseks.

Tabel 3. Määratlejad ja nende infostaatus laste narratiivides

Määratleja	Aktiveerituse aste	Kokku
POSS.PRO	uus	2
<i>sealt</i>	antud	1
<i>see</i>	antud	3
	uus	6
<i>teine</i>	saadaval	1
	uus	2
<i>üks</i>	saadaval	3
	uus	9
<i>sinna</i>	uus	1
Kokku		28

Määratlejate kasutus täiskasvanute narratiivides on mõneti erinev laste narratiivides. Tabelist 4 selgub, et sagedasemad määratlejad on täiskasvanutel hoopis possessiivpronoomen ja *teine*, samas kui *see* ja *üks* on neist poole võrra vähem kasutatud leidnud. Samuti ei leidu täiskasvanute jutustustustes proadverbe *sealt* ja *sinna* määratleja positsioonis. Seevastu esineb täiskasvanutel aga indefiniitne määratleja *mingi*. Erinevalt laste narratiivides ei ole täiskasvanute puhul üht selgelt esile tõusvat infostaatust, mille puhul määratlejaid kasutatakse, pigem on määratlejate kasutuse maht sarnane nii uute kui vanade referentide korral.

Tabel 4. Määratlejad ja nende infostaatus täiskasvanute narratiivides

Määratleja	Aktiveerituse aste	Kokku
<i>mingi</i>	uus	4
POSS.PRO	antud	4
	saadaval	2
	uus	2
<i>see</i>	antud	3
	uus	1
<i>teine</i>	antud	4
	saadaval	2
	uus	1
<i>üks</i>	saadaval	1
	uus	2
Kokku		26

Määratlejate kasutus laste narratiivides on huvitav, kuna materjalist selgub, et kui lapsed määratlejaid kasutavad, siis valdavalt referendi esimesel mainimisel. Näited (17)–(19) on määratleja *see* kasutuse kohta narratiivi alguses esimeses või teises lauses, kui referente *tüdruk* ja *poiss* mainitakse alles esimest korda.

- (17) läks niimoodi et **see tüdruk** sõidaks alla
- (18) **see tüdruk pani** sõitis kelguga vastu puud
- (19) *see see ta* hakkas laskma. *siis tuli* siis **see poiss** nägi siis ta kukkus

Näide (19) on omapärane, kuna selles esinevat esimest demonstratiivpronoomeni võiks tõlgendada kui eneseparandust või kohahoidjat (ingl *placeholder*, vt nt Keevallik 2011), mitte kui määratlejat. Laps otsib siin sobivat väljendit referendi (*tüdruk*) esmakordseks mainimiseks ning *see* ei kuulu ilmselt NP juurde. Samas järgnevas lausungis on *see* selgelt määratleja rollis, ehkki ka poissi mainitakse selles narratiivis esimest korda.

Kui võrrelda demonstratiivpronoomeni *see* ja indefiniitse pronoomeni *üks* kasutuskontekste, näeme, et need ei ole laste narratiivides nii selgelt erineva funktsiooniga, nagu võiks arvata. Täiskasvanute narratiivides seevastu on *see* ja *üks* kasutusel selgelt erinevates kontekstides – *see* kontekstis juba olemasolevale referendile viitamiseks, *üks* uue referendi diskursusesse toomiseks. Näited (20)–(22) on kõik laste jutustuste esimesed laused, mis toovad diskursusesse sisse uue referendi, mistõttu määratleja *üks* kasutus on neis lausetes igati kohane. Samas ei erine need laused oma funktsiooni poolest lausetest (17)–(19), mille eesmärgiks oli samuti uue referendi tutvustamine.

- (20) et **üks tüdruk** hakkas liugu laskma.
- (21) **üks tüdruk** laseb alla ja *siis* siis ta läks vastu puud peaga.
- (22) siin **üks tüdruk** lasi alla siis ta kukkus puu otsa

Ei saa siiski kindlalt väita, et lapsed ei tee vahet definiitsel ja indefiniitsel määratlejal. Näiteks on Michael Tomasello (2003: 209) arvanud, et inglise keelt kõnelevad lapsed ei käsitلة definiitsust kui kahe alamklassiga kategooriat, vaid kasutavad definiitsust ja indefiniitsust kui kaht eraldiseisvat kategooriat. Michèle Kail ja Maya Hickmann (1992) on aga uurinud prantsuse lapsi ning leidnud, et nooremad lapsed (nende uurimuses samuti kuueaastased) eelistavad kasutada pigem definiitset määratlejat indefiniitse asemel, ka seal, kus kontekst nõuaks indefiniitset määratlejat (st juhul, kui kuulajal ei ole definiitse referendi tuvastamiseks piisavalt infot). Ka inglise keelt kõnelevad lapsed kasutavad väga sageli definiitset artiklit *the* neis kontekstides, kus täiskasvanupärane keelekasutus eeldaks indefiniitse artikli *a* kasutamist (Wexler 2011).

Klassikaline seletus definiitse artikli liigkasutusele on Piaget' st lähtuv egotsentrilisuse teooria, mille kohaselt on lastel keeruline panna end kuulaja positsiooni. Kui lapse referentsiaalses raamistikus on üks esilduv referent, siis kasutab ta seda definiitsena ka siis, kui kuulaja jaoks sellist esilduvat referenti veel ei eksisteeri. (Piaget 1955, Maratsos 1976) Annette Karmiloff-Smith (1979: 72) on definiitse artikli kasutust indefiniitse asemel seletanud sarnaselt, lisades, et definiitse artikli kasutus laste puhul on deiktiline selles mõttes, et definiitne artikkel osutab objektile, mis parasjagu lapse tähelepanu fookuses on. On ka arvatud, et lapsed kasutavad artikleid pigem spetsiifilisuse kui definiitsuse markeerimiseks (Hickmann 1995: 205). Ent

inglise ja prantsuse keel on artiklikeeled, eesti keeles grammatilist artiklit sama grammatiseerunud kujul ei esine. Seega on võimalik, et eesti lapsed tajuvad keele grammatilise struktuuri erinevuse tõttu definiitsuse kateooriat hoopis teisiti. Pildiseeria alusel jutustatud narratiivide puhul on muidugi ka võimalik, et laps kasutas demonstratiivi *see* osutamaks konkreetsele entiteedile pildil, ehkki katse läbiviija (kuulaja) oli öelnud, et tema pilte hästi ei näe. Meie andmete põhjal ei saa öelda, et lapsed kasutaksid pigem definiitset määratlejat, aga ilmneb tendents, et lapsed eelistavad uute referentide nimetamisel kasutada pigem definiitset NP-d (nt ilma määratlejata NP-d, pärisnime).

7. Kokkuvõtte ja arutelu

Siinses artiklis analüüsisime referentsiaalsete NP-de kasutust laste narratiivides. Uurimismaterjaliks on pildiseeria alusel loodud 6–7-aastaste laste narratiivid, võrdlusena on käsitletud ka täiskasvanute narratiive sama pildiseeria põhjal. Uuringust selgub, et 6–7-aastased lapsed ei ole veel samasugused jutustajad nagu täiskasvanud. Juba esimesel vaatlusel on näha, et lapsed ei kirjelda ei tegelasi ega ümbrust, seega ei kasuta peaaegu üldse adjektiive ega relatiivlauseid. Samuti ei hinda lapsed olukorda ega anna muud taustainfot.

Konkreetsemalt keskendusime selles artiklis referentsiaalsete fraaside kasutusele nii üldise viitamisstrateegia kui konkreetsete fraasitüüpide kasutusele kaupa.

Üldises jutustamisstrateegias on näha kahte tüüpi erinevust täiskasvanulikust narratiivist: laps kas annab igale tegelasele nime (pärisnime või üldnime *poiss* ja *tüdruk*) ja kasutab samal kujul läbi narratiivi või ei kasuta üldse viitavaid nominaalvorme ja esitab kogu sündmustiku ainult verbi isikuliste vormide kaudu. Laste narratiivides on seega võrreldes täiskasvanute sama pildiseeria põhjal loodud narratiividega vähem näha infostaatuse mõju referentsiaalsete vahendite kasutusele: referendile viidatakse läbi narratiivi sama fraasiga. Tegelikult on 6–7-aastaste laste puhul erinevus täiskasvanulikust narratiivist eelkõige skalaarne: mõne lapse jutustus sarnaneb täiskasvanu jutustusele rohkem, mõnel vähem. Enamik lastest suudab luua arusaadava narratiivi, näha on ka selget seaduspära, et jutustuste algusosas on referentsiaalsed fraasid täiskasvanulikumad kui lõpuosas.

Esineb ka kontekste, kus lapsed täiskasvanulikule keelele iseloomulikke referentsiaalseid vahendeid sageli väldivad. Meie andmestikus ilmnes see eriti mitmuse 3. isiku pronoomeni puhul sündmustiku selles kohas, kus *poiss* ja *tüdruk* asusid koos tegutsema, seega tekkis uus kollektiivne referent. Lapsed aga jätkavad viiteahelat, nagu oleks referent sama, mis eelmises lauses, kus tegelikult kõneldi vaid ühest lapsest, ja esitavad sündmuse subjektitu verbi kaudu. Seevastu ainsuse 3. isiku pronoomeneid kasutavad lapsed palju, kuid ei suuda kahe tegelasega sündmustiku kirjeldamisel alati tagada seda, et pronoomenid moodustaksid sidusa viiteahela. Ka viitavate substantiivifraaside täpsem analüüs näitas, et referendi infostaatuse määratlemise vahendeid on jutustamisel kasutatud täiskasvanukeelest erinevalt. Eriti märgatav on see määratlejate (*see*, *üks*) kasutuses – eelkoolialised lapsed ei erista süstemaatiliselt uusi ja vanu referente ega suuda end veel samal määral kuulaja rolli panna nagu seda teevad täiskasvanud.

Käesolev uurimus erineb analoogilistest anafoorilise viitamise uurimustest indoeuroopa keelte põhjal selle poolest, et piltidel kujutatut ei võimalda kasutada

läbivalt ühte pronoomenit ühe tegelase kohta, sest poisile ja tüdrukule ei saa viidata erinevate pronoomenitega vaid tuleb leida mingi teine referentsiaalne strateegia. Seega nõuab pildidel kujutatut erinevat jutustamisstrateegiat kui varasemates indoeuroopa keeltes põhinevates uurimustes ja siinkohal ei ole päriselt võimalik öelda midagi selle kohta, kas temaatilise subjekti strateegia samal viisil ka eesti laste puhul kehtiks. Edaspidi tulekski sama tüüpi narratiive koguda ka ühe tegelasega sündmuste põhjal.⁴

Oleme siinses artiklis näidanud, et kuigi lapse grammatika areng on 6–7 aasta vanuses juba väga arenenud, et ole pragmaatilised oskused, muuhulgas referentsiaalsete fraaside kasutus, selles vanuses veel täiesti täiskasvanupärane. Selle põhjuseks on ilmselt ühelt poolt veel ebapiisav oskus panna end kuulaja olukorda ja teiselt poolt mõningane ebakindlus pidevalt vahelduvate referentide ja neile viitavate fraaside voos orienteerumises.

Lühendid

ADJ	adjektiiv	OBJ	objekt
ADP	adpositsioon	PN	pärinimi
ADV	adverb/adverbiaal	POSS	possessiivne
DET	määratleja	PRO	pronoomen
N	nimisõna	SUBJ	subjekt
NP	nimisõnafraas	T	katse läbiviija
NUM	numeraal		

Viidatud kirjandus

- Abbott, Barbara 2010. Reference. Oxford: Oxford University Press.
- Adamka, Aive 2008. 6–8 aasta vanuste laste suuliste jutustuste makrostruktuur. Magistritöö. Tartu: Tartu Ülikool.
- Berman, Ruth A. 2001. Setting the narrative scene: How children begin to tell a story. – Keith E. Nelson, Ayhan Aksu-Koç, Carolyn E. Johnson (Eds.). *Children's Language*, Vol. 10. Developing Narrative and Discourse Competence. Mahwah, NJ: Lawrence Erlbaum, 1–31.
- Chafe, Wallace 1994. Discourse, Consciousness and Time. The Flow and Displacement of Conscious Experience in Speaking and Writing. Chicago: The University of Chicago Press.
- Cornish, Francis 1999. Anaphora, Discourse, and Understanding. Evidence from English and French. Oxford: Clarendon Press.
- Chien, Yu-Chin; Wexler, Kenneth 1990. Children's knowledge of locality conditions in binding as evidence for the modularity of syntax and pragmatics. – *Language Acquisition*, 1 (3), 225–295. http://dx.doi.org/10.1207/s15327817la0103_2
- Foster-Cohen, Susan 1999. An Introduction to Child Language Development. London, New York: Longman.
- Fraurud, Kari 1990. Definiteness and the processing of NP's in natural discourse. – *Journal of Semantics*, 7 (4), 395–433. <http://dx.doi.org/10.1093/jos/7.4.395>
- Givón, Talmy 2005. Context as Other Minds. The Pragmatics of Sociality, Cognition and Communication. Amsterdam/Philadelphia: John Benjamins PC.
- Gundel, Jeanette; Hedberg, Nancy; Zacharski, Ron 1993. Cognitive status and the form of referring expressions in discourse. – *Language*, 69 (2), 274–307. <http://dx.doi.org/10.2307/416535>

⁴ Täname sellele tähelepanu juhtimise eest anonüümset retsensenti.

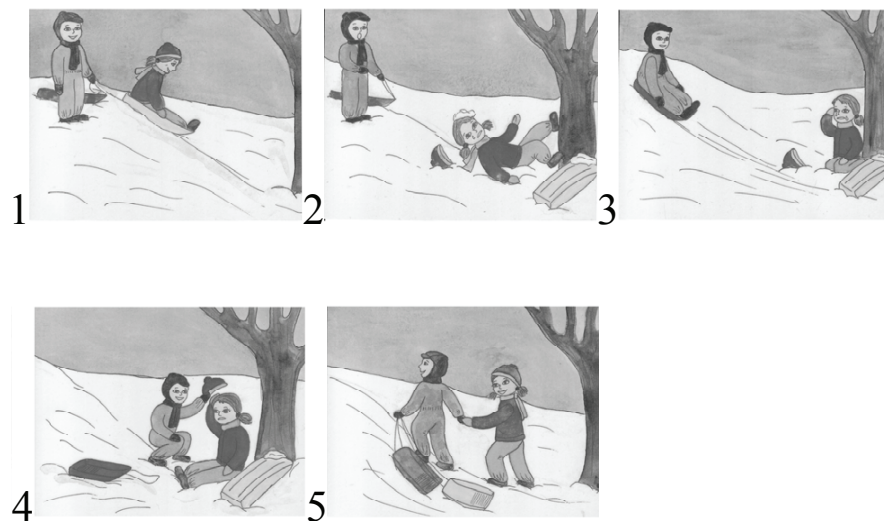
- Hickmann, Maya 1995. Discourse organization and reference. – Paul Fletcher, Brian MacWhinney (Eds.). *The Handbook of Child Language*. Oxford: Blackwell, 194–218.
- Hickmann, Maya; Hendriks, Henriëtte 1999. Cohesion and anaphora in children's narratives: A comparison of English, French, German, and Mandarin Chinese. – *Journal of Child Language*, 26 (2), 419–452. <http://dx.doi.org/10.1017/S0305000999003785>
- Huang, Chiung-chih 2011. Referential choice in Mandarin child language: A discourse-pragmatic perspective – *Journal of Pragmatics*, 43 (7), 2057–2080. <http://dx.doi.org/10.1016/j.pragma.2010.12.007>
- Kail, Michèle; Hickmann, Maya 1992. French children's ability to introduce referents in narratives as a function of mutual knowledge. – *First Language*, 12 (34), 73–94. <http://dx.doi.org/10.1177/014272379201203405>
- Laak, Anneli 2006. Pronoomeneid üks, mingi ja see sisaldavad fraasid laste keeles. Avaldamata bakalaureusetöö. Tartu ülikool, eesti ja üldkeeleteaduse instituut.
- Karmiloff-Smith, Annette 1979. *A Functional Approach to Child Language*. Cambridge Studies in Linguistics, 24. Cambridge: Cambridge University Press.
- Keevallik, Leelo 2011. Pro-Forms as Projective Devices in Interaction. – *Discourse Processes*, 48 (6), 404–431. <http://dx.doi.org/10.1080/0163853X.2011.559150>
- Maratsos, Michael 1976. *The Uses of Definite and Indefinite Reference in Young Children: An Experimental Study of Semantic Acquisition*. Cambridge: Cambridge University Press.
- Pajusalu, Renate 1997. Is There an article in (spoken) Estonian? – Mati Ereht (Ed.). *Estonian: Typological Studies II*. Publications of the Department of Estonian of the University of Tartu, 8. Tartu: Tartu University Press, 146–177.
- Piaget, Jean 1955. *The Language and Thought of the Child*. Meridian Books.
- Soodla, Piret; Kikas, Eve; Pajusalu, Renate; Adamka, Aive; Parm, Sirli 2010. Vahendamata ja vahendatud narratiiv laste kõnearengu hindamisel. – *Eesti Rakenduslingvistika Ühingu aastaraamat*, 6, 277–296. <http://dx.doi.org/10.5128/ERYa6.17>
- Soodla, Piret 2011. *Picture-Elicited Narratives of Estonian Children at the Kindergarten-School Transition as a Measure of Language Competence*. Dissertationas Pedagogicae Universitatis Tartuensis, 12. Tartu: Tartu University Press.
- Stein, Nancy L.; Christine G. Glenn 1979. An analysis of story comprehension in elementary school children. – Roy O. Freedle (Ed.). *New Directions in Discourse Processing*, 2. Norwood, NJ: Ablex, 53–120.
- Tomasello, Michael 2003. *Constructing a Language. A Usage-Based Theory of Language Acquisition*. London: Harvard University Press.
- Toolan, Michael 2001. *Narrative: A Critical Linguistic Introduction*. Routledge.
- Vija, Maigi 2007. Pronoomenid lapsekeeles: mõnda *mina* ja *sina* omandamisest. – *Rakenduslingvistika Ühingu aastaraamat*, 3, 373–384. <http://dx.doi.org/10.5128/ERYa3.24>
- Vija, Maigi; Renate Pajusalu 2009. Pronouns in Estonian child language. *Ad verba liberorum: Linguistics & Pedagogy & Psychology*, 1 (1), 23–31.

Helen Kirsipuu (Tartu Ülikool) uurimisvaldkonnad on referentsiaalsus ja viiteahelad, sealhulgas esimese keele omandamise aspektist.
helen_kirsipuu@yahoo.com

Piret Soodla (Tallinna Ülikool) uurimistöo valdkonnad on suulise ja kirjaliku kõne areng, kõnearengu ja lugemisoskuse hindamine.
piret.soodla@tlu.ee

Renate Pajusalu (Tartu Ülikool) uurimistöo valdkonnad on semantika ja pragmaatika, eriti referentsiaalsusega seotu. Uurinud ka esimese ja teise keele omandamist.
renate.pajusalu@ut.ee

Lisa 1. Uurimuse aluseks olnud pildiseeria



REFERENTIAL NOUN PHRASES IN ESTONIAN CHILDREN'S NARRATIVES

Helen Kirsipuu¹, Piret Soodla², Renate Pajusalu¹

¹University of Tartu, ²Tallinn University

The article discusses how Estonian preschool children (aged 6–7) use referential NPs for referring to entities when telling stories. We used the story-eliciting method to collect the narratives; the stimulus picture sequence was based on Stein and Glenn's (1979) story grammar model. In order to compare children's narratives to adult language we also recorded adult narratives based on the same picture sequence.

The purpose of the study is to determine which NPs are used in order to refer to entities and how new (indefinite) and old (definite) referents are distinguished in the discourse. More specifically we examine the structure of referring NPs and reference chains in children's narratives. The use of personal pronouns (3rd person) and determiners will be discussed, and to some extent we also describe grammatical roles.

The results show that Estonian children use two different story-telling strategies. In the first one children use substantives (with or without determiners) or proper names to refer to the main characters of the story. When the referent has already been introduced into the discourse, then children may switch from substantives to personal pronouns. This strategy is the default of adult-like narrative language, where new referents must first be introduced with a substantive phrase, and only then can the speaker begin to use pronouns. The second strategy (rather rare) is based on the extensive use of pronouns, where children even use the pronoun in the first utterance of the narrative. This also includes cases where children do not use overtly marked pronouns at all.

Our analysis also indicates that the means for marking the informational status of the referent are not yet consistent in children's narratives. Although children usually start their stories with fully informative NPs, there are many instances of pronoun usage which do not maintain coherent reference chains. Sometimes the pronoun is used to bring a new referent into the discourse. Nevertheless, most children are able to narrate in a way that the informational status of discourse referents is comprehensible to the listener.

Keywords: referentiality, language acquisition, pronouns, determiners, narrative, Estonian

KONVERSATSIOONIAGENDI MODELLEERIMINE ARGUMENTEERIMISDIALOGIS: SUHTLUS KUI INFOSEISUNDITE VÄRSKENDAMINE

Mare Koit

Ülevaade. Artiklis modelleeritakse sellist vestlusprotsessi, milles üks osaleja püüab mõjutada teist, et see otsustaks teha teatava tegevuse. Keskendutakse konversatsiooniagendi infoseisundite esitusele ja värskendamisreeglitele, mille rakendamise tulemusel liigub agent suhtluses ühest infoseisundist teise. Infoseisundi koosseisu kuulub partneri mudel, mis koosneb partneri oletatavatest hinnangutest vaadeldava tegevuse erinevatele aspektidele. Partneri mudel muutub suhtluse käigus, vastavalt argumentidele ja vastuargumentidele, mida osalejad esitavad tegevuse tegemise poolt või vastu. Konversatsiooniagendi mudeli praktilise realiseerimise kavandatakse dialoogsüsteemi, nn suhtlustreenerit. Artiklis arutatakse ka inimestevahelise suhtluse modelleerimise üle, tuues näiteid eestikeelsetest argivestlustest.*

Võtmesõnad: argumenteerimisdialog, dialoogi mudel, arutlusmudel, konversatsiooniagent, eesti keel

1. Sissejuhatus

Konversatsiooniagentide modelleerimise ja dialoogsüsteemide arendamise eesmärk on muuta inimeste ja arvuti vaheline suhtlus mugavamaks. Konversatsiooniagendid suhtlevad kasutajatega loomulikus keeles, inimestevahelise suhtluse reeglite alusel, aidates planeerida reise, vastates küsimustele näiteks ilma või spordivõistluste kohta, tegutsedes telefoniassistentidena või täites veelgi keerulisemaid ülesandeid (McTear 2004).

Igal dialoogsüsteemil on rida tüüpilisi komponente (Allen jt 2001). Kõnetuvastusmoodul leiab kasutaja sisendi tähenduse, tekstigenereerimis- ja kõnesünteesikomponent teisendavad dialoogsüsteemi väljundi tähenduse vastavalt tekstiks ja kõneks. Dialoogihaldur juhhib konversatsiooniagendi käitumist, tehes selleks koostööd ülesannete halduriga, mis omakorda kasutab teadmisi ainevaldkonna kohta.

* Artikli valmimist on toetanud Euroopa Regionaalarengufond Eesti Arvutiteaduse Tippkeskuse kaudu, Eesti Teadusfond (grandid 7503, 8558 ja 9124) ning Haridus- ja Teadusministeerium (projektid SF0180078s08, EKT5 ja EKT8). Autor tänab anonüümseid retsenseente.

Dialoogihaldurites rakendatakse nelja põhilist ülesehitusviisi ehk arhitektuuri (Ginzburg, Fernandez 2010). Varaseim, aga samas praktiliseks realiseerimiseks kõige raskem arhitektuur pärineb 1960-ndate aastate tehisingintellektist ja põhineb planeerimistehnikal (Allen 1995). Siin eeldatakse, et agendil on eesmärk/kavatsus, mis käivitab plaani koostamise selle eesmärgi saavutamiseks. Formaalselt võttes koosneb dialoog siis kahest otsinguprotsessist: planeerimine, kus rääkija kavatsustele seatakse vastavusse tegevused, ja plaani tuvastamine, kus kuulaja tuvastab rääkija kavatsused (Webber 2002). Plaanipõhised dialoogimudelid võtavad arvesse dialoogis osalejate eesmärgi ja nende eesmärkide saavutamise viise. See võimaldab paindlikku suhtlust arvutiga. Kuid nagu juba öeldud, on selliste mudelite konstrueerimine ja arvutil rakendamine keeruline.

Järgmised kaks arhitektuuri, mis põhinevad vastavalt olekuautomaatidel ja freimidel, on kõige lihtsamad ja kommertsrakendustes kõige populaarsemad (Wilks jt 2005). Olemasolevad dialoogsüsteemid ongi enamasti realiseeritud kui olekuautomaadid, mis kasutavad regulaaravaldisi, nende hulgas on mitmesugused infosüsteemid, ärikokkulepete sõlmimise süsteemid, abilauad (ingl *help desk*). Nii saavutatakse praktilistes rakendustes vajalik tõrkekindlus, sest igas dialoogi olekus on kasutaja sõnavara ja muud valikud piiratud. Sel viisil saab lihtsustada dialoogi juhtimist ja kõnedialoogsüsteemides automaatset kõnetuvastust. Paraku jääb sellistel süsteemidel puudu paindlikkusest ja funktsionaalsusest, mis aga on inimestevahelise suhtluse olulised tunnused (Jurafsky, Martin 2008).

Neljas, kõige keerulisem ja võimsam dialoogihaldurite arhitektuur põhineb infoseisundi mõistel (Traum, Larsson 2003). Iga infoseisund tuleneb dialoogis eelnevatest seisunditest ja selles seisundis sooritatavad tegevused määravad omakorda järgmise infoseisundi. Dialoogihalduri funktsioon saab formaliseerida kui infoseisundite ajakohastamisi vastavalt etteantud nn värskendamisreeglitele. Dialoogi modelleerimisel on mõnikord vastandatud dialoogi oleku mõistet kasutatavat lähenemisviisi (nt olekuautomaatides) plaanipõhisele lähenemisele. Infoseisundi mõiste võimaldab need kaks lähenemist ühendada, sest infoseisund võib sisaldada nii dialoogi oleku komponente kui ka dialoogis osalejate arvamusi, soove, eesmärgi jms, mida kasutab plaanipõhine lähenemine.

Tuleb veel märkida, et dialoogihaldurite ülesehitamisel on rakendatud lisaks eespool nimetatud reeglipõhiste meetoditele ka andmepõhiseid. Dialoogi on käsitletud kui Markovi otsustusprotsessi, kus on antud olekute hulk $S = \{s_i \mid i=1, \dots, n\}$ ja võimalike tegevuste (dialoogiaktide) hulk $A = \{a_j \mid j=1, \dots, m\}$. Iga tegevusega on seotud "hüvitis", mis on seda suurem, mida edukam on tegevus lõppeesmärgi saavutamise seisukohast. Igas olekus valib dialoogsüsteem sellise tegevuse, mis maksimeerib tõenäolise hüvitise (Levin jt 2000). Selle mudeli nõrgaks küljeks on Markovi eeldus: iga järgmine olek loetakse sõltuvaks ainult vahetult eelnevast olekust ja täidetavast tegevusest (mis tegelikkuses nii ei tarvitse olla). Üldisem ja võimsam mudel on nn osaliselt vaadeldav Markovi otsustusprotsess (ingl *partially observable Markov decision process*), kus olekuteks on kasutaja oletatavad kavatsused (Roy jt 2000, Williams, Young 2007). Iga selline nn arvamuselek tuletatakse tõenäosusjaotusest, mis määratakse tegelike dialoogiantmete põhjal. Andmepõhiste meetodite rakendamist piirab aga empiirilise materjali kättesaadavus ja kogus.

Käesolevas artiklis vaadeldakse dialooge, kus ühe osaleja (A) eesmärk on saavutada teise osaleja (B) otsus teha tegevus D . A alustab dialoogi, tehes partnerile

teatavaks oma suhtluseesmärgi. Kui B keeldub (olles eelnevalt läbi viinud arutluse selle üle, kas D teha või tegemata jätta), siis peab A teda suhtluse käigus mõjutama, analüüsis selleks, millisel oma arutluse sammul jõudis B negatiivse otsuseni, ja samas korrigeerides olemasolevat partneri mudelit.

Artiklis arendatakse edasi mudelit, mida on kirjeldatud artiklis (Koit 2010). Artikli ülesehitus on järgmine. 2. osas antakse ülevaade suhtlusprotsessi modelleerimisest. Esitatakse konversatsiooniagendi mudel, mille koosseisu kuulub arutlusmudel. 3. osas vaadeldakse konversatsiooniagendi suhtlust partneriga kui protsessi, kus pidevalt värskendatakse infoseisundeid. 4. osas käsitletakse selle mudeli rakendamist dialoogsüsteemis, 5. osas vaadeldakse mudeli kasutatavust inimestevaheliste dialoogide analüüsimisel ning 6. osas tehakse kokkuvõtteid.

2. Suhtlusprotsessi modelleerimine

Toimugu suhtlus kahe osaleja – konversatsiooniagendi A ja tema partneri B vahel, kus B võib olla kas inimkasutaja või hoopis teine konversatsiooniagent. A eesmärgiks on mõjutada suhtluse käigus partnerit B vastu võtma otsust teha teatav tegevus D . Suhtlusprotsessi võib esitada järgmiste hulkade abil (Koit, Õim 2003):

- 1) suhtluseesmärkide hulk G , millest kumbki osaleja saab valida oma algse eesmärgi (vastavalt G^A ja G^B). Antud juhul $G^A = “B$ otsustab teha $D”$.
- 2) Osalejate suhtlusstrateegiate hulk S . Suhtlusstrateegia on algoritm, mida osaleja kasutab oma suhtluseesmärgi saavutamiseks. Strateegia määrab osaleja tegevuse suhtluse igal sammul.
- 3) Suhtlustaktikate hulk T . Suhtlustaktika on osaleja järjekordse voo konstrueerimise algoritm, mis määrab suhtluspartneri mõjutamise viisi. Näiteks võib A oma suhtluseesmärgi G^A saavutamiseks ahvatleda, veenda või ähvardada partnerit tegema D , rõhutades selle tegevuse erinevaid aspekte.
- 4) Arutlusmodelite hulk R , mida osalejad kasutavad, arutlemaks tegevuse D tegemise või mittetegemise üle. Arutlusmudel on algoritm, mille tulemuseks on kas positiivne või negatiivne otsus arutelu objekti kohta (antud juhul on selleks tegevuse D tegemine).
- 5) Osalejate mudelite hulk P . Osaleja mudel on osaleja ettekujutus kas iseendast või oma partnerist:
$$P = \{P^A(A), P^A(B), P^B(A), P^B(B)\}.$$
- 6) Keeleteadmiste hulk, mida osalejad kasutavad oma lausungite sünteesimisel ja partneri lausungite analüüsimisel.
- 7) Maailmateadmiste hulk, mida osalejad kasutavad arutlusprotsessis.

2.1. Arutlusmudel

Arutlusprotsess, kus tuleb vastu võtta otsus tegevuse D tegemise kohta, kujutab endast kindlate sammude järjendit, kus kaalutakse tegevuse tegemiseks vajaminevaid/olemasolevaid ressursse ning tegevuse positiivseid ja negatiivseid aspekte. Tegevuse positiivsed aspektid on näiteks tema meeldivus ja kasulikkus, negatiivsed näiteks ebameeldivus ja kahjulikkus. Suhtluspartner ei saa otseselt osaleda selles

arutluses, sest see toimub teise osaleja peas. Ainus, mida ta saab teha, on oma lausungite abil püüda juhtida partneri arutlusprotsessi, andes (uut) infot tegevuse aspektide kohta ning püüdes rõhutada positiivseid ja alla suruda negatiivseid aspekte.

Siin kasutatav arutlusmudel koosneb kahest osast: 1) inimese motivatsioonisfääri mudelist ja 2) arutlusprotseduuridest. Subjekti motivatsioonisfääri mudelina kasutatakse vektorit, mille komponentideks on "kaalud", mida subjekt on oletatavasti omistanud arutlusobjektiks oleva tegevuse aspektidele:

$$w = (w(\text{ressursid}), w(\text{meeldiv}), w(\text{ebameeldiv}), w(\text{kasulik}), w(\text{kahjulik}), w(\text{kohustuslik}), w(\text{keelatud}), w(\text{karistus-keelatud-tegevuse-tegemise-eest}), w(\text{karistus-kohustusliku-tegevuse-tegematajätmise-eest})).$$

Siin tähistab $w(\text{ressursid})$ ressursside olemasolu/puudumist (vastavalt 1 või 0) ning $w(\text{meeldiv})$ jne D meeldivuse jt aspektide kaalu. Vektorit w^{AB} kasutab A kui partneri mudelit $P^A(B)$, s.t see on tema ettekujutus partnerist. Tegevuse D aspektide kaalud on A arvamus/oletused B hinnangute kohta (mille tegelikke väärtusi ta ei tarvitse teada). Suhtluse kestel teeb A vastavalt vajadusele korrekture partneri mudelis.

Arutlusmudeli teine osa on arutluskkeemid, mis oletatavasti reguleerivad inimese arutlusi tegevuste üle. Arutluskkeem kujutab endast nn arutlussammude järjendit, mida arutlev subjekt läbib arutlusprotsessis. Iga samm seisneb teatavate kaalude võrdlemises ja lõpptulemuseks on otsus: teha D või mitte teha D (vt Koit, Õim 2003). Motivatsioonisfääris saab eristada kolme olulist tegurit, mis reguleerivad subjekti arutlust tegevuse D üle. Esmalt, subjekt soovib teha D , kui D meeldivus ületab ebameeldivust. Teiseks, subjektil on vaja teha D , kui kasulikkus ületab kahjulikkust. Kolmandaks, subjekt peab (on sunnitud) tegema D , kui see on talle kohustuslik, s.t kui D tegematajätmine tooks kaasa karistuse. Neid tegureid võib nimetada vastavalt soov-, vaja- ja peab-teguriteks. Igaüks neist käivitab teatava arutlusprotseduuri, olgu need protseduurid vastavalt SOOV, VAJA ja PEAB.

Siin eeldatakse, et vektori w kõikidel komponentidel on arvulised väärtused ja arutlusprotsessis, kus kaalutakse argumente ja vastuargumente, liidetakse ja võrreldakse neid kaalusid. (Muidugi on see robustne eeldus, mis lihtsustab tegelikkust.)

Üldjoontes järgib see arutlusmudel BDI-mudeli (ingl *belief-desire-intention*) ideid (Allen 1995).

2.2. Arutlus suhtluses

Ühe osaleja (konversatsioonigendi A) eesmärkide baasis aktiveeritakse eesmärk G^A , mis on seotud partneri B tegevustega. A kontrollib partneri mudelit, mis koosneb D aspektide oletatavatest kaaludest, mida B on A arvates neile aspektidele omistanud. Selle mudeli alusel valib A taktika, mida kasutada B mõjutamisel (näiteks *veenda* B -d tegema D , s.t rõhutada D tegemise kasulikkust B jaoks). Seega püstitab A alameesmärgi – käivitada B -s kindel arutlusprotseduur (näiteks veenmise puhul käivitada arutlusprotseduur VAJA, mis lähtub *vaja*-tegurist). Konstrueerides oma esimest voo vr_1 , kavandab A selles voores esinevad dialoogiaktid ja määrab nende keelelise vormi. A voo vr_1 käivitab B -s arutlusprotsessi, kus ta kõigepealt interpreteerib A voo ja seejärel genereerib oma voo vr_2 . B vastus käivitab omakorda

A -s arutluse, mille käigus A kontrollib, kuidas on kulgenud tema suhtluseesmärgi saavutamine, ja vastavalt aktiveerib G^A uue alameesmärgi. Tsükkel kordub: A ehitab oma järgmise vooru vr_3 jne. Dialoog lõpeb, kui A on kas saavutanud oma eesmärgi või sellest loobunud.

3. Suhtlus kui infoseisundite värskendamine

3.1. Infoseisundite esitus

Infoseisundi oluliseks komponendiks on partneri mudel, mis suhtluse käigus pidevalt muutub.

Konversatsiooniagendi infoseisund koosneb kahest osast – privaatsest ja jagatud osast. Privaatne osa sisaldab infot, mis on kättesaadav ainult konversatsiooniagendile, seevastu jagatud osas olev info on kättesaadav ka partnerile.

Siin tehakse veel üks lihtsustav eeldus: konversatsiooniagendi jaoks on ette antud valmis laused, mille hulgest ta saab valida argumente partneri mõjutamiseks. Need laused on klassifitseeritud vastavalt vaatluse all oleva tegevuse D aspektidele ning nende aspektide kaalude muutmiseks kasutatavatele dialoogiaktidele (nt meeldivuse suurendamiseks, ebaseeldivuse vähendamiseks jne). Seega lihtsustub konversatsiooniagendi voorude genereerimine: tal on vaja üksnes valida laused õigetest klassidest.

(Jooksva) infoseisundi privaatne osa võib sisaldada näiteks järgmised komponendid:

- jooksev partneri mudel (kaalude vektor w^{AB} – A ettekujutus B hinnangutest tegevusele D). Suhtluse algul võib A konstrueerida selle vektori kas juhuslikult või – kui võimalik – võtta arvesse oma eelteadmisi B kohta.
- Suhtlustaktika t_i^A , mille A on valinud B mõjutamiseks (lähtuvalt partneri mudelist).
- Arutlusprotseduur r_j , mida A püüab partneris B käivitada ja seejärel juhtida positiivse otsuseni (see on määratud A taktikaga, nt kui A on valinud veenmise taktika, siis püüab ta B -s käivitada arutlusprotseduuri v_{AJA}).
- (Alam-)eesmärkide pinu (magasin), mis sisaldab püstitatud, kuid seni veel saavutamata eesmärgid. Alguses paigutab A pinusse oma lähte-eesmärgi (“ B otsustab teha D ”). Igas infoseisundis sisaldab pinu kindlasti ka D selle aspekti, mis on jooksva taktika nn nimiaspekt. (Näiteks kui A alustab B veenmist, siis on pinu tipus eesmärk “kasu”: suurendada D kasulikkust B jaoks.)
- Võimalike dialoogiaktide hulk $DA = \{d_1^A, d_2^A, \dots, d_n^A\}$. A kui suhtluse algataja saab kasutada näiteks järgmisi dialoogiakte: ettepanek, väited D erinevate aspektide kaalude suurendamiseks ja vähendamiseks jne.
- (Lõplik) hulk lausungeid, mida A saab kasutada tegevuse D aspektide kaalude suurendamiseks/vähendamiseks (s.t argumentid ja vastuargumentid D tegemisele): $U = \{u_{i1}^A, u_{i2}^A, \dots, u_{ik_i}^A\}$. Igale sellisele lausungile olgu samuti omistatud kaal (arvuline väärtus) nagu tegevuse D aspektidelegi, mis tähendab, et mõned argumentid on suurema ja teised väiksema kaaluga: $V = \{v_{i1}^A, v_{i2}^A, \dots, v_{ik_i}^A\}$, kus v_{i1}^A jne on vastavalt u_{i1}^A jne väärtus. Siin

eeldatakse, et A saab iga lausungit valida üheainsa korra. See tähendab, et kui argumendid lõpevad, siis peab A oma eesmärgist loobuma.

Infoseisundi jagatud osa sisaldab näiteks järgmised komponendid:

- arutlusmodelite hulk $R = \{r_1, \dots, r_k\}$,
- suhtlustaktikate hulk $T = \{t_1, t_2, \dots, t_p\}$,
- dialoogi ajalugu – dialoogis (seni) esinenud lausungite järjend koos osalejate tähistega ja vastavate dialoogiaktidega:

$$p_1:u_1[d_1], p_2:u_2[d_2], \dots, P_i:u_i[d_i],$$

kus p_1, p_2 jne on osaleja tähis A või B .

3.2. Värskendamisreeglid

Dialoogi käigus toimub ühest infoseisundist teise liikumine nn värskendamisreeglite abil, mida saab jagada kahte liiki: genereerimis- ja interpreteerimisreeglid.

I. Reeglid, mida A kasutab oma voorude genereerimiseks, saab omakorda liigitada vastavalt rakendatavale taktikale ja suhtlussituatsioonile. Siia kuuluvad reeglid erinevate juhtude käsitlemiseks, näiteks:

- 1) juhul, kui jooksva taktika “nimiaspekt” paikneb parajasti pinu tipus (näiteks kui jooksev taktika on veenmine, siis on nimiaspektiks “kasu”).
- 2) Juhul, kui nimiaspekti peal paikneb pinus mõni muu aspekt. (Näiteks kui A püüab B -d veenda, rõhutades kasu, aga B viitab hoopis tegevuse ebameeldivusele, siis asub “ebameeldivus” pinus “kasu” peal, sest A peab kõigepealt püüdma vähendada tegevuse ebameeldivust, mille töö esile B , ja alles seejärel saab suurendada kasulikkust.)
- 3) Juhul, kui jooksva taktika jätkamiseks puuduvad lausungid (ja A peab valima uue taktika, kui see on võimalik).
- 4) Juhul, kui A peab loobuma oma algsest eesmärgist.
- 5) Juhul, kui B on teinud positiivse otsuse ja A on seega saavutanud oma eesmärgi.

Suhtluse alustamiseks, s.t esmase infoseisundi värskendamiseks, on omaette reeglid.

II. Reeglid, mida A kasutab B lausungite interpreteerimiseks. Nende reeglite alusel saab A tuvastada B lausungi sisu (näiteks tegevuse kahjulikkuse osutamine) ja kasutatud dialoogiakti (näiteks keeldumine).

Kahest järgnevast värskendamisreegli näitest kuulub (1) liiki II ja (2) liiki I-2.

- (1) kui jooksev taktika on veenmine JA B viimane lausung käsitles D ebameeldivust, siis paiguta eesmärkide pinusse esmalt “kasu” ja seejärel “ebameeldivus”.
- (2) kui jooksev taktika on veenmine JA eesmärkide pinu tipus on “ebameeldivus” (s.t B ütles, et D tegemine on ebameeldiv) JA “kasu” paikneb eesmärkide pinu tipus oleva kaalu all JA leidub lausungeid ebameeldivuse vähendamiseks x ühiku võrra JA leidub lausungeid kasulikkuse suurendamiseks y ühiku võrra JA arutlusprotseduur VAJA korrigeeritud partneri mudelil (kus $w(\text{ebameeldivus})$ on x ühiku võrra väiksem kui seni ja $w(\text{kasulikkus})$

on y ühiku võrra suurem kui seni) lõpeb otsusega “teha D ”, siis vali need lausungid (ja vastavad dialoogiaktid) JA eemalda eesmärkide pinust “ebameeldivus” ning “kasu”.

Reegleid (2) ja (1) rakendatakse juhul, kui agent A on valinud partneri B veenmise taktika, mille puhul ta rõhutab tegevuse D kasulikkust, aga partner – seadmata küll kahtluse alla kasulikkust – viitab oma järjekordses voorus hoopis tegevuse ebameeldivusele. Agent peab reageerima partneri lausungile, näidates, et tegevus ei ole nii ebameeldiv, nagu arvab partner, aga – järgides valitud taktikat – ühtlasi rõhutama tegevuse kasulikkust.

Eesmärk eemaldatakse pinust niipea, kui agent leiab, et arutlus partneri mudeli järgi annab positiivse otsuse. Kui partneri järgmisest lausungist selgub, et eesmärk siiski ei ole saavutatud, siis pannakse see pinusse tagasi ja protsess jätkub.

4. Arutelu

Kui A püüab mõjutada B otsuseni jõudmist, siis kasutab ta mitmesuguseid väiteid käsitletava tegevuse D positiivsete aspektide kaalude suurendamiseks ja negatiivsete aspektide kaalude vähendamiseks, et tulemuseks oleks B otsus teha D .

Kui B viitab oma lausungis kindlale aspektile, mille tegelik kaal (liiga väike või liiga suur) ei võimalda tal vastu võtta otsust teha D , siis saab A lihtsalt valida ühe võimalikest argumentidest selle aspekti “tõrjumiseks”. Näiteks kui B ütleb, et tegevuse D tegemiseks puuduvad tal ressursid, siis saab A valida lausungi, mis näitab, et ressursid on hangitavad ja kuidas see saaks toimuda. Kui B ei põhjenda oma negatiivset otsust, siis saab A üksnes rõhutada kasutatava taktika nimiaspekti, näiteks veenmise korral kasu.

Dialoogi alustades moodustab konversatsiooniagent partneri mudeli w^{AB} (kaalude vektori) ja valib taktika, mis sellel mudelil viiks positiivse otsuseni. Teiste sõnadega, A eeldab, et B on valitud taktika abil mõjutatav. Dialoog algab sellega, et A väljendab oma suhtluseesmärki (näiteks teeb B -le ettepaneku teha tegevus D). See on agendi esimene voor vr_1 . Kui kasutaja keeldub (olles kaalunud tegevuse aspektide tegelikke, temal olemasolevaid väärtusi), siis leiab agent kasutaja lausungist vr_2 tegevuse D selle aspekti, mille kaalu tuleks muuta, et kasutaja mudelit tegelikusega kooskõlla viia. Siin eeldatakse, et igal lausungil (argumendil) on arvuline väärtus ja selle lausungi ütlemine muudab vastavat kaalu partneri mudelis. Oma vooru vr_3 konstrueerides valib arvuti olemasolevate lausungite hulgast lausungi selle kaalu muutmiseks ja ühtlasi muudab sellele lausungile omistatud väärtuse võrra vastavat kaalu mudelis w^{AB} . Eeldatavasti annab arutlusprotseduur muudetud mudelil positiivse tulemuse (teha D), sest A peaks valima just sellise lausungi, et tulemuseks oleks B positiivne otsus. Nüüd on kasutaja kord valida vastuslausung vr_4 ja protsess jätkub analoogilisel viisil.

Järgnevas lihtsas näites (3) on tegevuseks D “valmistada kartulisalat”: A eesmärk on, et B otsustaks teha D (vrd Koit 2011). A genereerib sellise partneri mudeli, millel arutlusprotseduur $VAJA$ annab positiivse tulemuse, A rakendab veenmise taktikat.

Olgu konversatsiooniagendi A algne (esmane) infoseisund järgmine.

Privaatne osa:

- algne partneri mudel
 $w^{AB} = (w^{AB}(\text{ressursid}) = 1, w^{AB}(\text{meeldiv}) = 3, w^{AB}(\text{ebameeldiv}) = 0, w^{AB}(\text{kasu}) = 7, w^{AB}(\text{kahju}) = 0, w^{AB}(\text{kohustuslik}) = 0, w^{AB}(\text{keelatud}) = 0, w^{AB}(\text{karistus-keelatud-tegevuse-tegemise-eest}) = 0, w^{AB}(\text{karistus-kohustusliku-tegevuse-tegematajätmise-eest}) = 0).$
- A valitud taktika – veenmine.
- A püüab käivitada B arutlusprotseduuri v_{AJA} (kuna selle eeltingimused on täidetud: $w^{AB}(\text{ressursid}) = 1; w^{AB}(\text{kasu}) > w^{AB}(\text{kahju})$).
- Eesmärkide pinu sisaldab ainult A algse eesmärgi (“B otsustab teha D”).
- A käsutuses olevate dialoogiaktide loend: {ettepanek, väited D erinevate aspektide kaalude suurendamiseks/vähendamiseks jne}.
- Lausungite hulk erinevate dialoogiaktide väljendamiseks (sh kaalude suurendamiseks/vähendamiseks, koos nende väärtustega): {Ma aitan sind – väärtus 1, Minu köögis on hea ventilatsioon – väärtus 5 jne}.

Esmase info seisundi jagatud osa sisaldab:

- arutlusprotseduurid: SOOV, v_{AJA} , PEAB.
- Taktikad: ahvatlemine, veenmine, ähvardamine.
- Dialoogi ajalugu on tühi hulk.

A esimene dialoogiakt on ettepanek, mida väljendab lausung *Palun valmista kartulisalat* (3):

(3)

A (arvuti): Palun valmista kartulisalat. [Ettepanek.]

B (kasutaja): Mul pole nii palju aega. [Keeldumine D tegemisest + väide, mis vähendab aspekti ressursi väärtust.]

Seega on aspekti $w^B(\text{ressursid})$ tegelik väärtus 0 ja arvuti peab korrigeerima komponendi $w^{AB}(\text{ressursid})$ väärtust partneri mudelis w^{AB} . Seepärast valib ta väite selle väärtuse suurendamiseks (ehk argumendi D tegemise kasuks):

A: Ma aitan sind. [Argumendiga mittenõustumine + väide komponendi $w^{AB}(\text{ressursid})$ kaalu suurendamiseks: koos tegutsemine võimaldab aega kokku hoida.]

B: Köögis on liiga palav. [Keeldumine D tegemisest + argumendiga mittenõustumine + väide, mis suurendab aspekti kahju väärtust.]

Kasutaja tõi esile tegevuse D kahjulikkuse. Järelikult peab arvuti korrigeerima komponendi $w^{AB}(\text{kahju})$ kaalu kasutaja mudelis:

A: Minu köögis on hea ventilatsioon. [Argumendiga mittenõustumine + väide aspekti kahju kaalu vähendamiseks.]

jne.

Tartu Ülikoolis on arendamisel dialoogsüsteem, mis kasutajaga suhtlemisel võib täita kas A või B rolli. Hetkel opereerib arvuti ainult keelelise sisendi ja väljundi semantiliste esitustega; nii arvuti kui ka kasutaja jaoks võimalike eelnevalt klasifitseeritud lausungite hulgad on ette antud. Sellisest programmist võiks olla abi argumenteerimise harjutamisel. Kui kasutaja täidab osaleja A rolli, siis võib ta püüda mõjutada oma “vastasmängijat” B (arvutit) näiteks sel teel, et esitab süstemaatiliselt ainult meeldivust rõhutavaid ja ebameeldivust allasuruvaid argumente, s.t rakendab

ahvatlemise taktikat ega “hüple” ühelt taktikalt teisele. Samas, kui kasutaja mängib *B* rolli ja arvuti rakendab järjekindlalt ühte fikseeritud taktikat, saab kasutaja harjutada kindlate argumentide tõrjumist kindlate vastuargumentidega. Sel viisil saab distsiplineerida suhtlemist ja argumentide esitamist oma seisukohtade kaitseks.

5. Kuidas inimesed argumenteerivad?

Kas eespool toodud mudel võimaldab kirjeldada ka tegelikke, inimeste vahel toimuvaid dialooge? Käesoleva artikli tarvis analüüsiti Tartu Ülikooli eesti dialoogikorpussest üheksat argivestlust, kus leidis argumenteerimist. Argivestlustes on tavaline, et arendatakse korraga mitut teemat, mistõttu jutt liigub siia-sinna ning järjekindlat argumenteerimist kohtab harva. Siiski on analüüsitud dialoogides löike, kus kõne all on teatava tegevuse tegemine ning sellele tuuakse argumente ja vastuargumente.

Järgnevas näites (4) püüab osaleja *A* saavutada partneri *B* nõusolekut, et too paigutaks oma nime ja muud andmed ühe instituudi veebilehele. *B* kuulub teatava nõukogu juhatusse, kuid ei pea vajalikuks oma andmete avalikku seostamist selle instituudiga. Dialoogide litereerimisel on kasutatud vestlusanalüüsi transkriptsiooni, dialoogiaktid on märgendatud TÜ aktitüpoloogia kohaselt (Hennoste, Rääbis 2004).

(4)

B: a:ga ma mõtlen oopis `teist asja, | YA: EELTEADE |
[võib]olla ma=ei:: `pane sinna `ültse mingeid andmeid. | SEE: ARVAMUS |
[Keeldumine.]

/---/

A: aga [kõik `teised] `panevad. | SEJ: MITTENÕUSTUMINE | | SEE: VÄIDE | [Argumentidiga mittenoustumine + väide aspekti kasu kaalu suurendamiseks: kasulik on olla samasugune nagu teised.]

/---/

B: jaahhhhh ((ohates)) | SEJ: NÕUSTUMINE |
ei=no ma=i ma=i `näe nagu suurt `vajadust `miks: `miks see peaks seal
isegi `olema. | SEE: ARVAMUS |
[Keeldumine *D* tegemisest + väide, mis vähendab aspekti kasu väärtust.]
(0.3) ausalt.=hh= | IL: MUU |

A: =noh ma=i=tea esiteks on see .hhhhhh nagu nimodi=hhhh .hhhhh
`otsustatud=et=hh et nõukogu `tuleb=ja:=ja q noh (1.0) .hhhhhhhhh
eeee ikkagi noh, see on ka `enda tutvustamise=hhhhhhh `mõttes kogu=se
.hhhhh > Kaerajaani instituudi kodu`lehekül.=hhh < .hhh | SEJ:
MITTENÕUSTUMINE | | SEE: VÄIDE | [Väide aspekti kasu kaalu suurendamiseks:
kasu pole isiklik, vaid institutsionaalne.]

B: ei seda `küll= | SEJ: PIIRATUD NÕUSTUMINE |
aga ma mõtlen mõtlen noh igasuguseit meili`aadresse=ja `vär[ke=ja]
| IL: TÄPSUSTAMINE | | SEE: ARVAMUS | [Keeldumine *D* tegemisest + väide, mis suurendab aspekti kahju väärtust: avalikustatakse personaalseid andmeid.]
jne.

Osalejate lausungid moodustavad seisukohavõtude naabrusepaare (SEE ja SEJ), kusjuures iga seisukohavõtt kujutab endast ühtlasi argumenti või vastuargumenti.

A rakendab vestluses veenmise taktikat, rõhutades tegevuse kasulikkust. B, püüdes edasi lükata otsuse vastuvõtmist, viitab enamasti (väikesele) kasule või (suurele) kahjule. Vestluse lõpuks ei teegi B selget otsust.

Analüüsitud dialoogide põhiosa võib kujutada (pärast mõnede dialoogiaktide, nt IL, YA jmt, “destilleerimist”, vt Jönsson, Dahlbäck 2000) enamasti järgmise dialoogiaktide järjendina (5). Lausungid on varustatud kahe märgendiga, moodustades “kokkusulavaid” naabrusepaare (eelneva paari järelliige, mis väljendab mittenõusolekut, on samal ajal järgneva paari esiliige). See näitab muuhulgas, et osalejate suhtluseesmärgid on vähemalt suhtluse algul vastandlikud.

(5)

A: DIE: ETTEPANEK / SOOV

B: DIJ: MITTENÕUSTUMINE / PIIRATUD NÕUSTUMINE + SEE: ARVAMUS / VÄIDE

A: SEJ: MITTENÕUSTUMINE + SEE: ARVAMUS / VÄIDE

B: SEJ: MITTENÕUSTUMINE / PIIRATUD NÕUSTUMINE + SEE: ARVAMUS / VÄIDE

...

A: SEJ: MITTENÕUSTUMINE + SEE: ARVAMUS / VÄIDE

B: SEJ: NÕUSTUMINE / PIIRATUD NÕUSTUMINE / MITTENÕUSTUMINE + DIJ: NÕUSTUMINE / PIIRATUD NÕUSTUMINE / MITTENÕUSTUMINE

Missuguseid taktikaid A ja B rakendavad ning kui järjekindlalt, see sõltub konkreetsete arvamuste või väidete sisust. Näiteks A taktika võib sellise dialoogiaktide järjendi puhul olla nii ahvatlemine, veenmine ja ähvardamine kui ka mõni selline, mida eespool toodud mudelis pole otseselt käsitletud.

Järgnev näide (6) illustreeribki ühte omapärast taktikat: A, kes soovib, et B keedaks talle kas piparmündi- või pärnaõieteed (see on siin tegevus D), kordab kogu vestluse jooksul ühtainsat (vastu-)argumenti (et mustsõstratee, mida kavatseb keeta B, on liiga mõru, s.t ebameeldiv), kuid lõpuks nõustub siiski B-ga. Niisiis jääb A-l saavutamata esialgne eesmärk, aga veelgi enam – ta nõustub B-ga, et see, mida ta algul pidas ebameeldivaks, on tegelikult hoopis meeldiv. Rollid justkui vahetuvad: A loovutab initsiatiivi B-le.

(6)

A: `tee mulle [ˈole=ea] | DIE: SOOV |

B: [ˈmustusõ]strateed.= | DIE: PAKKUMINE | [B soovib tegevuse D asemel teha tegevust D'.]

/---/

A: `piparmünt või [seda] | DIE: SOOV |

{B}: [hehe] | YA: MUU |

(0.5)

A: [ˈpärnaõit.] | PA: ENESEPARANDUS | | DIE: SOOV |

/---/

A: ta=on `minu jaoks on `liiga [siuke] `mõru. | SEE: VÄIDE | [A argument: D' tulemus – mustsõstratee – on mõru, s.t ebameeldiv.]

B: % [ei=ˈole `mõru.] % | SEJ: MITTENÕUSTUMINE | | SEE: VÄIDE | [B vastuargument: lihtsalt eitab A esitatud omadust, ei põhjenda. Sellised “argumendid” on inimestevahelises suhtluses võimalikud.]

(0.4)

A: mulle maitseb `jubedalt `mustsõstar aga `mustsõstra (.) [see] `tee on noh \$ siukse pagana `mõruda | SEJ: PIIRATUD NÕUSTUMINE | | SEE: VÄIDE | [A kordab oma argumenti: *D'* tulemus on ebameeldiv.]

{B}: [heheh] | YA: MUU |

(0.5)

{B}: \$.nhh \$ | YA: MUU |

(1.1)

A: \$ siukse: \$ | YA: MUU |

{B}: heh | YA: MUU |

(1.7)

A: no igal=jul `mõruda. | IL: ÜLERÕHUTAMINE | [A kordab oma argumenti: *D'* tulemus on ebameeldiv.]

/---/

A: [{-}] mina tahan öelda \$ mustsõstratee on `mõru onju i-ikka \$ | SEE: VÄIDE | [A kordab oma argumenti: *D'* tulemus on ebameeldiv.]

{-}: he[hehe] | YA: MUU |

A: \$ [ja siis] ja siis tulevadki sellised et @ mm. @ \$ | SEE: VÄIDE | (.)

B: väga `ea=ju sis ongi: | SEE: ARVAMUS | [Huvitav olukord: *B* väidab, et *A* esiletoodud negatiivne omadus on tegelikult positiivne (mõru tee ei ole ebameeldiv, vaid hoopis meeldiv).]

A: \$ jah \$ | SEJ: NÕUSTUMINE | (0.4) [Ja *B* nõustub sellega!]

Seega õnnestus hoopis *B*-l saavutada *A* nõusolek, et *B* teeb tegevuse *D* asemel tegevuse *D'*. *B* ei teinud selleks mitte midagi muud, kui ignoreeris korduvalt *A* argumenti (et *D'* tulemus on ebameeldiv) ja viimaks väitis, et see, mida *A* peab ebameeldivaks, on tegelikult meeldiv (“ongi ju hea”). Sellega *A* nõustus; sama naabruspaari SEE: ARVAMUS ja SEJ: NÕUSTUMINE korratakse veel kaks korda. Seega jõuavad *A* ja *B* vestluse lõpuks kokkuleppele, et *B* keedab mustsõstrateed.

Kõigis analüüsitud argivestlustes on suhtlejad omavahel tuttavad või koguni sõbrad, niisiis teavad teineteisest küllaltki palju. See lihtsustab taktikate valimist oma suhtluseesmärkide saavutamiseks, sest ettekujutus partnerist (partneri mudel) vastab üsna täpselt tegelikkusele ning ühest infoeisundist teise liikumine on seetõttu sujuv.

Tuleb siiski tõdeda, et reaalsed inimestevahelised dialoogid on palju rikkalikumad kui need, mida suudaks läbi viia eespool modelleeritud konversatsiooniagent isegi siis, kui lisada talle keele- ja maailmateadmised.

6. Kokkuvõte

Artiklis vaadeldi dialooge, kus ühe osaleja (*A*) eesmärk on, et partner (*B*) võtaks vastu otsuse teha tegevus *D*. Seda tüüpi dialoogid kuuluvad argumenteerimisdialoogide hulka. Kui *A* on vestluse algul väljendanud oma eesmärki (teinud ettepaneku), saab *B* vastata kas nõustumise või keeldumisega, sõltuvalt oma arutluse tulemusest. *B* võib põhjendada keeldumist, lisades sellele argumendi(d). *B* argumendid

annavad *A*-le infot *B* arutlusprotsessi kohta ning *A* saab edasise dialoogi käigus neid oma vastuargumentidega tõrjuda.

Artikkel kirjeldas arvutusmudelit, mida rakendatakse eksperimentaalses dialoogsüsteemis. Keskenduti konversatsiooniagendi infoseisundite esitusele ja värskendamisreeglitele. Infoseisund sisaldab partneri mudeli, mis koosneb tema oletatavatest hinnangutest kõnealuse tegevuse erinevatele aspektidele. Suhtluse käigus muutub partneri mudel pidevalt, vastavalt esitatavatele argumentidele ja vastuargumentidele. Praktilise rakendusena peetakse silmas nn suhtlustreenerit – programmi, millega kasutaja saab harjutada argumenteerimist.

Edasise töö suunad on järgmised. Detailiseerida mudelit, vaadeldes erinevaid stsenaariume: kui *A* ja *B* eesmärgid on vastandlikud ja üks osalejaist peab oma algsest eesmärgist loobuma (see on seni vaadeldud stsenaarium); kui *A* ja *B* teevad koostööd ühise eesmärgi nimel; kui *A* ja *B* on mõlemad konversatsiooniagendid, kumbki oma infoseisundite ja värskendamisreeglitega. Analüüsida erinevaid suhtlusstrateegiaid ja -taktikaid, võttes arvesse nende edukust osalejate algsete eesmärkide saavutamisel.

Viidatud kirjandus

- Allen, James 1995. *Natural Language Understanding*. Second Edition. Redwood City etc.: The Benjamins/Cummings Publ. Co.
- Allen, James; Ferguson, George; Stent, Amanda 2001. An architecture for more realistic conversational systems. – IUI'01: Proceedings of the 6th International Conference on Intelligent User Interfaces. Santa Fe, New Mexico, United States: ACM Press, 1–8. <http://dx.doi.org/10.1145/359784.359822>
- Ginzburg, Jonathan; Fernández, Raquel 2010. Computational models of dialogue. – A. Clark, C. Fox, S. Lappin (Eds.). *The Handbook of Computational Linguistics and Natural Language Processing*. Oxford: Wiley-Blackwell, 429–481. <http://dx.doi.org/10.1002/9781444324044.ch16>
- Hennoste, Tiit; Rääbis, Andriela 2004. *Dialoogiaktid eesti infodialoogides: tüpoloogia ja analüüs*. Tartu: Tartu Ülikooli Kirjastus.
- Jurafsky, Daniel; Martin, James H. 2008. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Prentice Hall.
- Jönsson, Arne; Dahlbäck, Nils 2000. Distilling dialogues – A method using natural dialogue corpora for dialogue systems development. – Proceedings of the 6th Applied Natural Language Processing Conference. Seattle, 44–51.
- Koit, Mare 2011. Conversational agent in argumentation: Updating of information states. – J. Filipe, J. L. G. Dietz (Eds.). Proceedings of the International Conference on Knowledge Engineering and Ontology Development: KEOD 2011, Paris, France, 26–29 October, 2011. Paris: SciTEC Publications Ltd, 375–378.
- Koit, Mare 2010. Eesti dialoogikorpus ja argumenteerimisdialogi arvutil modelleerimine. – Keel ja Kirjandus, 4, 241–262.
- Koit, Mare; Õim, Haldur 2003. Eestikeelse dialoogi modelleerimine. – Keel ja Kirjandus, 10, 721–735.
- Levin, Esther; Pieraccini, Roberto; Eckert, Wieland 2000. A Stochastic Model of Human-Machine Interaction for Learning Dialogue Strategies. – IEEE Transactions on Speech and Audio Processing, 8 (1), 11–23. <http://dx.doi.org/10.1109/89.817450>
- McTear, Michael 2004. *Spoken Dialogue Technology. Toward the Conversational User Interface*. London: Springer.

- Roy, Nicholas; Pineau, Joelle; Thrun, Sebastian 2000. Spoken Dialogue Management Using Probabilistic Reasoning. – Annual meeting of the Association for Computational Linguistics (ACL-2000).
- Traum, David; Larsson, Staffan 2003. The Information State Approach to Dialogue Management. – Jan van Kuppevelt, Ronnie W. Smith (Eds.). Current and New Directions in Discourse and Dialogue. Dordrecht: Kluwer, 325–353.
- Webber, Bonnie L. 2002. Computational perspectives on discourse and dialogue. – D. Schiffrin, D. Tannen, H. Hamilton (Eds.). The Handbook of Discourse Analysis. Oxford: Blackwell Publishers, 798–816.
- Wilks, Yorick; Webb, Nick; Setzer, Andrea; Hepp, Mark; Catizone, Roberta 2005. Machine learning approaches to human dialogue modelling. – C. J. van Kuppevelt, Laila Dybkjær, Niels Ole Bernsen (Eds.). Advances in Natural Multimodal Dialogue Systems. Text, Speech and Language Technology, 30. Dordrecht: Springer, 355–370. http://dx.doi.org/10.1007/1-4020-3933-6_16
- Williams, Jason D.; Young, Steve 2007. Partially observable Markov decision processes for spoken dialog systems. – Computer Speech and Language, 21 (2), 393–422. <http://dx.doi.org/10.1016/j.csl.2006.06.008>

Mare Koit (Tartu Ülikool) on uurinud dialoogi modelleerimist arvutil.
mare.koit@ut.ee

MODELLING OF CONVERSATIONAL AGENTS IN ARGUMENTATION: CONVERSATION AS UPDATING OF INFORMATION STATES

Mare Koit

University of Tartu

This article considers conversations where one participant is trying to lead his/her partner to a decision: to do an action. The concentration is on the representation of information states of the conversational agent and their update rules. In conversation, the agent moves from one information state to another, applying the update rules. An information state includes a partner model which consists of evaluations which the agent believes his/her partner has assigned to different aspects of the action under consideration. The partner model changes during conversation, according to arguments and counter-arguments presented by the participants. As a practical implementation of the model the present study has in view a dialogue system which is called a communication trainer. Estonian human-human dialogues are considered and the applicability of the model is discussed.

Keywords: argumentation dialogue, dialogue model, reasoning model, conversational agent, Estonian

ARGUMENTEERIMISVÕTTED RIIGIEKSAMIKIRJANDITES

Kersti Lepajõe

Ülevaade. Artiklis käsitletakse riigieksamikirjandite retoorikat. Teoreetiliseks aluseks on arusaam tekstist kui keelekasutuse valikutest, millel on oma eesmärk ja funktsioon. Keelelised valikud sõltuvad suhtlussituatsioonist ja kultuuris kinnistunud keelekasutustavadest. Riigieksamikirjand on kindla retoorilise struktuuriga tekst, millel on oma kindlaks kujunenud väidete esitamise ja tõestamise viisid. Artiklis jagatakse riigikirjandites kasutatud retoorilised võtted kaheks suuremaks rühmaks: fakti- ja tundepõhisteks argumentideks. Kirjutajad kasutavad mõlemat tüüpi argumente, kuid tundepõhiste hulk ületab tunduvalt faktipõhiste oma. Kirjandites on valdav enesekeskne vaatepunkt. Autoriteedina esitatakse iseend ja oma subjektiivset arvamust, mis vormistatakse üldkehtiva tõena.*

Võtmesõnad: süsteemfunktsionaalse grammatika teooria, retoorika, fakti- ja tundepõhised argumendid, eesti keel

1. Sissejuhatus

Artiklis analüüsitakse riigieksamikirjandeid nende retooriliste omaduste alusel. Uurimismeetodiks on kvalitatiivne tekstianalüüs. Kirjandite retooriliste omaduste käsitlemisel lähtutakse siinses artiklis funktsionaalse tekstianalüüsi ideest: keel ja keelekasutus konstrueerivad tegelikkust ja see moodustub keelekasutustavadest. Ka nendest tavadest, mis koolis kasutatava tekstiliigi – kirjandi kirjutamisega seostuvad. Riigikirjandite analüüsi teoreetiliseks aluseks on Michael Alexander Kirkwood Halliday **süsteemfunktsionaalse grammatika teooria**, eelkõige keele metafunktsioonide käsitus. Hallidayst (1985) lähtuvas tekstianalüüsis vaadeldakse keelekasutust paljufunktsioonilisena. Tekstil on kolm metafunktsiooni: tähistusfunktsioon, suhtlusfunktsioon ja tekstuaalne funktsioon. Teksti autor kasutab keelt alati valikuliselt. Kasutades teatud sõnavara, grammatikat paljude keelesüsteemi pakutavate variantide hulgast, loob kirjutaja teatud tunnuste ja

* Artikkel on kirjutatud Eesti Teadusfondi grandil nr 6236 toel.

tähendusega teksti (tähistusfunktsioon). Tekstis kasutatavad valikud on tingitud teksti eesmärgist ja funktsioonist. Samaaegselt maailma tähistamisega kinnistab tekst suhteid ja loob identiteete (suhtlusfunktsioon). Tekstianalüüsi meetoditega on võimalik vaadelda, kuidas tekstid toimivad suhtluses – tekstiga tehakse midagi, tekstil on alati mingi eesmärk. Teksti suhtlusfunktsioon avaldub teksti modaalsuse kaudu ehk nende tekstis väljendatud hoiakute kaudu, mida tekstis on esitatud teksti teema, selle vastuvõtja või kirjutaja enda kohta. Kolmas, tekstuaalne metafunktsioon keskendub sellele, kuidas ja milliste keelevahenditega saadakse tulemuseks teatud tunnustega sidus terviktekst, mis seatud eesmäärke kõige paremini täidab. Siinses artiklis jäetakse riigikirjandite tekstuaalne funktsioon vaatlusest kõrvale ja keskendutakse tähistus- ja suhtlusfunktsioonile. Lähtutakse sellest, et keeikasutus alati valikute tulemus, aga need valikud pole juhuslikud, vaid on seotud ühiskonnas parasjagu valitsevate sotsiokultuuriliste tavade ja suhetega.

Teiseks tekstianalüüsi teoreetiliseks aluseks funktsionaalse grammatika kõrvale on **retoorikaõpetus**. Enamasti mõistetakse retoorika all oskust mõjusalt kõnelda ja selle kaudu oma eesmäärke saavutada. Tänapäeval räägitakse lisaks sellele üha enam ka kirjalike tekstide retoorikast. Kui uurimisobjektiks on kirjalike tekstide retoorilised võtted, analüüsivad uurijad neid keelelisi valikuid, mille abil lugejat mõjutatakse, kuidas kujundatakse lugeja suhtumisi ja hoiakuid, kuidas pakutakse teksti retoorika najal suuniseid teksti tähenduse tõlgendamiseks (vt nt Perelman, Olbrechts-Tyteca 2000, Potter 1996).

Artiklis analüüsitakse riigieksamikirjandite argumentatsiooni ja vaadeldakse, milliste argumentide najal konstrueerivad gümnaasiumilõpetajad tegelikkust ja põhjendavad oma väiteid, muudavad need usutavaks ja lugejale vastuvõetavaks.

Artikkel on osa uurimusest, kus analüüsitakse kirjandite tekstuaalseid, interpersonaalseid ja diskursiivseid omadusi. Materjaliks on valik aastatel 1999 ja 2007 kirjutatud riigieksamikirjandeid, kokku 400 tööd. 1999. aastast pärit materjali hulk on väiksem – 60 tööd (nende abil katsetati esialgu analüüsimeetodi sobivust), suurem osa materjalist (340 kirjandit) on pärit 2007. aastast. Riigieksamikirjandite valim on juhuslik, aga materjali hulk on piisav selleks, et esindatud oleksid erinevatel teemadel ja eri tasemel kirjutatud tööd, kuigi artiklis pole vaadeldud kirjandeid teemade ega eksamitulemuste alusel.

Riigieksamikirjandeid analüüsitakse artiklis seetõttu, et kirjand on põhiline tekstiliik, mille kirjutamist Eesti koolis järjekindlalt õpetatakse ja harjutatakse. Viivi Maanso nimetab oma ülevaateartiklis, et kirjandi kirjutamise traditsioon Eesti koolis ulatub möödunud sajandi 30-ndate aastatesse. Siis hakati kirjandit kasutama koolide sisseastumiskatsetel. (Maanso 2006) Praeguse riigieksamikirjandi kui gümnaasiumi lõpueksami traditsioon ulatub 1997. aastasse, mil kirjandist sai kohustuslik riigieksam. Pea 15 aastat kestnud eksamivorm on kirjandi kui tekstiliigi traditsioone sedavõrd kinnistanud, et neist kõrvalekalduva teksti koostamine on raske, kui mitte võimatu. Väljakuunenud tavadest hälbiv kirjanditekst võib tekitada hindajates nõutust ja põhjustada negatiivset tagasisidet. Kirjandikirjutajalt oodatakse ükskõik missuguse teema puhul etteantud mahu ja üsna kindla retoorilise struktuuriga teemaarendust, kuhu kuuluvad probleemi sõnastamine sissejuhatuses, väidete esitamine ning nende põhjendamine kirjandusest või elust võetud näidete ja sobiva ulatusega kokkuvõte (vt nt Ehala 2000: 50–68; Hennoste 1995: 29–36). Kirjanditeksti argumenteerimisvõtteid õpitakse kasutama õppekirjanduse

vastavate teemakäsitluste najal. Gümnaasiumi tekstiõpetuse õpikutes on rohkesti materjali argumentatsioonivõtete õpetamiseks (vt nt Ehala 2000, Hage 2000, Aava 2003a, 2003b, Kaldjärv 2007).

2. Argumentatsiooni eesmärgid

Retoorilises analüüsis uuritakse tekste suhtlusprotsessis. Teksti retooriline põhi-eesmärk on saavutada teksti mõjus. Retoorilises analüüsis vaatlevad uurijad tähenduse loomise keelelisi protsesse sellest vaatenurgast, kuidas mõned tõlgendusversioonid on võimalik muuta veenvaks ja usutavaks. Huvi pakub see, kuidas kuulajaid, lugejaid või vestluspartnereid veenva retoorikaga mõjutada. Argumentatsiooni juures uuritakse retoorilisi vahendeid või tähendusi, aga samal ajal jälgitakse, mida antud argumentide või seletustega tehakse selles situatsioonis, kus neid kasutatakse. (Potter 1996: 103)

Argumentatsioon ei leia aset vaakumis, vaid see on alati osa laiemast kontekstist, mida võib nimetada argumentatsiooni kontekstiks. Nii retoorika kui ka argumentatsiooni põhiülesandeks on veenda kedagi esitatud seisukoha tõesuses ja mingite teiste seisukohtade paikapidamatuses. Argumenteerimist on võimalik analüüsida mitmest aspektist. Üks on varasest tekstilingvistikast ja teksti makrostruktuurist lähtuv formaalne analüüs, mis jälgib argumentatsiooni sisemist loogikat ja ülesehitust (Dijk 1972, 1980, Hoey 1983, 2001, Siitonen, Halonen 1997, Kakkuri-Knuutila 1998). Sellest aspektist on Eesti õpilaste argumenteerimisoskust mõnevõrra uuritud (Lepajõe 2002, Kaldjärv 2008, 2010). Siinse artikli teoreetiliseks aluseks on funktsionaalne tekstianalüüs, mis keskendub teksti eesmärkidele suhtlusprotsessis ja tekstidega konstrueeritavatele tähendustele, ning sellest lähtudes on järgnevas analüüsis keskendunud eeskätt argumentatsiooni sisulistele omadustele: mõjutamisvõtetele, diskursiivsete seoste loomisele ja argumenditüüpidele. Retoorilises lähenemises argumentatsioonile nähakse argumenteerimist mitte kui ettemääratud muustritel ja reeglitel põhinevat ratsionaalset analüüsi, vaid eelkõige kui retoorilist tegevust, mille eesmärgiks on oma seisukohtade usutavaks muutmine. Selles lähenemises põhineb argumentatsioon väärtus- ja uskumussüsteemidel ja on seotud kirjutaja ja lugeja interpretatsiooniprotsessidega (Isaksson-Wikberg 1999: 68).

Argumenteerimine on ühelt poolt verbaalne tegevus. Kõneleja või kirjutaja kasutab argumenteerimisel sõnu ja lauseid, et midagi väita või mõnd esitatud väidet vaidlustada või eitada, esitada küsimusi ja väidetule reageerida. Argumenteerimise eesmärgiks on õigustada või ümber lükata mingit arvamust, mis koosneb teatud väidete hulgast ja on suunatud auditooriumi heakskiidu saavutamisele. Teiselt poolt on argumenteerimine sotsiaalne tegevus, mis ilmneb vaidluses kahe või enama osapoole vahel. Ka vaidlus iseendaga, võimalike poolt- ja vastuväidete otsimine on sotsiaalne ja seotud inimese mõtlemisprotsesside eripäraga.

Argumenteeriva teksti ülesehitusele on omased kindlad tunnused. Selline tekst sisaldab arvamusi ja hoiakuid ning seondub tihti tekstiväliste ilmingutega, näiteks päevakajaliste probleemidega, mis arvatakse olevat lugejale tuttavad, või etteantud teemaga, nagu kirjandi puhul. Argumenteeriv tekst on dialoogiline – selles esitatakse vastandusi ja püütakse tuua esile vastandatavate asjaolude positiivseid ja negatiivseid külgi. Keelekasutuse kaudu avaldub kirjutaja suhtumine; eesmärgipäraste

keeleliste valikutega üritab kirjutaja konstrueerida oma teksti nii, et lugeja võtaks omaks autori seisukohad.

3. Argumentide jaotusest

Jonathan Potter (1996: 106–107) kasutab retoorikast rääkides sõja metafoori. Tema järgi kasutavad inimesed suheldes mitmesuguseid retoorilisi vahendeid – küsimusi, vastuseid, väiteid, jutustusi – nagu laskemoona, kaitstes oma seisukohti ja rünnates vastase omi. Sellega seoses räägib ta retoorilise argumentatsiooni kahest ulatusest – kaitsvast ja ründavast retoorikast. Riigikirjandid on sellest vaatepunktist üsna ühepoolse argumentatsiooniga, ründava retoorika võtteid ehk kujuteldava vastuargumentatsiooni ümberlükkamist neis ei leidu. Piirduakse oma seisukohtade põhjendamise ehk kaitsmisega ja see on ka loomulik, sest puudub suhtlusosaline, kes ründab ja kelle vastu end kaitsma peaks.

Teun A. van Dijk (2005: 132–135) nimetab, et sotsiaalsed uskumused võivad olla kas faktilised või hinnangulised. Ka keelekasutuses saab eristada kaht põhilist mõjutamistüüpi: faktipõhine ja tundepõhine argumentatsioon. Faktipõhised laused on informatiivsed väitlused, mille tõeväärtust saab tõestada: väide on kas õige või vale. Faktipõhiste argumentidega esitatakse kuulajale või lugejale teavet, mis on tõestatav. Tundepõhised laused väljendavad kirjutaja suhtumist, arvamusi ja soove. Need ei ole tõesed ega väärad. Tundepõhised argumendid sisaldavad kirjutaja isiklikku arvamust. Argumendid toetuvad sel juhul inimeste kogemustele, uskumustele ja tavaarusaamadele. Esitav väide seostatakse millegi muuga, millesse lugeja eeldustekohaselt suhtub tundepõhiselt: kas negatiivselt või positiivselt. Dijk juhhib tähelepanu sellele, et alati ei ole arvamust ja fakti kerge eristada. Fakte kiputakse eitama, kui nad toetavad autori seisukohast erinevat positsiooni. Teksti kirjutaja saab fakte, mida ta esitab, valida nii, et nad teenivad soovitud eesmärki. Argumentidena saab esitada selliseid asjaolusid, mis autori hoiakuid toetavad, ning nimetamata saab jätta vähem soodsad. (Kasik 2007: 94–98, Leiwo jt 1995: 93, Pääkkönen, Varis 2000: 108–112)

Kirjandeid analüüsid on otsitud tekstidest põhiväiteid ja neid toetavaid argumente. Keskendutakse sellele, millist tüüpi argumente riigikirjandites kasutatakse. Õpilaste argumenditüübid on jaotatud kahte põhirühma: **faktipõhisteks** ja **tundepõhisteks argumentideks**. Sellist jaotust võib pidada tinglikuks, sest näiteks faktipõhine informatiivlause võib mingis suhtlussituatsioonis sisaldada ka kirjutaja suhtumist. Jaotuses on peetud faktipõhisteks argumentideks selliseid argumente, mille õigsust saab tõestada – need on kas tõesed või valed. Kui väitlause pole võimalik tunnistada tõeseks ega valeks, on siinses analüüsis argumendid rühmitatud tundepõhisteks. Lisaks fakti- ja tundepõhiste argumentidele on lühidalt välja toodud ka muud kirjandites sagedamini ettetulevad retoorilised võtted, mille eesmärk on lugejat aktiveerida või mõjutada, kuid mida ei saa seostada mingi konkreetse väite esitamise või põhjendamisega.

3.1. Faktipõhised argumendid

Diskursuseanalüüsis vaadeldakse veenvat retoorikat sageli kui faktide konstrueerimist. Fakti konstrueerimise strateegiatele keskendub näiteks Potteri teos "Representing reality" (1996). Faktiline kirjeldus on veenva retoorika omamoodi äärevorm, faktide abil püütakse lugeja panna uskuma, et kirjutaja on õigel teel, ta teab, kuidas asjad tegelikult on (Jokinen 2002: 129). Tüüpilised faktiargumendid on arvulised andmed, toetumine väidet toetavale statistikale. Riigikirjandit kirjutades pole abiturientidel kasutada muid allikaid kui õigekeelsussõnaraamat. Seega võiks eeldada, et faktilise materjali esitamine riigikirjandites on tagasihoidlik. Eesti keele riigieksamitöödest ilmneb siiski, et faktide esitamine tekstis on levinud argumenteerimisevõte. Õpilased on riigikirjandiks valmistudes viinud end kurssi avalikkuses aktuaalsete arutlusteemadega, jätnud meelde arvulisi andmeid, fakte ja tsitaate, et põimida neid võimalust mööda valitud kirjanditeemasse ja rajada neile oma väiteid ja põhjendusi. Oskuslikult teemaga seotud faktid on heaks veenmisvahendiks. Fakti argumendina esitades õnnestub kirjutajal näidata, et tema positsiooni mingis küsimuses saab ratsionaalselt seletada.¹

- (1) Tänapäeva eesti noored tarbivad liiga palju alkohoolseid jooke, rikkudes sellega oma tervist. Viimastel andmetel on alkoholi tarbimine 16 liitrit inimese kohta aastas. Selle näitaja poolest oleme Euroopa riikide seas esimeste hulgas. Alkoholitarbimist tuleb vähendada alkoholipoliitika karmistamisega. (758 039 Eesti mured on ka minu mured)

Kirjutaja alustab tekstilõiku väitega, et noored tarbivad liiga palju alkoholi. Oma väidet tõestab ta faktiga eestlaste keskmise alkoholitarbimise kohta aastas, aga sama fakt toetab ka direktiivset hoiakut: *alkoholitarbimist tuleb vähendada*. Järeldus on loogiline ja sellele toetudes pakub kirjutaja teadja positsioonilt ka lahenduse – *alkoholipoliitikat tuleb karmistada*. Faktide esitamise ja sõnavalikuga on konstrueeritud endale teadja roll. Üldiselt on retoorikas teadja rolli konstrueerimine levinud võte, millega kirjutaja püüab rõhutada oma sõnumi usaldusväärtust (autori usaldusväärsuse kohta vt pikemalt Kasik 2007: 101–104). Teadja rollile viitab teksti esitamine kindlas kõneviisis väitlausestena, sest just kindla kõneviisi ülesandeks on millegi tõdemine. Näites (1) antakse lugejale informatsiooni, kuid infoallikale on viidatud ebamääraselt (*viimastel andmetel*). Argumendina kasutatavat fakti tugevdab viitamine mingile infoallikale, nagu on tehtud järgmistes näidetes.

- (2) Euroopa Liidu eesmärgiks on dünaamilise ja konkurentsivõimelise majanduse väljatöötamine ning põhiväärtuseks heaolu ühiskond ja õigusriigi põhimõte. 2003. aasta andmete põhjal on elatustase Euroopa Liidus, kuhu kuulub 457 miljonit inimest, aina tõusnud. Kõige kõrgem oli see 2003. aasta andmete põhjal Luksemburgis ja madalaim Lätis. Euroopa Liitu kuulumine annab Eestile võimaluse oma elatustaset tõsta. (758 038 Eesti mured on ka minu mured)

Mõlemas näites on lõigu kokkuvõtva lausena esitatud faktidel põhinev järeldus. Näites (1) on see sõnastatud olukorra kirjeldusena, näites (2) tulevikuvisionina.

¹ Näited riigieksamikirjanditest on esitatud muutmata kujul. Iga näite järel on sulgudes kirjandi kood ja töö pealkiri.

Autor ei esine tuleviku teadjana, vaid esitab oma järelduse esitatud faktidest tulevane võimalusena.

Üksikjuhtudel on fakti allikas esitatud veelgi detailsemalt, järgmises näites on autor osanud oma teksti põimida konkreetsest ajalehenumbri meelde jäänud spetsialisti seisukoha.

- (3) Ohter laenuvõtmine, kõrge inflatsioonimäär ning ebanormaalselt kallis kinnisvara on majandusteadlane Jeffrey Sommersi sõnul märgid halvasti reguleeritud turumajandusest (9.03.2007 Eesti Päevaleht) (756 061 Eesti mured on ka minu mured)

Allikaviited pole kirjanditekstis kohustuslikud ja see on ka ainus näide valimist, kus kirjandis nii täpset viitamist on kasutatud.

Autoriteetidele viitamine on arvandmete kõrval tavalisemaid faktidel põhineva retoorika võtteid. Selleks aga, et argument oleks usutav, peab autoriteedil, kellele toetutakse, olema kaalukust, tuntust, prestiiži (Perelman, Olbrechts-Tyteca 2000: 305). Kõige enam kasutavad õpilased autoriteedina teadlasi ja kirjanikke. Teadlase seisukohad arvatakse olevat tõesed ja teadlasele toetumine on seega tugev argument. Kirjanikele ja kirjandusteoste toetumine kuulub aga eesti keele ja kirjanduse riigiüksami loomulikku konteksti.

- (4) Suurim viga, mida üks eestlane võib endast leida, on ükskõiksus isamaa ja tema kultuuripärandi vastu. Sigmund Freudi psühhoanalüüsi teooria väidab, et ükskõiksus on kohutavam vihastki, sest viha eeldab emotsioonide olemasolu, ükskõiksus on aga tühjus, mis lämmatab kõik. (587 030 “Meie suurim ja hiilgavaim meistritöö on õigesti elatud elu.” (Michel de Montaigne))
- (5) Rousseau on öelnud: “Ükskõik kui vaimselt arenenud me ka poleks, oleme me oma tõeliste vajaduste mõistmisel äärmiselt viletsad.” Kuulame teiste soovitusi, et teada saada, mis teeb meid õnnelikuks ja toob hingele kosutust. Käitume niimoodi, et kardame erineda üldsusest. (215 082 Kõigel on hind, aga vähesel väärtus)
- (6) Tundub, et tänapäeva ühiskonnas hakkavad traditsioonilised väärtused nagu perekond ja armastus kaduma. Tuntud kirjaniku K. Kenderi teoses “Yuppiejumal” puudub sõna “armastus”, selle asemel on seal juttu intiimvahekorra, kus tunded ei mängi mingit rolli. (175 079 Kõigel on hind, aga vähesel väärtus)

Ka järgmises lõigus on toodud oma väite põhjenduseks näitena Freudi ja Einsteini teooriad. Ka ainuüksi teooriate nimetamist neid lähemalt avamata saab tõlgendada autoriteetsusvõttena (Mäntynen 2003: 106).

- (7) Suuri ja murrangulisi teooriaid ei võta inimkond omaks. Sigmund Freud oli oma teooria õigsuses veendunud, kuid kaaskodanikud ei uskunud teda. Paljud ei usu praegugi. Ka Albert Einstein ei kohanud palju positiivset vastukaja, kui ta esitles oma relatiivsusteooriat. Ta jätkas avalikkuse vastupanust hoolimata ja tänapäeval teab igauks, kes Einstein oli ja mida ta tegi. (175 074 Mida suudab avalik arvamus?)

Emotsionaalse sõnavara kasutamine (*suuri ja murrangulisi teooriaid*) lisab argumendile jõudu. Näites (7) kasutab kirjutaja eitust *kaaskodanikud ei uskunud, paljud ei usu* ja vastandab nii teadusliku tõe ja tavainimeste käibetõed. Eitusega saavutab kirjutaja tekstilõigu dialoogilisuse, ta vastandab eri seisukohti ja tõendab veenvalt, et tal on õigus. Keeleliste valikutega kujundab ta endale teadja rolli: kirjutaja teab, kuidas asjad on, millised üldised seaduspärased ja tendentsid maailmas kehtivad, ja nii esitatakse neid ka lugejale. Autor ei ütle, et tegemist on tema isikliku arvamusega, vaid esitab neid seisukohti kui üldkehtivat tõe.

Paljudes kirjandites on olulisel kohal kirjutaja oma kogemus. Suhtudes isiklikku kogemusse kui fakti, saab seda argumenteerimisel edukalt kasutada. Isiklikul kogemusel põhinev argumentatsioon on mõjus, sest seda on raske ümber lükata, kuna teine pool ei saa kogemusse sekkuda (Jokinen 2002: 134–135). Nii viidataksegi kirjandites oma kogemusele, isiklikult läbielatule ja kujundatakse endale autoriteedi roll.

- (8) Eesti toodab kõrgharidusega inimesi, kellele kõigile siin tööd ei jätku. On toimumas nn ajude äravool. **Tean endagi tutvusringkonnast paljusid, kes on kodumaalt lahkunud.** (197 093 Eesti mured on ka minu mured)
- (9) Mõned päevad tagasi vaatasin Ilmar Raagi filmi “Klass”, mis põhineb koolivägivallal. Filmi üks peategelasi on klassi poolt tõrjutud Joosep. Film annab üsna täpselt edasi vägivalla tekkemehhanisme. **Võin seda julgelt väita, sest olen ka ise vägivaldal lähedalt kogenud.** (738 020 Haridus – kas vahend või eesmärk?)

Kui kirjutajal kirjandi teemaga seostatavaid konkreetseid üksikfakte või allikavii- teid meeles ei ole, kasutatakse riigikirjandites sagedase argumendina tuginemist üldteada tõsiasiadele, mida vastavalt konstrueerituna saab kasutada kui väidet toetavaid fakte. Konkreetse allika puudumist kompenseeritakse viitega fakti tuntusele: *kõik me teame, on üldtuntud tõsiasi, on tavaline, nagu maailmas on tavaks, alati on arvatud, paljud arvavad, inimkond on ikka arvanud* jne.

- (10) **Kõik me teame**, et maavarad on ammendumas ja nende eest käib tõeline monopolivõitlus, mis toob endaga sageli kaasa konflikte. See teadmine ei aita õigupoolest mitte kedagi, sest meil on vaja lahendusi. (197 094 Kõigel on hind, aga vähesel väärtus)

Tuntud infona esitatud olukorrajeldusega polemiseerides ja selle olulisust eitades (*see teadmine ei aita kedagi*) ning modaalsuse lisamisega (*õigupoolest*) distantseerib kirjutaja end esitatud seisukohast ja tugevdab vastandust kasutades niiviisi oma väidet (*meil on vaja lahendusi*). Ühtlasi konstrueerib ta selliste keeleliste valikutega endale teadja positsiooni: kõlama jääb, et kirjutaja teab, mida tuleks teha. Argumentatsiooni muudab veenvamaks *meie*-positsioonilt kirjutamine, millega püüab kirjutaja luua lugejaga ühist identiteeti.

Üldtuntud tõsiasiade kasutamine oma väidet toetavate faktidena (10)–(12) on kirjandites tavaline.

- (11) **Paljud arvavad**, et inimese saatuse on juba sündides paika pandud. (720 059 “Lõbusalt saeb inimkond oksa, millel ta istub” (Mati Unt))

- (12) Inimkond on eksisteerinud **tarkade inimeste väidete järgi** mitu miljonit aastat, mida on enam, kui oskame ette kujutada. (368 038 “Lõbusalt saab inimkond oksa, millel ta istub” (Mati Unt))

3.2. Tundepõhised argumendid

Ratsionaalsed argumendid üksi inimestele tavaliselt ei mõju, sageli saavutatakse palju suurem mõjususe tundepõhiste argumentide abil. Viimaseid kasutatakse veenvas retoorikas sageli. Veenvas retoorikas on auditooriumi eeldatavad väärtused ja arvamused omamoodi eelkõkkulepped. Nende olemus meenutab süsteemfunktsionaalsest keeleteadusest pärit ennustatavuse ja loomulikustumise mõistet: kui teatakse teksti põhiväidet ja kirjutise konteksti, võib ennustada, milliseid argumente tekstis kasutatakse (Pääkkönen, Varis 2000: 110). Tundepõhised argumendid toimivad väärtuste, hoiakute ja tunnete pinnal. Teksti emotsionaalse mõjususe määr sõltub sellest, milliseid tundeid kuulajas-lugejas tekitatakse. Tundepõhised argumendid põhinevad kuulaja-lugeja tunnetel, aga esindavad ka sarnaseid vajadusi, uskumusi, lootusi või kasu. Argumenteerimine põhineb sel juhul üldistel arvamustel, kirjutaja usaldusväärsusel ja asjatundlikkusel. Olulist rolli mängivad autori maine ja ühiskondlik staatus. Fakti- ja tundepõhiste argumenteerimist võib olla raske eristada, sest need on sageli põimunud: ka faktiargument võib lisaks informeerivale funktsioonile väljendada ühtlasi ka kirjutaja suhtumisi. Tundeargumente tugevdatakse sageli intensiivsussõnadega nagu *väga, alati, iga kord, mitte ükski, täiesti*.

Tundepõhiste argumentide kaalukust riigikirjandites nõrgestab see, et lugeja ei tõlgenda kirjandi autori sotsiaalset positsiooni asjatundjana ega ühiskondliku autoriteedina, mis vähendab tema väidete usaldusväärsust. Sellegipoolest toetuvad kirjandi kirjutajad oma väiteid esitades ja põhjendades sageli üksnes isiklikule arvamusele, kasutades julgelt n-ö subjektiivsuse objektiveerimist.

Üks keelelise suhtlemise ja argumentatsiooni eeldusi on saatja usaldusväärsuse põhimõte. Kirjutaja esitab kindlas kõneviisis väite ja loodab, et lugeja usaldab tema arvamust ja interpreteerib seda kui üldkehtivat tõde.

- (13) Andmise rõõm ongi kõige suurem. Õigesti elatud elu on oskus rahul olla ja leppida nende asjadega, mida ei ole võimalik muuta. (955120 “Meie suurim ja hiilgavaim meistritöö on õigesti elatud elu.” (Michel de Montaigne))
- (14) Kõige rohkem kasutatakse ära kodutuid ja alkoholismi küüsi langenuid. (298 048 Haridus – kas vahend või eesmärk?)
- (15) Inimene on sedavõrd “tark”, et ta ei piirdu enesehävitamisega, vaid kaasab ka loomad ja taimed. (368 038 “Lõbusalt saab inimkond oksa, millel ta istub” (Mati Unt))

Enamasti kirjandikirjutajad siiski oma arvamusi nii enesekindlalt ei objektiveeri, vaid viitavad subjektiivsusele, kasutades vastavaid keelelisi konstruktsioone *minu arvates, mina arvan, minu meelest, mulle tundub*. Pikk pronoomenivorm (*mina, minu pro ma, mu*) tõstab isikliku arvamuse rõhutatult esile (16), (17).

- (16) Önn on **minu arvates** kõigi peaesmärk – kõik pingutused ja sehken-
dused taanduvad järele mõeldes selle poole püüdlemisele. (600 029
“Meie suurim ja hiilgavaim meistritöö on õigesti elatud elu.” (Michel de
Montaigne))
- (17) **Minule tundub**, et kõikidest religioosetest õpetustest paremini on
inimese tegelikku olemust osanud seletada inglise härrasmees Darwin
(210 034 Paremasse homsesse viivad meid ...)

Palju kasutatakse ka üldistusel põhinevaid argumente. Need on äratuntavad selle järgi, et oma arvamus vormistatakse mingi rühma arvamuseks või mingi üksik-
juhtum üldistatakse reegliks. Kui õpilased teevad üldistusi, siis sageli on tegemist hoopiski lihtsustamisega. Need on väited, kus üksiknäited või juhtumid esitatakse kui korduvad ja reeglipärased ilmingud.

- (18) Kuna meie põlvkonnale on juba väikesest peale sisendatud, et igal asjal on oma hind ja tasuta lõunaid pole olemas, on ka Eestis **väga palju ükskõikseid inimesi**. (175 072 Kõigel on hind, aga vähesel väärtus)

Näites (18) kirjutatakse oma põlvkonna nimel, aga sellegipoolest on tegemist subjektiivse tundepõhise argumendiga, sest väita, et *Eestis on väga palju ükskõikseid inimesi* on meelevaldne ja põhineb kirjutaja isiklikul hinnangul.

Üldistamisega on tegemist ka järgmises argumentatsioonis (19).

- (19) **Enamus inimesi järgib** nõutud kohustusi, aga on ka neid, keda ei huvita seadused ja kes seavad endale omad reeglid. Peaaegu iga päev võib lehest lugeda või uudistest näha, et taas on tabatud teatud arv kiiruse ületajaid või alkoholi juubes autojuhte. (166 035 Kirjutatud ja kirjutamata seadustest)

Kuigi näite (19) esimeses lauses väidetakse üldistatult, et *enamus inimesi järgib kohustusi*, siis lause teisest poolest võib järeldada, et nii see tegelikult pole. Õpilane on teinud üldistuse, mis sisuliselt on pigem lihtsustamine. Samuti on see väide kirjutaja subjektiivne otsustus.

Rohkesti edastatakse kirjandites ka hinnanguid, mis esindavad kirjutaja otsustusi selle üle, mis on hea, mis halb, mis on tähtis ja vajalik või mis seda ei ole. Sellistel juhtudel on tegemist modaalsuse ühe alaliigi – väärtushinnanguga. Väärtushinnang seisneb sündmuse positiivses või negatiivses väärtustamises. Väärtushinnangut väljendavad tavaliselt adjektiivid, aga ka nimisõna fraas, verbid või adverbid. (Vt EKG II: 190–191, Keevallik 2010) Kirjutaja esitab mingi asja kohta oma arvamuse, hinnates, kui hea või halb miski on. Hinnangu andjast lähtudes on hinnangud peamiselt isiklikud, kuigi mõjutatud kultuurilisest ja sotsiaalsest taustast, järelikult põimuvad hindaja isiklik arvamus ning ühised normid ja väärtused. (Fairclough 2003: 172–173) Hinnang antakse tavaliselt kas hetkeolukorrale või minevikus toimunud sündmustele. Hinnangud näitavad, et kirjutaja ei edasta ainult oma teadmisi, vaid tal on kirjutatuga ka isiklik suhe. Kirjandites kasutatakse hinnangute andmiseks sageli väljendeid *on halb, on hea, on kahetsusväärne, pole kiita, on ohtlik, on taunitav, on mõistmatu*. Ka hinnangud skaalal *õnneks ja kahjuks* on tavalised. Suurem osakaal kirjandites on negatiivsetel hinnangutel.

- (20) Meil on palju õppida ajaloost, kuid olenemata lugematutest õppetundidest, ei **ole me kahjuks seda ikkagi teinud**. (605 072 “Lõbusalt saeb inimkond oksa, millel ta istub” (Mati Unt))
- (21) Haridus on kindlasti väga suur mõjutusvahend. Heade teadmistega inimesed võivad kergesti teelt kõrvale juhtida inimesi, kel puudub haridus. Nad on kergesti mõjutatavad ja **see on väga halb**. (298 048 Haridus – kas vahend või eesmärk?)

Enamasti vormistatakse hinnangud öeldistäitega, mille juures hinnatavat situatsiooni väljendab alusena toimiv kõrvallause (22). Näites (21) on öeldistäitele lisatud intensiivsussõna *väga*. Hinnangusõnad ja -väljendid on erineva intensiivsusega. Hinnanguid andvaid adjektiive on otstarbekas vaadata koos intensiivsussõnadega. Mati Erelt (1986: 73) jaotab adjektiividega tähistatud omaduste intensiivsuseastmete väljendamise kaheks põhitüübiks: relatiivseks ja absoluutseks gradatsiooniks. Relatiivse gradatsiooni korral hinnatakse objekti omaduse intensiivsusest mõne teise konkreetse objekti suhtes, absoluutse gradatsiooni puhul aga kogu klassi ulatuses, kuhu objekt kuulub. Hinnangu tugevdamiseks on uuritavates kirjandites kasutatud absoluutset superlatiivi (vt Erelt 1986:111–114): adjektiivi laiendatakse adverbiga nagu *väga* (21) või *äärmiselt* (24).

- (22) On olemas elementaarsed asjad, mida iga inimene vajab eluks. **On hea**, kui on olemas toit, elukoht, perekond, tervis jne. (465 028 Kõigel on hind, aga vähesel väärtus)
- (23) Kui tahetakse saada poliitikuks, siis peab olema haridus. Hariduseta kõrgpoliitik **oleks kohutav**. (738 022 Haridus – kas vahend või eesmärk?)
- (24) Paljud noored ei tunne tänapäeval oma kodumaa ajalugu, mis **on äärmiselt kahetsusväärne**. (605 079 Eesti mured on ka minu mured)

Palju viidatakse kirjandites tulevikule, mis kirjutaja arvates on võimalik ühel või teisel tingimusel. Kui arvamusi ja väiteid põhjendatakse kirjandites võimalikkusega, siis enamasti tahetakse näidata, mis juhtub siis, kui mingile kirjutaja esitatud probleemile ei pöörata tähelepanu või ei tehta midagi, et asju parandada. Situatsiooni ja selle tegelikkusele vastavuse hinnang seisneb sündmuse võimalikuks või mitte-võimalikuks pidamises. Seda saab väljendada nii leksikaalsete kui ka grammatiliste vahenditega (EKG II: 183). Võimalikkuse väljendamiseks kasutatakse kirjandites modaalverbe *võima*, *saama*, *pidama*, *tahtma*, aga ka tingivat kõneviisi.

Näites (25) loob interpersonaalse suhte lause algus (*ma ei tahaks*), tuleviku-prognoos sisaldub kõrvallauses (25).

- (25) Ma ei tahaks, et **paarikümne aasta pärast hakkaks meie kultuur kaduma ja elu oleks nagu Soomes** – kirju teistest kultuuridest. Riik peaks juba praegu kõik tegema, et see nii ei läheks. (738 020 Haridus – kas vahend või eesmärk?)

Tingimuslausega põimlause (26) on kasutatud modaalverbi *võima* ja impersonaali. Hinnangut on tugevdatud superlatiivse adverbiga *väga kergelt*.

- (26) Harrastajast **võib** väga kergelt **saada** tippsportlane, kui hakataks tõsisemalt tegelema selle spetsiaalse alaga (298045 “Meie suurim ja hiilgavaim meistritöö on õigesti elatud elu.” (Michel de Montaigne))

Näites (27) põimuvad mitmesugused modaalsuse väljendamise vahendid. Ene-sekindlat väidet on tugevdatud väljendiga *on selge*. Teises lauses on tingivasse kõneviisi pandud modaalverb *võima*. Tingiv kõneviis väljendab siin võimalikkust – kirjutaja tahab näidata, mis juhtub, kui riik tagaks tasuta lasteaiakohad, kuid kuna ta pole kindel, kas kõik ikka nii läheb, pehmemdab ta võimalikkust tingiva kõneviisi abil. Näites kasutatakse tingivat kõneviisi tagajärgede väljatoomiseks tulevikus. Tuleviku ennustamine on endale teadja rolli võtmine. Kirjutaja teab, et kui riik teeks nii, siis suureneks kindlustunne.

- (27) On selge, et üks takistus laste saamiseks on raha. Riik **võiks ette võtta** ka järgmise sammu: tagada tasuta lasteaiakohad ja seeläbi ka suurem kindlustunne tulevikus. (756 061 Eesti mured on ka minu mured)

3.3. Muud retoorilised võtted

Lisaks fakti- ja tundepõhiste argumentide esitamisele oma kirjandites kasutavad õpilased lugeja aktiveerimiseks ja mõjutamiseks ka muid retoorilisi vahendeid. Üks selliseid on küsimuste esitamine, mis on kirjandites kõige tavapärasem sissejuhatavas tekstilõigus. Lisaks avalõigule leidub küsimusi, kuigi tunduvalt vähem, ka muudes tekstiosades. Küsimuste kasutamine on lugejat aktiveeriv võtte. See on viis esitada mõnd probleemi, kui kirjutaja eeldab, et lugeja hakkab tekstis esitatuga kaasa mõtlema, sellega polemiseerima. Kuna kirjalikus tekstis pole esitatud küsimustele võimalik vastust saada, siis enamasti pole sellistel küsilausest vormistatud tekstielementidel otsest küsimise funktsiooni. Sellised küsimused võivad väljendada hoopis ebakindlust, imestust või rõhutada isiklikku arvamust (Mäntynen 2003: 83). Küsilausest eriliik on retoorilised küsilausest, mis väljendavad esiletõstatuna väidet, mis on jaatuse-eituse poolest vastupidine väidetule. (EKG II: 174)

Siinsast käsitlusest jäetakse kõrvale kirjandite alguslõikude küsimused (kirjandite alustamisest vt Lepajõe 2005) ja vaadeldakse, milliseid küsimusi kasutatakse teksti muudes osades seoses teemaarendusega.

- (28) Kui riigis töötajaid vähemaks jääb, toob see kaasa uued probleemid. nende probleemidega tundub Reformierakonna lubadus viieteist aasta-ga viie jõukaima riigi hulka pürgida suure muinasjutuna. **Kes hakkab maksuma pensione, kui tööaliste osakaal rahvastikus pidevalt väheneb?** Kui maksumaksjaid on vähem, siis pole eriti tõenäoline, et alandatud maksudega meie riik rikkaks saaks. (197 093 Eesti mured on ka minu mured)

Näite (28) retooriline küsimus ei eelda vastust, vaid selle funktsiooniks on eeskätt probleemi tõstatamine. Kaudselt väljendatakse siin küsimusega hoopis kirjutaja kahtlust, et pensionimaksjaid tulevikus ei piisa. Kirjutaja küll täpselt ei ütle, mis ta asjast arvab, ta jätab järelduse tegemise lugejale.

(29) **Mis kasu on inimese elust, kui ta seda normaalselt elada ei saa?**

Võib-olla ma mõtleks teisiti, kui mul mõni füüsiline puue oleks, kuid praegu arvan küll, et ei taha kellelegi koormaks olla. (456 028 Kõigel on hind, aga vähesel väärtus)

Ka näites 29 on küsimus oma arvamuse väljendamise vorm. Siin on kirjutaja järeldus selgem: kui normaalselt elada ei saa, pole elul mõtet. Järgmise lausega tehakse mööndus (*võib-olla ma mõtleks teisiti*), selleks kasutatakse modaalsõna *võib-olla*, tingivat kõneviisi. Vastandava seosega (*kuid praegu ma arvan küll*) tuuakse oma arvamus möönduse taustal veelgi rõhutatumalt esile.

(30) Varsti oli kogunenud pronkssõduri juurde kaks vastasleeri – eestlased ja venelased, kes üksteist solvasid ja sõimasid. Meedia oli samuti kohal ja filmis ning tegi sellest suure loo ning nüüd on see Eestis üks suuremaid probleeme. Tekib küsimus, et **mis siis lõpuks saab, kas jäävad eestlased või venelased peale?** (166 033 Mida suudab avalik arvamus?)

Teist laadi küsimusega on tegemist näites (30). Küsimus on tulevikku suunatud – *mis siis lõpuks saab* – ja seda konkretiseeritakse alternatiivsete võimaluste sõnas-tamisega. Kirjutajal ei ole vastust. Sellise retoorilise võttega püütakse lugejat panna kaasa mõtlema, probleemi üle arutlema.

Kirjandites kasutatud retoorilisi küsimusi tõlgendades võib neid funktsionaal-sest aspektist seostada tundepõhiste argumentidega – enamasti sisaldub retoorilises küsimuses kirjutaja arvamus küsimuse taustal oleva probleemi kohta, aga küsilause vorm vähendab väite enesekindlust ja taandab kirjutaja teadja positsioonilt kahtleja positsioonile.

Üks lugeja mõjutamise ja võimalike seoste konstrueerimise viise on kasutada metafoore retoorilise võtena. Metafoor kui tähenduse ülekanne muudab teksti väljendusrikkamaks, eripärasemaks, meeldejäävamaks. See on viis võita lugeja tähelepanu. Lisaks sellele on metafoori funktsiooniks tekitada emotsioone, üllatust, seletada või rõhutada midagi, näitlikustada või üldistada. Metafoore kasutatakse väite ilmestamiseks. Analüüsitud riigikirjandites oli metafoore kasutatud vähe ja need vähesedki olid sageli laenatud kirjandi etteantud pealkirjast. Järgmine näi-telause on kirjandist pealkirjaga “Meie suurim ja hiilgavaim meistritöö on õigesti elatud elu” (Michel de Montaigne).

(31) Elu on suurim ja hiilgavaim **meistritöö**, kui me suudame ellu viia oma unistused. (641 012 “Meie suurim ja hiilgavaim meistritöö on õigesti elatud elu.” (Michel de Montaigne))

Õnnestunud metafoor võib luua üldistava võrdlusseose ja tugevdada niiviisi kir-jutaja väidet.

(32) Alati ei ole **muru teisel pool rohelisem**. (507 034 “Meie suurim ja hiilgavaim meistritöö on õigesti elatud elu.” (Michel de Montaigne))

Näites (32) on metafoori kasutatud seoses noorte Eestist välismaale lahkumise teemaga rõhutamaks, et lahkujate ootused ei pruugi täituda, pole kindel, et elu on mujal parem. Kinnistunud väljendist – *taevas sinisem ja rohi rohelisem* – on õpilane teinud oma variandi (*muru rohelisem*). Väite kategoorilisust pehmendab selle piiramine (*alati ei ole*).

Metafoorid ei ole neutraalsed keeleväljendid. Kinnistunud metafooridega kaasneb sageli teatud stiilivärving, mis võib sisuliselt adekvaatse tähendusega metafoori konkreetseesse tekstiliiki sobimatuks muuta. Noored kirjutajad ei taju alati kujundliku väljendi stiilinüansse.

- (33) Noored lahkuvad välismaale **parematele jahimaadele** ning vähesed nendest siirduvad koju tagasi. (026 071 Kõigel on hind, aga vähesel väärtus)

Argikeelest laenatud metafoori (*parematele jahimaadele*) on kirjutaja tahtnud kasutada väite ilmestamiseks, kuid argikeele väljendi ja asjaliku arutleva teksti vahele tekib stiililine ebakõla.

Metafooride kasutamise nappusest võib järeldada, et kirjutajad ei valda kujundliku väljenduse võimalusi ja rikkusi või siis pole kujundlik keelekasutus nende meelet kirjandile kui asjalik-arutlevale tekstiliigile omane. Kirjandit peetakse pigem asjalikuks tekstiks, kus pole põhjust keele kujundlike võtetega liialdada. Vähesel määral metafoore muidugi leidub, aga nendegi kasutus pole igakord sobiv.

Suhtluseesmärgil ja lugejaga dialoogi konstrueerimiseks on kirjandites vahel harva kasutatud ka otseseid lugeja poole pöördumisi. Käske, palveid ja üleskutseid vormistatakse tavaliselt käskiva kõneviisi või jussiivi abil.

- (34) **Saagem aru**, et ühiskond peab olema ise aktiivne, mõistmaks hukka halvad teod. (450 021 Paremasse homsesse viivad meid ...)

Näites (34) on üleskutsega hõlmatud nii kirjutaja kui lugeja. Otsese käsu andmine on tavaline suulises suhtlussituatsioonis, kui kuulajat püütakse õhutada midagi tegema. Tegevussituatsioonis mitteosalevale isikule, nagu see kirjalikus suhtluses tavaline, pole põhimõtteliselt võimalik otsust käsku anda. Seega on ka kirjanditektis otsuste käskude andmine põhjendamatu, ometi leidub üksikutes töedes selliseidki väljendeid (35)–(37).

- (35) **Vali** kleit, mis peab vastu su lõpetamise, olgugi, et ta ei ole nii uhke, kui sa seda soovisid. (456 026 Kõigel on hind, aga vähesel väärtus)
- (36) **Ära tee seda teistele, mida sa endale ei taha!** (507 034 “Meie suurim ja hiilgavaim meistritöö on õigesti elatud elu.” (Michel de Montaigne))
- (37) Niisiis, **kaalu** enne pikalt sellise tähtsa otsuse tegemist. (019 013 Kirjutatud ja kirjutamata seadustest)

4. Kokkuvõte

Riigikirjandite analüüsi põhjal võib tõdeda, et õpilaste argumenteerimisevõtted on suhteliselt piiratud. Sellisele tulemusele on jõudnud ka argumentide formaalloomalisi struktuure uurinud Merle Kaldjärv (2008). Argumentatsioonis kasutatakse nii fakti- kui ka tundepõhiseid argumente, kuid viimaste hulk ületab tunduvalt esimeste oma. Konkreetsete faktide esitamine oma väidete toetuseks pole kuigi sage, pigem toetatakse ka faktiargumente kasutades üldtuntud tõsiasiadele ja isiklikule kogemusele. Tugevamate kirjandites tuuakse teksti faktipõhiste argumentidena statistilisi andmeid või muid kontrollitavaid fakte, aga sagedamini toetatakse autoriteetidele.

Autoriteetideks on teadlased, kirjanikud, harvemini tuntud poliitikud – nendele viitamisega näitab kirjutaja, et ta on vastavate teadusteooriate, kirjanike arvamuste, teoste või poliitiliste seisukohtadega tuttav.

Tundepõhine argumenteerimine on faktipõhisest sagedasem, see on üldse riieksamikirjandite soosituim argumenteerimisvõte. Kirjandites on valdav enese-keskne vaatepunkt. Autoriteedina esitatakse iseend ja oma subjektiivset arvamust, mis vormistatakse üldkehtiva tõena. Nii konstrueeritakse teksti keeleliste valikutega endale teadja roll, informeerides lugejat, kuidas asjad tegelikult on. Kirjutaja võib väite subjektiivsusele tekstis eksplitsiitselt viidata (*mina arvan, minu arvates*), aga sageli esitatakse enda nimel enesekindlaid tõdemusi, kasutades kindlas kõneviisis väitlauseid, viitamata allikatele või kasutamata väite kategoorilisust pehmendavaid modaalsõnu. Kirjanditest ilmneb, et diskursuse rollidest kirjutavad õpilased ennast enamasti just teadja rolli, luues niiviisi vastuolu oma sotsiaalse rolli ja diskursuserolli vahele. Tundepõhiseid argumente täpsemalt liigitades võib nende hulgas eristada arvamusi, üldistusi, hinnanguid ja võimalikkusele viitavaid tulevikuennustusi.

Kirjanditest võib leida näiteid ka muude retooriliste võtete, nagu küsimuste, käskude, soovide, üleskutsete, metafooride kasutamise kohta. Viimaste osakaal on vägagi tagasihoidlik ja see on teatud mõttes ootuspärane, sest kirjanditelt ilmselt ei eeldata kujundlikku keelekasutust. Kirjandite hindamisjuhendiskil on tüüpiliste lausestusvigade all toodud välja kujundlikkusega liialdamine. See tähendab, et metafoorset keelekasutust ei peeta sobivaks sellisesse tekstiliiki nagu kirjand. Argumenteerimist seevastu soositakse ja õpetatakse. Tekstiõpetuse õpikutes on esitatud palju mitmekesisemaid argumenteerimisvõtteid, kui kirjanditest tegelikult leida on. Vaatamata õpikumaterjalidele, mis argumenteerimise eri viise käsitlevad, pole õpilaste argumenteerimisoskus kiita. Sellele viitab hulk kirjandeid, mis koosnevad peamiselt põhjendamata väidetest. Ometi on argumenteerimisoskus vajalik mitte ainult kirjandi jaoks, vaid tulevase elu jaoks, sotsiaalses keskkonnas toimetulemiseks. Tõenäoliselt peaks argumenteeriva teksti ülesehitust ja kasutatavaid argumentatsioonitehnikaid ja muid retoorilisi võtteid õpetama koolis põhjalikumalt ühtpidi selleks, et õpilane argumenteerivat teksti teha oskaks, ja teisalt selleks, et ta sellise teksti kas või väliste tunnuste põhjal ära tunneks. Meie igapäevaelus kasutab argumenteerivat teksti ajakirjandus, see on ka mõjus retooriline võte reklaamis. Seega seondub argumentatsiooni tundmine argivajadustega, rääkimata keerukamate tekstidest eri ametites ja valdkondades, millega noorel inimesel tulevasel tööelul kokku puutuda tuleb.

Viidatud kirjandus

Aava, Katrin 2003a. Veenmiskunst. Õpik gümnaasiumile. Tallinn: Avita.

Aava, Katrin 2003b. Kõnekunst. Tallinn: Avita.

Dijk, Teun A. van 1972. Some Aspects of Text Grammars: A Study in Theoretical Linguistics and Poetics. The Hague: Mouton.

Dijk, Teun A. van 1980. Macrostructures: An Interdisciplinary Study of Global Structures in Discourse, Interaction and Cognition. New York: Erlbaum.

Ehala, Martin 2000. Kirjutamise kunst. Tekstiõpetuse õpik gümnaasiumile. Tallinn: Kün-
nimees.

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

- Erelt, Mati 1986. Eesti adjektiivisüntaks. Emakeele Seltsi toimetised, 19. Tallinn: Valgus.
- Fairclough, Norman 2003. Analysing Discourse. Textual Analysis for Social Research. London: Routledge.
- Hage, Maaja 2000. Tekstiõpetus. Aineraamat. Tallinn: Koolibri.
- Halliday, Michael Alexander Kirkwood 1985. An Introduction to Functional Grammar. London: Arnold.
- Hennoste, Märt 1995. Tekstiõpetuse õpik. Tallinn: Avita.
- Hoey, Michael 1983. On the Surface of Discourse. London: Allen & Unwin.
- Hoey, Michael 2001. Textual Interaction. An Introduction to Written Discourse Analysis. London: Rotledge.
- Isaksson-Wikberg, Maria 1999. Negotiated and Committed Argumentation a Cross-Cultural Study of American and Finland-Swedish Student Writing. Åbo: Åbo akademis förlag.
- Jokinen, Arja 2002. Vakuuttelevan ja suostuttelevan retoriikan analysoiminen. – Arja Jokinen, Kirsi Juhila, Eero Suoninen (Toim.). Diskurssianalyysi liikkeessä. 2. painos. Tampere: Vastapaino, 126–159.
- Kakkuri-Knuutila, Marja-Liisa 1998. Argumentti ja kritiikki. Lukemisen, keskustelun ja vakuuttamisen taidot. Helsinki: Gaudeamus.
- Kaldjärv, Merle 2007. Argumendi jõud. Õpik gümnaasiumile. Tallinn: Koolibri.
- Kaldjärv, Merle 2008. Probleemist lahenduseni riigieksamikirjandite näitel. – Haridus, 11–12, 34–40.
- Kaldjärv, Merle 2010. Riigieksamikirjandi kui argumenteeritud teksti kirjutamine. – Tekstid ja taustad VI. Tekstiuurimus ja kool. Emakeeleõpetuse Infokeskuse toimetised, 5. Tallinn: Tallinna Ülikool, 69–94.
- Keevallik, Leelo 2010. Hinnangu grammatikast. – Eesti ja soome-ugri keeleteaduse ajakiri, 1–2, 147–161.
- Kasik, Reet 2007. Sissejuhatus tekstiõpetusse. Tartu: Tartu Ülikooli kirjastus.
- Leiwo, Matti; Luukka, Minna-Riitta; Nikula, Tarja 1995. Pragmatiikan ja retoriikan perusteita. Jyväskylä: Jyväskylän Yliopisto.
- Lepajõe, Kersti 2002. Argumenteerimisoskusest eksamikirjandite põhjal. – Keel ja Kirjandus, 4, 269–274.
- Lepajõe, Kersti 2005. Riigikirjandite alustamise retoorilisi võtteid. – Reet Kasik (Toim.). Tekstid ja taustad IV. Tekstiliigivaatlusi. Tartu Ülikooli eesti keele õppetooli toimetised, 29. Tartu: Tartu Ülikooli Kirjastus, 53–67.
- Maanso, Viivi 2006. Gümnaasiumi lõpukirjand: kas hea traditsioon või iganenud kontrollivorm? – Keel ja Kirjandus, 3, 231–237.
- Mäntynen, Anne 2003. Miten kielestä kerrotaan. Kielijuttujen retoriikkaa. Suomalaisen Kirjallisuuden Seuran toimituksia, 926. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Perelman, Chaim; Olbrechts-Tyteca, Lucie 2000. The New Rhetoric. A Treatise on Argumentation. Notre Dame: University of Notre Dame Press.
- Potter, Jonathan 1996. Representing Reality. Discourse, Rhetoric and Social Construction. London: Sage.
- Pääkkönen, Irmeli; Varis, Markku 2000. Kriittinen lukutaito. Helsinki: Finn Lectura OY.
- Siitonen, Arto; Halonen, Ilpo 1997. Ajattelu ja argumentointi. Porvoo, Helsinki, Juva: WSOY.

Kersti Lepajõe (Tartu Ülikool) on uurinud eelkõige kooliga seotud tekste.
Kersti.Lepajoe@ut.ee

ARGUMENTATION METHODS IN THE ESTONIAN HIGH SCHOOL MATRICULATION EXAMINATION ESSAYS

Kersti Lepajõe

University of Tartu

This article discusses the rhetoric in Estonian high school matriculation examination essays. The theoretical basis is the idea of text as a set of language use choices, each with their own aim and function. The linguistic choices depend on the communication situation and language use practices established in the culture. The high school matriculation examination essay is a text with a particular rhetorical structure and established ways of presenting and proving one's arguments. The article divides the rhetorical methods into two major groups: factual and feeling-related arguments. The essay writers use both types of arguments in their discussion, but the number of feeling-related arguments is considerably higher than that of factual arguments. A self-centered viewpoint is prevalent in the essays. The students present themselves and their subjective opinion as authoritative and formulate this as a generally accepted truth.

Keywords: systemic functional grammar, rhetoric, factual and feeling-related arguments, Estonian

EESTI WORDNET'I STRUKTUURI ANALÜÜSIST

Ahti Lohk, Leo Võhandu

Ülevaade. Artikkel pakub üldise lähenemisviisi relatsiooniliste süsteemide suletud hulkade leidmiseks ja korrastamiseks ning demonstreerib selle meetodi kasutamist Eesti Wordnet'il. Lahatakse Eesti Wordnet'i struktuuri, lähtudes semantilistest suhetest. Selgitatakse analüüsiks kasutatava infotöötlusmeetodi ideed ja sõnastatakse lahendatav probleem mitteformaalselt. Esitatakse meetodi rakendamise järjestikused sammud. Andmetöötluse tulemusena tekivad visuaalsed analüüsi objektid/pildid, mis avavad Wordnet'i struktuuri viisil, mis võimaldab leksikograafil hinnata struktuurides peituvaid eripärasid. Artikli lõpuosas antakse näidete põhjal vihjeid võimalikele probleemidele ja nende lahendustele.

Võtmesõnad: teaurus, semantilised suhted, Wordnet'i visualiseerimine, suhete järjestamine, eesti keel

1. Üldist

Kevadel 2011 toimunud rakenduslingvistika konverentsil autorite esitatud ettekanne "Eesti sünovara visualiseerimisest" tegi sisekaemuse kahte, Internetist kättesaadavasse, eesti keele mõistelisse sõnaraamatusse ehk teaurusse: Filosoft OÜ koostatud teaurus¹ ja Tartu Ülikooli eesti keele teaurus (TEKSaurus) ehk Eesti Wordnet². Eesmärgiks oli uurida sõnastike sisestruktuure ja esitada neid tabelite, jooniste ning graafikute kaudu. Samuti pakuti välja võimalusi Eesti Wordnet'i andmete struktuurseks esitamiseks (Lohk, Võhandu 2011).

Käesolevas artiklis piirdume siiski vaid ühe sõnastiku, Eesti Wordnet'iga. Kohe alguses tasuks mainida, et Eesti Wordnet'i loomisel on võetud eeskujuks nii mitmekeelne EuroWordNet³ kui ka algne ingliskeelne, Princetoni ülikoolis loodud Wordnet⁴, mis oma olemusliku struktuuri poolest on eeskujuks eri keelte *wordnet*-tüüpi sõnastikele (Orav jt 2011). Selliste sõnastike loomine ja arendamine on üleilmsete mõõtmetega ettevõtmine, millest annavad tunnistust

¹ Filosoft OÜ teaurus vt http://www.filosoft.ee/thes_et/ (1.10.2011).

² Tartu Ülikooli TEKSaurus ehk Eesti Wordnet vt <http://www.cl.ut.ee/ressursid/teksaurus/teksaurus.cgi.et/> (1.10.2011).

³ EuroWordNet vt <http://www.illc.uva.nl/EuroWordNet/> (1.10.2011).

⁴ Wordnet vt <http://wordnet.princeton.edu/> (1.10.2011).

iga kahe aasta tagant toimuvad konverentsid pealkirjaga “The Global Wordnet Conference”⁵.

Nii Eesti Wordnet kui teised *wordnet*-tüüpi tesaurused omavad sõnade organiseeritust sünohulkadesse, kuhu on koondatud ühte mõistet väljendavad sünonüümsed sõnad ja sõnaühendid. Sünohulgad võivad olla ka üheliikmelised, sest mõistet võib esindada keeles ka ainult üks sõna.⁶ Seega tingib sünohulga tesaurusesse lisamise mõiste eksisteerimine mentaalses leksikonis, mitte sünonüümiaste mõistet esindavate sõnade vahel keeles.

Sünohulkades olevate sõnade eristamiseks on igale sõnale lisatud tähendusindeks ja sõnaliik. Kolmik {*sõna_tähendusindeks_sõnaliik*} annab sõnale n-ö tähendusliku unikaalsuse. Kõigis illustreerivates joonistes arvestatakse just selle kolmest osast koosneva tervikuga.

Sünohulgad on omavahel seotud semantiliste suhete kaudu, mida Eesti Wordnet'is on 43 erinevat liiki, nt hüperonüümiasteos, hüponüümiasteos, antonüümiasteos jne.

Eesti Wordnet'i süvastruktuurne uurimine tõi esile mõningaid ebakõlasid. See on ka loomulik, sest *wordnet*-tüüpi tesauruse puhul on tegu vägagi keerulise ja paljurelatsioonilise süsteemiga.⁷

Esialgused tulemused näitasid kätte tee *wordnet*-tüüpi süsteemide kiireks ja usutavasti leksikograafide kasulikuks analüüsiks. Tegemist on infoteoreetilises mõttes keeruka probleemiga, mille jaoks ei ole siiani olemas täielikku formaliseeritud lahendusmeetodit. Graafiteooria meetodeid kombineerides õnnestus luua programmipakett, mis teeb *wordnet*'i mahuga süsteemide süvaanalüüsi mõistliku ajaga ära. Tulemuste esitluseks kasutatakse sellist visualiseerimismeetodit, mis usutavasti annab leksikograafide kergesti käsitletava vahendi *wordnet*-tüüpi sõnastike hetkeolukorda kajastava struktuuri uurimiseks, hindamiseks ja ebakõlade kõrvaldamiseks.

Vahemärkusena tuleb öelda, et ka Eesti Wordnet'i sisestusliides Polaris võimaldab mõnesugust visualiseerimist, kuid jälgib seejuures leksikograafi valitud rada, tegemata süsteemi sõltumatut üldkontrolli, mis on selle artikli peaesmärk.

Järgnevas esitatakse mitteformaalne ülevaade kasutatud matemaatilisest metoodikast, demonstreeritakse paari sõnastikus olevate erirelatsioonide kinniste alamhulkade eraldamise tüüpilist tulemust ja näidatakse, kuidas neid tulemusi edasiseks tõlgendada.

Artikli piiratud mahu tõttu piirdume ainult nimisõnade hüperonüümiasteosega määratud allsüsteemi uurimisega. Eraldasime Eesti Wordnet'ist suure (40099 x 5887) tabeli, mille olemusest annab aimu allpool esitatud väike väljavõte (vt tabel 1).

⁵ The Global WordNet Association vt <http://www.globalwordnet.org/> (1.10.2011).

⁶ Üheliikmelised sünohulgad moodustavad tesauruses hetkel 57% kõigist sünohulkadest ning kõige rohkem on ühes sünohulgas 20 liiget (Orav jt 2011: 97).

⁷ Muu hulgas selgus näiteks, et autorid olid heauskselt kasutanud Eesti Wordnet'i veebivarianti, mis osutus viisi versiooni värskeimast vanemaks. Tänu Neeme Kahuski lahkele vastutulekule õnnestus peagi saada sõnastiku viimane versioon (versioon 60, 25.04.2011) Eesti Wordnet'i arendusvahendina kasutatava sisestusliidese Polaris exportfaili txt-formaadis. Edasine koostöö tesauruse kollektiiviga ongi kulgenud õige tulemusrikkalt.

Tabel 1. Suhtetabel, milles hüponüüm-sünohulkade esimesed liikmed (sõnad) on ridades ja hüperonüüm-sünohulkade liikmed (sõnad) on veergudes

	ehitustöörist	nimene	põhijoon	seisund	tegevus	tegu	osa
galeri
asi	■	.
alkuüdvärv
osa	■
alkuüdemal
emailvärv
koht	■
kord	.	.	.	■	.	.	.
lakkvärv
lõpp	.	.	.	■	.	.	■
märk
paviljon
päev
alus
elu
keel
konsool
koor
korraldus
käik	■	.	■
maa	■
tee	■	.
uretaanaküü

2. Infotöötlusmeetodi idee selgitus

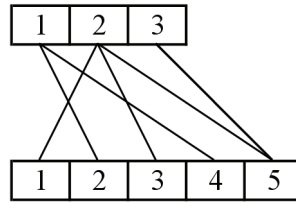
Selgitame kasutatava meetodi põhiolemust lihtsa kunstliku näitega.

Olgu antud objektide ja neid kirjeldavate tunnuste suhtemaatriks. Selle maatriksi ridadeks olgu Eesti Wordnet'i hüponüüm-sünohulgad ja veergudeks hüperonüüm-sünohulgad, s.t me kasutame selles näites vaid ühte suhet tesauruse paljudest suhetest.

Näitemaatriks on joonisel 1a (1 näitab mingi hüponüüm-sünohulga seotust teise, hüperonüüm-sünohulgaga).

1	0	1	0
2	1	0	0
3	0	1	0
4	1	0	0
5	0	1	1
	1	2	3

Joonis 1a. Suhtemaatriks



Joonis 1b. Kahealuseline graaf

Seda maatriksit saab esitada kahealuselise graafina, kus ühel joonel on objektide (ridade) järjenumbrid ja teisel joonel on tunnuste järjenumbrid (vt joonis 1b). Joonisel 1b on ülemisel joonel veerunumbrid ja alumisel joonel ridade numbrid.

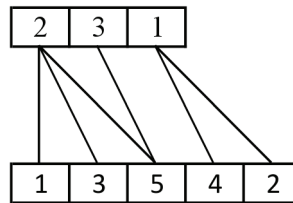
Joonisel 1b on kerge loendada, et joonte ristumisi on parajasti kolm. Kui objekte ja tunnuseid on palju (kümned, sajad, tuhanded, sajad tuhanded jne), siis muutuks sellise joonise koostamine peaaegu võimatuks.

Püstitame endale küsimuse: kas tabeli ridu ja veerge ümber järjestades ei oleks võimalik ristumiste arvu vähendada? Selline ekstreemumprintsipiidest lähtuv mõtteviis on matemaatikule omane.

Praeguse triviaalnäite korral on kerge otsese kontrolliga veenduda, et muutes ülemisel joonel tunnuste järjestuse (123) asemele (231) ja alumisel joonel ridade (12345) asemele (13542), saame korrastatud suhtemaatriksi ja vastava graafi (vt joonis 1c).

1	1	0	0
3	1	0	0
5	1	1	0
4	0	0	1
2	0	0	1
	2	3	1

Joonis 1c. Korrastatud suhtemaatriks ja kahealuseline graaf



Joonise 1c parempoolse osa põhjal on kohe näha, et esiteks ei ole enam ühtegi ristumist ja teiseks lagunes algne joonis 1b kaheks seotud tükiks.

Need lihtsad faktid on tähtsad, sest nad näitavad, et algne andmemaatriks laguneb kaheks sidusaks alamhulgaks (kinniseks hulgaks). Teiseks tekkis tunnuste ja objektide jaoks loomulik lineaarne järjestus, millele tavaliselt praktikas vastab samuti mingi loomulik semantiliselt sisuline järjestus.

Nüüd on kerge ka algmaatriksi ümber järjestada uude järjestusse (vt joonis 1c vasakpoolne osa). Selline ribamaatriks on psühholoogias tuntud multidimensioonalise järjestusskaalana (MDSCALE). Praegu me selle tulemuse juures ei peatu, aga Eesti Wordnet'i sisestruktuuri uurimisel võib sellest kasu olla.

3. Infotöötlusprobleemi mitteformaalne sõnastus

Järgnevalt püstitame endale ülesande leida sobiv meetod ja koostada algoritm ning programm, mille abil saada algandmete tabeli kahealuselisest esitusest minimaalse võimaliku joonte ristumiste arvuga esitus (reaalsetes andmetes ei pääse me lõpptulemuses ristumistest!). See probleem on graafiteoorias tuntud nime all “kahealuselise graafi ristumiste arvu minimeerimine” ning see on tõestatud NP-keerukas probleem.⁸ See tähendab, et lõplikku polünoomiaga töötavat algoritmi, mis garanteeriks vajaliku miinimumi, pole olemas (Garey, Johnson 1979). Mida siis teha?

Esimese vahesammuna – veel enne, kui asuda ristumiste minimeerimise juurde – saab lihtsa võttega andmetabelist eraldada kõik seotud kinnised hulgad. Nii suure andmekogu juures nagu Wordnet, on see väga suur ajavõit ja kergendab ühtlasi ka saadud tulemuste hilisemat tõlgendamist.

Meie pisinäites kasutatud maatriksesitus (vt joonis 1c) võtab suurte tabelite korral liiga palju mälu ja teeb ka algoritmid aeglaseks. Seepärast kasutasime Eesti Wordnet'i andmete jaoks nimistulist esitust (ehk *list*-esitust), kus iga sõna jaoks antakse nimekiri, millised sõnad on algsõnaga vastavas suhtes, hoides märkimisväärselt mälumahtu kokku. Selline esitus on tingitud Polarise eksportfaili andmetest, kus ükski sünohulk ei ole otseselt seotud ühegi teise sünohulga identifitseeriva numbriga ega sõnade loendiga, vaid on viidud vastavusse (on semantilises seoses) teise sünohulga esimese liikmega (sõnaga).

Teine väga oluline võte seisnes selles, et teisendasime kahealuselise graafi lihtsal viisil nn intervallgraafiks, viies ülemisel joonel olevad elemendid alumisele joonele seal juba paiknevate elementide järele. Loomulikult tuleb muuta nende elementide järjenumbreid. Liidame lihtsalt igale ülemise joone elemendi numbrile juurde arvu N (s.o objektide arvu). Selle teisenduse järel on tegu standardse intervallgraafiga (seosed jäävad muidugi alles!). Knuthi (1968) programmeerimisentsüklopeedias ja arvukate algoritmide raamatus (Flannery jt 2009: 346) leidub mõnus algoritm, mis võimaldab eraldada kõik algse andmetabeli kinnised hulgad väga kiiresti. Eesti Wordnet'i kõigi kinniste hulkade eraldamine hüperonüümsete suhete korral võttis (tavalisel sülearvutil) aega ainult 2,5 minutit.

Alles kõigi kinniste hulkade eraldamise järel kasutame Niermanni (2005: 41–46) kirjeldatud evolutsioonilise optimeerimise ideed, et ristumiste arv viia igas eraldatud kinnises hulgas miinimumini.

Selline samm-sammuline lähenemine on osutunud väga efektiivseks suurte kahendmaatriksite sisemise struktuuri uurimisel. Muidugi tuleb paraku märkida, et infotöötlusliku käsitluse täielik esitus jääb vastavasisulise ajakirja teemaks.

4. Näiteid eraldatud kinnistest hulkadest

Toodud valik eraldatud kinnistest hulkadest (vt tabel 2; joonis 2, 3, 4, 5) on illustreeriva iseloomuga, et anda aimu meetodiga saadavatest tulemustest. Tabeli 2 igal real on järgmine sisu: taustatoonita lahtrites on hüponüüm-sünohulkadest pärinevad sõnad ja tooniga lahtrites hüperonüüm-sünohulki esindavad sõnad, millel on semantiline seos hüponüümidega. Et ideaaljuhul on igal sünohulgal vaid üks hüperonüüm (Vider 2001), annavad hüperonüüm-sünohulka esindavad

⁸ NP (ingl *non-deterministic polynomial*) – mittedeterministlikult polünoomiaalne.

sõnad⁹ leksikograafide kiire ülevaate, kui paljude kõrgema taseme sünohulkadega on hüponüümid seotud. Nii on nt tabeli 2 esimesel real sünohulka esindavateks sõnadeks “hotell” ja “ujuvrajatis”. See tähendab, et kogu real kui kinnisel hulgal on kaks hüperonüüm-sünohulka. Millisest või millistest sünohulkadest pärinevad “flotell”, “pontoonkai” ja “ujuvkai”, ei ole siin tähtis teada. Kõikidel valge taustaga ridadel on ühe või mitme sünohulga sõnad täies mahus.

Tegelikus töös koostab programm KLASSID kõigepealt vastavalt uurija soovile suurest Eesti Wordnet'i andmefailist töötlusse mineva suhetabeli (vt tabel 1). Järgmise sammuna koostatakse kinniste sidusate alamhulkade loetelu. Tükike sellest tabelist on nähtav tabelis 2.

Tabel 2. Fragment kinniste alamhulkade tabelist

flotell_1_n	pontoonkai_1_n	ujuvkai_1_n	hotell_1_n	ujuvrajatis_1_n
frakk_1_n	vatt_3_n	kuub_1_n	peigmehe rõivastus_1_n	ülirikond_1_n
grammofon_1_n	plaadimasin_1_n	vinüülplaadi- mängija_1_n	fonograaf_1_n	taas- esitusseade_1_n
kabuhärg_1_n	kabuvärss_1_n	kohihärg_1_n	härg_1_n	kohiloom_1_n
muusikahall_1_n	black box_1_n	hall_2_n	teatrisaal_1_n	

Kõige suurem hüperonüümiasuhte abil Eesti Wordnet'ist eraldatud sidus kinnine alamhulk on mõõtmetega 4945 x 457 (hüperonüüm-sünohulka esindavat sõna). Kinniseid alamhulki tekkis üldse 6051. Jaotumine vastavalt: 4854 nimisõna, 1178 tegusõna ja 19 omadussõna hulka.

Kõigi semantiliste suhete ja liikide samaaegsel kasutamisel tekib kinniseid alamhulki ca 500.

Kolmanda sammuna valime nendest ribadest (vt tabel 2) suvalise meid huvitava seotud hulga ja laseme programmil MINCROSS luua minimaalsete ristumiste arvuga pildi. Tüüpiline väljatrükk tehtud tööst on niisugune, kus nii vasakpoolse ümberjärjestatud tabeli kui ka parempoolse joonise äärtel olevad numbrid osutavad sellele analüüsile eelnenud riba väljatrüki järjenumbritele (vt joonised 2, 3, 4, 5). Sellised minimaalsete ristumiste arvuga joonised on aluseks tekkinud kinniste hulkade edasisele sisulisele analüüsile.

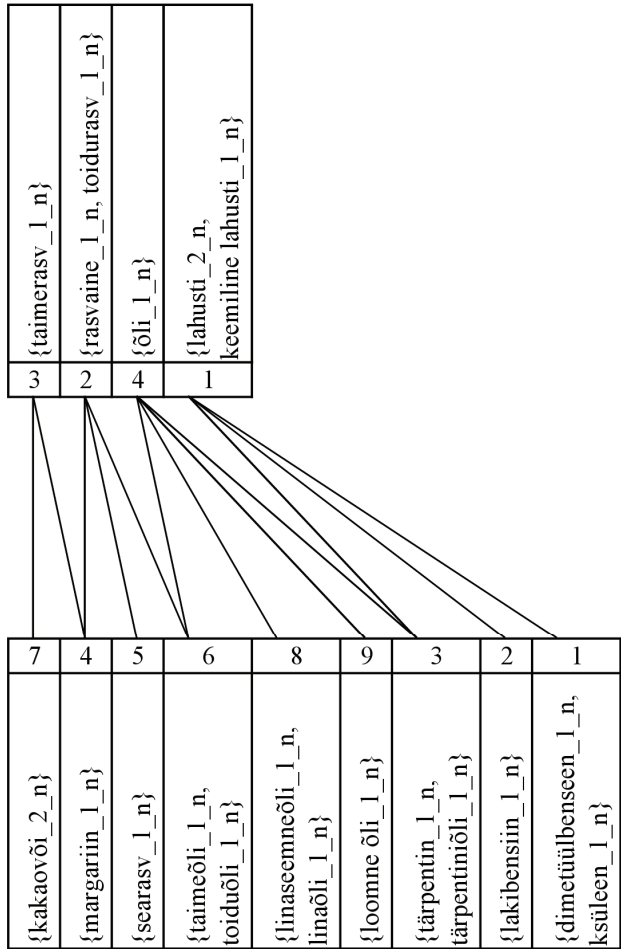
Wordnet-tüüpi sõnastikes on semantilised seosed esitatud sünohulkade vahel. Kuigi tutvustatava meetodi puhul ei ole tähtis, millised objektid millistes seostes esinevad, on antud artiklis jäädud siiski vaate “seosed sünohulkade vahel” juurde (vt joonised 2, 3, 4, 5).

Pisipilt joonisel 4 on hea näide (leksikograafi) mingil põhjusel pooleli jäänud töö kohta. Pildilt puudub rida teisi spordialasid või vähemalt hüppeid.

Niisuguse pildi nägemine peaks leksikograafi ergutama muude spordiga seotud sõnade ülesotsimisele, nende seospiltide loomisele ja siis ühendklastrite loomisele. Nii võib Eesti Wordnet'i lähemal uurimisel leida, et “kolmikhüpe” on sidumata kergetõustikuga, samas kui teised samalaadsed kergetõustikualad (teivashüpe, kaugushüpe, kõrgushüpe jt) on sellega seotud. Probleemi olemasolu kinnitab ka kergetõustiku kui sünohulga definitsioon: “üldisem spordiala, mille alla kuuluvad jooksud, kiirkäimine, hüpped, heited ja mitmevõistlus”.

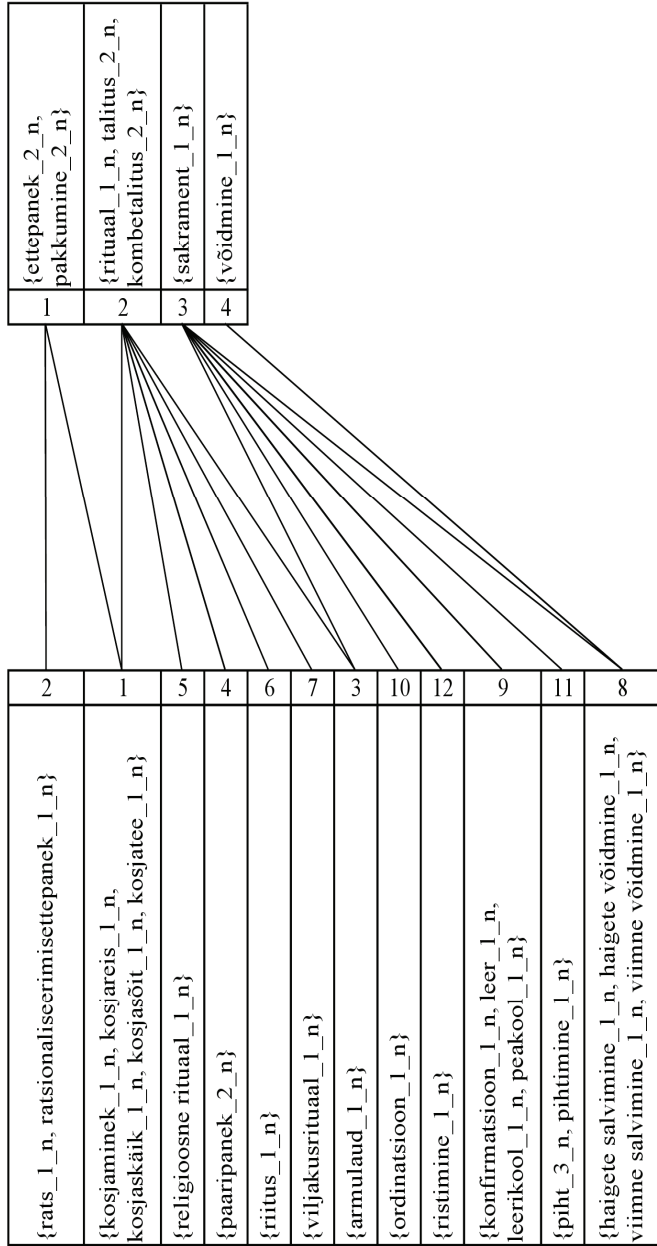
⁹ Eesti Wordnet'i veebirakenduses kasutatav formaat esitab samuti sünohulga hüperonüümiks vaid ühe (esimese) hüperonüüm-sünohulka kuuluva sõna. Sealt on artikli autorid ka eeskuju saanud.

7	1	0	0	0
4	1	1	0	0
5	0	1	0	0
6	0	1	1	0
8	0	0	1	0
9	0	0	1	0
3	0	0	1	1
2	0	0	0	1
1	0	0	0	1
	3	2	4	1

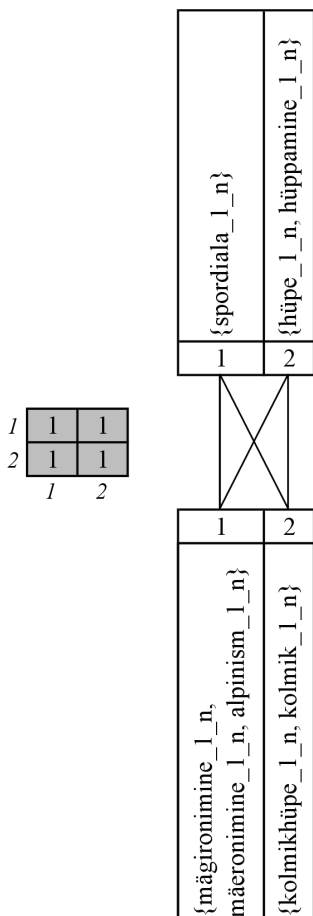


Joonis 2. Näide (1), ümberjärjestatud suhtematriks ja minimaalse ristumiste arvuga kahealuseline graaf

2	1	0	0	0
1	1	1	0	0
5	0	1	0	0
4	0	1	0	0
6	0	1	0	0
7	0	1	0	0
3	0	1	1	0
10	0	0	1	0
12	0	0	1	0
9	0	0	1	0
11	0	0	1	0
8	0	0	1	1
	1	2	3	4



Joonis 3. Näide (2), ümberjärjestatud suhtemaatriks ja minimaalse ristumiste arvuga kahealuseline graaf



Joonis 4. Näide (3), ümberjärjestatud suhtemaatriks ja minimaalse ristumiste arvuga kahealuseline graaf

5. Vihjeid saadud piltide edasiseks analüüsimiseks

Vaatleme näitena pilti, kus peaks huvi äratama hüperonüümide “võidmine” ja “sakrament” vahekord (vt joonis 3). Programmi MINCROSS töö tulemusel saadud pildilt on näha, et hüperonüümi “võidmine” kõik seosed kuuluvad “sakramendi” alla. Järelikult on oodata, et “sakrament” oleks “võidmise” hüperonüüm. Edasine ebakoht seoste skeemis on seotud sõnaga “armulaud”. Põhimõtteliselt õigesti on “armulaud” seotud nii “sakramendi” kui “rituaaliga”. Ometi on ju kõik “sakramendi” hüponüümid “talitused” (toimingud). Need seosed on ilmselt süsteemist puudu.

Lisaks veel üks pisipeensus. Kuigi abielu kuulub sakramendi hulka, on Eesti Wordnet’is ta liigitatud vaid rituaaliks ja mitte otseselt, vaid kaudselt sõna “paaripaneku” kaudu!

Joonisel 2 on sattunud kokku “lakibensiin” ja “margariin”. Esimene neist on lahusti, teine aga Eesti Wordnet’i andmetel lahustite hulka ei kuulu. Margariini ja lakibensiini seostatus oleks ilmselt põhjendatud, kui esimene neist omaks (semantilist) seotust toiduainekeemiaga.

Äsjane informaatikute tehtud lühianalüüs on ilmselt üpris ebatäielik, aga ometi peaks see osutama MINCROSS programmi pakutavatele huvitavatele võimalustele Eesti Wordnet'i sisestruktuuri üksikasjalikuks süvauurimiseks.

Lõpuks veel üks lisavõimalus, mille pakub juursünohulga kasutusele võtmine.

Juursünohulk on siin defineeritud kui sünohulk, mis rahuldab samaaegselt kahte tingimust: 1) on vähemalt ühele sünohulgale hüperonüüm-sünohulgaks ja 2) talle endale ei ole vastavusse seatud ühtegi hüperonüüm-sünohulka. Ehk sünohulk, millel on alluvad, kuid puuduvad ülemused. Neile tingimustele vastavaid sünohulki on Eesti Wordnet'is 185. Jagunemine vastavalt: 142 nimisõna, 24 tegusõna ja 19 omadussõna sünohulka.

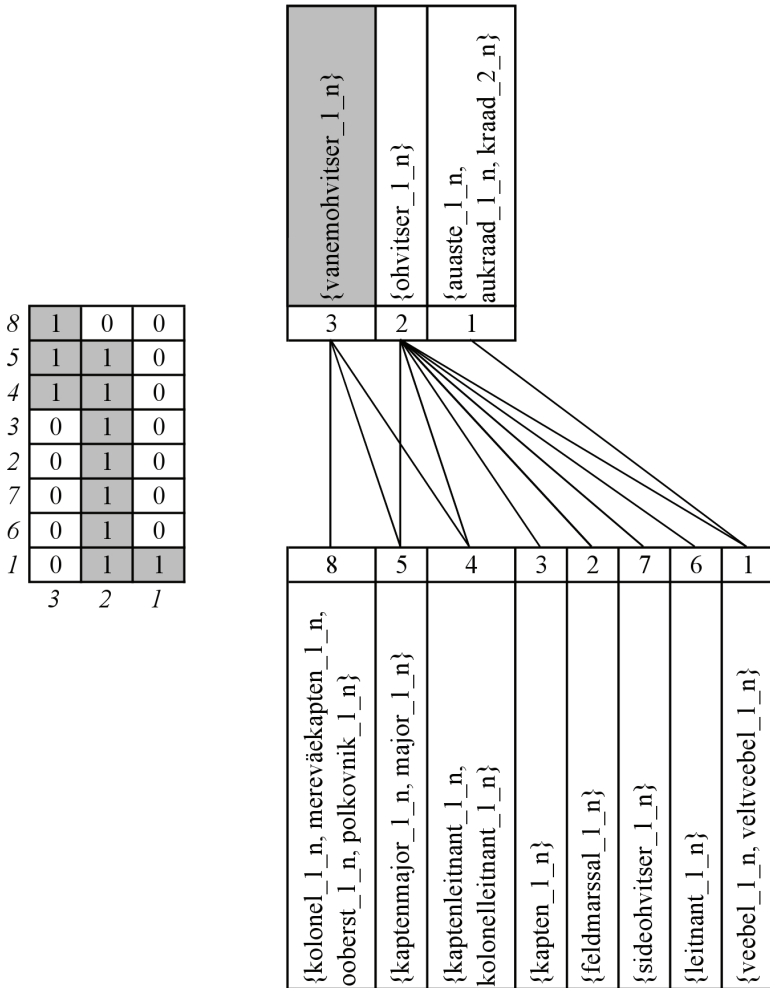
Juursünohulkade läbivaatamine oleks muidugi omaette teema. Meid aga huvitab kahealuseliste graafide puhul see, kas esineb selliseid graafe, mille hüperonüümiosa (kahendgraafi ülemine rida) sisaldab muuhulgas juursünohulki. Kinnise hulga hüperonüümsünohulgale saab tuua juurde uued hüperonüümiaseosed. Niimoodi järjest üles liikudes mõistame selgesti, et sünohulgad paiknevad oma hüperonüümiapuus eri tasemetel. Juursünohulkade esinemine kahealuselise graafi hüperonüümiosa osas võiks sellisel juhul viidata kas võimalikule probleemile eri taseme sünohulkade kasutamise kohta, mis ei pruugi küll alati probleemiks osutuda, kuid mis võib juhatada pooleli jäänud (semantilisele) seostamisele. Sellist näidet illustreerib joonis 5: {*vanemohvitser_1_n*} on siin juursünohulk (eristatud halli taustaga). S.t sellel sünohulgal endal ei ole ühtegi hüperonüüm-sünohulka. Täpsemal uurimisel selgubki, et ohvitseri ülemmõisteks on sõjaväelane, mis on ju ka "vanemohvitser". Antud näite puhul hakkab veel silma, et {*auaste_1_n*, *aukraad_1_n*, *kraad_2_n*} on sidumata auastmeid sisaldavate sünohulkadega.

Lõpetuseks, kui vaadata näiteks hüperonüüm-sünohulki (vt joonis 3) {*ettepanek_2_n*, *pakkumine_2_n*}, {*rituaal_1_n*, *talitus_2_n*, *kombetalitus_2_n*}, {*sakrament_1_n*}, {*võidmine_1_n*}, siis pärast seda, kui leksikograaf on sõltuvused korrastanud, võib see korrastus aidata kergemini käsitleda taksonoomilist lisatükeldust paljude alluvatega ülemmõistete korral (Orav jt 2011: 101, Kozareva, Hovy 2010: 1110-1111)¹⁰. Näiteks joonisel 2 esitatud ülemmõisted "taimerasv" ja "lahusti" viitavad selgele vajadusele täpsustada taksonoomilist struktuuri. Pildil eksisteeriv seoste süsteem ise annab aga võimaluse luua korralik side ka hüperonüümide vahel ning saada kõrgema taseme seosgraaf, mis n-õ ronib antud relatsiooni abil juurmõistele (juursünohulgale) lähemale. Sellise automaatse või interaktiivse pealisehituse loomine on muidugi omaette tõsine töö.

6. Kokkuvõtte

Elu on näidanud, et sageli saavutatakse mingis valdkonnas edu, kui sekkuvad "mitteeksperdid". Seetõttu esitavadki autorid – informaatikud – Eesti Wordnet'i jaoks spetsiifilisi andmetöötluse ja visualiseerimise meetodeid. Piltide ja tabelitena esitatud tagasisidemehhanismid sobivad iga *wordnet*-stiilis loodud süsteemi jaoks, võimaldades semantilist analüüsi ja seaduspärasuste uurimist. Leksikograaf saab kasutada uudset lähenemist, et uurida süsteemi varjatud struktuure ja teha vajalikke parandusi.

¹⁰ Hoolimata paljudest katsetest ei ole täisautomaatset taksonoomilist tükeldust/liigitamist veel teostatud, *wordnet*-tüüpi sõnastikes koostatakse need aga käsitsi. Nii on loomulik, et tekivad korrastust vajavad vead.



Joonis 5. Näide (4) – juursünohulk, ümberjärjestatud suhtemaatriks ja minimaalse ristumiste arvuga kahealuseline graaf

Artiklis kasutatud mitmed keerukad andmeanalüüsi meetodid (kahealuselise graafi klasterdamine, intervallgraafi servade ristumiste arvu minimeerimine) on eelkõige sobilikud suvaliste relatsiooniliste süsteemide kinniste hulkade eraldamiseks ja nende korrastamiseks. Seega sobib hästi neid kasutada ka *wordnet*-tüüpi sõnastike hetkeolukorra kajastamiseks.

Positiivne tagasiside Eesti Wordnet'i tegijatelt ja valdajatelt lubab olla kindel, et esitatud meetodid oleksid kasulikud ka teiste keelte *wordnet*-süsteemidele.

Viidatud kirjandus

- Flannery, P. Brian; Press, H. William; Teukolsky, A. Saul; Vetterling, T. William 2009. Numerical Recipes in C. The Art of Scientific Computing. South Asia: Cambridge University Press India.
- Garey, Michael R.; Johnson, David S. 1979. Computers and Intractability: A Guide to the Theory of NP-Completeness. New York: W. H. Freeman.
- Knuth, Donald Ervin 1968. Fundamental Algorithms. The Art of Computer Programming, Vol. 1. Reading, MA: Addison-Wesley.
- Kozareva, Zornitsa; Hovy, Eduard 2010. A semi-supervised method to learn and construct taxonomies using the web. – Conference on Empirical Methods in Natural Language Processing, MIT, Massachusetts, USA, 9-11 October 2010, 1110–1118.
- Lohk, Ahti; Võhandu, Leo 2011. Eesti sünovara visualiseerimisest. – Ettekanne 10. rakenduslingvistika kevadkonverentsil, 28.-29. aprill 2011, Tallinn.
- Niermann, Stephan 2005. Optimizing the Ordering of Tables With Evolutionary Computation. – The American Statistician, 59 (1), 41–46.
- Orav, Heili; Kerner, Kadri; Parm, Sirli 2011. Eesti Wordnet'i hetkeseisust. – Keel ja Kirjandus, 1, 96–106.
- Vider, Kadri 2001. Eesti keele teaurus – teooria ja tegelikkus. – M. Langemets (Toim.). Leksikograafiaseminar “Sõna tänapäeva maailmas” / Leksikografinen seminaari “Sanat nykymaailmassa”. Ettekannete kogumik. Eesti Keele Instituudi toimetised, 9. Tallinn: Eesti Keele Sihtasutus, 134–156.

Võrgumaterjalid

- EuroWordNet <http://www.illc.uva.nl/EuroWordNet/> (1.10.2011).
- Filosoof OÜ teaurus http://www.filosoof.ee/thes_et/ (1.10.2011).
- Princetoni Wordnet <http://wordnet.princeton.edu/> (01.10.2011).
- Tartu Ülikooli eesti keele teaurus (TEKSaurus) ehk Eesti Wordnet <http://test.cl.ut.ee/ressursid/teksaurus/> (01.10.2011).

Ahti Lohk (Tallinna Tehnikaülikool), uurimisvaldkond on teksti ja keeruliste suhtesüsteemide uurimine. ahti.lohk@ttu.ee

Leo Võhandu (Tallinna Tehnikaülikool), uurimisvaldkond on suurte andmesüsteemide süvastruktuuride uurimine. leov@staff.ttu.ee

STRUCTURAL ANALYSIS OF THE CURRENT ESTONIAN WORDNET

Ahti Lohk, Leo Võhandu

Tallinn University of Technology

Control of any expanding and developing system requires a feedback control mechanism to evaluate the normal trends of the system and also the unsystematic steps. Experience has shown that often more progress is achieved in a field if non-specialists intervene. For this reason the authors (from the field of informatics) suggest special data processing and visualization methods for the Estonian Wordnet. The images presented are suitable as a feedback mechanism for every Wordnet-style linked system, for semantic analysis and examination of regularities. They allow the lexicographer to use an innovative approach to study the hidden structures of the system and make corrections if needed. Several complex data analysis methods have been used (bipartite graph clustering, minimization of interval graph crossing numbers) and the results presented in this article. Positive feedback from the makers and maintainers of the Estonian Wordnet allows us to be sure that the methods presented would be useful for other language Wordnets also.

Keywords: Thesaurus, semantic relations, visualization of Wordnet, ordering of relations, Estonian

AUTOMAATNE AJAVÄLJENDITE TUVASTAMINE EESTIKEELSETES TEKSTIDES

Siim Orasmaa

Ülevaade. Artikkel käsitleb eestikeelsete tekstide arvutianalüüsi alamprobleemi: ajaväljendite automaatset tuvastamist tekstist. Ülesanne on püstitatud kaheosalisena: tekstist tuleb üles leida ajaväljendid (piiritleda ajaväljendifraasid) ning normaliseerida leitud ajaväljendite semantika (st esitada semantika eeldefineeritud märgenduskeele raamid). Artiklis kirjeldatakse ajaväljendite tuvastamisel kasutatavat märgenduskeelt ning piiritletakse vaadeldavate ajaväljendite hulk lähtuvalt märgenduskeele (aga ka praktilise analüüsi) võimalustest. Antakse ülevaade loodud reeglipõhise ajaväljendite tuvastaja tööpõhimõtetest ajaväljendite leidmisel ning semantika normaliseerimisel kasutatavatest strateegiatest. Programmi testimiseks moodustatakse Tartu Ülikooli koondkorpuse tekstidest u 70 000-sõnaline korpus, millel parandatakse käsitsi automaatse tuvastamise vead ning hinnatakse tuvastaja töö kvaliteeti.*

Võtmesõnad: arvutilingvistika, reeglipõhine keeletötlus, semantiline märgendus, annoteerimine, eesti keel

1. Sissejuhatus

Ajaväljendite tuvastamine on tekstide automaatse analüüsi alamprobleem, mis seisneb ajaväljendite leidmises tekstist ning nende semantika kirjeldamises fikseeritud märgendusviisi alusel. Probleemi lahendamise aitab kaasa mitmete keeletehnoloogia rakenduste (nt automaatne küsimustele vastamine, sisukokkuvõtete tegemine, dialoogisüsteemid) arengule. Samuti võib ülesannet käsitleda laiemas kontekstis, tekstide ajasemantilise analüüsi alamülesandena, kus laiemaks probleemiks on sündmuste tuvastamine tekstides ning sündmuste ajalise järgnevuse määramine.

Viimase aastakümne jooksul on tekstide ajasemantiline analüüs pärvinud automaatse infoeraldamise (ingl *information extraction*) kontekstis laialdast tähelepanu. Välja on töötatud ajasemantika märgendamise keeli, nt spetsiaalselt

* Artikli valmimist on toetanud Euroopa Regionaalarengufond Eesti Arvutiteaduse Tippkeskuse kaudu. Autor tänab Tarmo Vainot nõu eest koondkorpuse märgendamise küsimustes ja anonüümseid retsensente kasulike kommentaaride eest.

ajaväljendite märgendamiseks loodud keel TIMEX2 (Ferro jt 2005) ning ajaväljendite, sündmuste ja nende vaheliste seoste märgendamiseks loodud keel TimeML (Pustejovsky jt 2003). Inglise ja prantsuse keele jaoks on loodud ka TimeML märgendusega tekstikorpused (nn TimeBank korpused¹).

Eesti keeles on ajaväljendite tuvastamise ülesannet käsitletud kõige põhjalikumalt dialoogisüsteemide kontekstis (Treumuth 2008), kus eesmärgiks on olnud kasutajasisendist ajaväljendite leidmine ning nendele vastavate infopäringute tegemine (nt kinokavast etteantud kuupäeva(de)l linastuvate filmide leidmine).

Artikkel toetub autori magistritööle (Orasmaa 2010) ning käsitleb ajaväljendite tuvastamist tekstide (ajasemantilise) märgendamise ülesandena. Esmalt antakse ülevaade kasutatavast märgenduskeelest ning tutvustatakse lühidalt ajaväljendite tuvastamist teistes keeltes. Seejärel esitatakse ülevaade loodud süsteemist, pöörates eraldi tähelepanu ajaväljendite eraldamisele ning normaliseerimisele. Artikli lõpuosas kirjeldatakse süsteemi hindamiseks valitud korpust ning analüüsitakse tuvastaja töö kvaliteeti selle korpuse põhjal.

2. Märgendatavad ajaväljendid ja märgenduskeel

Ajaväljendite märgendamisel on siin võetud aluseks märgenduskeel TimeML (Pustejovsky jt 2003)². Ajaväljendite kirjeldamise vahendid keeles TimeML on robustsed: need on orienteeritud eelkõige kalendrilise ja absoluutse semantika märgendamisele ning võimaldavad relatiivse semantikaga³ väljendite lahendamist teatud määral edasi lükata (st esitatakse vaid osaline semantika lahendus ning lõplik interpreteerimine jäetakse lõppkasutaja hooleks). Järgnevalt antakse märgenduskeelest lühiülevaade, põhjalikuma käsitluse võib leida TimeML märgendusjuhistest (Saurí jt 2006).

Antud märgenduskeele järgi jagatakse ajaväljendid järgmisteks liikideks:

- **kalendrilised toimumisajad** (ingl *date*), mille alla kuuluvad ajaväljendid võib paigutada ajateljele ning need sisaldavad *aasta, kuu, nädala* või *päeva* detailsusega ajalist informatsiooni, nt ajaväljendid *1999. aastal, eelmine kuu, kolmandal nädalal* ning *22. veebruariks*. Siia alla ei kuulu kellaaja detailsusega ajalist informatsiooni sisaldavad ajaväljendid;
- **kellaajalised toimumisajad** (ingl *time*), mille alla kuuluvad ajaväljendid võib paigutada ajateljele ning need võivad sisaldada kuupäevalise detailsusega informatsiooni ja peavad sisaldama kellaaja detailsusega ajalist informatsiooni (nt *reedel kell 13.45, poole kolmeks*) või päevaosa detailsusega ajalist informatsiooni (nt *järgmise hommikuni*);
- **ajalised kestused** (ingl *duration*), mille alla kuuluvad ajaväljendid avalduvad eelkõige intervallidena, nt *kolm päeva, 8 kuu jooksul*. Siia alla loetakse ka nn suunaga kestused (nt *eelneva 6 kuu jooksul, järgmised kolm aastat*), kuna esmalt on selliste väljendite semantika normaliseerimisel võimalik välja tuua kestus (kuigi ka ajateljele paigutamine võib olla teatava analüüsi järel võimalik);

¹ Vt <http://timeml.org/site/timebank/timebank.html>, <http://www.linguist.univ-paris-diderot.fr/~abittar/french-timebank/> (26.09.2011).

² Teatud erijuhtudel minnakse käesolevas töös TimeML märgendusviisist lahku. Need erijuhud on dokumenteeritud ja huvi korral autori kaudu kättesaadavad.

³ Absoluutse semantikaga ajaväljendid on üheselt ajateljel piiritletavad (nt 2009. aastal, 22. oktoober 2011); relatiivse semantika korral nõuab väljendi ajateljele paigutamine täiendavat kontekstiinformatsiooni (nt reedel, järgmisel kuul).

- **ajalised korduvused** (ingl *set of times*), mille alla kuuluvad ajaväljendid viitavad (sündmus-)aegade kordumisele, nt *igal aastal, kolm korda nädala jooksul*.

Selline liigitusviis ühtib osaliselt ka “Eesti keele grammatikas” (EKG II: 76–86) kasutatud ajamääruste liigitusviisiga. Seal jagatakse ajamäärused nelja rühma: 1) toimumisaega väljendavad ajamäärused (nt *homme*), 2) ajapiiri väljendavad ajamäärused (nt *hommikust, lõunani*), 3) kestust näitavad ajamäärused (nt *seitse tundi*) ja 4) korduvust väljendavad ajamäärused (nt *öösi*). Kuigi TimeML ei näe ette ajapiiride väljatoomist eraldiseisva ajaväljendiliigina, on seal vahendid ajavahemike märgendamiseks hierarhiliselt, kombineerides toimumisaegade ning kestuste märgendusi (sellest edaspidi täpsemalt).

Erinevalt EKG II käsitlusest piirdub käesolevas töös vaadeldavate ajaväljendite hulk väljenditega, mis rahuldavad vähemalt üht kahest kriteeriumist:

- a) ajaväljend sisaldab *aasta, kuu, nädala, päeva, tunni* või *minuti* detailsusega ajalist informatsiooni. Vaatluse alt jäävad välja näiteks mittekonkreetsed ajaväljendid (*omal ajal, lühikese aja jooksul, aeg-ajalt*) ning valdkonnaspetsiifilised ajaväljendid (nt *sel hooajal, teiseks poolajaks, varase kiviaja lõpuosas, nõukogude ajal*);
- b) ajaväljend viitab minevikule, olevikule või tulevikule, lähtuvalt teksti kirjutamise hetkest (nt *hiljuti, praegu, varsti* prototüüpses kasutuses) või mõnest märgendamisele kuuluvast ajaväljendist (nt *varem, tookord, hiljem* prototüüpses kasutuses). Seega jäävad vaatluse alt välja tekstis kirjeldatud sündmustele toetuvad ajalised viited.

Toodud kitsendused tingib eelkõige praktiline nõue: märgendatav väljend peaks olema ka normaliseeritav antud märgenduskeele raamides ning normaliseerimine ei tohiks nõuda olulist valdkonna- või taustateadmiste rakendamist.

Märgendus näeb ette väljendi ümbritsemist TIMEX-märgenditega⁴, mille atribuutides tuuakse välja ajaväljendi unikaalne identifikaator (atribuudis *tid*), liik (atribuudis *type*), normaliseeritud kujul semantika (atribuudis *value*) ning olemasolul semantika mõjutajad (atribuudis *mod*). Näited märgendatud lausetest:

- (1) **1929. aastal** toodeti USAs 5,4 miljonit autot.
<TIMEX tid=“t1” type=“DATE” value=“1929”>1929. aastal</TIMEX>
toodeti USAs 5,4 miljonit autot.
- (2) **1999. aasta lõpul** esilinastunud Macbeth on teeninud meedias erakordset tähelepanu.
<TIMEX tid=“t2” type=“DATE” value=“1999” mod=“END”>1999. aasta lõpul</TIMEX> esilinastunud Machbeth” on teeninud meedias erakordset tähelepanu.

Semantika normaliseerimiskujud (atribuudi *value* väärtused) baseeruvad rahvusvahelisel kalendriaegade esitamise standardil (ISO-8601) ning jagunevad kolmeks:

- **kuupõhised toimumisaja esitused** (nt *15. veebruaril 2009* esitatakse kujul 2009-02-15);

⁴ Rangelt TimeML formaati järgides peaks märgendi nimi olema TIMEX3. Käesolevalt kasutatakse nimetust TIMEX, kuna käesolev märgenduskeel ei ühildu täielikult TimeML formaadiga.

- **nädalapõhised toimumisaja esitused** (nt eeldades, et kõnehetk viitab 2010. aasta 35. nädalale, võib ajaväljendi *järgmise nädala teisipäeval* normaliseerida kujul 2010-W36-2);
- **kestuste esitused** (nt *kolm aastat* esitatakse kujul P3Y).

Kuupõhist ja nädalapõhist normaliseerimiskuju võib paremast otsast pikendada või lühendada vastavalt sellele, millise ajalise detailsusega informatsiooni ajaväljend sisaldab. Näiteks väljendi *2009. aasta maikuus* puhul tuuakse välja vaid aasta ja kuu: 2009-05, samas väljendi *2009. aasta 20. mai hommikul kell 6* puhul saab normaliseerimisega minna kuni tunni täpsuseni kellaajas: 2009-05-20T06.

Lisaks näeb märgendusformaad ette erimärgiste kasutamist aastaegade, kvartalite, poolaastate, nädalalõppude ning päevaosade semantika normaliseerimisel, nt *2009. aasta suvel* esitatakse kujul 2009-SU (SU = ingl *summer*) ning *2011, 6. mai õhtu* esitatakse kujul: 2011-05-06TEV (EV = ingl *evening*).

Ajavahemiku normaliseerimisel märgendatakse eraldi vahemiku otspunktid ning nendega määratud kestus. Selle järgi võib eristada ilmutatud otspunktidega vahemikke (3) ning varjatud otspunktidega vahemikke (4).

- (3) Maailma Kirikunõukogus töötas Tutu **aastail 1972–1975.**

Maailma Kirikunõukogus töötas Tutu
 <TIMEX tid="t3" type="DATE" value="1972">aastail 1972</TIMEX>–
 <TIMEX tid="t4" type="DATE" value="1975">1975.</TIMEX>
 <TIMEX tid="t5" type="DURATION" value="P3Y" beginPoint="t3"
 endPoint="t4"/>

- (4) Flaami Bloki populaarsus on märkimisväärselt kasvanud just **viimase kümne aasta jooksul.**

Flaami Bloki populaarsus on märkimisväärselt kasvanud just <TIMEX tid=t6 type="DURATION" value="P10Y" beginPoint="t7" endPoint="to">viimase kümne aasta jooksul.</TIMEX>
 <TIMEX tid="t7" type="DATE" value="1990" />

Näites (3) on ajavahemiku otspunktid tekstis kujul *1972* ja *1975* ning seda vahemikku kattev kestus tuuakse välja ilma tekstilise sisuta märgendis, mille identifikaator on *t5*. Kestus viitab otspunktidele atribuutide *beginPoint* ja *endPoint* kaudu. Näites (4) on tekstis ilmutatud kujul kestus (*viimase kümne aasta jooksul*) ning *beginPoint* ja *endPoint* kaudu viidatakse selle varjatud otspunktidele. Otspunkt identifikaatoriga *t7* on esitatud kui ilma tekstilise sisuta märgend ning otspunkt identifikaatoriga *to* viitab vaikumisi alati teksti kirjutamise ajale (näite 4 tekst on kirjutatud kuupäeval 2000-10-10).

Hägusa või mittetäieliku semantika esitamiseks on kolm viisi. Esiteks, kui ajaväljend sisaldab semantikat mõjutavat sõna või fraasi (nt *1990ndate alguses, aprilli lõpus, rohkem kui 5 päeva*), tuuakse semantika mõjutaja ingliskeelne tähis välja atribuudis *mod*, vt (2). Teiseks, kui ajaväljend näitab ainult ajalise viite suunda, täpsustamata ajalist detailsust (kriteeriumi *b* rahuldavad väljendid), kasutatakse atribuudi *value* väärtustena eritähiseid PAST_REF (minevikuviide), PRESENT_REF (olevikuviide) ja FUTURE_REF (tulevikuviide). Kolmandaks, kui ajaväljendis avaldub mingi ajaline detailsus, aga selle väärtus on mittekongkreetne,

kasutatakse atribuudi *value* esituses konkreetsete väärtuste asendamist tähtedega X (nt *ühel päeval* esitatakse atribuudi *value* väärtusena XXXX-XX-XX ning *aastateks* atribuudi *value* väärtusena PXY).

Ajaliste korduvuste märgenduslaad tugineb kestuse märgenduslaadile. Korduvust hõlmava perioodi pikkus tuuakse välja atribuudis *value* ning atribuudis *freq* täpsustatakse täisarvuline kordumissagedus, millega võib olla seotud ajaline detailsus (5).

- (5) Seda vahemaad läbib Heimonen **vähemalt neli korda päevas**.

Seda vahemaad läbib Heimonen

<TIMEX type="SET" value="P1D" mod="EQUAL_OR_MORE" freq="4X">

vähemalt neli korda päevas</TIMEX>.

3. Ajaväljendite tuvastamisest teiste keelte arvutianalüüsis

Automaatset ajaväljendite tuvastamist on sageli vaadeldud kahest alamülesandest koosnevana: ajaväljendite eraldamine tekstis (ingl *recognition of temporal expressions*) ning ajaväljendite semantika normaliseerimine (ingl *normalization of temporal expressions*).

Lähenemised, mis keskenduvad ainult ajaväljendite eraldamisele, kasutavad sageli statistilist masinõpet. Juhendatud masinõppe puhul õpitakse käsitsi märgendatud treeningkorpusest statistiline ajaväljendite keelemudel, mida on seejärel võimalik ennustavalt rakendada uute tekstide märgendamisel (Kolomiyets, Moens 2009). Osaliselt juhendatud masinõppe puhul kasutatakse väikest arvu positiivseid näiteid, et õppida märgendamata korpusest uute ajaväljendite tekstilist kuju ja fraasstruktuuri kirjeldavad mustrid (Craveiro jt 2009).

Lähenemised, mis lisaks ajaväljendite eraldamisele seavad eesmärgiks ka ajaväljendite kalendriline semantika normaliseerimise, on seni valdavalt kasutatud reeglipõhiseid meetodeid. Ühe näitena sellistest lähenemistest võib vaadata süsteemi Chronos (Negri, Marseglia 2004), kus kasutatakse ligi 1000 põhireeglit ajaväljendite eraldamiseks ja kontekstiinfo kogumiseks, väikest arvu lisaregleid märgenduse ülekattuvuse probleemide lahendamiseks ning heuristilisi eeskirju ajaväljendite ankurdamisel ja lõpliku normaliseerimisstrateegia valimisel.

4. Ajaväljendite tuvastamine eestikeelses tekstis

4.1. Ülevaade süsteemist

Ajaväljendite tuvastaja on programm, mis saab sisendiks loomuliku keele teksti ning teksti loomise aja (nn kõnehetke). Programmi väljundiks on tekst, milles on märgendatud ajaväljendid, juhindudes eelnevalt kirjeldatud märgenduskeelest.

Programmi detailse kirjelduse leiab tööst (Orasmaa 2010). Selles ja järgnevatel alapeatükkides kirjeldatakse olulisemaid aspekte programmi toimeloogikas, laskumata tehnilistesse detailidesse.

Siinne ajaväljendite tuvastaja on üles ehitatud kahte liiki reeglitele. Primaarsed reeglid on **tuvastamisreeglid**, mis kirjeldavad ühelt poolt ajaväljenditele vastavaid fraase⁵ tekstis **fraasimustrite** abil ning teiselt poolt annavad edasi operatsioonide jada, mis tuleb ajaväljendi semantika leidmiseks läbi viia. Sekundaarsed reeglid on **liitumisreeglid**, mis täpsustavad, kuidas kõrvuti paiknevad ajaväljendid ühendatakse pikemateks fraasideks.

Tuvastamisreeglite abil eraldatakse tekstist võimalikult lühikesed terviklikku semantikat kandvad fraasid (näiteks *järgmisel neljapäeval, aprillis, kell pool kuus*). Enamasti annavad sellised fraasid edasi ajalist informatsiooni ühe detailsuse (ajaühiku) piires. Samuti kasutatakse tuvastamisreegleid semantikat mõjutavate fraaside eraldamiseks (nt *rohkem kui, umbes*). Liitumisreeglid näitavad, kuidas tuvastamisreeglite poolt eraldatud fraase võib liita pikemate ajaväljendifraaside moodustamiseks (nt *järgmisel neljapäeval + kell pool kuus*). Liitumisreeglite kasutamine võimaldab vähendada tuvastamisreeglite hulka (ei tule defineerida eraldi tuvastamisreegleid kõikvõimalike ajaliste täpsustuste jaoks, nt *mullu, mullu aprillis, mullu 26. aprillil* jne) ning samuti pakub osalist⁶ lahendust sõnajärje probleemile (ei tule defineerida eraldi reegleid kõikvõimalike fraasijärjestuste nagu *täna kell 8 hommikul, kell 8 täna hommikul, täna hommikul kell 8* tabamiseks).

Iga eraldatud ajaväljendiga seotakse semantika kirjeldus, mis on defineeritud **operatsioonide** (arvutuskäskude) jadana. Semantika normaliseerimisel rakendatakse neid operatsioone ettemääratud järjekorras. Võimalikud operatsioonid võib jagada kolmeks liigiks: 1) kalendriaritmeetika operatsioonid, 2) ankurdamise operatsioonid, 3) märgenduse atribuutide täitmise operatsioonid. Kalendriaritmeetika operatsioonid võimaldavad teha kalendrimudelil arvutusi (nt liita või lahutada ajaühikuid). Vaikimisi rakendatakse kalendriarvutusi programmile sisendiks antud kuupäeval (teksti loomise aeg), ankurdamise operatsioonid võimaldavad aga muuta kalendriarvutuste alust: näiteks võib arvutamisel aluseks võtta tekstis eelneva ajaväljendi normaliseeritud semantika. Märgenduse täitmise operatsioonid võimaldavad seada ajaväljendi atribuutide väärtuseid (nt atribuudi *mod* väärtust).

Ajaväljendiga seotud semantikaoperatsioonide defineerimisel tuleb arvestada ka ajaväljendite liitumisega, st semantika lahenduskäik on oluliselt erinev tekstis üksikult esineva ajaväljendi puhul (nt *esmaspäeval*) ning teise ajaväljendiga liitunud ajaväljendi puhul (nt *järgmise nädala + esmaspäeval*). Probleemi lahenduseks saab iga semantikaoperatsiooniga siduda rea **kontekstikitsendusi**, mis võimaldavad enne operatsiooni rakendamist kontrollida ajaväljendite liitumise ja ankurdamise tulemusi.

4.2. Ajaväljendite eraldamine

Ajaväljendite eraldamisele eelneb teksti eeltöötlus. Selles etapis viiakse läbi sisendteksti automaatne osalausestamine, morfoloogiline analüüs ja ühestamine (Filosoofi programmi *t3mesta* abil) ning leitakse järgmistel analüüsietappidel olulised tunnused (nt arvsõnafraaside paiknemine ja normaliseeritud semantika, verbide grammatilised ajad).

⁵ Fraasi all on siinses kirjutises mõeldud ühest või mitmest sõnast koosnevat sõnade järjendit, esitamata sellele mingeid täpsemaid süntaktilisi kriteeriume.

⁶ Lahendus on osaline, kuna ei aita juhtudel, kui sõnajärg varieerub normaliseerimise seisukohast terviklikes fraasides, nt *viimase kolme kuu* ja *kolme viimase kuu*.

Ajaväljendite eraldamiseks tekstis kasutatakse tuvastamisreeglite alla kuuluvaid **fraasimustreid**. Fraasimuster sisaldab üldistavat fraasikirjeldust ning sarnaneb toimeloogika poolest lõplikule mittedetermineeritud automaadile. Fraasimustrile antakse ükshaaval sisendteksti sõnu ette ning iga kord, kui järjestikku etteantud sõnad vastavad mingile mustri poolt kirjeldatud fraasile (terves fraasi ulatuses), teostatakse uue **ajaväljendikandidaadi** eraldamine.

Fraasimustris on sõnad kirjeldatud **sõnamallide** abil. Käesolevas töös eristatakse nelja sõnamalliliiki: **regulaaravaldisega** määratud sõnamall, **algvormiga** määratud sõnamall, **arvsõnafraasi** mall ning **sõnaklass**. Regulaaravaldisega määratud sõnamalle kasutatakse eelkõige numbrimustrite või vähese arvu erinevate sõnavariantide kirjeldamiseks. Algvormiga määratud sõnamallid on mõeldud kasutamiseks juhtudel, kui tahetakse tabada sõna kõiki võimalikke vorme. Arvsõnafraasi malli abil saab kirjeldada sõnadega edasi antud täis- või murdarve, seega võib malli poolt tabatav olla nii üksiksõna kui ka fraas. Kirjeldamisel antakse ette arvu liik (põhiarvsõna, järgarvsõna või murdarvsõna) ning vajadusel ka võimalike väärtuste vahemik. Sõnaklass võimaldab kasutada teisi sõnamalle alamosadena (elementidena) ning seeläbi koondada eri sõnade kirjeldused ühe malli alla. Näiteks võib koostada sõnaklassi NÄDALAPÄEV, mille elementideks on eri nädalapäevanimetusi algvormide kaudu kirjeldavad sõnamallid: {*esmaspäev, teisipäev, kolmapäev, neljapäev, reede, laupäev, pühapäev*}.

Ajaväljendite kirjeldamisel on lisaks positiivsete juhtude kirjeldamisele tarvis määratleda ka negatiivsed juhud: juhud, mil üleliigselt eraldatud ajaväljendikandidaat tuleb kustutada. Kustutamine järgneb ajaväljendikandidaatide eraldamisele ning koosneb kahest alametapist: ülekaetud kandidaatide kustutamine ning kandidaatide kustutamine negatiivsete mustrite alusel. Ülekaetud kandidaatide kustutamine viiakse läbi alati, kui üks kandidaat on täielikult üle kaetud pikema kandidaadi poolt. Näiteks kuupäevana eraldatud ajaväljendikandidaadiga *30. jaanuar* paralleelselt eraldatakse ka ainult kuunimetusest koosnev kandidaat *jaanuar* ning just viimane kustutatakse kui täielikult ülekaetud kandidaat. Negatiivsete mustrite alusel kustutamine viiakse läbi juhul, kui fraasimustriga on seotud mitte-ajaväljendilise konteksti kirjeldus (sisuliselt sõnu kirjeldavate regulaaravaldiste jada) ning eraldatud fraasi lähiümbrus rahuldab seda kirjeldust. Näiteks kasutatakse negatiivseid mustreid, et piirata ajaväljendite (kuunimetuste) eraldamist isikunimedest (*Karl August Hermann, August Mälk*).

Kui ajaväljendikandidaatide seast on välja jäetud üleliigsed kandidaadid, viiakse läbi kõrvuti paiknevate kandidaatide liitmine. See toimub kahel tasemel: **fraasi** tasemel ning ajavahemike tasemel. Fraasi tasemel liitmist kontrollivad liitumisreeglid, mis kirjeldavad, kuidas liita kõrvuti paiknevad erineva ajalise detailsusega kandidaadid (nt *järgmise nädala + esmaspäeval*) või kuidas liita ajaväljendikandidaadiga semantika täpsustus (nt *2009. aasta + lõpus*). **Ajavahemike** tasemel liitmine toetub kahe eelmise taseme (kandidaatide eraldamine ja fraasi tasemel liitmine) tulemustele ning toimub heuristiliste eeskirjade alusel. Esimene eeskiri eraldab vahemiku seestütlevas ja rajavas käändes sõnade olemasolu põhjal (nt *reedest+pühapäevani*), teine eeskiri tuvastab vahemiku, kui kandidaadist paremal või vasakul esineb lühikujul vahemiku teine otspunkt (nt *aastatel 2007 +kuni 2009, 1.– + 3. juunil*).

4.3. Ajaväljendite normaliseerimine

Ajaväljendite normaliseerimise seisukohast on oluline toimumisaegade jagunemine absoluutseteks ja relatiivseteks ajaväljenditeks:

- **absoluutsete ajaväljendite** normaliseerimine on triviaalne ning seisneb kalendriväärtuste ümberkirjutamises märgenduskeele esituskujusse. Sellised väljendid sisaldavad *aasta* detailsusega absoluutset ajalist informatsiooni, nt *20. mai 2009*, *12.03.2011*;
- **relatiivsete ajaväljendite** normaliseerimine nõuab kalendriaritmeetika rakendamist: ajaväljendi normaliseeritud väärtus tuleb leida mingi teise ajapunkti suhtes. Käesolevas töös eristatakse kaht liiki relatiivseid ajaväljendeid:
 - **teksti loomise aja järgi lahenduvad väljendid** – sellisteks on vahetult deiktisised väljendid, nt *täna*, *homme*, *järgmisel nädalal*, *mullu* ning ilma täpsustava *aasta*-detailsuseta väljendid nt *24. aprillil*, *mai lõpus*;
 - **teise ajaväljendi külge ankurduvad ajaväljendid** – nt ajaväljend *aasta varem* ankurduv eelneva aastaarvu külge lauses **1996. aastal oli puudujääk 5% väiksem kui aasta varem.**

Toimumisaegade semantika normaliseerimisel võetakse vaikselt aluseks teksti loomise aeg (enamasti antud kuupäeva täpsusega) ning eesmärgiks on semantikaoperatsioonide rakendamise teel teisendada alusaega seni, kuni on jõutud ajaväljendi semantikale vastava kalendriajani.

Lihtsaim kalendriaritmeetika operatsioon on **omistamisoperatsioon**, mis sisuliselt kirjutab alusaja mingi kalendrivälja⁷ väärtuse üle uue väärtusega. Omistamisoperatsioone kasutatakse absoluutsete ajaväljendite semantika kirjeldamisel.

Liitmis- ja **lahutamisoperatsioonid** lubavad alusaja mingit kalendrivälja suurendada või vähendada etteantud arvu ajaühikute võrra, võimaldades nn suunaga ajaväljendite (nt *eelmisel nädalal* ja *järgmisel nädalal*, *viis aastat tagasi* ja *viie aasta pärast*) lahendamist. Teatud suunaga ajaväljendite puhul muutub aga liitmis- ja lahutamisoperatsioonide rakendamine problemaatiliseks. Näiteks väljend *eelseisval reedel* võib viidata nii *selle nädala* kui ka *järgmise nädala reedele*, seega ei saa anda ühest eeskirja, kas alusaja väljale *nädal* tuleks üks nädal liita või mitte.

Saksa keele ajaväljendite tuvastamist uurinud Frank Schilder ja Christopher Habel (2001) on pakkunud välja liitmis- ja lahutamisoperatsioonidele alternatiivse strateegia, mida võib tinglikult nimetada *otsimisoperatsiooniks*. Sisuliselt võimaldab otsimisoperatsioon leida etteantud suunalt (minevik või tulevik) alusajale lähima, nõutud kalendriväljade konfiguratsioonile vastava kalendriaja. Eelneva näite puhul võib otsimisoperatsiooni defineerida selliselt, et leitakse alusajale lähim reede tulevikust, seega lahenduskäik järgib märksa täpsemalt keeles väljendatavat tähendust. Käesolevas töös kasutatakse otsimisoperatsioone peamiselt selliste ajaväljendite normaliseerimisel, kus suund on edasi antud oleviku kesksõna kujulises täiendis (nt *eelneval reedel*, *eelseisval nädalavahetusel*, *tuleval kevadel*).

Suunaga relatiivsete väljendite juures on problemaatilised veel juhud, kus väljendi kasutamisel ollakse mainitud ajatsükli alguses või lõpus. Näiteks kui pühapäeval kasutatakse väljendit *viimasel nädalal*, ei ole selge, kas mõeldakse seda

⁷ Kalendriviäljade all mõeldakse kalendriaja erinevaid tahke: *aasta*, *kuu*, *nädal*, *nädalapäev*, *kuupäev*, *tund*, *minut*.

või eelmist nädalat, samas nt esmaspäeval antud väljendit kasutades mõeldakse tõenäoliselt eelmist nädalat.

Suunaga väljendite kasutus eeltoodud probleemide lõikes vajab ilmselt põhjalikumat uurimist, mis aga ei mahu käesoleva töö raamidesse.

Omaette probleemina võib käsitleda ilma täpsustava suunata relatiivsete väljendite, s.o tekstis üksikult esinevate nädalapäevade, kuude ja kuupäevade⁸ lahendamist. Selliste väljendite lahendamiseks on välja pakutud kaks strateegiat. Esimene strateegia näeb ette ajaväljendile (lause piires) lähima verbi leidmist ning suuna otsustamist verbi grammatilise aja põhjal. Sellist strateegiat on inglise keele suunata ajaväljendite lahendamisel kasutanud näiteks Inderjeet Mani ja George Wilson (2000) ning Jannik Strötgen ja Michael Gertz (2010). Alternatiivne strateegia (Baldwin 2002) näeb ette alusaja ümber kalendrivalja unikaalsete väärtuste akna loomist ning otsitava väärtuse valimist akna seest. Strateegiat illustreerib kõige paremini nädalapäevade lahendamise näide (6): kui kõnehetk on 2010-03-28 (*pühapäev*) ning otsitakse lahendust väljendile *teisipäeval*, võib moodustada seitsme unikaalse nädalapäeva akna järgmiselt (tärn märgib kõnehetke):

(6)	N	R	L	P	E	T	K
	25	26	27	28	29	30	31
				*			

Toodud näite puhul on lahenduseks akna sisse jääv teisipäev: 2010-03-30. Strateegia rakendamisel üksikute kuude või kuupäevade lahendamiseks jääb üks kuu alati aknast välja, seega tuleb kasutada mingit varustrateegiat (nt võib eeldada, et väljajääv kuu on käesoleva aasta kuu).

Kahte strateegiat on inglise keele üksikute nädalapäevade lahendamisel võrrelnud Pawel Mazur ja Robert Dale (2008), kes leidsid, et transkribeeritud tekstide korpusel (telefonivestluste, jutusaadete, uudiste jm transkribeeritud) andis Jennifer Baldwini 7 päeva aken korrektse tulemuse 94,28% juhtudel ning verbi grammatilisele aja heuristik viis korrektse lahenduseni 92,64% juhtudel. Käesoleva töö autor on neid strateegiaid võrrelnud eestikeelsete ajakirjandustekstide (täpsemalt: päevalehtede) korpusel (Orasmaa 2010) ning on leidnud, et üksikute nädalapäevade lahendamisel andis kõrgemaid tulemusi verbi grammatilise aja heuristik (90,2% korrektseid tulemusi), samas üksikute kuude ja kuupäevade lahendamisel osutus valitud korpusel parimaks Baldwini akna heuristik (vastavalt 94,7% ja 92,7% korrektseid tulemusi). Süsteemi praegune reeglikonfiguratsioon juhindub nendest tulemustest, kuigi probleem iseenesest vajab edasist uurimist.

Kui valdava osa relatiivseid ajaväljendeid võib ankurdata teksti loomise aja külge (eeldus, mis näib kehtivat vähemalt suure osa ajakirjandustekstide puhul), siis teatud võtmesõnade olemasolu ajaväljendifraasis nõuab ankurdamisstrateegia muutmist. Näiteks võtmesõnade *following* 'järgmine', *previous* 'eelmine', *same* 'sama', *that* 'too', *before* 'varem' ja *later* 'hiljem' olemasolu inglise keele ajaväljendifraasides nõuab väljendi ankurdamist tekstis eelneva toimumisaja külge. Matteo Negri ja Luca Marseglia (2004) kasutavad seda tüüpi väljenditele ankurdamisstrateegia valimisel ajalise detailsuse kitsendust: võrreldes ankurdatavaga peab ankruks valitav väljend olema suurema või sama ajalise detailsusega. Näiteks võib väljend *three days later* 'kolm päeva hiljem' ankurduda küll sama detailsusega väljendi *on*

⁸ Laiemas plaanis mõeldakse siin relatiivseid väljendeid, mille puhul ei saa ainuüksi ajaväljendifraasi põhjal otsustada, millisest suunast (minevik, olevik, tulevik) tuleks mainitud nädalapäeva, kuud või kuupäeva otsida. Siia kuuluvad nii ühest sõnast koosnevad ajaväljendid (nt *teisipäeval*, *märtsis*) kui ka suuremat detailsust sisaldavad liitväljendid (nt *teisipäeva õhtul*).

Monday 'esmaspäeval' külge, aga mitte väiksema ajalise detailsusega väljendi *on this month* 'sellel kuul' külge.

Käesolevas töös kasutatakse toimumisaja tüüpi ajaväljendite ankurdamisel samuti heuristilisi strateegiaid, mis tuginevad autori senisele (ajakirjandus)korpuste uurimise kogemusele ning mille efektiivsust eraldi mõõdetud pole. Tekstis eelneva ajaväljendi külge ankurdatakse võtmesõnu *varem*, *hiljem*, *sama* ja *too* sisaldavad ajaväljendifraasid, ajalise detailsuse kitsendusi ankurdamisel ei rakendata. Tekstis eelneva, *päeva* detailsusega ajaväljendi külge ankurdatakse kellaaja ja päevaosa detailsusega ajaväljendid. Kõikidel ankurdamise juhtudel ei vaadata sobiva ankrutsemisel kaugemale kui kolm eelnevat lauset.

5. Süsteemi testimine

5.1. Testimiseks valitud korpus

Ajaväljendite tuvastaja hindamiseks kasutati Tartu Ülikooli koondkorpuse tekste. Koondkorpuse kirjakeele osa kogumaht on umbes 250 miljonit sõna, millest 84% moodustavad ajakirjandustekstid, 5,9% riigikogu stenogrammid, 4,7% seadustekstid, 2,9% ilukirjandus ning 2,5% teadustekstid. Mitmed koondkorpuse alamosad on heterogeensed ning representatiivse testkorpuse koostamine seetõttu problemaatiline: paratamatult tuleb teha ka subjektiivseid valikuid. Heterogeensuse tõttu on vaatluse alt välja jäetud ilukirjanduse ja teadustekstide alamkorpused; seadustekstide ja ajakirjanduse korpustest on vaadeldud vaid teatud kriteeriumidele vastavaid alamkorpuseid.

Ajakirjanduskorpuses püüti eristada alamkorpuseid vastavalt sellele, millistesse rubriikidesse olid artiklid jagatud. Paraku kasutavad erinevad meediaväljaanded erinevaid rubriikidesse jagamise viise ning isegi ühe väljaande piires võib rubriigiline jaotus oluliselt muutuda aastate lõikes. Kitsendades valikut kolmele väljaandele – Postimees (aastatest 1995–2000), Eesti Päevaleht (aastatest 1995–2007) ning ajakiri Luup (aastatest 1996–2002) –, oli võimalik eristada sagedasemaid rubriike: “Eesti uudised”, “Välisuudised”, “Arvamused”, “Sport”, “Majandus” ja “Kultuur”⁹.

Eesmärgiga hinnata süsteemi toimimist ajalugu käsitlevate tekstide märgendamisel moodustati eraldi alamkorpus ajakirja Horisont ajalooteemalistest artiklitest. Seadustekstide korpusest võeti vaatluse alla Eesti seaduste alamkorpus.

Testkorpust suurendati terviktekstide kaupa, kuna ajaväljendite normaliseerimisel tuleb lähtuda laiemast kontekstist kui üks lause. Ajakirjanduskorpuses oli terviktekstiks üks artikkel, riigikogu stenogrammid loeti terviktekstiks ühe päevakorrapunkti arutamine ning seadustekstides moodustas tervikteksti üks seadus.

Juhuslikult valitud tekstidest moodustati 70 209 sõna suurune korpus, mille proportsioonid esitab tabel 1. Ajaväljendite märgendamine korpuses viidi läbi poolautomaatselt: esmalt rakendati ajaväljendite tuvastajat ning seejärel parandati käsitsi automaatsel tuvastamisel tekkinud vead. Vigu parandas üks inimene (käesoleva töö autor).

Pärast vigade parandamist oli testkorpuses 1900 tekstilise sisuga ajaväljendimärgendit; ajaväljendite jagunemist alamkorpuste vahel kirjeldab tabel 2. Kuigi

⁹ Autori poolt pandud üldistavad nimed, rubriikide täpsed nimetused erinevad väljaannete lõikes.

tuvastaja väljundis on ka ilma tekstilise sisuta märgendid (nagu eespool näidetes (3) ja (4)), on sellise märgenduse arvestamine hindamisel mõnevõrra keerukas ning siinses töös seda ei tehta.

Tabel 1. Testkorpuse alamkorpuste proportsioonid

	Alamkorpus	Sõnu¹⁰	Tekste	Osakaal kogu korpusest
Ajakirjandus	Eesti uudised	17271	31	24,6%
	Välisuudised	7724	15	11,0%
	Arvamused	7024	15	10,0%
	Sport	6981	16	9,9%
	Majandus	5205	12	7,4%
	Kultuur	5102	12	7,3%
	Ajalugu (ajakirjast Horisont)	7088	6	10,1%
	Riigikogu stenogrammid	6950	5	9,9%
	Eesti seadused	6864	3	9,8%

Tabel 2. Ajaväljendite arv testkorpuses

	Alamkorpus	Ajaväljendeid
Ajakirjandus	Eesti uudised	553
	Välisuudised	155
	Arvamused	130
	Sport	160
	Majandus	177
	Kultuur	142
	Ajalugu (ajakirjast Horisont)	229
	Riigikogu stenogrammid	109
	Eesti seadused	245

5.2. Süsteemi hindamine

Käesoleva artikli autor oli ajaväljendite tuvastajat varem arendanud ja testinud ajakirjanduskorpuste peal; ajakirjanduskorpustel väljatöötatud reegleid ja heuristikuid kasutati ka käesolevas eksperimendis. Kuna aga varem ei olnud tehtud süstemaatilist alamkorpuste eristamist, oligi üheks huvipakkuvaks küsimuseks, mil määral erinevad tulemused alamkorpuste lõikes.

Programmi töö hindamisel mõõdeti ajaväljendite eraldamise saagist ja täpsust (rangelt ja mitterangelt fraasipiire arvestades) ning atribuutide *type* ja *value* määramise täpsuseid. Mitterangel fraasipiiride arvestamisel loeti automaatne eraldamine korrektseks ka juhul, kui eraldatud fraas kattis käsitsi eraldatud ajaväljendifraasi vaid osaliselt (vähemalt ühe sümboli ulatuses). Rangel fraasipiiride arvestamisel

¹⁰ Sõnade hulka loetakse ka punktatsioon.

nõuti automaatse ning käsitsi märgendatud ajaväljendi fraasipiiride täielikku katumist. Ajaväljendite eraldamise saagis ja täpsus defineeriti järgmiselt:

$$\text{saagis} = \frac{\text{korrekselt eraldatud ajaväljendite arv}}{\text{käsitsi eraldatud ajaväljendite arv}}$$

$$\text{täpsus} = \frac{\text{korrekselt eraldatud ajaväljendite arv}}{\text{automaatselt eraldatud ajaväljendite arv}}$$

Atribuutide normaliseerimise täpsust hinnati vaid korrekselt eraldatud ajaväljendite puhul, mitterangelt fraasipiire arvestades. Normaliseerimise täpsus arvutati järgmise valemi järgi:

$$\text{täpsus} = \frac{\text{korrekselt määratud atribuudiväärtuste arv}}{\text{automaatselt määratud atribuudiväärtuste arv}}$$

Tabelis 3 on esitatud hinnangud programmi töö tulemustele ajaväljendite eraldamisel.

Tabel 3. Ajaväljendite eraldamise tulemused

Alamkorpus		Saagis	Täpsus	Saagis (range)	Täpsus (range)
Ajakirjandus	Eesti uudised	447/553 (80,8%)	447/447 (100%)	421/553 (76,1%)	421/447 (94,2%)
	Välisuudised	135/155 (87,1%)	135/139 (97,1%)	130/155 (83,9%)	130/139 (93,5%)
	Arvamused	104/130 (80%)	104/107 (97,2%)	95/130 (73,1%)	95/107 (88,8%)
	Sport	146/160 (91,2%)	146/149 (98%)	138/160 (86,2%)	138/149 (92,6%)
	Majandus	137/177 (77,4%)	137/141 (97,2%)	127/177 (71,8%)	127/141 (90,1%)
	Kultuur	112/142 (78,9%)	112/116 (96,6%)	105/142 (73,9%)	105/116 (90,5%)
Ajalugu (ajakirjast Horisont)		166/229 (72,5%)	166/168 (98,8%)	130/229 (56,8%)	130/168 (77,4%)
Riigikogu stenogrammid		99/109 (90,8%)	99/103 (96,1%)	96/109 (88,1%)	96/103 (93,2%)
Eesti seadused		212/245 (86,5%)	212/214 (99,1%)	204/245 (83,3%)	204/214 (95,3%)
Kogu korpus		1558/1900 (82%)	1558/1584 (98,4%)	1446/1900 (76,1%)	1446/1584 (91,3%)

Võib öelda, et ajaväljendite eraldamisel on koostatud reeglid on kaldu täpsuse kasuks – kogu korpusel mõõdeti täpsuseks 98,4% (rangelt fraasipiire arvestades 91,3%), samas kui saagis oli vaid 82% (rangelt 76,1%).

Üldist madalat saagist aitavad selgitada mõned reeglite koostamisel tehtud valikud. Reeglite loomisel keskenduti konkreetsetele ajaväljenditele, madala prioriteediga oli nn lühiväljendite ja umbmääraste väljendite kirjeldamine. Lühiväljenditeks loeti ühest sõnast koosnevaid väljendeid nagu aasta ja kuus (tähdendusega 'ühes kuus'), üksikuid aastaarve ja minimalistlikke kuupäevi (nt 6.3.). Selliste väljendite puhul on problemaatiline mitmetähenduslikkus, sh ka võimalus, et tegemist võib olla mitteaajaväljenditega. Umbmäärasteks loeti väljendid, mis jätavad kvantiteedi või kalendrilise informatsiooni lahtiseks, nt mõni aeg tagasi, mõneaastane, aastakümnete jooksul, aasta-paari pärast ja mitu päeva.

Kõige madalam saagis (mitterangelt 72,5%) mõõdeti alamkorpuses "Ajalugu", kus ühe põhjusena võib välja tuua täpsustusega *enne Kristust* aastaarvude eraldamata jätmise. Üksikute ajakirjanduses esinenud näidete alusel koostatud reeglid ei suutnud katta selliste väljendite variatiivsust.

Alamkorpusespetsiifiliseks ilminguks võiks lugeda ka kümnenditele viitavate lühiväljendite (nt *seitsmekümnendate*) suhteliselt sagedase esinemise alamkorpuses "Kultuur"; selliste väljendite eraldamata jätmise põhjustas antud korpuses keskmisest madalamat saagist (mitterangelt 78,9%).

Peatükis 2 seatud kriteeriumid ei võimaldanud alati tõmmata selget piiri märgendatava ning mittemärgendatava väljendi vahele. Näiteks alamkorpustes "Majandus" ja "Eesti seadused" tuli kriteeriumi (a) järgi lugeda märgendatavateks ka valdkonnaspetsiifilised väljendid *viimase nelja börsipäevaga* ja *eelarveaasta alguseks*. Samas ei kuulunud alamkorpuses "Sport" sagedasti esinevad valdkonnaspetsiifilised väljendid nagu *teise poolaja alguses* või *viimane hooaeg* selle kriteeriumi järgi märgendamisele, mis selgitab ka keskmisest kõrgemat eraldamise saagist (mitterangelt 91,2%) antud korpuses.

Üleminekul mitterangelt mõõtmiselt rangele langesid eraldamise saagis ja täpsus keskmiselt 6–7%. Ajaväljendifraasi piiride määramisel olid sagedased vead *arv ja pool* -tüüpi fraaside poolik eraldamine (nt *kolm ja [pool aastat]*¹¹, *ligemale kahe- ja [pooletunnises]*) ning umbmäärasusele viitavate sõnade väljajätmine fraasist (nt *mitu [tuhat aastat]*).

Tabelis 4 on toodud hinnangud programmi töö tulemustele ajaväljendi liigi määramisel (atribuudi *type* täitmine) ning semantika normaliseerimisel (atribuudi *value* täitmine). Ajaväljendi liigi määramine üldiselt probleeme ei tekitanud (täpsus kogu korpuses 97,2%); sagedased olid eksimused vaid alamkorpuses "Ajalugu" (täpsus 87,3%), kus liigi valesti määramine oli põhjustatud fraasipiiride mittetundmisest (nt *u. [7500 aastat] e.Kr.*).

Semantika normaliseerimisel mõõdeti keskmiseks täpsuseks 87,4%. Sagedasti põhjustas vigu suutmatus eristada väljendi kasutust üldises ja konkreetsetes tähenduses. Nt *täna* konkreetsetes tähenduses viitab käesolevale kuupäevale, üldises tähenduses aga *tänapäevale*. Sagedasti eksiti veel pooliku eraldamise ning vale ankurdamisstrateegia tõttu. Suhteliselt suur vigade arv alamkorpuses "Ajalugu" (täpsus 61,8%) oli peamiselt tingitud poolikust eraldamisest.

Kõige kõrgemad täpsused semantika normaliseerimisel mõõdeti alamkorpustes "Riigikogu stenogrammid" (93,9%), "Välisuudised" (93,3%) ning "Eesti seadused" (92,9%). Riigikogu stenogramme uurides võis täheldada suhteliselt lihtsalt lahenda-

¹¹ Kandilised sulud märgivad automaatselt eraldatud ajaväljendi piire.

tavate kuupäevaliste toimumisaegade (nt *1. juuli*) domineerimist (ligi 42% kõigist märgendatud ajaväljenditest olid normaliseeritavad kuupäevaks).

Alamkorpuses “Eesti seadused” olid eranditult kõik relatiivsed väljendid (nt *järgneva aasta 1. veebruariks*) absoluutsel ajaskaalal mitteajastatavad, seega ajakirjanduses kasutatavad normaliseerimisstrateegiad seal ei toiminud. Kuna aga korpuses olid ülekaalus absoluutsed väljendid (seaduste jõustumise kuupäevad ning aastaarvud Riigi Teataja allikaviidetes), siis eksimused lõplikku tulemust oluliselt ei mõjutanud.

Välisuudiste korpuses võis täheldada suhteliselt suurt olevikuviidete (*praegu, nüüd*) osakaalu ajaväljendite seas (ligi 20%, võrdluseks Eesti uudistes oli olevikuviidete osakaal ligi 12%); selliste väljendite lihtsa lahendatavuse võib lugeda üheks kõrgema täpsuse põhjuseks.

Table 4. Ajaväljendite liigitamise ning semantika normaliseerimise tulemused

Alamkorpused		Ajaväljendite liigitamise täpsus ¹²	Semantika normaliseerimise täpsus
Ajakirjandus	Eesti uudised	441/447 (98,7%)	405/447 (90,6%)
	Välisuudised	132/135 (97,8%)	126/135 (93,3%)
	Arvamused	103/104 (99%)	89/104 (85,6%)
	Sport	140/145 (96,6%)	125/145 (86,2%)
	Majandus	134/137 (97,8%)	124/137 (90,5%)
	Kultuur	108/110 (98,2%)	97/110 (88,2%)
Ajalugu (ajakirjast Horisont)		144/165 (87,3%)	102/165 (61,8%)
Riigikogu stenogrammid		99/99 (100%)	93/99 (93,9%)
Eesti seadused		208/211 (98,6%)	196/211 (92,9%)
Kogu korpused		1509/1553 (97,2%)	1357/1553 (87,4%)

6. Kokkuvõte

Artikkel tutvustab ajaväljendite automaatse tuvastamise ülesannet ning kirjeldab eesti keele jaoks loodud ajaväljendite tuvastajat. Ajaväljendite tuvastamise ülesanne jagati kaheks alamosaks: ajaväljendite tekstist eraldamine ning ajaväljendite semantika normaliseerimine valitud märgenduskeele raamides. Märgenduskeele võimalustest lähtuvalt kitsendati ka vaadeldavat ajaväljendite hulka ning keskenduti eelkõige kalendriliselt normaliseeritavatele väljenditele. Kasutatav märgenduskeel tugineb sündmuste ja ajaväljendite märgenduskeelele TimeML.

Kasutatud on reeglipõhiselt üles ehitatud ajaväljendite tuvastajat. Tuvastamisreeglite abil kirjeldatakse võimalikult lühikesi terviklikku semantikat kandvaid ajaväljendifraase ning liitumisreeglite abil kontrollitakse pikemate fraaside (sõnajärjest sõltumatut) moodustamist. Fraaside kirjeldamisel kasutatakse lõpliku automaadi tehnikat, üksiksõnade kirjeldamine toetub morfoloogilise analüüsi ja ühestamise tulemustele, arvsõnade tuvastamisele ning regulaaravaldistele. Aja-

¹² Teatud juhtudel esines süsteemi väljundis märgendeid, millel puudus liiki määrav atribuut *type*. Neid juhte arvestades tuleks välja tuua ka liigi määramise saagis. Täpsustava hinnangu saab lisada järgmistes alamkorpustes: “Kultuur” (saagis 108/112; 96,4%), “Ajalugu” (saagis 144/166; 86,7%), “Eesti seadused” (saagis 208/212; 98,1%) ning kogu korpuses (saagis 1509/1557; 96,9%).

väljendi semantika esitatakse normaliseerimiseks vajalike operatsioonide jadana. Artiklis on käsitletud erinevaid heuristilisi meetodeid mitmese semantikaga väljendite normaliseerimiseks.

Tuvastajat arendati ajakirjandustekstidel. Programmi hindamiseks moodustati u 70 000-sõnaline testkorpus, mis koosnes Tartu Ülikooli koondkorpuse kirjakeele osast juhuslikult valitud ajakirjandustekstidest, ajalugu käsitlevatest artiklitest, riigikogu stenogrammidest ning Eesti seadustest. Programmi koondtulemuseks ajaväljendite eraldamisel oli saagis 82% ning täpsus 98,4% (fraasipiire mitterangelt arvestades), ajaväljendite liigitamisel saavutati täpsus 97,2% ning semantika normaliseerimisel täpsus 87,4%.

Programmi arendamisel on keskendutud konkreetsete ajaväljendite kirjeldamisele, ent edaspidi tuleks rohkem tähelepanu pöörata ka mitmese semantikaga lühiväljendite ning umbmääraste väljendite kirjeldamisele, kuna nende eraldamata jätmine on üheks oluliseks madala saagise põhjuseks. Semantika normaliseerimise seisukohast vajab mitmete ajaväljendite kasutus eraldiseisvat ja süstemaatilist uurimist; laiemateks probleemideks on relatiivsete ajaväljendite korrektne ankurdamine ning konkreetse ja üldise tähenduse eristamine.

Tulevikus rakendatakse ajaväljendite tuvastajat Tartu Ülikooli kirjakeele korpuse terviklikul märgendamisel ning märgendatud ajaväljenditega korpus muutub kättesaadavaks ka Keeleveebi¹³ kaudu.

Viidatud kirjandus

- Baldwin, Jennifer A. 2002. Learning temporal annotation of French news. Magistritöö. Graduate School of Arts and Sciences. Georgetown University, Washington, DC. www.jenniferbaldwin.com/jb/documents/Thesis.pdf (24.09.2011).
- Craveiro, Olga; Macedo, Joaquim; Madeira, Henrique 2009. Use of Co-occurrences for Temporal Expressions Annotation. – Jussi Karlgren, Jorma Tarhio, Heikki Hyvärö (Eds.). Proceedings of the 16th International Symposium on String Processing and Information Retrieval. Berlin, Heidelberg: Springer-Verlag, 156–164.
- EKG II = Erelt, Mati; Kasik, Reet; Metslang, Helle; Rajandi, Henno; Ross, Kristiina; Saari, Henn; Tael, Kaja; Vare, Silvi 1993. Eesti keele grammatika II. Süntaks. Tallinn: Eesti Teaduste Akadeemia Keele ja Kirjanduse Instituut.
- Ferro, Lisa; Gerber, Laurie; Mani, Inderjeet; Sundheim, Beth; Wilson, George 2005. TIDES 2005 Standard for the Annotation of Temporal Expressions. http://timex2.mitre.org/annotation_guidelines/2005_timex2_standard_v1.1.pdf (24.09.2011).
- Kolomiyets, Oleksandr; Moens, Marie-Francine 2009. Meeting TempEval-2: Shallow Approach for Temporal Tagger. – Proceedings of the Workshop on Semantic Evaluations: Recent Achievements and Future Directions (DEW 2009). Stroudsburg, USA: Association for Computational Linguistics, 52–57.
- Mani, Inderjeet; Wilson, George 2000. Robust temporal processing of news. – Proceedings of the 38th Annual Meeting on Association for Computational Linguistics (ACL 2000). Stroudsburg, USA: Association for Computational Linguistics, 69–76.
- Mazur, Paweł; Dale, Robert 2008. What's the date?: High accuracy interpretation of weekday names. – COLING 2008 Proceedings of the 22nd International Conference on Computational Linguistics, Vol. 1. Stroudsburg, USA: Association for Computational Linguistics, 553–560.

¹³ <http://www.keeleeveebi.ee/> (26.09.2011).

- Negri, Matteo; Marseglia, Luca 2004. Recognition and Normalization of Time Expressions: ITC-irst at TERN 2004. Tehniline raport. ITC-irst, Trento. <http://www.lsi.upc.edu/~nlp/meaning/documentation/3rdYear/WP3.6.pdf> (20.09.2011).
- Orasmaa, Siim 2010. Ajaväljendite tuvastamine eestikeelses tekstis. Magistritöö. Tartu Ülikool, matemaatika-informaatikateaduskond.
- Pustejovsky, James; Castaño, José M.; Ingria, Robert; Saurí, Roser; Gaizauskas, Robert J.; Setzer, Andrea; Katz, Graham; Radev, Dragomir R. 2003. TimeML: Robust Specification of Event and Temporal Expressions in Text. – Mark T. Maybury (Ed.). *New Directions in Question Answering. Papers from 2003 AAAI Spring Symposium*, Stanford University. Stanford, CA: AAAI Press, 28–34.
- Saurí, Roser; Littman, Jessica; Knippen, Bob; Gaizauskas, Robert; Setzer, Andrea; Pustejovsky, James 2006. TimeML Annotation Guidelines. http://www.timeml.org/site/publications/timeMLdocs/annguide_1.2.1.pdf (07.09.2011).
- Schilder, Frank; Habel, Christopher 2001. From temporal expressions to temporal information: Semantic tagging of news messages. – *Proceedings of the ACL-2001 Workshop on Temporal and Spatial Information Processing, ACL-2001*. Toulouse, 65–72.
- Strötgen, Jannik; Gertz, Michael 2010. HeidelTime: High quality rule-based extraction and normalization of temporal expressions. – *Proceedings of the 5th International Workshop on Semantic Evaluation (SemEval 2010)*. Stroudsburg, USA: Association for Computational Linguistics, 321–324.
- Treumuth, Margus 2008. Normalization of temporal information in Estonian. – Petr Sojka, Aleš Horak, Ivan Kopecek (Eds.). *Proceedings of the 11th international conference on Text, Speech and Dialogue*. Brno, Czech Republic, September 8-12. *Lecture Notes in Computer Science*, 5246. Springer, 211–218. http://dx.doi.org/10.1007/978-3-540-87391-4_28

Siim Orasmaa (Tartu Ülikool) on erialalt informaatik. Uurimisvaldkondadeks on infootsingu meetodid ning tekstide automaatne ajasemantiline analüüs.
siim.orasmaa@gmail.com

AUTOMATIC RECOGNITION AND NORMALIZATION OF TEMPORAL EXPRESSIONS IN ESTONIAN LANGUAGE TEXTS

Siim Orasmaa

University of Tartu

In this article, I give an overview of the task of *temporal expression recognition* and normalization in natural language texts and describe a rule-based system, which was designed to solve this task in Estonian language texts.

A TimeML-based markup language was used to annotate temporal expressions. Following the markup possibilities in TimeML, I define two rough criteria for deciding whether a temporal expression is markable or not: 1) a markable expression should contain temporal information of *year, month, week, day, hour* or *minute* granularity, or 2) a markable expression should contain a reference to the past, present or future, anchored to the time of creation of a document or to some other markable expression.

I use a finite state technique for *temporal expression recognition* and define recognition rules as phrase patterns consisting of templates for words (describing a word by lemma or by regular expression) and descriptions of numeral phrases. Recognition rules were mostly used to extract the smallest phrases that carried separate temporal meanings and had unchangeable word order. Another set of rules was combined with built-in heuristics to join extracted phrases into longer temporal expressions.

The semantics of an expression were defined as a sequence of instructions, which needed to be executed in order to normalize the expression. Three types of instructions were distinguished: 1) calendar arithmetic instructions, 2) anchoring instructions, and 3) markup changing instructions. I also discuss special heuristics, which can be used to interpret semantically ambiguous expressions.

The system was tested on a 70,000-word subcorpora of the Estonian Reference Corpus. The test corpus consists of newspaper texts from six subgenres, legal texts, parliamentary transcripts and articles describing historical events or periods. I measured a recall of 82% and a precision of 98.4% as overall performance on recognition of temporal expressions. Normalization accuracy of the attribute “type” was 97.2% and for the attribute “value” was 87.4%.

Keywords: computational linguistics, semantic markup, annotation, temporal expressions, Estonian

LEXICON-BASED DETECTION OF EMOTION IN DIFFERENT TYPES OF TEXTS: PRELIMINARY REMARKS

Hille Pajupuu, Krista Kerge, Rene Altrov

Abstract. Paragraphs of four genres are analysed to detect their emotional colouring, while a lexicon-based approach of linguistic analysis is weighed against reader opinion. The aim is to find out the prospects of automatic detection of emotion in any text by using a very small lexicon of about 600 frequent emotion words.*

Keywords: analysis of emotion, paragraph, reader opinion, text-linguistic analysis of emotion, word frequency

1. Introduction

The aim of the study is to make preparations for the creation of an automatic tool enabling the provision of a preliminary idea of the emotionality of any written text (including factual); that is, whether this or that text may affect the reader as positive or negative. The theoretical background relies on the cognitivist understanding of the inseparability of linguistic and other experience (see e.g. Langacker 2000 and other articles in Barlow, Kemmer 2000 on usage-based models of language) and on views of the recent 30 years on the importance of text reception. According to the latter we define text as a unit of communication to be interpreted by the receiver – relying, of course, on the norms of genre and usage (see e.g. de Beaugrande, Dressler 1981/2002, Kecskes 2008). Therefore the study is centred on reader opinion, whether or not the reader has any expert knowledge of what the emotive effect might be due to (certain linguistic expressions, personal or general background of knowledge and cognition, a fact, argument, emotion or opinion contained in the text).

Determining the emotionality of texts with the help of a dictionary, we rely on the principles of lexicon-based detection of emotion as used in sentiment analysis (opinion mining) for various systems, which means that the text words carrying an emotional colouring are labelled accordingly using dictionaries or word lists where the words have been provided with emotion tags either automatically or manually

* We are thankful to the anonymous referees for their vital remarks and the Projects EKT1, SF0050023s09 and ETF8605.

(see Taboada et al. 2011). In different dictionaries word emotion may be described in layers; that is, using different dimensions of emotions, one of which is polarity, defining words as positive or negative. In systems focused on text polarity, texts are searched for positive and negative words, tagging them accordingly, while the final score of text emotionality is received by adding up the respective polarities (Missen et al. 2009).

Our study aims to test the prospects of determining the positive or negative colouring of any written text by using a very small lexicon containing frequent words with a clearly sensed polarity. Unlike in sentiment analysis that seeks to pinpoint the personal attitude of the text writer, our aim is to find out to what extent the text polarity established by the lexicon-based method might coincide with the reader's impression of the emotionality of the text.

One of the problems is the optimal length of the text subjected to an emotionality measure. As different units of a full text, such as sentences or paragraphs, may carry conflicting polarities, a full-text score need not be a too precise measure of emotionality (Pang, Lee 2008). Proceeding sentence by sentence, however, need not yield any better results, because in some cases even a human reader may find it difficult to decide over the polarity of this or that sentence (e.g. *I know he is not a good boy but he is not that bad too*; Missen et al. 2009).

In the present study the basic text unit subjected to emotionality measurement is an orthographic paragraph as a computer cannot be expected to identify substantial paragraphs. We assume that orthographic paragraph is an important functional and meaningful unit of text, one within which conflicting emotions seldom occur. However, interpretation of the results requires qualitative attention to be paid to the focus of the paragraph, making sure that the emotion of the final sentences coincides with that of the whole paragraph, because presumably the emotion perceived at the end of the paragraph has a vital effect on the reader's perception of its general polarity.

2. Procedure, material and method

The research procedure consists of three stages. In stage one we check if orthographic paragraph is an appropriate text unit for emotion detection. So we searched the Estonian Emotional Speech Corpus for such journalistic passages that had been included in the corpus only after being subjected to a reading test and getting a unanimous (either positive or negative) polarity score from all readers participating (see Altrov, Pajupuu 2008). In addition each sentence of the paragraphs was isolated from the context and was subjected to another reading test where the testers were asked to label the emotional colouring of the sentence choosing from *positive*, *negative*, and *neutral*. The results were analysed to learn the emotions dominating the paragraphs chosen and the possible change of emotion within the paragraphs.

The following stages involved analysis – using different methods – of paragraphs representing the following four different genres: literary diary (Ristikivi 2008); economy news; editorials of a daily; and the weekly horoscope of *Arter*, a weekly attachment to the *Postimees* daily. The preliminary genre selection is motivated by the conviction that one should start with studying most different kinds of

media texts, because although many text studies tend to take journalistic language as something homogeneous, in our case genre differences may affect the results. In addition it could be useful to take, as a parallel test case, a text sample from a different domain of language use where emotional polarity is supposedly manifested. As diaries looked promising in this respect we turned to Varrak publishers who provided research access to digitised parts of Ristikivi’s diary, proportionally representative of each period involved.

The aim of stage two was to ascertain the emotionality of the paragraphs by using the lexicon-based method; that is, to decide for each paragraph whether its general emotionality is positive, negative, or ambivalent (an equal number of positive and negative words) or neutral (no positive or negative words). For the material analysed see Table 1.

Table 1. Material

Material	Passages from the Emotional Speech Corpus		Passages of different genres			
	Positive	Negative	Literary diary	Economy news	Horoscope	Editorials
Number of passages	18	18	17	15	12	14
Number of sentences	143	140	59	53	41	71
Number of tokens	1,189	1,322	651	812	547	1,193

In the lexicon-based detection of paragraph emotion we used *The Basic Estonian Dictionary* (Kallas, Tuulik 2011) containing the 3,015 most frequently used Estonian words; 639 of those have been labelled as having an emotional meaning, falling into 317 positive and 322 negative words. In tagging the words, not only semantics (e.g. *rõõm*^{pos} ‘joy’, *kuritegu*^{neg} ‘crime’, *ilus*^{pos} ‘beautiful’) but also the cultural aspect has been considered, because a word evoking positive or negative feelings in one language need not do the same in another (e.g. in Estonian words such as *kodumaine*^{pos} ‘homemade’, *leib*^{pos} ‘bread’, *sõltumatu*^{pos} ‘independent’, *tagasihoidlik*^{pos} ‘modest’, *hapu*^{neg} ‘sour’, *suitsetamine*^{neg} ‘smoking’ have an emotional colouring; see also Balahur, Montoyo 2008; on the principles of labelling emotion words in the Estonian Base Word Dictionary, see Vainik forthcoming). Our decision to use frequent emotion words only is based on the knowledge that ~60% of the words of any Estonian written text belong to the 3,000 most frequent Estonian words (Pajupuu et al. 2010). In addition, an analysis of the text corpora used by the compilers of a frequency dictionary also show that the 3,000 most frequent words cover about 64% of texts (Kaalep, Muischnek 2002). Hence we assume that every emotional paragraph would probably contain a couple of frequent emotion words.

Accordingly, each word of a paragraph was compared with the entry words of the afore-mentioned base word dictionary and tagged as positive or negative according to the emotion marker in the dictionary. The rest of the words were regarded as neutral.

As negation changes word polarity the following rules were applied (see Li et al. 2010, Pajupuu et al. forthcoming):

- by negation a neutral word (without an emotion tag) is turned negative; for example *hindama* ‘to appreciate’ (o), *ei hinda* ‘do/does not appreciate’ (-);

- by negation the following positive word is turned negative; for example, *rõõmustama* ‘to be happy (about)’ (+), *ei rõõmusta* ‘is/are not happy’ (–);
- by negation the following negative word is turned positive; for example, *valetama* ‘to lie’ (–), *ei valeta* ‘do/does not lie’ (+).

Next the positive and the negative words of a paragraph were added up. The emotion score of the paragraph depended on whether it was dominated by positive or negative words. If, for example, there was one positive and four negative words, the paragraph’s emotion score was –3; that is, negative. If the number of positive and negative words was equal, the paragraph was classified as ambivalent. If there were no positive or negative words, the paragraph was regarded as neutral.

- (1) Lexicon-based analysis of a weekly (*Arter*) horoscope (see full translation under example (2)):

Isegi kui üritad endast parima ‘the best’ (+) anda, ei pruugi ‘need not’ (–) kõik su soovide kohaselt laabuda. Nii mõnigi asi ei sõltu ‘does not depend’ (–) ainult sinust, need, kelle käes on võim ja otsustusõigus, ei pruugi ‘need not’ (–) olla sinuga ühel meel. Teiste inimeste tegude eest ei saa ‘cannot’ (–) sina vastutust enda õlgadele võtta, ometi võib nende tegevus sulle nüüd üksjagu mõju avaldada. (Polarity –3, negative)

The aim of stage three was to test the correctness of the results of the economical, lexicon-based method described earlier and the prospects of its use in automatic analysis requiring the most optimal solutions. Here the results of lexicon-based analysis were compared with those of linguistic expert analysis, while the comparison was conducted by a native Estonian specialist of linguistic text analysis. There follows a comparison of the two methods based on reader opinion (for details see below).

The qualitative linguistic text analysis (expert analysis) was carried out in order to find out whether identification of the ‘human’ semantic orientation (see Taboada et al. 2011) by relying on all levels of language use revealed in the paragraphs in question could possibly yield results that are considerably better than the results of using a rather minimalistic frequency-based emotion dictionary. Thus a native specialist in linguistic text analysis read each paragraph twice, with a reasonable time interval, marking the units (words, expressions, morphological forms, phrases, or sentences) bearing polarity; that is, having a positive or negative colouring in that particular context, and finally checking whether there was sufficient coincidence between the two readings.¹ The factors that were considered also included the intensity of the emotional colouring, which may be carried by an intensifier (cf., e.g. halb ‘bad’ and väga halb ‘very bad’), a higher degree of comparison (cf. ilus ‘beautiful’ and ilusam ‘more beautiful’), reference to an authoritative opinion (e.g. paljude arvates ‘according to general opinion’), coordination of polarity-bearing linguistic units (e.g. tüütu ja väsitav ‘tedious and tiresome’), a certain type of a clause (e.g. isegi kui üritad endast parima anda ‘even if you are doing your best’ amplifying a negative statement), or markers of subjective modality (e.g. kindlasti ‘certainly’, framing the main clause mind ajas vihale ‘I hated’). As a result the emotion-bearing unit may score +1 (positive emotion) or even +3 (positive emotion twice amplified). The results were added up for each paragraph.

¹ The whole procedure has been described in detail by Krista Kerge, who is the expert for the present study, in her paper “Ristikivi päevaraamatu tundetoon” (‘The emotional tonality of Ristikivi’s diary’) delivered at a joint conference on text issues (see Kerge 2010), but the topic is beyond the limits and relevance of the present article. For one tester, the coincidence of two tests separated by a time interval should be at least 80%. In our case the double procedure has yielded sufficient coincidence. See for example resources related to content analysis and text analysis: <http://www.content-analysis.de/> (accessed 24.09.2011).

For a better comparison of the paragraphs the emotion score was calculated on a different basis than with the lexicon, namely, as a rate of summed polarity to length of paragraph (the percentage of the total emotionality reading of the number of textual words in the paragraph was rounded to the nearest whole number). Such calculation – however conditional – was expected to give a better idea of the extent of the influence of the polarity-bearing words or expressions. In the observational results of the expert a score of qualitative analysis under 10 meant, conditionally, a weak polarity, 10–29 signalled of a medium one, while 30 and over meant high polarity.

For example, the Estonian version of example 2 below was a paragraph of 53 text words with a total of one positive marker and five negative ones, thus scoring –8; that is, ‘slightly negative’:

$$[(+1-5)/53] \times 100 = -7.547; \text{ emotion score } -8.$$

Note that, unlike in the lexicon-based analysis, the expert opinion is affected by two different contexts: first, the personal cognitive context of the expert, and second, that of the preceding and following text (for more on context see e.g. Kecskes 2008). In some cases the expert found the context ambivalent, resulting in alternative scores (0/–1, or neutral/slightly positive), as seen in the tables 2–5 below, the choice depending on interpretation.

The results of the two methods can be compared on example (1) (above) and example (2) (below).

(2) Linguistic text analysis of a weekly (*Arter*) horoscope:

Even if you try your best, you need not succeed in everything (–2). There are things that just do not depend on you (–1); those having power and authority need not agree with you (–1). Although you cannot take responsibility for other people’s actions (+1) they may affect you quite considerably now (–1). (Polarity –4, sentiment score –8, slightly negative)

Next the results of the lexicon-based and expert analyses are compared to reader opinion, which provides, of course, the most reliable reference, but is hardly accessible to automatic analysis.

To extract reader opinion, at least five native Estonian readers were asked to read the paragraphs independently and decide by intuition whether a paragraph was positive, negative or neutral. The final result for each paragraph was gained by using the dominant opinion (Pennebaker et al. 1997). In the case of three alternative options (positive, negative, neutral) the dominant opinion was the one expressed by the readers more times than the other two together. For example, if four out of six readers assessed a passage as positive, one as negative and one as neutral, the dominant opinion defined the passage as positive. If no opinion dominated over the sum of the rest, the paragraph was regarded as emotionally ambivalent. (For example, there was no dominating opinion if three of the six readers thought that a certain passage was positive, while two marked it as negative and one as neutral.)

Both the lexicon-based and linguistic expert analyses were validated against reader opinion as it is, after all, up to the reader to interpret the meaning of a written text. If the assessments gained by different methods diverged, the material was returned to qualitative analysis, trying to guess the reasons (possibly applicable in later research). In the case of ambivalent reader opinion it becomes irrelevant

whether or to what extent the lexicon-based or expert results coincide with it. Emotionality assessment by means of any other method can be considered correct only if its results coincide with the opinion of part of the readers of the passage (for details see part 4). In this case all that needs explaining at the interpretation of the results is the mutual coincidence or divergence of the lexicon-based and linguistic text analyses.

3. Stage one: Emotion in an orthographic paragraph

The aim of this stage was to ascertain whether an orthographic paragraph was emotionally consistent enough to qualify as our research object.

According to an analysis of the paragraphs drawn from the Emotion Corpus, most of the sentences bear the same emotion as the paragraph in general. If a corpus paragraph had been marked as negative, most of its sentences were marked as negative in a context-free test (see Figure 1). If a corpus paragraph had been marked as positive, most of its sentences were marked as positive in a context-free test as well (see Figure 2). For a summary survey of all of the paragraphs analysed see Figure 3.

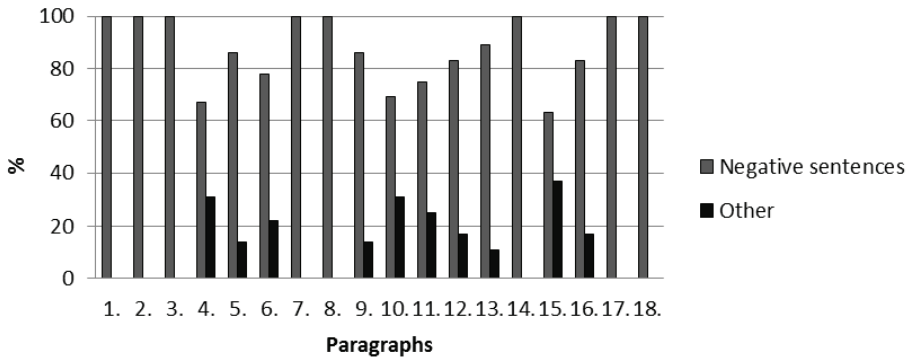


Figure 1. Negative vs. other sentences in paragraphs classified as negative

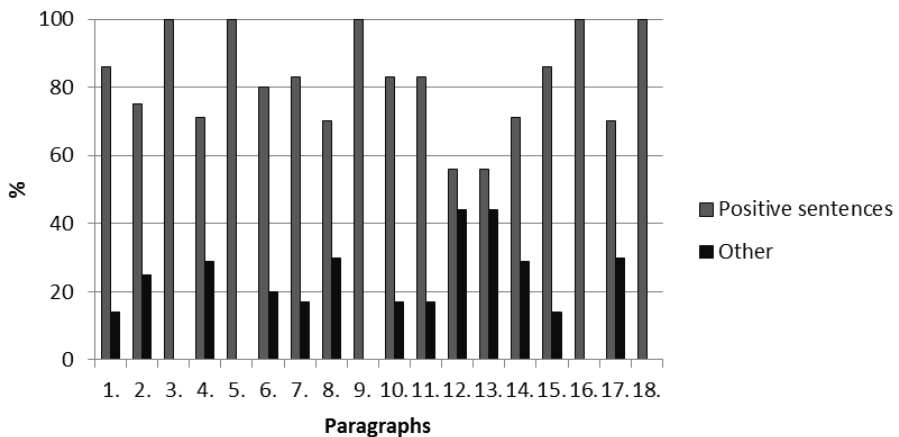


Figure 2. Positive vs. other sentences in paragraphs classified as positive

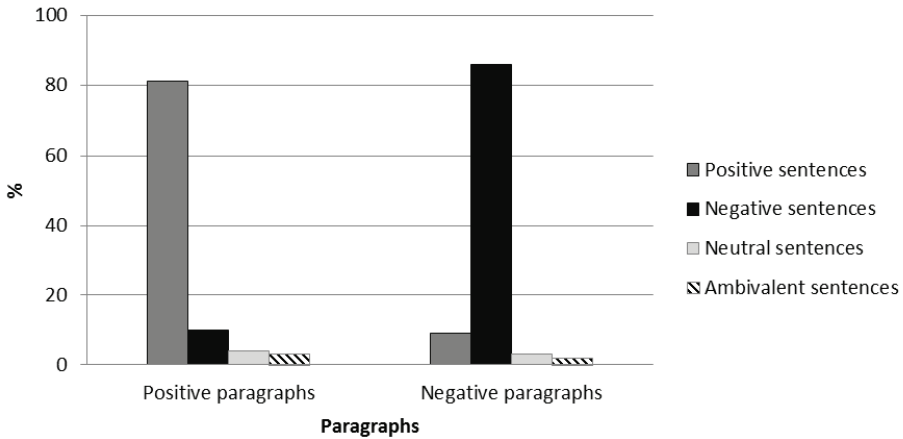


Figure 3. Assessment of sentence emotion in positive vs. negative paragraphs

The results enable the conclusion that, as far as emotional colouring is concerned, a paragraph is united enough to be treated as a unit in our analysis.

4. Stages two and three: Results of the emotion analysis of paragraphs of different genres

The results of the emotionality analysis of paragraphs of different genres are presented in Tables 2, 3, 4 and 5 as compared with reader opinion. In this analysis reader opinion is regarded as the reliable (correct) reference value (see the arguments earlier). Thus the results obtained by the lexicon-based method as well as those of linguistic text analysis were evaluated against reader opinion as correct. If the reader opinion has been marked in the table as ambivalent, it means that no reader opinion really dominates over the others, as different readers have assessed the emotionality of the paragraph differently; either: 1) some negative, some positive, some neutral, while none of the three opinions surpasses the sum of the rest; or 2) the number of negative opinions equals that of the positive; or 3) the number of negative opinions equals that of the neutral; or 4) the number of positive opinions equals that of the neutral.

In the case of ambivalent reader opinion the results obtained by either of the other two methods is regarded as correct if the ambivalence is due to polar assessments (points 1 and 2 above) because, in this case, the assessment depends on the reader, not on the lexical or some other linguistic content of the text. If the ambivalent opinion is the result of a combination of neutral assessments and those of just one pole (points 3 and 4), the results obtained by either of the other two methods is regarded as incorrect if those are opposite to the reader's opinion (i.e. if the reader's opinion is ambivalent because the text has scored an equal number of neutral and positive (resp. negative) assessments, but another method results in a 'negative' (resp. 'positive') total score, the latter result is considered incorrect as it does not match the reader opinion even partly).

The lexicon-based emotion of the literary diary (*Ristikivi päevaraamat*) coincided with reader opinion in 76.5% of the cases analysed, while the coincidence of the opinions of the text expert and non-specialist readers was 88.2%. The mutual coincidence of the lexicon-based and linguistic text analyses was 76.5% (see Table 2). The divergent opinions depend on a wider context (such as recurrence of a neutrally worded fact of everyday life in diary entries of different days (e.g. *Raadioreparaatorit täna ei tulnud* ‘The radio repairman did not come today’; *Raadioreparaatorit ei tulnud ka täna* ‘The radio repairman did not come today either’), which need not be available to a paragraph tester.

The grey background in the tables set off the results differing from reader opinion, while the white script against black is used for those reader opinions that coincide with the emotion determined by means of the two other tests.

Table 2. Emotion analysis of paragraphs from a literary diary

Paragraph	Lexicon-based approach		Reading test	Linguistic text analysis	
	Lexicon-based assessment	Score	Reader opinion	Linguist's assessment	Score
1.	ambivalent	0	positive	slightly positive	+5
2.	positive	+3	ambivalent	slightly positive	+6
3.	negative	-3	negative	negative	-17
4.	negative	-2	negative	highly negative	-32
5.	negative	-1	negative	slightly negative	-2
6.	negative	-2	negative	negative	-23
7.	positive	+1	positive	highly positive	+30
8.	negative	-4	negative	slightly negative	-3
9.	negative	-1	neutral	negative	-10
10.	negative	-1	negative	negative	-16
11.	ambivalent	0	negative	negative	-22
12.	negative	-1	negative	slightly negative	-7
13.	neutral	0	negative	slightly negative	-9
14.	negative	-3	negative	negative	-10/11
15.	negative	-2	negative	negative	-14
16.	negative	-1	negative	slightly positive / neutral	+1/0
17.	positive	+1	positive	slightly positive	+6

For economic news, the lexicon-based assessment coincided with reader opinion in 73.3% of the cases, while coincidence between the assessment of a text expert and reader opinion was 80.0%. The results of the lexicon-based and linguistic text analyses coincided in 46.7% of the cases (see Table 3). Besides one ambivalent paragraph assessment, differences appeared in the case of some paragraphs where critical analysis (see e.g. Fairclough 2010: 121–154, 164 ff.) revealed reader manipulation or irony. Although the 46.7% coincidence between the results of the last two methods

is not very high, the 73.3% coincidence between the lexicon-based detection (which is the method being tested) and reader opinion as the most important touchstone seems good enough, all the more so because a specific kind of news could easily use some words that do not belong to a basic dictionary.

Table 3. Emotion analysis of economic news

Paragraph	Lexicon-based method		Reading test	Linguistic text analysis	
	Lexicon-based assessment	Score	Reader opinion	Linguist's expert opinion	Score
1.	positive	+1	positive	positive	+14
2.	ambivalent	0	positive	slightly positive	+4/+5
3.	neutral	0	positive	positive	+13
4.	positive	+1	negative	negative	-13
5.	neutral	0	neutral	slightly negative	-8
6.	positive	+1	positive	positive	+11
7.	positive	+5	negative	neutral	0
8.	negative	-1	ambivalent	slightly positive	+4
9.	negative	-6	negative	negative	-10
10.	neutral	0	neutral	neutral / slightly positive	0/+5
11.	positive	+6	positive	slightly negative	-9
12.	positive	+8	positive	positive	+11
13.	negative	-2	negative	negative	-22
14.	positive	+1	ambivalent	slightly negative	-9
15.	positive	+2	positive	positive	+14

For the horoscope, the lexicon-based result coincided with reader opinion in 75.0% of the cases, while the expert–reader agreement was 58.3%. The results of the lexicon-based method and linguistic text analysis coincided in 50.0% of the cases (see Table 4).

Our material suggests that, in some genres, the objective analysis of an expert linguist may yield results differing from reader opinion. Horoscope recommendations, for example, are neither negative nor positive from a linguist's point of view, but they nevertheless rely on the partly negative premise that the recommended action is not typical of the reader (e.g. the instruction to *plan systematically* assumes that this is not something the reader usually does). In some cases the expert's opinion was biased by the fact that the general emotional colouring of the paragraph was neutralised by the strong opposite polarity of the focus at the end of the paragraph. Possibly, the reader of a relatively manipulative horoscope text will rather be guided just by certain keywords whose polarity can also be found easily by using the lexicon-based method.

Table 4. Emotion analysis of a weekly horoscope

Paragraph	Lexicon-based method		Reading test	Linguistic text analysis	
	Lexicon-based assessment	Score	Reader opinion	Linguistic assessment	Score
1.	positive	+4	positive	slightly negative	-2
2.	positive	+3	ambivalent	slightly negative	-5
3.	negative	-1	negative	neutral	0
4.	negative	-3	negative	slightly negative	-8
5.	negative	-2	positive	slightly negative	-4
6.	positive	+6	positive	highly positive	+35
7.	negative	-1	neutral	negative	-12
8.	ambivalent	0	positive	slightly negative	-4
9.	negative	-4	negative	negative	-10
10.	positive	+2	ambivalent	slightly negative	-2
11.	positive	+1	ambivalent	negative	-24
12.	negative	-3	ambivalent	slightly negative	-8

For an editorial, both the lexicon-based and expert assessment coincided with reader opinion in 71.4% of the cases. The mutual coincidence between the results of the lexicon-based method and linguistic text analysis was 85.7% (see Table 5).

Table 5. Emotion analysis of paragraphs of a daily's editorial

Paragraph	Lexicon-based method		Reading test	Linguistic text analysis	
	Lexicon-based assessment	Score	Reader opinion	Linguistic assessment	Score
1.	negative	-5	negative	slightly negative	-8
2.	negative	-7	negative	negative	-10
3.	positive	+4	positive	positive	+10
4.	negative	-4	negative	negative	-13
5.	negative	-6	negative	negative	-14
6.	neutral	0	negative	slightly positive	+2
7.	negative	-4	negative	slightly negative	-4
8.	neutral	0	positive	neutral / slightly negative	0/-1
9.	negative	-4	positive	neutral	0
10.	negative	-4	negative	negative	-10
11.	negative	-1	negative	slightly negative	-3
12.	positive	+1	negative	slightly positive	+2
13.	negative	-6	negative	slightly negative	-7
14.	negative	-11	negative	negative	-15

According to qualitative analysis the few divergent results for the editorial can be accounted for by irony or double negation (*ei ütle, et ei ole* ‘will not say it is not’) noticed by in-depth analysis only.

5. Discussion and conclusions

This study proves that automatic detection of the emotion of Estonian paragraphs by using a lexicon of about 600 frequent emotion words gives rather good results. For all four genres analysed the coincidence of the lexicon-based method and reader opinion was over 70%. This is comparable with other lexicon-based detectors of emotion where, depending on the lexicon and text type, the rate of correct detection can be anything between 53% and 80% (Taboada et al. 2011).

The lexicons used in emotion detection can be very different in size, with most of them containing anything between 5,000 and 70,000 words. Analyses have shown that bigger lexicons do not necessarily mean better detection rates – rather the opposite. Hitherto the best results of lexicon-based detection of emotion (78% on average) have been reported for the SO-CAL system (Semantic Orientation CALculator; see Taboada 2011: 270 ff.), which uses part-of-speech lists considering both frequency and domain and combines them into a lexicon with a total of under 5,000 entries (Brooke et al. 2009, Taboada et al. 2011). Our success with a considerably smaller lexicon can be accounted for by the fact that most of the frequent emotion words are monovalent, so that their emotional connotation is seldom changed by context (e.g. *koostöö* ‘cooperation’ – which is no. 349 in the frequency dictionary of Kaalep, Muischnek (2002) – is invariably positive, whereas the relatively rare word *vähenõudlik* ‘modest; frugal; indulgent’ can be positive or negative depending on the context).

Comparison with linguistic text analysis showed that, although in-depth analysis may be more precise, the difference is hardly big enough to discard the simple lexicon-based method and start automatising multi-level linguistic analysis. Moreover, we have observed that, for some genres, expert analysis may diverge from reader opinion even more than lexicon-based detection of emotion. Linguistic analysis of text emotion may nevertheless yield interesting results in comparative studies of genres, texts and authors. The present study indicates that differences in emotion perception are relatively big for the consciously manipulative horoscope genre (many ambivalent reader assessments and under 50% coincidence in computer vs. human assessments), whereas computer–human agreement was strongest for the diary.

The lexicon-based method should be automatised and subsequently tested on considerably more material of more genres. In addition it should be considered whether just one summary score is a sufficient parameter to characterise text emotion. According to some studies the emotion of a text generally seeming positive or negative can be reversed by the emotion of its first or last sentence, while in some other cases the summary assessments of sentence emotion diverge from that given to the whole text for other reasons (see Bestgen 1994: 11 ff., Polanyi, Zaenen 2006). The material used in our experiment turned out to contain paragraphs where the focus emotion contrasted with the emotional colouring of the rest of

the text. Therefore it might be useful to show the emotional dynamics of the whole paragraph from beginning to end.

References

- Altrov, Rene; Pajupuu, Hille 2008. The Estonian emotional speech corpus: Release 1. – František Čermak, Rūta Marcinkevičienė, Erika Rimkutė, Jolanta Zabarskaitė (Eds.). Proceedings of the Third Baltic Conference on Human Language Technologies: The Third Baltic Conference on Human Language Technologies. Vytauto Didžiojo Universitetas, Lietuvių Kalbos Institutas, Vilnius, 9–15.
- Balahur, Alexandra; Montoyo, Andres 2008. Applying a culture dependent emotion triggers database for text valence and emotion classification. – *Procesamiento del Lenguaje Natural*, Revista 40, 107–114.
- Barlow, Michael; Kemmer, Suzanne (Eds.) 2000. Usage-based Models of Language. Stanford, California: CSLI Publications, Centre for the Study of Language and Information.
- Bestgen, Yves 1994. Can emotional valence in stories be determined from words? – *Cognition & Emotion*, 7 (1), 21–36.
- Brooke, Julian; Tofiloski, Milan; Taboada, Maite 2009. Cross-linguistic sentiment analysis: From English to Spanish. – International Conference RANLP 2009, 50–54.
- de Beaugrande, Robert-Alain; Dressler, Wolfgang Ulrich 1981/2002. Introduction to Text Linguistics. London: Longman, 1981. Digitally reformatted 2002. http://www.beaugrande.com/introduction_to_text_linguistics.htm (2.01.2012).
- Fairclough, Norman 2010. Analysing Discourse. Textual Analysis for Social Research. Oxon, New York: Routledge.
- Kaalep, Heiki-Jaan; Muischnek, Kadri 2002. Eesti kirjakeele sagedussõnastik. Tartu: TÜ kirjastus.
- Kallas, Jelena; Tuulik, Maria 2011. Eesti keele põhisõnavara sõnastik: ajalooline kontekst ja koostamis põhimõtted. – Eesti Rakenduslingvistika Ühingu aastaraamat, 7, 59–75. <http://dx.doi.org/10.5128/ERYa7.04>
- Kecskes, Istvan 2008. Dueling contexts: A dynamic model of meaning. – *Journal of Pragmatics*, 40 (3), 385–406. <http://dx.doi.org/10.1016/j.pragma.2007.12.004>
- Kerge, Krista 2010. Ristikivi päevaraamatu tundetoon. – Ms. Ettekanne TÜ ja TLÜ teksti-päeval 2010, 9.12.2010. Tallinn: Tallinna Ülikool.
- Langacker, Ronald W. 2000. A dynamic usage-based model. – Michael Barlow, Suzanne Kemmer (Eds.). Usage-Based Models of Language. Stanford, California: CSLI Publications, Centre for the Study of Language and Information, 1–63.
- Li, Shoushan; Lee, Sophia Y. M.; Chen, Ying; Huang, Chu-Ren; Zhou, Guodong 2010. Sentiment classification and polarity shifting. – Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010). Beijing: Coling 2010 Organizing Committee, 635–643.
- Missen, Malik Muhammad Saad; Boughanem, Mohand; Cabanac, Guillaume 2009. Challenges for sentence level opinion detection in blogs. – Eighth IEEE/ACIS International Conference on Computer and Information Science, 347–351. <http://dx.doi.org/10.1109/ICIS.2009.190>
- Pajupuu, Hille; Kerge, Krista; Altro, Rene (forthcoming). Detecting emotional valence of text by using a small dictionary. – *Empiricism and Analytical Tools for Applied Linguistics in the 21st Century – Empirismo y herramientas analíticas para la Linguística del Siglo XXI*. Salamanca: Universidad de Salamanca.
- Pajupuu, Hille; Kerge, Krista; Meister, Lya; Asu, Eva Liina; Alp, Pilvi 2010. Natural speaking and how to assess it. – *Trames: Journal of the Humanities and Social Sciences*, 59 (2), 120–40. <http://dx.doi.org/10.3176/tr.2010.2.02>

- Pang, Bo; Lee, Lilian 2008. Opinion mining and sentiment analysis. – Foundations and Trends® in Information Retrieval 1–2, 1–135. <http://dx.doi.org/10.1561/1500000001>
- Pennebaker, James W.; Mayne, Tracy J.; Francis, Martha E. 1997. Linguistic predictors of adaptive bereavement. – Journal of Personality and Social Psychology, 72 (4), 863–871. <http://dx.doi.org/10.1037/0022-3514.72.4.863>
- Polanyi, Livia; Zaenen, Annie 2006. Contextual valence shifters. – James G. Shanahan, Yan Qu, Janyce Wiebe (Eds.). Computing Attitude and Affect in Text: Theory and Applications. The Information Retrieval Series, 20. Dordrecht: Springer, 1–10.
- Ristikivi, Karl 2008. Päevaraamat 1957–1969. Tallinn: Varrak.
- Taboada, Maite; Brooke, Julian; Tofiloski, Milan; Voll, Kimberly; Stede, Manfred 2011. Lexicon-based methods for sentiment analysis. – Computational Linguistics, 37 (2), 267–307. http://dx.doi.org/10.1162/COLI_a_00049
- Vainik, Ene 2012. Kuidas määrata eesti keele sõnavara tundetooni? – Eesti Rakenduslingvistika Ühingu aastaraamat, 8, 257–274. <http://dx.doi.org/10.5128/ERYa8.17>

Hille Pajupuu (Institute of the Estonian Language). Fields of research: speech acoustics, speech communication, national stereotypes, language testing and assessment, emotional speech and text. hille.pajupuu@eki.ee

Krista Kerge (Tallinn University). Fields of research: linguistic varieties; discourse, genre, and text analysis; applied linguistics (e.g. L1 and L2 teaching and testing, legal and administrative communication and linguistic norms, speech synthesis, etc.). krista.kerge@tlu.ee

Rene Altrov (Institute of the Estonian Language). Fields of research: intercultural communication, speech perception, emotional speech and text. rene.altrov@eki.ee

ERI TÜÜPI TEKSTIDE EMOTSIONAALSUSE TUVASTAMINE LEKSIKONIPÕHISEL MEETODIL: ESMASED KATSETUSED

Hille Pajupuu¹, Krista Kerge², Rene Altrov¹

¹Eesti Keele Insituut, ²Tallinna Ülikool

Artiklis analüüsitakse leksikonipõhisel, teksti lingvistilise analüüsi ja domineeriva lugejahinnangu meetodil neljast žanrist tekstilõikude emotsionaalsust: kirjanduslik päevaraamat (Ristikivi 2008), majandusuudised, nädalalehe Arter horoskoop, päevalehe juhtkirjad (vt tabel 1). Uurimuse eesmärk on välja selgitada, kui perspektiivikaks võiks osutada millise tahes teksti emotsionaalsuse automaatne määramine väga väikse leksikoni abil, mis sisaldab vaid u. 600 sagedast emotsioonisõna.

Esimeses etapis tehti kindlaks, kas tekstilõik sobib uurimise objektiks. Sel eesmärgil lasti eesti emotsionaalse kõne korpuse varem määratud emotsiooni kandvate (s.o positiivsete ja negatiivsete) lõikude lausetes kontekstivabalt määrata emotsiooni poolus (positiivne, negatiivne või neutraalne). Osutus, et lõik on sobiv uurimisüksus: lausete emotsioonimäärang kattub suuresti selle lõigu emotsiooniga, millesse laused kuuluvad (vt jooniseid 1–3).

Teises etapis määrati eelnimetatud žanriti tekstilõikude emotsionaalsus eesti keele põhisonavara sõnastiku abil (Kallas, Tuulik 2011), mille 3015 kõige sagedamast sõnast on 317 märgendatud positiivset ja 322 negatiivset emotsionaalset tähendust kandvaks. Tekstisõna sai pluss- või miinuspunkti kattumisel sõnastiku vastava emotsioonisõnaga, kuid nii, et eitus muutis neutraalse sõna negatiivseks, polaarse sõna aga vastandmärgiliseks. Lõigu emotsionaalsusskoori arvutamisel summeeriti pluss- ja miinusmärgiga sõnad.

Kolmandas etapis kasutati kaht meetodit. Asjatundja teostatud lingvistiline tekstianalüüs arvestas erinevalt sõnastikust kõigi tasandite keelendeid, nende intensiivsust, lingvistilist ja laiemat konteksti. Emotsionaalsusskoor arvutati taas pluss- ja miinuspunkte summeerides, kuid ühtlasi võeti arvesse lõigu pikkust, mis andis võimaluse rääkida lõigu nõrgast, keskmisest ja tugevast emotsioonist. Lugejahinnangu andis iga žanri tekstilõikudele vähemalt viis lugejat. Seejärel võrreldi nii sõnastikupõhise meetodi kui ka lingvistilise tekstianalüüsi tulemuste kattuvust lugejahinnanguga – viimane on kõige usaldusväärsem ehk õige, sest just lugeja annab tekstile tähenduse. Muudel meetoditel saadud hinnangud loeti õigeks siis, kui lugejahinnang oli osutunud vasturääkivaks (st ükski hinnang ei domineerinud teiste üle) ja muul meetodil saadud hinnang kattus osa lugejate hinnanguga. Võrdluse tulemused kajastuvad ülal mainitud žanrite järjestuses tabelites 2–5.

Selgus, et väike emotsioonisõnastik on tekstilõikude emotsiooni määramisel tõhus vahend: sõnastiku abil saadud määrang kattus lugejahinnanguga sõltuvalt žanrist 71,4–76,5 protsenti, mis pole halvem teiste uurijate tulemustest (Taboada jt 2011 järgi 53–80%). Lingvistilise asjatundjaanalüüsi erinevus sõnastikupõhisest analüüsist polnud nii märkimisväärne, et tasuks keeruka ja mitmetasandilise analüüsi automatiseerimisele mõelda.

Võtmesõnad: emotsioonianalüüs, lõik, lugejaarvamus, teksti lingvistiline analüüs, sõnasagedus

STUDY DIARIES IN TRANSLATION EDUCATION AT TALLINN UNIVERSITY

Triin Pappel, Tiia Falk

Abstract. In translator education, different introspection and retrospection methods are used. In this article, we present a retrospection method called study diaries. The goal of study diaries should be to guide translation students towards learning how to find adequate solutions to problems on their own and to make students aware of the different aspects of their action (or the process of translation, etc.), which could then help them be aware of the lifelong learning aspect, as well as of the importance of being able to reflect, analyse, and be self-critical.

Keywords: translation, translation education, self-assessment and self-analysis

1. Introduction

In recent years, a number of studies on the academic environment have suggested ways of assessing both the product and the process of translation. It is hoped that by making the process of learning transparent, we can better judge and predict the quality of the end product and, therefore, be in better control of the efficiency of training. It has been emphasized in previous research that through feedback and self-assessment learners develop their self-concept and self-confidence, which can be considered as crucial for academic progress and successful careers.

The role of a translator has changed rapidly over recent years – the image of the translator, the ways and methods of the translator’s work, have changed dramatically. As described by Daniel Gouadec:

The qualified professional translator is a vital player, both economically and technically: professional translators are highly skilled technical experts, both on account of the contents they translate and of the various ever more sophisticated IT tools and software they must be able to use. Languages are essential, but insufficient. What is needed beyond absolute linguistic proficiency is a perfect knowledge of the relevant cultural, technical, legal,

commercial backgrounds, and a full understanding of the subject matter involved. (Gouadec 2007: xiii).

Translators may be called upon to translate just about anything. Any text, message, fragment of a message or code element may need to be translated. (Gouadec 2007: 4)

The consequences of these changes are now reflected in one of the main difficulties that has arisen in translator education – preparing students for the constantly changing market. On top of all this, terminological issues in Translation Studies have often proved to be a problem, as has been discussed in various articles (e.g. Marco 2009, Snell-Hornby 2009).

According to Andrew Chesterman, “the general goal of understanding a phenomenon means that we need access to appropriate concepts, to an ability to describe and explain” (Chesterman 2007: 2). One solution in preparing the students for a demanding market is to help students understand that they need to learn how to make decisions and how to be critical as well as self-critical. To achieve this, different methods have been adopted in translation education, one of which is the process-oriented approach. Daniel Gile believes that “the process-oriented approach has a number of advantages over the traditional product-oriented method in the earlier stages of Translation training. The idea is that in the classroom, trainers should focus on the Translation process, not on the end product.” Gile (2009: 14). As Gile suggests:

More specifically, rather than giving students texts to translate or speeches to interpret, commenting on the Translation produced by saying what is “right” and what is “wrong”, suggesting appropriate solutions and counting on the accumulation of such indications to guide trainees up the learning curve, in the process-oriented approach, trainers attempt to identify problems in the process followed by the students, raise their awareness of problems and suggest good principles, methods and procedures. (Gile 2009: 14)

A research study was carried out at Tallinn Pedagogical University in 2004 “to find out how to facilitate the process through which the translators join the profession” (Sass 2004: 42). The result of this study was the conclusion that: “Translator education is most efficient for all parties in the learning process when self-assessment and feedback are seen as the content as well as the method for organizing learning activities” (Sass 2004: 43).

Following this study and other similar studies, methods of self-assessment, self-analysis and self-criticism have become an important part of the curriculum at Tallinn University, and the students are encouraged constantly to analyse and criticize their own work. One of the aims of using self-assessment methods (study diaries, project work and similar methods) is to help the students to start thinking about the process of translation/interpreting, about their own conceptions of translation by discussing their work and the problems they come across in texts they have translated, and to make them aware of the conceptions of translation, as well as of the role of the translator.

2. Method

In this article, we would like to present ways of implementing study diaries in translator education and hope to offer practical guidelines on what works and what is effective to anyone planning to implement study diaries as a part of translator education.

The overall purpose of the study was to find out:

- Whether using study diaries would help students to start thinking about their own conceptions of translation and to appreciate the connections between practice and theory,
- Whether using study diaries helps students map the essence of a translator's work and role in society,
- Whether working in a team helps students analyse themselves and their work as a translator better,
- Whether using study diaries is a useful learning tool and if it is an adequate means of self-assessment.

In translation studies, different introspection/retrospection methods can be used, e.g., Think Aloud Protocols (TAP) (e.g. Kussmaul, Tirkkonen-Condit 1995), and other retrospection methods are also popular. Daniel Gile has used systematic reporting of problems and decisions encountered by translation students in their translation assignments for 25 years (Gile 2004). In 2004, Gyde Hansen carried out a project at the Copenhagen Business School, called the Copenhagen Retrospection Project. "It is a systematic comparison of several methods for introspection, the Integrated Problem and Decision Report (IPDR) by Gile (2004), Retrospection with Replay with Translog (R+Rp) and Retrospection with Replay combined with cognitive clarification via an immediate dialogue (ID) between the subject and the observer (R+Rp+ID)" (Hansen 2006: 2).

The aim of Hansen's study was to investigate "*Störquellen*, which are sources of disturbance in translation processes (SDs), i.e. the sources of translation problems that are either overlooked in translation teaching or that cannot be coped with properly because of lack of time" (Hansen 2006: 2).

In our study, we have used a retrospection method, namely study diaries. The application of this method takes place after the translation process, and although it is sometimes regarded as less reliable than, for example, TAPs (Krings 1986: 68), it is still a widely used method in translation education. The goal of the study diaries is to guide the students towards learning to find adequate solutions to problems on their own, and to make students aware of the different aspects of their action.

3. Description of the translation project

Second-year Translation MA students were observed in the translation classroom, while they were working on a project with an authentic text. The students were translating a special report of the Estonian Court of Auditors, and it was later assessed and discussed by Estonian translators working at the Court of Auditors (i.e., by a so-called "real client").

The students had to work as a team, and they had to solve the problems they encountered during the project work together as a group. There were 21 students altogether, working in three different groups of 7 students. Every student had his or her own personal task, but the final translation was the result of teamwork. In every group, there was one project manager, two proofreaders who were also responsible for the coherence of the translated text, and four translators.

During the project, the students were also asked to keep a study diary. At the end of the project, after they had received feedback, they were also asked to fill out a questionnaire. In this particular course, the purpose of the study diaries was to make the students reflect on their own activities, to make them aware of the process of translating and also the main factors that influenced their work, e.g. factors influencing decisions made in the translating process (without taking any stand on the suitability of those decisions at this stage), and to find out which steps they had taken to find possible solutions or strategies. After the translation project, when the students had received feedback from the “real client” and the teacher, they were also asked to fill in a questionnaire, to find out whether this exercise helped the students to start thinking about their own conceptions of translation, whether it was useful as a learning tool, and whether it was adequate as a means of self-assessment.

4. Analysis of the diaries

The guidelines given to the students were not particularly rigid, because the primary goal was to encourage them to start reflecting upon themselves and not to focus on strict guidelines. There were 21 study diaries in total, all quite different in their approaches, but the students were asked to concentrate on three main issues:

- Factors that might help/disturb/preclude translating;
- Comments and feelings about their work;
- What they learned.

4.1. Factors that might help/disturb/preclude translating

The students brought forth very different factors that had helped, disturbed or precluded their work in translating the text. Below are the most common factors the students mentioned:

- Daily routine and work – most students were translating in the evening when they were tired;
- Sometimes they had to translate too many hours in a row;
- In some cases, translating took more time than planned;
- An important factor is the emotional balance, e.g. some days are better than others;
- Students were constantly worried about the deadline (they were scared that there was not enough time to finish the project);
- The students admitted their lack of knowledge in translation technology and in computer skills (formatting the text and the tables was very difficult, time consuming and frustrating);

- Sometimes there were problems with accessing the internet or the computer was very slow;
- Family problems, e.g. children were ill;
- Factors distracting their attention, e.g. the TV was on, children were restless or partner was interrupting;
- Students had difficulties in finding the adequate terminology or background information (they were not sure how long it takes to find an appropriate solution or where to look to find the adequate terminology, etc);
- Problems in communication with other group members (trusting other group members);
- Some students felt more responsible for the task because it was group work;
- Some students admitted a lack of background knowledge (they had problems understanding the text);
- Some felt a lack of experience;
- Some students were not motivated enough (only interested in getting a good grade);
- They had problems inside the group (some translators were cutting corners, some not keeping to the deadline, proofreaders felt they had to do too much and be responsible more than the translators, the project manager was too arrogant, etc).

4.2. Main problems and aspects students were concentrating on

The main problems that occurred during the translation process and self-analyses were the following:

- Students had problems in formatting the text and tables;
- There were many problems with finding the adequate terminology;
- There were many discussions about the semantics and the coherence in the translation;
- Students were also concentrating a lot on trusting each other and on communication problems inside the group;
- One of the main problems they had to concentrate on was the lack of time.

4.3. Comments and feelings about their work

The students made different comments and reached conclusions that describe their work:

- Students realized that the more they read their text, the less they saw their own mistakes;
- Students understood that it is useful to discuss everything with the proof-reader;
- Students now understood better the problems with coherence in the text and in the terminology;
- An important conclusion was made – the text was easy to understand, but not easy to translate;

- Time management was essential;
- Students admitted that it was useful to organize such a task in the so-called “safe environment” of the classroom;
- Students also commented when they were not happy or very happy with the final result;
- The most important conclusion that was drawn was the understanding that if it had been a “real translation job”, some students would have rejected it because it was too difficult for them.

4.4. What did the students learn?

The students also had to comment on what they learned in the process of the translation project and the self-assessment and analyses:

- Discussing the problems and finding adequate solutions together in the group is useful and essential;
- Always trust only yourself;
- The translator has to be responsible for his or her own translation and not rely on the proofreader;
- The proofreader or the project manager should not accept translations where corners have been cut;
- A project manager’s job is very difficult;
- Computer skills are essential for translators (although a couple of students denied this need – according to them a translator does not need that kind of skills, because the translator only has to translate).

5. Analysis of the questionnaire

After the translation project and after receiving feedback from the “real client”, Estonian translators working at the Court of Auditors, the students were asked to fill out a questionnaire. The questionnaire was designed as part of a larger research project and consisted of eighteen open-format questions concerning the issues – besides the described translation project and used study diaries –, of different teaching methods used in the Translation MA programme at Tallinn University, as well as issues concerning theoretical and academic background for practising translators. The first eight questions were about different teaching methods and the aim was to find out which of the listed methods best helped the students prepare themselves for the real translation market. Questions 13 to 15 concerned another retrospection method students had to use in the classroom: reviewing and commenting on their own translations and on the translations of their peers. The aim of questions 16 to 18 was to find out whether the students find an academic and theoretical background essential and useful for a practising translator.

For the purposes of this article, only three questions were pertinent (questions 10 to 12, see Table 1). There were 21 students in the project and 21 students answered the questionnaire. Because the research method we used was qualitative and not quantitative, figures in the table and the analysis below are merely illustrative, providing a better overview of the overall results.

Table 1. Illustrative results of the questionnaire (questions 10 to 12)

Answers	Question 10: Did analysing your own work (in the form of the study diary) help you start thinking about your own conceptions of translation and realize the connections between practice and theory?	Question 11: Did analysing your own work (in the form of the study diary) help map the essence of a translator's work and role in society?	Question 12: Did analysing your own translation and your peers' translation help you analyse yourself and your work as a translator better?
Yes, it did	9	6	17
It did help to some degree	4	4	2
It did help only a little	1	5	–
It did not help	6	5	1
Difficult to say	1	1	–
The question was not answered	–	–	1

As the table shows, most of the students (9 out of 21) found using the study diaries useful (Question 10), because it made them start thinking about their own conceptions of translation and also see better the links between practice and theory in translation. Four students said it was helpful to some degree, one said it did help only a little, six didn't find it helpful at all and one was not sure.

Those who found it helpful said, for example, that it is always very useful to analyse your own work. The idea of putting down their own thoughts helped them analyse every little detail better and that helped them to start thinking about what translation and translating is.

Those who found it helpful to some degree said, for example, that the process of self-analysing is constantly happening anyway, and using study diaries for that is probably a good exercise for beginners but not necessarily for an experienced translator.

Of those who didn't find it helpful at all, some offered no further explanation, but those who made a comment indicated that they didn't believe in this kind of method, because according to them it was time-consuming and distracting.

Answers to Question 11 (Did analysing your own work (in the form of the study-diary) help map the essence of a translator's work and role in the society?) were more diverse. Only six students out of 21 found it to be very useful; according to four students it was useful to some degree, and five found it only slightly useful. Five students didn't find the study diary useful at all in this context and one was not sure.

Those students who found using study diaries helpful didn't comment on their answers. Those who found it helpful to some extent or a little mostly said that it was useful, but what was most useful was the project work itself, not the analysis. Here students unfortunately failed to see the link between the actual translating and commenting on their work in the diaries.

For Question 12 (Did analysing your own translation and your peers' translation help you analyse yourself and your work as a translator better?), the most popular action was still working in a team, being able to correct and analyse their peers'

translations and proofreading. Seventeen students out of 21 found it very useful, two students found it useful to some degree, only one didn't find it useful at all and one chose not to answer this question.

According to those who found the exercise helpful: it helped them realize what their own strengths and weaknesses are; it was interesting to see the solutions of their peers; it helped to them start reading their own work with some distance; it made them realize that practice makes perfect; and they also found that there is still a lot they don't know. Those students who didn't find the exercise to be helpful commented that only real practice is important and not the analysis itself.

6. Conclusions

Altogether, we could still say that the majority of the Translation students considered using study diaries as a method for self-assessment and self-analysis to be a useful and helpful tool. Many students did admit the importance of this kind of exercise in the Translation classroom, and many students started to realize the importance of self-analysis through the study diaries. They also started to think about the conceptions of translation, the essence of a translator's work and the role of the translator in society.

Further study on use of the diaries will consider feedback from students who have already used them, from which it can be deduced that instructions for what to record perhaps need to be more explicit. There is definitely potential here for further studies, with a larger set of data and the final results of the students who have completed the two-year programme.

References

- Chesterman, Andrew 2007. On the idea of a theory. – *Across Languages and Cultures*, 8 (1), 1–16. <http://dx.doi.org/10.1556/Acr.8.2007.1.1>
- Gile, Daniel 2004. Integrated problem and decision reporting as a translator training tool. – *Journal of Specialised Translation*, 2, 2–20.
- Gile, Daniel 2009. *Basic Concepts and Models for Interpreter and Translator Training*. Revised Edition. Amsterdam, Philadelphia: John Benjamins Publishing Company.
- Gouadec, Daniel 2007. *Translation as Profession*. Amsterdam, Philadelphia: John Benjamins Publishing Company.
- Hansen, Gyde 2006. Retrospection methods in translator training and translation research. – *Journal of Specialised Translation*, 5, 2–41.
- Krings, Hans-Peter 1986. *Was in den Köpfen von Übersetzern vorgeht*. Tübingen: Narr.
- Kusmaul, Paul; Tirkkonen-Condit, Sonja 1995. Think-aloud protocol analysis in translation studies. – *TTR: traduction, terminologie, rédaction*. Association canadienne de traductologie, 8 (1), 177–199.
- Marco, Josep 2009. The terminology of translation. – Yves Gambier, Luc van Doorslaer (Eds.). *The Metalanguage of Translation*. Amsterdam, Philadelphia: John Benjamins Publishing Company, 65–81.
- Pinazo, Encarnación Postigo 2008. Self-assessment in teaching interpreting. – *TTR: traduction, terminologie, rédaction*. Association canadienne de traductologie, 21 (1), 173–209.
- Sass, Eve 2004. Enesehindamine ja tagasiside tõlkijakoolituse tulemuslikkuse tagamisel. Magistritöö. Tallinna Pedagoogikaülikool: Tallinn.

Snell-Hornby, Mary 2009. What's in a name? On metalinguistic confusion in Translation Studies. – Yves Gambier, Luc van Doorslaer (Eds.). *The Metalanguage of Translation*. Amsterdam, Philadelphia: John Benjamins Publishing Company, 123–135.

Triin Pappel (Tallinn University) teaches translation. Her research interests include: written translation, translation education, translation process.

tpappel@tlu.ee

Tiia Falk (Tallinn University) teaches conference interpreting and is a freelance conference interpreter and translator. Her research interests include: simultaneous and consecutive interpreting, interpreter/translator training methods, practical training tools.

tiiafalk@tlu.ee

TÕLKEPÄEVIKUD TÕLKIJAKOOLITUSES

Triin Pappel, Tiia Falk

Tallinna Ülikool

Tõlkepäevikuid kasutatakse õppevahendina palju. Päevikute peamine eesmärk on aidata tudengitel analüüsida seda, kuidas nad õpivad ja miks nad õpivad. Kirjaliku tõlke õpetamisel on tõlkepäevikud samuti üsna levinud õppevahendiks ja siin on nende üheks kõige olulisemaks eesmärgiks see, et tudengid õpiksid eneseanalüüsi kaudu leidma adekvaatseid lahendusi töös ette tulevatele tõlkeprobleemidele: Samuti on oluline, et tulevased tõlkijad õpiksid teadvustama oma tegevust tõlkijana ehk tõlkimise protsessi.

Tallinna Ülikoolis on tõlkepäevikuid kasutatud kirjaliku tõlke magistriõppes suuremate tõlkeprojektide juures eneseanalüüsi ja enesehindamise vahendina.

Antud uurimuse ja artikli eesmärk oli saada teada:

- kas tõlkepäevikute kasutamine aitab kaasa sellele, et tudengid hakkaksid mõtlema ja arutlema tõlke olemuse üle ning näeksid seost teooria ja praktilise tegevuse vahel;
- kas tõlkepäevikute kasutamine aitab tudengitel kaardistada tõlkija töö olemust ning tõlkija sotsiaalset rolli;
- kas projektitöö, ehk töö rühmas, aitab kaasa sellele, et tudengid õpiksid paremini analüüsima ennast ja oma tööd tõlkijana.

Tõlkepäevikute ja läbiviidud küsitluse analüüsi tulemusena võib öelda, et suurem osa kirjaliku tõlke tudengeid pidas tõlkepäevikut heaks eneseanalüüsi vahendiks, paljude meelest aitas see neil teadvustada eneseanalüüsi vajalikkust, pani nad mõtlema tõlke olemuse ja tõlkija sotsiaalse rolli üle.

Võtmesõnad: tõlkimine, tõlkijaharidus, eneseanalüüs

AJASÕNAD VARASEL PERIOODIL EESTI LAPSEKEELES

Sirli Parm

Ülevaade. Artiklis käsitletakse ajasõnade ilmumist lapsekeelde ja nende kasutust varasel perioodil ehk enne lapse kolmeaastaseks saamist. Analüüsitav materjal pärineb andmebaasi CHILDES eesti laste lindistustest ja kolmikute korpuse lindistustest. Ajakategooria omandamise algetapil on repertuaaris esialgu vaid olevikuaja väljendamise vahendid, kuid enne kolmeaastaseks saamist hakkavad lapsed rääkima ka neist sündmustest, mis minevikus aset leidsid või mis alles ees on. Siinse uurimuse eesmärgiks on vaadelda, missugused ajasõnad ilmuvad esimestena ning millises tähenduses need esinevad. Tulemused näitavad, et esimesed ajasõnad on deiktikud, mis a) osutavad olevikule (*praegu, nüüd, täna*), minevikule (*eile, ammu*) ja tulevikule (*homme, kohe, varsti*); b) iseseisvalt ei osuta ühelegi ajale, kuid viitavad korduvusele või järgnevusele, n.o hädusama tähendusega ajasõnad: *juba, veel, alles, jälle* ning c) väljendavad kirjeldatava tegevuse ajalist järjestuslikkust: *enne, pärast ja siis*.*

Võtmesõnad: esimese keele omandamine, sõnavara, semantika, ajakategooria, eesti keel

1. Sissejuhatus

Varase lapsekeele uurimisel keskendutakse enamasti sõnavara ja grammatilise struktuuri omandamise kirjeldamisele. Kuigi lapsed suudavad sõnu produtseerida ja järele korrata suhteliselt vara (juba enne üheaastaseks saamist), võtab nende täiskasvanupärane kasutus veel aega. Alates umbes kolmandast eluaastast alustavad lapsed komplekssemate väljendite loomisega (Diessel 2004) ning aste-astmelt areneb nende keelekasutus täiskasvanupäraseks. Kuid millised on sõnad, mis ilmuvad lapse keelde esimestena, ja miks just need? Need küsimused on uurijaid köitnud juba pikemat aega ning erinevatest lähtekohtadest on püütud neile vastuseid leida (nt Bloom 2000, Clark 1973, 2009, Slobin 1985, Tomasello 2003 jt).

* Uurimust on toetanud Eesti Teadusfondi grant nr 7492.

Uurijad pole siiani üksmeel, milliste printsiipide järgi omandatakse sõnavara, kuid ollakse veendunud selles, et esimesed sõnad luuakse sildistamise otstarbel ehk need on sellised, mis tähistavad lapsele olulisi asju, inimesi ja tegevusi. Eesti keele omandamise uurimustest (Argus 2004: 26) selgub, et esimesed sõnad ilmuvad enne üheaastaseks saamist, need on nimetavakujulised ja kuuluvad enamikus pragmaatilis-leksikaalsesse klassi, mis tähistavad "lapsele olulisi inimesi". Sõnavara omandamise uurimused eesti keele kohta keskenduvad peamiselt nimisõnadele (vt nt Argus 2004, Salasoo 2005), verbidele (vt nt Salo 1995, Argus 2007, Vija jt 2009, Pajusalu jt 2011) ja pronoomenitele (vt nt Vija 2005, 2007). Uurimused, mis keskendusid eesti lapse varases keeles ilmuvate ajaväljendite kasutusele ja tähendusele, esialgu puuduvad.

Siinse uurimuse eesmärk on selgitada, millised ajasõnad ilmuvad ajasüsteemi omandamise algetapil ehk enne lapse kolmeaastaseks saamist ning millises tähenduses neid kasutatakse. Käesolevas artiklis mõistetakse ajasõnade all kõike, mis kuulub leksikaalsesse ajakategooriasse – sõnu, millega tähistatakse ajasuhet või viidatakse sellele. Ajasõnade ilmumine keelde on otseselt seotud vajadusega väljendada kõnealuse sündmuse aega. Ajasõnu ei saa "õppida" mingi objekti või tegevuse ja selle kohta käiva tähistuse kokkuviimise teel. Pigem võiks ajasõnade väljendamise ja mõistmise arengut vaadelda koondatud tähelepanu idee raames (seda tutvustatakse lühidalt allpool osas 2). Senistes uurimustes on andmeid selle kohta, et kaheaastase eesti lapse kõnes esinevad nii olevikule, minevikule kui ka tulevikule osutavad ajasõnad (vt Argus 2008: 51, Vider 1995), ning et grammatilistest aegadest on kolmeaastaselt omandatud vähemalt olevik ja lihtminevik (vt Vihman, Vija 2006).

2. Ajasüsteemi omandamine

Sündmused on seotud ajaga, mil need toimuvad, on toimunud või alles tulevad. Neist lugusid jutustades peab lapsel olema omandatud ajasüsteem. On väidetud (Weist 2002b: 198), et ajasüsteemist arusaamine ja leksikaalse ajakategooria omandamine pole neljandaks eluaastaks veel kaugeltki täielik. Kuid keeltevaheline uurimus on näidanud, et täiskasvanupärane ajasüsteem, kus verbi ajavormide ja ajaväljendite tähendused on enamasti vastavuses, kujuneb välja enne viieaastaseks saamist (vt Weist jt 1991: 69). Michael Tomasello (2003: 223) järgi valmistab lastele raskusi selliste ajaväljendite mõistmine ja omandamine, mille tähendus viitab mingile muule ajale kui käesolev hetk, ning komplitseeritud on paljude ajaväljendite kasutamine seoses verbidega, mis väljendavad tuleviku- või minevikuaega. Seda saab seletada üldise kognitiivse arenguga ajast arusaamisel, mille järgi paigutatakse siin ja praegu toimuv keskmesse ning sellest lähtub ajas osutamine edasi- või tagasisuunal (Tomasello 2003: 223). Sellest tulenevalt võib oletada, et lapse kõnesse ilmuvad minevikule ja tulevikule osutavad ajasõnad hiljem kui olevikule osutavad ajasõnad. Ajadeiksise (aja mõistestamine olevikukesksest nullpunktist lähtudes) arengut eesti lapsekeeles pole varem uuritud, kuid vihje selle kohta võib leida muutemorfoloogia omandamise käsitlusest. Nähtub, et kui võrrelda üldise deiksise (ruumi-, isiku- ja ajadeiksise) arengut ja muutemorfoloogia omandamist, siis varem areneb välja deiksise väljendamine leksikaalsete vahenditega. (Argus 2008: 34)

Olevikukeskset ajasüsteemi arengut võib seletada ka ajadeiksise olemuse ja koondatud tähelepanu seostega keele omandamisel. Psühholingvistilisest vaatenurgast lähtudes omandatakse keele sõnavara ja grammatika koondatud tähelepanu raames, mis hõlmab imiteerimist (eriti uute sõnade ja vormide puhul), suhtluseesmärkide tajumist ning kõige aluseks olevat ühiste vaatlusaluste objektide või tegevuste koos märkamist ja jagamist (Tomasello 2003: 22–25). Suhtluseesmärkide elluviimiseks rakendavad lapsed lalisemisperioodil ja isegi varem žestilist viitamist. See tähendab, et tähelepanu koondub viidatavale üksusele, laps ja täiskasvanu võtavad hoiaku viidatava suhtes ning keelelise arengu seisukohalt on oluline, millise väljendiga täiskasvanu sellele reageerib (nt *see, näe see seal* jmt). Viitamine on alati kõnelejakeskne (vt Levinson 1983), ajalise viitamise puhul viidatakse kõnehetkele ning kujuteldaval ajateljel sellest edasi või tagasi suunal. Deiksise funktsiooni keeles võib kirjeldada samuti kui tegevust, millega tõmmatakse tähelepanu kõnealusele üksusele (Elliot 1981: 141). Koondatud tähelepanu printsiibile tuginedes võib oletada, et ajalistest sündmustest rääkimisel asetab täiskasvanu end lapse poolt mõistetavasse deiktillise nullpunkti.

4–7-aastaste eesti laste ajakategooria omandamise uuringust selgub, et opositsioon mineviku ja mittemineviku vahel areneb välja alles kuueaastaselt – olevikuvormid on mõistetavad küll juba vähemalt nelja-aastaselt, kuid mineviku- ja tulevikuvormide mõistmine ning kasutamine arenevad välja alles viie-kuueaastaselt. Eksperimentaalse uurimuse tulemustest ilmneb veel, et ajasõnad hõlbustavad 4–7-aastastel lastel verbi ajavormide õigesti tajumist ja kasutamist. (Argus, Parm 2010: 38) Sellest saab järeldada, et enne neljandat eluaastat omandatav ajasüsteem hõlmab nii leksikaalse (ehk ajasõnade) kui ka grammatilise aja arengu. Varase sõnavara uurimustes (nt Bloom 2000) räägitakse tihti sõnavara hüppelisest kasvust, mis leiab aset lapse vanuses umbes 1;5–1;8. Sellest tulenevalt saab eeldada, et enne lapse kolmeaastaseks saamist on lapse kõnes muu hulgas olemas ka piisav hulk ajasõnu, et osutada olevikule, minevikule ja tulevikule.

Richard Weisti (1986, 2002a, 2002b) järgi saab varase aja omandamise jaotada tinglikult kolmeks etapiks, mis pole oma periodiseeringult selgepiirilised, vaid pigem üksteisesse sulanduvad. Reichenbachi (1947) ajamudelile ja mitme keele uurimisandmetele tuginedes on Weist (2002a: 84) esitanud ajasüsteemi arengu etappide kirjelduse, mille kohaselt saab ajasüsteemi omandamine alguse hetkest, kui laps hakkab väljendama **kõnehetke aega** (KH). Sellele järgnevat etappi nimetatakse **sündmuse ajaks** (SA), mil lapsed hakkavad kirjeldatava sündmuse ajale viitama suhteliselt (u kaheaastaselt): nad hakkavad rääkima olnust või tulevast (minevik/tulevik) ja vastandavad kõnehetke nendele. **Relatiivse aja** (RA) etapi algus on piiritletud umbes lapse kolmeaastaseks saamisega, kui nad võimelised sündmuste toimumise aega mõistma kesksest deiktillisest nullpunktist eemal, st kasutatakse ajasõnu, mis viitavad sellest minevikule või tulevikule (Weist 2002a: 84).¹ Alates neljandast eluaastast läheneb ajasüsteemi omandamine täiskasvanupärasele ajast arusaamisele, kuid see pole selleks ajaks veel täielik (Weist 2002b: 196). Weisti väide ebataieliku ajasüsteemi kohta põhineb William J. Friedmani (1982) uurimisel, mille järgi kujuneb ajasüsteemi mõistmine kahe erineva, kuid omavahel selgelt suhestuva kultuurispetsiifilise dimensiooni kaudu: ühikulise aja (nt päevad/aastaajad) ja kestuse aja (nt tunnid/kuud) kaudu. Nende täielik mõistmine

¹ Vt ajakategooria omandamise perioodide kohta ka Weist 1986: 371.

tuleb lastel umbes kümnendaks eluaastaks – päevaseid sündmusi suudavad lapsed järjestikulisust tähistavaid sõnu kasutades kirjeldada alates neljandast eluaastast, iga-aastaseid sündmusi umbes kuueaastaselt (Friedman 1982: 179). Seda väidet toetavad mitmed teisedki mõistmiskatsetel põhinevad uurimused (vt nt Winskel 2004: 333–336), mis ütlevad, et lapsed interpreteerivad ettevaatlikult just selliseid lauseid, millest arusaamiseks on oluline konventsionaalsete ajasõnade (nt *sügisel, kevadel*) mõistmine.

Lapse esimesed sõnad ilmuvad juba lalisemisperioodi lõpus ja ei pruugi olla ühetähenduslikud ning nende puhul tuleb arvestada ka seda, et tähendused, mis võivad tähistada tegijat, tegevuse objekti, tegevust, olukorda, kohta jms, on põimunud paralingvistiliste ja mitteverbaalsete teguritega (Leiwo 1993: 86). Semantilised piirid objektide, nende omaduste ja tegevuste vahel on lapse varaste sõnade puhul ähmased (vt nt Bloom 2000).

Eesti lapsekeele sõnavara uurimisel on Kadri Vider (1995)² andnud leksikostatistilise ülevaate eesti keelt emakeelena kõnelevate 2–3-aastaste laste sõnavarast. Sõnaliikide sagedussõnastiku andmetel on enim kasutatavad ajasõnad *juba, alles, enne, pärast, homme, nüüd, praegu*. Tuginedes verbimorfoloogia arengu andmetele (Argus 2008) ning eelnimetatud leksikostatistilisele uurimusele, võib eeldada, et grammatilise aja ja leksikaalse aja väljendamise vahendid ilmuvad lapsekeelde samaaegselt. Teiste keelte uurimused on näidanud, et ajaline viitamine lapsekeeles algab kõnehetke ja sündmuse aja deiktiliste suhete väljendamisest (vt Weist 2002a, 2002b). “Ajadeiktikud viitavad ajalistele referentidele lähtuvalt kõnelemishetkest, nt *täna, homme, eile, nüüd, praegu*. Nendega viidatud ajahetke saame tuvastada ainult siis, kui teame, millal nad on öeldud.” (Pajusalu 2009: 189)

3. Andmestik

Uurimuse materjaliks on andmebaasi CHILDES eesti lapsekeelekorpus³ olevad nelja lapse ja kolmikute korpus⁴ olevate laste ning nende hoidjate vahelised spontaanse kõne lindistused. Analüüsitud keelematerjalist (kokku 90 tundi lindistusi) moodustab suurema osa Andriase alamkorpus⁵, mis sisaldab 74 tundi materjali (lapse vanuses 1;7.24–3;1.13). Väiksema osa moodustavad Mariliisi, Carlose ja Hendriku alamkorpus⁶, mis sisaldavad kokku 11 tundi materjali lastelt vanuses 1;7–2;9. Kolmikute korpus⁷ olev materjal hõlmab kolme lapse (vanuses 2;6–3;0) ja hoidjate vahelist dialoogi, mida on lindistatud 15 tunni ulatuses. Ülevaate materjalist annavad tabelid 1 ja 2. Erinevus materjali hulgas on tingitud lindistuste pikkusest: kui Andriase puhul on tegemist regulaarse pikiuuringuga, siis teiste puhul mõnetunniste lindistustega.

Keelematerjali automaatseks analüüsimiseks on kasutatud programmi CLAN⁷, mis võimaldab transkriptsioonis CHAT kirja pandud andmete hulgast selekteerida uurimisel sõnad koos kontekstiga. Seega on võimalik vaadelda ka sisendkeelt, mis ümbritseb sõna sisaldavat lausungit. Sisendkeele infot olen analüüsimisel arvesse võtnud järgmistel juhtudel: a) kui on olnud tarvis kindlaks teha, kas lapse

² See uurimus põhineb kümne eesti lapse keelematerjalil, mis hõlmab ühtekokku 14339 sõnet.

³ <http://childes.psy.cmu.edu/data/Other/Estonian> (20.02.2012).

⁴ Kolmikute korpus⁷ on koostanud Reili Argus, materjal pole veel lastekeelekorpusesse sisestatud, kuid on vormistatud vastavalt CHILDES-i transkribeerimis- ja kodeerimisreeglitele.

⁵ Andriase alamkorpus⁵ on koostanud Maigi Vija.

⁶ Mariliisi ja Taimo alamkorpus⁶ on koostanud Kaja Kohler, Hendriku alamkorpus⁶ on koostanud Reili Argus.

⁷ <http://childes.psy.cmu.edu/clang/> (20.02.2012).

öeldud sõna esineb vahetu imitatsioonina või mitte; b) kui lapsekeele ajasõna tähendus täpsustub sisendkeele konteksti kaudu. Sisendkeele ajasõnade esinemis-sagedust (vt tabelid 1 ja 2) toetavad ka suulise kõne korpuse sagedustabeli andmed. Sagedustabeli koostajad väidavad, et suuline argivestlus on igasuguse kultuuri suht-luse tuum ja esindab seega prototüüpset suulist suhtlust, mida kasutatakse kõige enam ning mis omandatakse kõige varem. Näiteks suulise argikõne sagedustabeli andmetel on täiskasvanute keeles sagedasemad ajasõnad *nüüd, juba, täna, kohe, jälle, pärast, praegu*. (Hennoste jt 2000: 264–283)

Lapsekeele esimesed sõnad on tinglikult määratavad (vt nt Leiwo 1993: 84), seepärast ei saa ka siinse analüüsi puhul tähelepanuta jätta mõningaid kitsaskohti varase lapsekeele ajasõnade identifitseerimisel ja analüüsimisel. Polüsemsete sõnade puhul olen analüüsimiseks valinud vaid aja tähenduses esinevad juhud (nt *veel* esineb lapsekeeles sagedasti pigem hulga tähenduses). Veel on kõne omanda-mise algetapile iseloomulik häälikulise kuju varieerumine (Argus 2005: 24), mis-tõttu võib ühe lapse spontaanse kõne materjalist leida sõnale mitu hääldusvarianti (nt *praegu: praegu, päegu, peegu, paegu, prägust; jälle: lälle; homme: omm* jm). Alati jääb võimalus, et kõiki hääldusvariante pole olnud võimalik sõnaks määrata ning sellest tulenevalt ei pretendeeri siinne artikkel niivõrd kvantitatiivsete kuivõrd kvalitatiivsete tulemuste näitamisele.

Üksikutel juhtudel esines analüüsitavas materjalis ka konventsionaalseid ajaväljendeid (nt *sügisel, kevadel*) ja mõõdunimisõnu (nt *päev, aasta, minut*), mis toetuvad konventsionaalsele ajaüksuste järgnevusele, ning ajaliskestust väl-jendavaid sõnu (*kaua, kuni*). Need sõnad on aga siinsest analüüsist välja jäetud, kuna neid esines lapse poolt iseseisvalt öelduna vaid mõnel korral, samuti annavad varasemad uurimused (nt Weist 2002b: 198) alust oletada, et konventsionaalse aja mõistmine kujuneb välja alles 6–10-aastaselt.⁸

4. Esimesed ajasõnad – deiktikud

Deiktilised ajasõnad ilmuvad lapse kõnemes suhteliselt vara – ajasüsteemi oman-damise algetapil räägivad lapsed sündmustest, mis kõnehetkel toimuvad (Weist 2002b: 265). Ann Veismann (Pajusalu jt 2004: 93–94) jagab eesti keele deiktilised ajaväljendid kahte põhitüüpi: piiritlevat aega märkivateks sõnadeks ja väljendi-teks, mis hõlmavad mingite konventsioonidega kehtestatud kindla ajavahemiku (*eile, eelmisel nädalal*), ning ähmaste piiridega ajale viitavad sõnad, mis osutavad suhteliselt ebamäärasele ajahetkele minevikus, olevikus või tulevikus (*äsja, varsti, kohe*); viimased jagunevad omakorda ajas kaugemale ja ajas lähemale osutavateks.

Järgnevas analüüsis leiavad käsitlemist ajasõnad, mis esinevad kuni lapse kolmeaastaseks saamiseni (täpsemalt vanuseni 3;1). Esmalt käsitletakse olevikule (*praegu, nüüd, täna*), seejärel minevikule (*eile, ammu*) ja tulevikule (*varsti, kohe, homme*) osutavaid ajasõnu, mis vaadeldud materjalis esinesid. Eraldi rühmana on välja toodud samuti deiktikutena käituvad ajasõnad *juba, veel, alles ja jälle*. Need on aja mõistestamisega seotud sõnad, mille interpreteerimiseks vajatakse ajalisk mõõdet, ometi ei osuta nad iseseisvalt (nt ilma verbita) ei olevikule, minevikule ega tulevikule. Deiktiliste ajasõnade hulgas omakorda erinevad oma kasutus-

⁸ Võib eeldada, et alla kolmeaastasele lapsele suunatud kõnes on suhteliselt vähe selliseid ajaväljendeid, mis nõuavad mõistmiseks konventsionaalse aja tundmist. Ometigi oleks siinse uurimuse jätkuna huvitav järele uurida, kas konventsionaalsete ajaväljendite hulk suureneb hoidjakeeles lapse vanuse kasvades.

funktsioonilt ajamäärsõnad *siis*, *enne* ja *pärast*. Tegemist on ajalisele järjestusele viitavate sõnadega, millest vähemalt *siis* kannab jutustatavas tekstis ka sidususe funktsiooni. Neist sõnadest tuleb juttu analüüsiosa lõpus.

Tabel 1. Ajasõnad CHILDES-i Andreease alamkorpuses
(lindistusi kokku 74 tundi; lühendid: L = Andreas; E = Andreease ema)

Ajasõna		1;11–2;0		2;0–2;5		2;5–3;1		Kokku	
		L	E	L	E	L	E	L	E
Olevik	<i>praegu</i>	2	13	124	98	159	135	285	246
	<i>nüüd</i>	–	37	254	299	788	372	1042	708
	<i>täna</i>	–	19	29	81	49	67	78	167
Minevik	<i>eile</i>	–	4	5	22	7	14	12	40
	<i>ammu</i>	–	–	–	2	6	4	6	6
Tulevik	<i>kohe</i>	–	–	15	53	42	64	57	117
	<i>varsti</i>	–	1	28	16	30	12	58	29
	<i>homme</i>	–	1	16	24	9	14	25	39
Vihje	<i>juba</i>	1	5	7	85	67	68	75	158
	<i>jälle</i>	7	11	37	48	107	49	151	108
	<i>veel</i>	–	77	20	122	34	108	54	307
	<i>alles</i>	–	–	3	13	39	15	42	28
Järjestus	<i>enne</i>	–	–	7	24	35	30	42	54
	<i>pärast</i>	–	6	21	39	11	46	32	91
	<i>siis</i>	–	67	353	565	1210	1018	1563	1650

Tabel 2. Ajasõnade esinemine CHILDES-i alamkorpustes ja kolmikute korpuses
(lindistusi kokku 26 tundi; lühendid: L = laps, E = ema, R = kolmikute kõne lindistaja)

Ajasõna		Carlos		Hendrik		Mariliis		Mihkel		Liisa		Mariann	
		1;8–2;6		2;5–2;9		2;6–2;9		2;6–3;0		2;6–3;0		2;6–3;0	
		L	E	L	E	L	R	L	R	L	R	L	R
Olevik	<i>praegu</i>	–	1	3	8	–	8	1	23	3	–	–	–
Minevik	<i>eile</i>	–	–	–	2	1	8	–	–	–	–	–	–
	<i>ammu</i>	–	–	6	5	–	–	–	–	–	–	–	–
Tulevik	<i>homme</i>	–	–	6	7	1	1	–	–	1	–	1	–
	<i>varsti</i>	3	3	–	–	–	3	–	14	–	–	–	–
	<i>kohe</i>	5	2	–	5	5	4	2	10	1	–	2	–
Vihje	<i>juba</i>	–	2	1	11	1	5	8	11	2	–	–	–
	<i>jälle</i>	3	8	–	4	–	7	–	–	–	–	–	–
	<i>veel</i>	–	21	3	23	1	44	3	45	2	–	4	–
Järjestus	<i>enne</i>	–	2	1	3	–	–	–	7	–	–	–	–
	<i>siis</i>	1	16	1	69	–	77	14	135	24	–	20	–

4.1. Olevikule osutavad ajasõnad

Esimeste ajasõnade hulgas on eesti lastel kõnehetke aega märkivad ehk olevikule osutavad **praegu**, **nüüd** ja **täna**. Seega saab esimeste ajasõnade tuleku paigutada lapse kaheaastaseks saamise piirile ning öelda, et leksikaalne aeg hakkab arenema esmalt olevikukesksena.

Ajasõnad *praegu* ja *nüüd* on sünonüümsed, tähistades kõnelemise hetke hõlmavat pikemat või lühemat ajalõiku. Mõlemad on lapsekeeles sagedasti esinevad sõnad, kuid neid kasutatakse siiski mõnevõrra erinevalt. Täiskasvanupärase kasutuse kohta on Veismann välja toonud sõnade *nüüd* ja *praegu* semantilise erinevuse: *praegu* ajaviide väidetakse olevat üldisem ja neutraalsem, mistõttu lauses võib sõna *nüüd* enamasti asendada sõnaga *praegu*, kuid vastupidi paraku on ebasobivam. Sõna *nüüd* tähendus on oma osutuselt kitsam ja kontrasteeriv, st leidub veel mingi ajahetk, millega sõna *nüüd* tähendus end suhestab. (Pajusalu jt 2004: 95) Siinse uurimuse materjalist nähtub, et olevikule osutavate sõnade *nüüd* ja *praegu* kasutus sageneb vanuse kasvades järsult, ehk kui enne kaheaastaseks saamist on lapse kõnes vaid mõni nimetatud oleviku ajasõna, siis juba kolmeaastaselt on need lapse sõnavaras ühed sagedasemad ajasõnad üldse (vt tabel 1).

Valdavalt esineb ajasõna **praegu** samas tähenduses, nagu täiskasvanudki seda kasutavad, ehk tähistamaks kõnelemise hetke hõlmavat lühemat või pikemat ajavahemikku. Sõna *praegu* kasutatakse juba ajasüsteemi omandamise algetapil, kui ilmuvad esimesed verbi olevikuvormidki. Märkimisväärne on *praegu* sage esinemine eitavas lauses (1), selline kasutus ilmub juba kaheaastaselt ehk kõnehetke väljendamise perioodil ning on valdav vähemalt lapse kolmeaastaseks saamiseni (näiteks Andreasel esineb eituse mõjualas *praegu* 82 korral 285-st). Eitavas lauses esinev *praegu* väljendab nimetatud tegevuse ja/või sündmuse mittetoimumist kõnehetkel ning mõnel juhul aitab implitseerida selle võimalikku toimumist kõnehetke eel või järel (2). Teiste laste lindistused on lühemad, kuid neiski leidub andmeid, et *praegu* esineb sagedasti eituse mõjualas. Hendriku alamkorpuse ja kolmikute korpuse andmetes esineb *praegu* kõigil juhtudel eitavas lauses.

(1) Andreas 2;1.0

*CHI: **päägu** [: praegu] ei laula.

(2) Andreas 3;1.04

*CHI: mina alles hakkam tulema, ma **praegu** ei tule.

(3) Andreas 2;1.0

*CHI: **päägu** [: praegu] lindistas see äla [: ära].

Võib oletada, et eitavas kontekstis implitseerib *praegu* mingi muu aja, millal eitatud sündmus aset võiks leida (nt lausest *praegu ei lähe poodi* võib järeldada selle tegevuse hilisema võimaliku toimumise). Andrase alamkorpuse sisendkeele andmetest nähtub, et *praegu* esineb eitavas lauses 109 korral 246-st ehk lapsele suunatud kõnes umbes pooltel juhtudel. See omakorda osutab, et *praegu* kasutus lapsel on täiskasvanupärane juba varases eas. Veel leiab ühetähenduslik *praegu* varases lapsekeeles kasutamist perfektiivsuse väljendamise osana (3). Sellistes konstruktsioonides esineb verb lihtmineviku vormis ning *praegu* edastab tähenduse, et mingi tegevuse lõppemine on kehtiv kõnehetke ajal.

Olevikule üldisemalt ning vahetu kõnehetke ajale viitav ajasõna **nüüd** ilmub kõnehetke väljendamise perioodil lapse kaheaastaseks saamise piiril (4) ning esiti näikse seda sõna kasutatavat ilma verbita. Võib oletada, et ajasõna **nüüd** tähendus on n-ö laiahaardeline, sellest järeldub kõnehetke ajal toimuv ning igal tingimusel täpsustavat verbi ei lisatagi (vrd **nüüd teletupsud** = **nüüd tulevad teletupsud**). Siiski lisandub vanuse kasvades nende väljendite hulk, kus esinevad koos nii ajasõna **nüüd** kui ka mingit situatsiooni või tegevust tähistav verb (5), ning vähenevad ilma verbita juhud.

(4) Andreas 2;0.03
*CHI: **nüüd** teletupsud.

(5) Andreas 3;0.27
*CHI: aga sealt **nüüd** läksid mutid ära, sealt läksid putukad ära.

Tuginedes ajasõnade *praegu* ja *nüüd* eelkirjeldatud semantilisele erinevusele, mis osutab, et *nüüd* kasutus täiskasvanukeeles on kitsam ja kontrasteerivam kui *praegu* kasutus, tuleb tõdeda, et ka lapsekeele *praegu* ja *nüüd* erinevad oma kasutuselt. Ajasõna *nüüd* abil väljendab laps vahetut kõnehetke aega, mis implitseerib tegevuse või situatsiooni muutuse või mingi tegevuse algushetke (6), (7). Sellest tulenevalt esineb ajasõna *nüüd* lapsekeeles järjestust hõlmavates mängusituatsioonides, kus laps näiteks asetab klotse, puhub seebimulle, asetab mänguasju ritta vms (8). Neis järjestust hõlmavates tegevustes esinev *nüüd* implitseerib samuti olukorra muutuse ja/või algushetke.

(6) Andreas 2;0.13
*CHI: **nüüd** auto päliseb [: pläriseb]

(7) Andreas 3;0.23
*CHI: mina keeran **nüüd** lahti.

(8) Andreas 2;0.10
*CHI: nii **nüüd** korki ka värviks, **nüüd** selle, **nüüd** selle, **nüüd** sellega, **nüüd** Adu [: Andreas] puhkab.

Olevikule osutav ühikuline ajasõna **täna** esineb lapsele suunatud kõnes küllaltki sageli (vt tabel 1) ning lapsel ilmubki see esialgu imitatsioonina ehk täiskasvanu poolt öeldu kordamisena, kuid KH+SA väljendamise perioodi alguses esineb ka juba iseseisvalt (9)–(11). Esialgu tähistab laps ajasõna **täna** abil vaid käesolevat hetke, mis ulatub ajas edasi (12), kuid umbes alates vanusest 3;0 ehk KH+SA+RA väljendamise perioodil leiab **täna** kasutust ka sellistes lausetes, kus laps viitab verbi mineviku ajavormi abil ajas tagasi (13), (14).

(9) Andreas 2;0.07
*CHI: **täna** jutustan sellise loo.

(10) Andreas 2;0.22
*CHI: **täna** issi sünnipäev.

(11) Andreas 2;0.26
*CHI: mis tuleb **täna** külla?

(12) Andreas 2;8.13

*CHI: mängime **täna** tähemängu ära.

(13) Andreas 3;0.22

*CHI: aga mina **täna** lasteaias tegin õhupalle.

(14) Andreas 3;1.09

*CHI: ütle emmele, kes meil **täna** tädi külas käis, ütle issile.

Kokkuvõtvalt saab öelda, et lapse sõnavarasse ilmuvad esimesed olevikule viitavad ajasõnad on *praegu*, *nüüd* ja *täna*. Sünonüümsed *praegu* ja *nüüd* erinevad lapsekeeles oma kasutuse poolest. Ajasõna *praegu* tähendus hõlmab laiemat ajavahemikku olevikus, veel esineb see märkimisväärselt sagedasti eituse mõjualas, mille kaudu implitseeritakse kõnealuse sündmuse võimalik toimumine kõnehetkele eelnevalt või järgnevalt. Ajasõna *nüüd* leiab pigem kasutust sündmuse hetkelise muutuse tähistamisel ning selle tähendus hõlmab enamasti lühikest ajahetke. Ühikuline oleviku ajasõna *täna* on lapsekeeles sagedasti esinev sõna, mille abil väljendatakse kõnehetke ajal käesolevat päeva.

4.2. Minevikule osutavad ajasõnad

Esimeste mineviku ajasõnadena ilmuvad lapse repertuaari *eile* ja *ammu*. Selleks ajaks on lapse kõnes leida juba ka esimesi verbi lihtminevikuvorme (Argus 2007, 2008).

Minevikule osutavatest ajasõnadest ilmub esimesena ühikuline *eile* (15), (16) ning lapsekeele kasutuses tähistab see määratlemata aega minevikus. Spontaanse kõne andmete põhjal on keeruline määrata, kas sõna tähendus on omandatud või mitte. Küll aga nähtub andmetest, et lapsed kasutavad sõna *eile* minevikuaja tähenduses ning vähemalt esimese kolme eluaasta jooksul seda täiskasvanupäraselt enamasti ei kasutata (17). Varases eas kasutatakse ajasõna *eile* küll minevikuaja täpsustamiseks, kuid see võib hõlmata ükskõik millist aega minevikus.

(15) Andreas 2;0.01

*CHI: **eile** maal.

(16) Andreas 2;0.07

*CHI: kuidas käis **eile** maal?

(17) Andreas 2;8.13

*CHI: issi oli ka siin **eile** [= eelmine] kord seal vannis oli tudumas.

Kahes alamkorpuses (Andreas, Hendrik) leidis ka mineviku ajasõna *ammu* kasutust. See ilmub lapse kõnesse alles kolmeaastaseks saamise piiril; *ammu* abil hakkavad lapsed väljendama ajahetke kaugemas minevikus, vt allpool (51).

Esimeste mineviku ajasõnade *eile* ja *ammu* esinemine lapsekeeles on võrdlemisi harv. Uurimus verbi ajamorfoloogiast (Argus 2007) näitab, et umbes kaheaastasest peale sageneb minevikuvormis verbide hulk, mis omakorda võib selgitada seda, et minevikule osutavaid ajasõnu on varases lapsekeeles väga vähe. Järelikult on lapsel minevikulistest sündmustest rääkimiseks vajalikud grammatilised vahendid keeles juba olemas.

4.3. Tulevikule osutavad ajasõnad

Tulevikule osutavatest ajasõnadest ilmuvad esimestena **homme**, **kohe** ja **varsti**. Need ilmuvad lapse keelde KH+SA väljendamise perioodil, umbes kaheaastaselt; vanuse kasvades läheneb ka ühikulise ajasõna *homme* kasutus täiskasvanupärasele.

Ühikuline ajasõna **homme** ilmub samaaegselt vektoriaalsete tulevikule osutavate ajasõnadega *kohe* ja *varsti*, kuid selle kasutus erineb täiskasvanute omast. Lapsed kasutavad ajasõna *homme* nii tulevikuajale (18), (19), (22), (48) kui ka minevikuajale viitamiseks (20), (21).

- (18) Andreas 2;0.15
*CHI: **omme** [: homme] Atsu sünnipäev.
- (19) Hendrik 2;1
*CHI: **homme** tudi [= magab], see **homme** tudi.
- (20) Andreas 2;1.06
*CHI: **homme** sadas vihma.
- (21) Andreas 3;0.11
*CHI: ma tegin seda **homme**.
- (22) Liisa 3;0.20
*LII: **homme** lähme õue.

Tuleviku ajasõnad **kohe** ja **varsti** märgivad lapsekeeles määratlematus kauguses ees olevat ajahetke, osutades siiski pigem kõnehetkele vahetult järgnevale ajale. Lapsekeeles pole nimetatud kahe ajasõna esinemissagedusel märkimisväärseid erinevusi, kuid sisendkeeles on sagedasem *kohe* (vt tabelid 1 ja 2). Tulevikule osutav ajasõna *kohe* ilmub lapse repertuaari kaheaastaselt (23). Lisaks sellele, et lapsekeeles väljendab ajasõna *kohe* kõnehetkele vahetult järgnevat aega (24), (25), omab see mõnel juhul tähendusvarjundit 'kiiresti' (26), (27). Sama nähtub sisendkeelest (nt *ei tule toast ära, mine sinna tagasi kohe*); lastele suunatud kõnest võib leida veel ka *kohe* partiklistunud kasutust (nt *osta kohe suur auto; ahah, sul on kohe sellised soovid*).⁹

- (23) Andreas 2;0.14
*CHI: **kohe** hakkab.
- (24) Liisa 3;0.20
*LII: need kukuvad **koge** [: kohe] maha.
- (25) Andreas 3;0.07
*CHI: ma näitan sulle **kohe**.
- (26) Mariann 2;10.04
*CHI: paneme peale **kohe**.
- (27) Andreas 2;1.02
*CHI: issi napsas kinni selle **kohe** Atsu [: Andrease].

⁹ Andrease alamkorpuse materjalis leidub *kohe* tähenduste nn piiripealset ehk ka partiklilaadset kasutust (nt vanuses 3;0.22 *nüüd ma kasvan kohe* hästi suureks), kus ajaline tähendus on taandumas, kuid analüüsis on sellised juhud arvestatud ajalise tähenduse hulka.

Tulevikule osutav ajasõna **varsti** on selgepiirilise tähendusega kui *kohe* ning esineb lapse- kui ka sisendkeeles kõnehetkele järgneva ajavahemiku märkimiseks. See ilmub siinse analüüsimaterjali põhjal juba enne lapse kaheaastaseks saamist (28).

(28) Carlos 1;8.27

*CAR: tuh-tuh **vaarti** [: varsti] sõidab.

(29) Andreas 2;0.03

*CHI: **vasti** [: varsti] tuleb tibuke.

(30) Andreas 2;0.14

*CHI: **vasti** [: varsti] pime, siis paneme lambid pelema.

Võib oletada, et ajasõna *varsti* tähendus on laste jaoks hõlpsasti arusaadav (30), kuna selle tähendus on selgepiiriline. Sisendkeeles leidub ajasõnade *praegu* ja *varsti* vastandamist (31), mis kindlasti aitab lapsel mõlemate ajasõnade tähendusi mõista. Kõnehetke ja sellele järgneva aja vastandamist nende sõnadega esineb lapse kõneski (32).

(31) Andreas 2;1.06

MOT: **praegu** on veel valge, aga **varsti** kohe läheb pimedaks.

(32) Andreas 3;0.22

*CHI: meie oleme metsas **praegu**, **varsti** lähme koju tagasi.

Tuginedes eesti keele deiktiliste ajaväljendite liigitusele (Pajusalu jt 2004), saab siinse materjali põhjal öelda, et varase lapsekeele repertuaaris on tulevikule osutavateks ajasõnadeks üks piiritlevat ajavahemikku märkiv (*homme*), kaks ähmast ajavahemikku märkivat ja tulevikuajas lähemale osutavat sõna (*varsti*, *kohe*). Lähtudes siinkohal vaid lapsekeele materjalist, saab öelda, et tulevikule osutatakse ka järjestuslikkust väljendavate sõnadega *enne* ja *pärast* (vt allpool jaotis 4.5).

4.4. Teised aja mõistestamisega seotud deiktikud

Järgnevalt leiavad käsitlemist deiktilised ajasõnad, mis ei osuta iseseisvalt minevikule, olevikule ega tulevikule. Lauses kvantifitseeriva määrusena esinevatest ja korduvusele või järgnevusele viitavatest hāgusama tähendusega ajasõnadest **juba**, **veel**, **alles** ja **jälle** järeldeb kirjeldatava tegevuse eeldatav toimumisaeg või -ruum (nt lausest *ma juba õpin* järeldeb, et õppimine on ootuspärasest varasem tegevus).

Ajasõna **juba** ilmub lapse keelde kõnehetke väljendamise perioodil ning KH+SA väljendamise perioodi alguses (Andreasel juba vanuses 1;10), see esineb esmalt vaid kas üksikult (33) või koos verbi olevikuvormiga (34), (35). Ajasõna *juba* tähendusest võib jäeldada sel juhul kirjeldatud sündmuse või situatsiooni ootuspärasest varasema aja – näitelausest (33) võib jäeldada, et kõneleja seisukohast on pimedaks minek ootuspärasest varasem. Samas pole kindlaid andmeid, et lapsed selle sõna tähendust kirjeldatud viisil tajuvad. Verbi minevikuvormiga koos esinev *juba* (36), (37) osutab kirjeldatud tegevuse lõpetatusele.

- (33) Andreas 2;0.24
*CHI: õues **juba** pime.
- (34) Andreas 2;1.01
*CHI: notsul on **juba** heina.
- (35) Mihkel 3;0.20
*CHI: me sõidame uba [: **juba**].
- (36) Andreas 2;6.12
*CHI: Kriimsilm **juba** vaatasin.
- (37) Liisa 3;0.20
*CHI: **juba** tuttu kits läks.

Adverb **veel** on varases lapsekeeles üks sagedasemaid sõnu (vt ka Vider 1995), see ilmub lapse kõnesse esmalt küll lisahulga tähenduses (nt *veel piima*), kuid enne kolmandat eluaastat kasutavad lapsed seda ka aja väljendamiseks. Näiteks Andrease lindistustes esineb *veel* kokku 491 korral, millest aja tähenduses 54 korral. Aja tähenduses kasutavad lapsed seda sõna käesoleva sündmuse aja väljendamiseks (38). Eitavas lauses esinev ajaline *veel* on varases lapsekeeles üsna sage, näiteks Andrease materjalis 16 juhul 54-st ning kolmikute korpuse materjalis kõigil kasutusjuhtudel. Eituse mõjujalas võib *veel* implitseerida kirjeldatud sündmuse toimumise kõnehetke ajale järgnevalt ehk osutada tuleviku ajale (39), (40).

- (38) Andreas 3;0.09
*CHI: kutsu on **veel** haige.
- (39) Andreas 2;8.13
*CHI: ma **veel ei võta**; natukene on veel kuum
- (40) Liisa 3;0.20
*LII: **ei lähe veel** õue.

Ajasõna **alles** esineb siinse uurimuse materjalis vaid Andrease alamkorpuses ning ilmub KH+SA väljendamise perioodi alguses (41). See sõna esineb vaadeldud materjalis koos verbi olevikuvormiga kahes funktsioonis: viidates kõnehetke ajale (42), kui selle sõna tähendusest on järeldatav kirjeldatud situatsiooni ootuspärasest hilisem aeg (vrd *olen alles koolis = olen veel koolis*); üksikutel juhtudel ka viidates kõnehetkele vahetult eelnenud ajale (43).

- (41) Andreas 2;1.03
*CHI: emme tudub veel **alles**.
- (42) Andreas 3;0.04
*CHI: see auto **alles** tuleb siit.
- (43) Andreas 2;0.20
*CHI: Atu [: Andreas] **alles** pani.

Korduvust väljendav ajasõna **jälle** ilmub suhteliselt vara, juba kõnehetke väljendamise perioodil (44). Ajasõna *jälle* kasutab laps mingi varem toimunud tegevuse või situatsiooni kordumisele viitamisel (45), (46) ning samas kasutuses koos ajasõnadega *homme* (48) ja *siis* (47) ka tuleviku viitamisel.

- (44) Andreas 1;08.25 [Situatsioon, kus ema palub lapsel müdistamine lõpetada, kuid laps teeb seda veel kord]
*CHI: **Lälle** [: jälle].
- (45) Andreas 1;09.11
*CHI: **jälle** uppi.
- (46) Andreas 2;0.03
*CHI: näe **jälle** suul [: suur] auto.
- (47) Andreas 2;5.30
*CHI: **siis jälle** korjame kartuleid.
- (48) Andreas 2;0.24
*CHI: **ome** [: homme] **lälle** [: jälle] puhastame.

Korduvusele ja järgnevusele viitavad hägusama tähendusega ajasõnad *juba*, *veel*, *alles* ja *jälle* on lapsekeeles sagedased ning ilmuvad kõnesse juba enne lapse kaheaastaseks saamist. Need on sõnad, mis võivad esineda nii verbi oleviku- kui ka minevikuvormiga, aidates teiste võimaluste kõrval väljendada ka tulevikuaega. Üldistavalt saab öelda, et ajasõnu *juba*, *veel*, *alles* ja *jälle* esinevad lapsekeeles kirjeldatud situatsiooni või sündmuse (ootuspärasest varasemat või hilisemat) muutuse märkimiseks.

4.5. Ajalisi järgnevussuhteid väljendavad sõnad

Järgnevalt leiavad käsitlemist ajasõnad **enne**, **pärast** ja **siis**. Need on sõnad, mille abil lapsed hakkavad kirjeldama sündmuste ajalist järjestust. Nii lapsekeeles kui ka sisendkeeles on **siis** sagedasim ajasuhet väljendav sõna.

Ajasõna **enne** ilmub lapse kõnesse KH+SA perioodi algul (49), kuid selle kasutus on harv ja tähistab pigem kirjeldatava tegevuse vahetut eelnemist. Sõna kasutus sageneb oluliselt KH+SA+RA väljendamise perioodil, kui materjalist võib leida **enne** abil nii tegevuste järjestumise raames eelnevusele viitamist (50) kui ka minevikus toimunud sündmusele viitamist (51). Neist esimene tähendus on sünonüümne sõnaga *kõigepealt* ja teine sõnaga *varem*. Lapsele suunatud kõneski esineb sõna *enne* nii minevikus toimunud sündmusele viitamisel (nt *loe see sõna, mis sa ütlesid enne*) kui ka tegevuse eelnemist tähistavas kasutuses (nt *söö suu tühjaks enne*).

- (49) Andreas 2;0.02
*CHI: kots [: klots] läheb **enne**, siis lusikas.
- (50) Andreas 3;0
*CHI: ei, **enne** peab soolase ära sööma.
- (51) Andreas 3;0.08 [Arsti-mängu situatsioon]
*CHI: aga ma **enne** tahtsin nii **ammu** neid [= arstimütsi ja näomaski].

Ajalist järgnevust¹⁰ väljendav **pärast** (52) on määratlematu pikkusega eesolevale ajavahemikule osutav sõna. Esimesed sõna *pärast* kasutusjuhud on lapsekeelele iseloomulikud täiskasvanu poolt öeldu imitatsioonid KH+SA väljendamise perioodil, kuid vanuse kasvades sõna kasutus sageneb. Tulevat aega hakkab laps väljendama koos verbi olevikuvormi ja ajasõnaga *pärast* (53). Teistes vaadeldud korpusetes pole sõna *pärast* esinemisjuhte ei lapse- ega sisendkeeles leida, sest need lindistused on lühemad, mistõttu on neis ka keelematerjali vähem.

(52) Andreas 2;0.23

*CHI: **pärast** võtame maast ära.

(53) Andreas 2;5.02

*CHI: ma panen selle raamatu sinna alla, siis **pärast** loevad, praegu on tuduaeg.

Lapse repertuaaris on **siis** esmalt kasutusel ajale ja järgnevusele viitavana (sõna *siis* semantilist analüüsi vt Jansons 2001). Esimesed esinemisjuhud on täheldatavad juba kõnehetke väljendamise perioodil, kui seda sõna kasutatakse sarnaselt ajalisele järgnevusele viitava sõnaga *pärast* (54), (55). Samuti haakub sõna *siis* järjestumise tähendust kandev kasutus sõnaga *enne* (56), mille abil väljendatakse kõnehetkel toimuva ja kavandatava tegevuse järjestamist.

(54) Andreas 2;0.01

*CHI: **siis** helistab manale [: vanaemale].

(55) Mihkel 3;0.20

*CHI: **siis** saate kommi.

(56) Andreas 2;0.02

*CHI: kots [: klots] läheb **enne**, **siis** lusikas.

Tegemist on lapse igapäevase suhtluse loomuliku osaga, kus ema ja laps on koonandanud tähelepanu samale tegevusele ning see annab põhjust järjestikku toimuvast rääkida. Järjestumist väljendavad sõnad *enne*, *pärast* ja *siis* esinevad ka lapsele suunatud kõnes sellises situatsioonis, kus tegevuse sooritamine sõltub nn vaheetappide läbimisest (nt mingi toidu valmistamine, pusle ja klotside kokkupanek vms). Relatiivset aega hakkavad lapsed esialgu väljendama adverbidega *siis* ja *kui*, esimene selline kasutus ilmub juba KH+SA väljendamise perioodi alguses (57), kuid see pole sage. Alates vanusest 2;5 võib märgata relatiivse aja väljendamise kasvu sõna *siis* (ja *kui*) abil (58), (59), kuid sõna *siis* valdavaks funktsiooniks jääb siiski ajalisele järjestusele viitamine.

(57) Andreas 2;0.22

*CHI: **siis kui** see läpeb [: lõpeb] äla [: ära] see.

(58) Andreas 3;0.09

*CHI: **kui** kutsul tuleb sünnipäev, **siis** sööme seda;

(59) Mariann 2;10.04

*MAN: **kui** uue kodu teevad ilusaks, **siis** saame.

¹⁰ Andraese materjalist nähtub, et põhjuslikkuse tähenduses esineb *pärast* esmakordselt vanuses 3;0.02 (*nalja pärast ma panin*), seega on valdav *pärast* kasutus vaadeldud keelematerjalis ikkagi ajaline.

Varases lapsekeeles ilmuvad ajasõnad *enne, pärast* ja *siis* osutavad kirjeldatava tegevuse ajalisele järjestussuhtele. Lapse vanuse kasvades tekib neile sõnadele ka laiem tähendus, mis osutab ajas kaugemale minevikule ja/või tulevikule.

5. Kokkuvõtteks

Varasemad uurimused (Weist 2002a) on näidanud, et ajaline viitamine lapsekeeles algab kõnehetke ja sündmuse aja deiktiliste suhete väljendamisest. Siinse uurimuse tulemused toetavad seda väidet – kuni kolmeaastase lapse repertuaar hõlmab valdavalt deiktilisi ajasõnu, mis osutavad kõnehetke ajale ning kõnehetkest lähtuvalt ajas edasi või tagasi. Ajasõnad ilmuvad lapse keelde kahe sõnalause perioodil ning mõni aeg hiljem kui näiteks esimesed deiktilised ruumisõnad (vt Pajusalu jt 2004: 114).

Ajalist viitamist alustab laps sõnadega, mis ka täiskasvanukeeles sagedasti esinevad, ning räägib esialgu sündmustest, mis toimuvad siin ja praegu. Siinse uurimuse tulemused näitavad, et varase keele ajasõnad on deiktikud, mida võib jagada kolme rühma: esiteks deiktilised ajasõnad, mis osutavad olevikule (*praegu, nüüd, täna*), tulevikule (*homme, kohe, varsti*) ja minevikule (*eile, ammu*); teiseks deiktikutena käituvad ajasõnad *juba, veel, alles, jälle*, mis ei osuta iseseisvalt minevikule, olevikule ega tulevikule – need on lauses kvantifitseeriva määrusena esinevad ja korduvusele või järgnevusele viitavad hägusama tähendusega ajasõnad; kolmandaks ajalist järjestuslikkust väljendavad *enne, pärast* ja *siis*.

Olevikule viitavate ajasõnade sagedas kasutus annab tõestust olevikukesksest aja-süsteemist lapseea varasel perioodil. Oleviku ajasõnad ilmuvad lapse kaheaastaseks saamise piiril ning samal perioodil esinevad lapse kõnes ka verbi olevikuvormid. Seega võib öelda, et aja väljendamine algab grammatilise aja kõrval samaaegselt leksikaalsete vahenditega. Esimesed oleviku ajasõnad (*praegu, nüüd, täna*) täpsustavad viidatavat oleviku ajahetke. Oleviku ajasõnadest väärrib esiletõstmist *praegu*, mis esineb lapsekeeles sageli koos sündmuse eitusega, lastes järeldada selle toimumist edaspidi.

Esimeste mineviku ajasõnade *eile* ja *ammu* esinemine lapsekeeles on harv. Minevikuvormis verbide hulk aga aina sageneb umbes kaheaastasest peale. Järelikult on lapsel vahetult enne kõnehetke toimunud sündmustest rääkimiseks vajalikud grammatilised vahendid keeles juba olemas. Mineviku ajasõnade vähesus lapsekeeles võibki sellest johtuda. Võrdluseks olgu siinkohal välja toodud tulevikule viitavad ajaadverbid lapseea varasel perioodil: *homme, varsti, kohe, siis, pärast*; lisaks eituse mõjualas tulevikuaega implitseerivad *veel* ja *praegu*. Tulevikku ei saa eesti keeles grammatiliselt väljendada, seega vajab see ajatähendus enam leksikaalseid vahendeid.

Esimesed ajasõnad on deiktikud, mis viitavad ajasuhetele lähtuvalt kõnehetkest. Leksikaalse aja väljendamine algab olevikukeskselt – kõigepealt ilmuvad olevikule osutavad ja olevikuaega täpsustavad ajasõnad. Seejärel sageneb nii mineviku väljendamine grammatiliste ja leksikaalsete vahenditega kui ka tuleviku väljendamine ajaadverbidega.

Lühendid ja sümbolid

CHI, MIH, CAR, LII, MAN lapse nime lühend lapsekeele korpuses

KH kõnehetk

MOT ema

RA relatiivne aeg

SA sündmuse aeg

[] nurksulgudes on lisatud kommentaarid

(*) täiskasvanupärasest keelest hälbiv vorm

Materjali allikad

Andmebaasi CHILDES eesti keele korpuse. <http://chilides.psy.cmu.edu/data/Other/Estonian>
(20.02.2012)

Kolmikute korpuse. (Reili Arguse koostatud andmekogu koostaja valduses.)

Viidatud kirjandus

Argus, Reili 2004. Eesti keele käändesüsteemi omandamine: esimestest sõnadest mini-paradigmadeni. – Emakeele Seltsi aastaraamat, 49 (2003), 24–49.

Argus, Reili 2005. Imitatiivide kohas lapsekeeles: reduplikatsioonist, morfoloogiast ja sõnaliigilisest ambivalentsusest. – Eesti Rakenduslingvistika Ühingu aastaraamat, 1, 19–34.

Argus, Reili 2007. Eesti keele verbi ajamorfoloogia ja aspektilisuse omandamisest. – Emakeele Seltsi aastaraamat, 52 (2006), 7–32.

Argus, Reili 2008. Eesti keele muutemorfoloogia omandamine. Tallinna ülikooli humanitaarteaduste dissertatsioonid, 19. Tallinn: TLÜ Kirjastus.

Argus, Reili; Parm, Sirli 2010. Eesti keele ajakategooria omandamisest – ajavormid ja ajasõnad. – Eesti Rakenduslingvistika Ühingu aastaraamat, 6, 25–41. <http://dx.doi.org/10.5128/ERYa6.02>

Bloom, Paul 2000. How Children Learn the Meanings of Words. Cambridge, MA: MIT Press.

Clark, Eve Vivienne 1973. How children describe time and order. – C. A. Ferguson, D. I. Slobin (Eds.). Studies of Child Language Development. New York: Holt, Rinehart & Winston, 585–606.

Clark, Eve Vivienne 2009. First Language Acquisition. 2nd ed. Cambridge: Cambridge University Press.

Diessel, Holger 2004. The Acquisition of Complex Sentences. Cambridge: Cambridge University Press. <http://dx.doi.org/10.1017/CBO9780511486531>

Elliot, Alison J. 1981. Child Language. Cambridge: Cambridge University Press.

Friedman, William J. 1982. Conventional Time Concepts and the Children's Structuring of Time. – W. Friedman (Ed.). The Developmental Psychology of Time. New York: Academic Press, 171–208.

Hennoste, Tiit; Lindström, Liina; Rääbis, Andriela; Toomet, Piret; Vellerind, Riina 2000. Eesti suulise kõne korpuse ja mõnede allkeelte võrdluse katse. – T. Hennoste (Toim.). Arvutuslingvistikalt inimesele. Tartu Ülikooli üldkeeleteaduse õppetooli toimetised, 1. Tartu: Tartu Ülikool, 245–283.

Jansons, Airi 2001. Partikkel siis eesti suulises kõnes: üldiseloostus ja funktsioonid. Bakalaureusetöö käsikiri Tartu Ülikooli eesti keele õppetoolis.

Leiwo, Matti 1993. Lapse keeleline areng. Tallinn: Tallinna Pedagoogikaülikool.

Levinson, Stephen 1983. Pragmatics. Cambridge: Cambridge University Press.

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

Pajusalu, Renate; Tragel, Ilona; Veismann, Ann; Vija, Maigi 2004. Tuumsõnade semantikat ja pragmaatikat. Tartu Ülikooli üldkeeleteaduse õppetooli toimetised, 5. Tartu: Tartu Ülikool.

- Pajusalu, Renate; Tõugu, Pirko; Vija, Maigi; Tulviste, Tiia 2011. Konditsionaali omandamisest eesti lapsekeeles. – Eesti Rakenduslingvistika Ühingu aastaraamat, 7, 141–155. <http://dx.doi.org/10.5128/ERYa7.09>
- Reichenbach, Hans 1947. *Elements of Symbolic Logic*. New York: Free Press.
- Salasoo, Tiiu 2005. The initial vocabulary of children learning Estonian – What do they talk about? – *Summaria acroasium in sectionibus*. *Linguistica*. Yoshkar-Ola, 225–227.
- Salo, Age 1995. Eesti keele verbivormistiku omandamine vanuses 1;5-3;11. Magistritöö. Käsikiri Tartu Ülikooli läänemeresoome keele õppetoolis. Tartu: Tartu Ülikool.
- Slobin, Dan Isaac 1985. Crosslinguistic evidence for the language-making capacity. – D. I. Slobin (Ed.). *The Crosslinguistic Study of Language Acquisition*. Vol. 2. *Theoretical Issues*. Hillsdale, NJ: Lawrence Erlbaum Associates, 1157–1256.
- Tomasello 2003. *Constructing a Language. A Usage-Based Theory of Language Acquisition*. Cambridge: Cambridge University Press.
- Vider, Kadri 1995. 2–3-aastaste laste sõnavara. Tartu Ülikool. Diplomitöö käsikiri Tartu Ülikooli eesti keele õppetoolis.
- Vihman, Marilyn May; Maigi Vija 2006. The acquisition of verbal inflection in Estonian. – N. Gagarina, I. Gülzow (Eds.). *The Acquisition of Verbs and their Grammar: The Effect of Particular Languages*. *Studies in Theoretical Psycholinguistics*, 33. Dordrecht: Springer, 269–295.
- Vija, Maigi 2005. Asesõnad ja isikuvõtamine lapsekeeles. Magistritöö. Käsikiri Tartu Ülikooli eesti keele õppetoolis.
- Vija, Maigi 2007. Pronoomenid lapsekeeles: mõnda MINA ja SINA omandamisest eesti laste näitel. – Eesti Rakenduslingvistika Ühingu aastaraamat, 3, 371–384.
- Vija, Maigi; Torn-Leesik, Reeli; Pajusalu, Renate 2009. Tegumood eesti lapsekeeles. – Eesti Rakenduslingvistika Ühingu aastaraamat, 5, 329–344. <http://dx.doi.org/10.5128/ERYa5.22>
- Weist, Richard 1986. Tense and aspect. – P. Fletcher, M. Garman (Eds.). *Language Acquisition: Studies in First Language Development*. Cambridge: Cambridge University Press, 356–374.
- Weist, Richard 2002a. Space and time in first and second language acquisition: A tribute to Henning Wode. – P. Burmeister, T. Piske, A. Rohde (Eds.). *An Integrated View of Language Development: Papers in Honor of Henning Wode*. Trier: Wissenschaftlicher Verlag Trier, 79–108.
- Weist, Richard 2002b. Temporal and spatial concepts in child Language: Conventional and configurational. – *Journal of Psycholinguistic Research*, 31 (3), 195–210. <http://dx.doi.org/10.1023/A:1015522711549>
- Weist, Richard; Wysocka, Hanna; Lytinen, Paula 1991. A cross-linguistic perspective on the development of temporal systems. – *Journal of Child Language*, 18, 67–92. <http://dx.doi.org/10.1017/S0305000900013301>
- Winkel, Heather 2004. The acquisition of temporal reference cross-linguistically using two acting-out comprehension tasks. – *Journal of Psycholinguistic Research*, 33 (4), 333–355. <http://dx.doi.org/10.1023/B:JOPR.0000035105.65440.b8>

Sirli Parm (Tartu Ülikool) uurimisvaldkondadeks on esimesese keele omandamine (ajasüsteemi omandamine) ja leksikaalne semantika.
sirli.parm@ut.ee

EARLY DEVELOPMENT OF TIME ADVERBS IN ESTONIAN CHILD LANGUAGE

Sirli Parm

University of Tartu

This article discusses the acquisition of the category of time by Estonian children between one and three years of age. The discussion is based on a study of recordings of Estonian children in the CHILDES database and in the triplets corpus. At the beginning of the acquisition of the category of time there exists only means for expressing present time but before turning three years of age children will also start talking about events that have happened in the past or will happen in the future. A previous study (Argus, Parm 2010) has showed that Estonian children develop a clear distinction between past and non-past; the present tense forms are already understood by four-year-olds while the past and future forms only by six-year-olds. The aim of the study is to find out which time adverbs appear in the speech of a child before the age of three and describe the functions in which these time adverbs occur in the child's speech.

The results of the study show that: a) when children begin to utilize tense morphology, they are capable of expressing present tense forms at the same time as some time adverbs; b) time adverbs appear in the child's speech before the age of three and many of them resemble adult usage; c) the time adverbs used the most frequently in order to express the future tense are: *varsti* 'soon', *kohe* 'immediately', *siis* 'then', *pärast* 'after'; d) the sense of time adverbs that express conventional time are not fully acquired (for example, *eile* 'yesterday' means any time in the past); e) between 2;5 and 3;1 years the amount of *then*- and *when*-clauses increase, consistent with the rapid development of talking ability; f) in early children's speech, the adverbs *siis* 'then', *enne* 'before' and *pärast* 'after' denote the sequence of time. When children are aged between about 1;7 and 3;0, they productively code relations between speech time and event time, as shown in the longitudinal Estonian data.

Keywords: first language acquisition, first words, semantics, tense, Estonian

DIREKTIIVISEKVENTSID ISA JA TÜTRE SUHTLUSES: JUHTUMIANALÜÜS

Andriela Rääbis

Ülevaade. Artiklis uuritakse direktiivisekventse isa ja viieaastase tütre vestluses. 42-minutilise suhtluses, mille käigus meisterdati paberist ninasarvikut, esitas isa 92 ja laps 15 direktiivi. Uurimuses järgitakse suhtluslingvistika põhimõtteid ning kasutatakse vestlusanalüüsi mõisteid. Analüüsitakse direktiivivooru konstrueerimisel kasutatavaid lingvistilisi ja mitteverbaalseid vahendeid ning uuritakse, missugused tegevused direktiivile eelnevad ja järgnevad.

Naabruspaari esiliikmena väljendavad direktiivid käske, palveid, keelde, juhiseid, ettepanekuid ja pakkumisi. Kesksed vormistusvõtted on modaalsed konstruktsioonid ja käskiv kõneviis. Artiklis keskendutakse sellele, kuidas direktiivi vormistus mõjutab järelliikme valikut: erinevalt vormistatud esiliikmed kutsuvad süstemaatiliselt esile eri tüüpi järelliikmed. Käskivas kõneviisis direktiiv täidetakse mitteverbaalselt, muul viisil vormistatud direktiivide eelistatud järelliikmeks on nõustumine.

Direktiivid on järelliikmeks küsimusele või ettepanekule, väljendades nõustumisi, lubamisi, palveid, keelde, juhiseid ja ettepanekuid.*

Võtmesõnad: suhtluslingvistika, vestlusanalüüs, lapsekeel, naabruspaar, direktiiv, eesti keel

1. Sissejuhatus

Artiklis uuritakse direktiivisekventse isa ja viieaastase tütre suhtluses, kasutades vestlusanalüüsi baasmõisteid ja järgides eeskätt suhtluslingvistika põhimõtteid. Tegu on juhtumianalüüsiga (ingl *case study*; vt Schegloff 1987). Direktiivi all mõistetakse tegevust, mille eesmärk on vestluskaaslase mõjutamine tegutsema teatud viisil või ühise tegevuse kavandamine. Vaadeldakse, kuidas kõnelejad direktiivivooru konstrueerivad ja direktiivi kui tegevuse läbi viivad. Keskendutakse sellele, kuidas vestluskaaslane direktiivile reageerib, ning otsitakse seoseid direktiivi

* Artikli valmimist on toetanud Euroopa Regionaalarengute Fond Eesti Arvutiteaduse Tippkeskuse kaudu, Eesti Teadusfond (grandid nr 7492, 7503 ja 8558), Haridus- ja Teadusministeerium (sihtfinantseeritav teema SF0180078s08).

vormi ning järellikme tüübi vahel. Analüüsitakse nii verbaalset kui mitteverbaalset suhtlust.

Uurimismaterjaliks on 42 minuti pikkune videosalvestus suhtlusest, mille käigus isa ja tütar meisterdavad paberist ninasarvikut. Ninasarvik tuleb välja lõigata teatri kavaraamatust, antud on ka tööjuhised. Samas toas viibib ka ema, kes korraks vestlusele sekkub. Dialog on transkribeeritud vestlusanalüüsi transkriptsiooni kasutades.¹

Direktiive on lingvistikas palju käsitletud, kuid enamik uurimusi lähtub kõneaktiteooriast, analüüsides üksikuid voore ilma konteksti arvesse võtmata. Vestlusanalüüs näitab, et suulise suhtluse oluline omadus on organiseeritus järjenditena ehk sekventsidenähtena (ülevaateid vestlusanalüüsi kohta vt nt Liddicoat 2007, Sidnell 2010, eesti keeles Hennoste jt 2001, Kasterpalu, Gerassimenko 2006). Järjendiliigenduse keskne mõiste on naabruspaar – esiliikmest ja järellikmest koosnev kahe vooru järjend. Esiliige ja järelliige on eri kõnelejate esitatud ning paiknevad ideaaljuhul kõrvuti (Schegloff, Sacks 1973: 295–296). Järellikme puudumisel osutab esimene kõneleja mingil viisil selle ootusele, nii võib ta pidada pausi, korrata esiliiget, lisada sellele midagi juurde, pakkuda oletatavat vastust jne (Drew 1984, Schegloff 1990).

Järelliige võib olla valitud kahe võimaluse hulgast, nt palvele võidakse vastata nõustumise või keeldumisega. Seda valikut juhib eelistusjärjestus: üks neist valikutest on eelistatud, teine mitte. Vestlusanalüüsis ei viita eelistuse mõiste suhtleja psühholoogilistele motiividele, vaid pigem konkreetsete tegevustega seotud voorude kujundamisele. Eelistatud ja mitte-eelistatud järellikme keeleline vormistus erineb süstemaatiliselt. Mitte-eelistatud voor lükatakse edasi, kujundatakse mitmeti tõlgendatavaks või keerukamaks, varustatakse seletusega (vt Pomerantz 1984, Sacks 1987 [1973], Schegloff 2007: 63–73). Mitte-eelistatud võib olla ka naabruspaari esiliige, nt palve, mida samuti põhjendatakse, pehmendatakse või lükatakse edasi (Schegloff 2007: 81–96).

Suhtluslingvistika (ingl *interactional linguistics*) on kujunenud lingvistilise antropoloogia, vestlusanalüüsi ja funktsionaalse lingvistika baasil (vt nt Couper-Kuhlen, Selting 1996, Ochs jt 1996, Hakulinen, Selting 2005, eesti keeles Keevallik 2002). Suhtluslingvistik huvitab see, missuguseid keelelisi vahendeid kasutatakse konkreetsete vestlusstruktuuride väljendamiseks ja suhtlusfunktsioonide täitmiseks; teisalt aga see, missugune suhtlusfunktsioon või vestlusstruktuur on väljendatud konkreetsete lingvistiliste vormide ja nende kasutusviisi kaudu (Couper-Kuhlen, Selting 2001). Vestlusanalüüsiga võrreldes on suhtluslingvistika rohkem keskendunud keelele.

Lapsekeelt ning lapsega suhtlemist ei ole Eestis seni vestlusanalüüsi ega suhtluslingvistika meetodil kuigivõrd uuritud. Mariliis Malahhov (2011) on analüüsinud lasteaiaste suhtlust, esimeses versioonis valminud on algklassitunde käsitlev uurimus (Mihkels 2011).

2. Direktiivide uurimisest

John Searle'i (1976, 1979) kõneaktide klassifikatsiooni järgi on direktiivid aktid, millega ütleja soovib panna adressaati midagi tegema. Direktiivide kategooriasse kuuluvad Searle'i (1979: 5) järgi palved, korraldused, käsud jms. Ka Marjorie Goodwin (2006: 517) defineerib direktiive kui väljendeid, mis on suunatud sellele, et panna keegi midagi tegema. Nii Searle'i kui Goodwini järgi on palved direktiivide üks liik. Mõned uurijad on palved ja direktiivid eristanud, nt Alexandra Craven ja Jonathan Potter (2010) iseloomustavad palvet kui tegevust, millega üks osaleja palub teisel midagi teha, ning direktiivi kui tegevust, millega üks osaleja käsib teisel midagi teha. Niisuguse lähenemise järgi projitseerib direktiiv järgmise relevantse tegevusena täitmise, palve aga aktsepteerimise.

Paljude direktiive käsitlevate uurimuste eesmärk on viisakusstrateegiate uurimine ja võrdlemine eri kultuurides, lähtudes Browni ja Levinsoni (1987) viisakusteooriast (nt Blum-Kulka 1987, Fukushima 1996, Márquez Reiter 2000, Law 2009). Sealjuures ei analüüsita kõnesituatsiooni kui tervikut, vaid üksikuid voore. Uurimismaterjal on sageli kogutud rollimängudes. Need uurimused näitavad, et kõneakti sooritamine põhineb universaalsetel pragmaatilistel printsiipidel, mille rakendusviisi ja kasutuskonteksti võib kultuuriti erineda.

Susan Ervin-Tripp (1976: 29) esitab direktiivide hierarhia otseste–kaudsete direktiivide skaalal ja väidab, et valik oleneb kõneleja ja adressaadi võimusuhetest. Niisugust lähenemist on edasi arendanud ka nt Marjorie Goodwin (1990) ja Leah Wingard (2006).

Palvete otsesemaid ja kaudsemaid variante on käsitlenud Traci Curl ja Paul Drew (2008), kes ei lähtu viisakusest ega võimusuhetest, vaid otsivad seoseid vooru ehituse ja lokaalse konteksti vahel. Nad väidavad, et palve tüüp valitakse lähtudes esiteks sellest, kas palve esitajal on õigus eeldada palve täitmist, ning teiseks sellest, kuidas palve esitaja hindab vastuvõtja võimalust palvet täita. Palvete lingvistilisele ehitusele on keskendunud ka Trine Heinemann (2006). Carmen Taleghani-Nikazm (2006) on uurinud palvesekventse saksakeelses suhtluses.

Ka lapsekeeles esinevaid direktiive on uuritud nii kõneaktiteooriast lähtudes kui ka empiirilisel tasemel kaasnevate tegevust arvesse võttes.

Browni ja Levinsoni viisakusteooriast lähtuv laste direktiivide uurimine on näidanud, et tüdrukud püüavad nägu ohustavaid akte pehmedada, poisid kalduvad aga otsesemad olema (Andersen 1990, Gleason 1987). Karin Aronsson ja Mia Thorell (1999) on analüüsinud direktiive laste rollimängus, keskendudes sellele, kuidas direktiivid kajastavad soolisi stereotüüpe ja vanuselisi hierarhiid. Erica Hulsi, Anne-Lies van de Peppeli ja Carel van Wijki (2007) ning Cheung-Shing Sam Leungi ja Lornita Wongi (2009) uurimused näitavad, kuidas laps omandab direktiivide kasutamise.

Järjest rohkem ilmub uurimusi perekondlike argivestluste kohta, milles käsitletakse terviklikke direktiivijärgendeid ja nende seoseid kaasneva tegevusega (nt Wootton 1981, Goodwin 2006, Aronsson, Cekaite 2007, Craven 2010, Craven, Potter 2010).

Eesti dialoogiaktide tüpoloogias EDiT (Hennoste, Rääbis 2004), mis baseerub reaalse suhtluse mikroanalüüsil ning mida on kasutatud institutsionaalsete dialoogide märgendamisel, on direktiivsete naabruspaaride esiliikmeteks soov, ettepanek, pakkumine, palve oodata ning muu (direktiivsetest naabruspaaridest

infodialoogides vt Hennoste jt 2003, Gerassimenko jt 2010). Soovid (mis hõlmavad ka käsud ja palved) väljendavad kõneleja vajadust või tahet midagi saada või tahet, et partner midagi teeks. Infodialoogides ootab soov vastuseks infot. Ettepanekud ja pakkumised ootavad partnerilt nõustumist või tagasilükkamist. Pakkumine lähtub endast (*tulen homme läbi*), ettepanek partnerist või mõlemast osalejast (*helistage homme; võime nimes kirja panna*). Soovid nõuavad alati vahetut täitmist ja on seotud dialoogi edasiarenemisega. Pakkumised eeldavad üldjuhul vahetut täitmist, ettepanekud aga ootavad tüüpiliselt mingit tegutsemist tulevikus, väljaspool antud dialoogi.

Direktiive eesti argivestlustes on analüüsinud Helle Metslang (2004), kes on leidnud, et käske ja teisi direktiivseid akte väljendatakse enamasti imperatiivvormis verbi ning käsklause abil, ilma pehendamata. Direktiivsuse väljendamiseks on aga ka arvukalt muid viise, nt konditsionaal, modaalverbid, küsilause jms. Konditsionaali kasutamist ettepanekute, soovide ja palvete pehendamiseks on uurinud Karl ja Renate Pajusalu (2004: 264–266). Helena Teemets (2010) on EDiT-ist lähtudes käsitlenud ettepanekuid ja pakkumisi eesti argidialoogides. Tiia Tulviste (2001) on uurinud direktiive emade ja kuueaastaste laste suhtluses kahe konteksti näitel, söögilauavestluses ja pusle kokkupanemisel, ning leidnud, et esimeses kontekstis kasutasid emad oluliselt rohkem käitumist reguleerivaid direktiive (nt *ära söögi ajal laula*), teises aga tähelepanu juhtivaid direktiive (nt *vaata see käib sinna*). Käesolevas töös käsitletav situatsioon on võrreldav pusle kokkupanemise kontekstiga: lapse tegevusel on kindel eesmärk.

3. Analüüs

Analüüsitava vestluses on 108 direktiivi, neist 92 esitas isa, 15 laps ja ühe ema. Paljud direktiivid tulenevad tööjuhendist: isa loeb seda, uurib jooniseid ning juhendab last, vahendades sellega juhendi koostaja antud direktiive. Samal ajal täidab ta ka ise juhendis olevaid direktiive. Osalusraamistikus (vt Goffman 1981 [1979]) on seega peale vahetute osalejate ka autor, kelle valitud mõtteid ja sõnu esitatakse.

Vaadeldavad direktiivid väljendavad käske, palveid, ettepanekuid, pakkumisi, keelde, juhiseid, lubamisi, nõustumisi. Nad esinevad vestluses nii naabruspaari esi- kui järelliikmena.

3.1. Direktiivid naabruspaari esiliikmena

Naabruspaari esiliikmena esitas isa 81, laps 11 ja ema ühe direktiivi.

Vaatleme esmalt, kuidas need direktiivid on vormistatud (tabel 1). Kesksed on modaalised konstruktsioonid (*s=sa pead liimi tooma.*) ja käsklused (*nii, sina võta siin niikaua see.*). Vähem kasutatakse modaliseerimata väitlauseid, mis on mitmuse 1. isikus (*liimime kohe.*), ainsuse 2. isikus (*pärast paned siia liimi vahele.*), ainsuse 1. isikus (*hhhhh tead mina lähen ja lõikan ned noaga: ned augud siia sisse.*). Kolm direktiivi on vormistatud küsilauseks (*kas nüüd paneme (0.7) sarve kokku või.*), üks verbita fraasina (*see ka.*). Direktiivid on käskivas ja kindlas kõneviisis, tingivat kõneviisi (mis on sage institutsionaalsetes dialoogides, vt Gerassimenko jt 2007) ei kasutata.

Tabel 1. Naabruspaari esiliikmena esinevate direktiivide vormistus

Vormistusviis	Isa	Laps	Kokku
Modaalset konstruktsiooni sisaldav väitlause	37	1	38
Käsklause	25	5	30
Väitlause mitmuse 1. isikus	11	5	16
Väitlause ainsuse 2. isikus	2	–	2
Väitlause ainsuse 1. isikus	2	–	2
Küsilause	3	–	3
Verbita fraas	1	–	1
Kokku	81	11	92

3.1.1. Käsklause (käskiv kõneviis)

Helle Metslang (2004) on leidnud, et eesti argivestlustes on imperatiiv peamine vahend käsu väljendamiseks. Ka minu materjalis oli käskiv kõneviis direktiivide sagedane vormistusvahend: isa kasutas seda 25 korda, laps 5 korda. Enamasti esitatakse direktiiv paralleelselt verbaalselt ja mitteverbaalselt: isa osutab sõrmega kujundile, paneb raamatu lapse ette, ulatab liimitava detaili lapse poole, laps tõstab näidatava asja üles jms. Käskivas kõneviisis direktiivid väljendavad käske, palveid (nende vahel ei ole selget piiri) ja keelde. Eelistatud järelliikmeks on kõikidel juhtudel tegevus.

Vaatleme kõigepealt isa algatatud direktiivisekventse. Laps reageerib isa direktiivile järgmistel viisidel.

Keskne on, et laps teeb soovitud tegevuse, järelliige on seega mitteverbaalne.

Mõnikord eelneb järelliikmele vahesekvents: kui laps ei saa direktiivist aru, algatab ta paranduse ning teeb soovitud tegevuse pärast paranduse läbiviimist. Näites (1) annab isa lapsele korralduse järgmine detail välja lõigata. Käsk *`sina võta siin `niikaua `see* ei ole üheselt mõistetav ning laps algatab paranduse deiktikuga *see*, osutades samal ajal detailile raamatus. Isa viib paranduse läbi ja jätkab direktiivi: *`lõika `see `välja*. Laps hakkab laualt kääre otsima, osutades valmisolekule direktiivi täita. Isa esitab küsimuse direktiivi täitmise võimalikkuse kohta (*`käärid sul `on*). Laps ei vasta küsimusele, ohkab ja hakkab lõikama, täites seega direktiivi mitteverbaalselt.

(1) I: **nii, `sina võta siin `niikaua `see**, ((I pöörab raamatu lehte))

(6.2) ((mõlemad pööravad raamatu lehti))

L: `see. ((L näitab sõrmega raamatus olevale detailile))

I: `see, (.) **`lõika `see `välja**.

(1.8) ((L vaatab lauall otsivalt ringi, liigutab kätt laua kohal, võtab käärid))

I: `käärid sul `on

(3.2)

L: ((ohkab))

(0.7)

I: `nii, `mina käin ((L hakkab lõikama)) (.) noaga natukene `lõikamas.

Kui laps ei täida direktiivi isa soovitud viisil, loob ta konteksti järgmisele direktiivile (näide 2). Isa kordab direktiivi tugevamal kujul (*hoia=oia`kõvasti*) või lisab seletuse (*vajuta`siit, kahelt`poolt*) (korratud direktiivide kohta vt Craven, Potter 2010). Kui laps on direktiivi täitnud, annab isa tema tegevusele hinnangu.

(2) I: **`nii, (.) näed, (1.8)** ((I näitab L-le oma tükki)) **nüd vajuta need`kinni.**

(8.1) ((I hoiab paberitükki käes, L vajutab, võtab kääre ära))

I: **vajuta`siit, kahelt`poolt.**

(7.3) ((L vajutab, võtab tüki enda kätte))

I: **`vot nimodi.**

(3.9) ((L hoiab tükki enda käes, I vaatab seda))

I: **hoia=`kinni. (0.4) hoia=oia`kõvasti.** (0.4) voh, (0.8) nüd`jääb.

Järgnevalt vaatleme lapse algatatud direktiivisekventse.

Laps kasutas käskivas kõneviisis ainult verbi *vaata*. *vaata* on suulises suhtluses sageli partiklistunud, sel juhul on ta tihti ka lühenenud (Keevallik 2003, 2005). Analüüsitava vestluses kasutatakse verbi *vaata* otseses tähenduses: see ei ole lühenenud (vahel on isegi rõhuline) ning samal ajal näidatakse ja/või seletatakse, mida vestluskaaslane peaks vaatama. Kui isa *vaata* on alati seotud juhustega edasiseks tegevuseks, siis laps näitab midagi, mida ta on teinud.

Sekvents jätkub järgmistel viisidel:

- isa reageerib nii verbaalselt kui mitteverbaalselt: vaatab näidatavat asja ning ütleb *jah*. Järgneda võib täpsustav küsimus, mis osutab isa soovile teemat edasi arendada;
- isa reageerib ainult verbaalselt: ütleb *jah*, aga ei vaata lapse poole. Lapsele piisab niisugusest reaktsioonist;
- üks lapse *vaata* on suunatud emale (isa on toast ära läinud), kes reageerib ainult mitteverbaalselt, pilku tõstes.

3.1.2. Modaalsete konstruktsioonide sisaldavad väitlused

Modaalsete konstruktsioonide kasutas isa 37 korda, laps ja ema kumbki üks kord. Modaalverbi sisaldav direktiiv on (vähemalt formaalselt) orienteeritud vastuvõtja võimalusele või soovile direktiivi täita.

27 juhul ei viidata otseselt agendile, kasutades konstruktsioone *ei tohi, tuleb, (on) vaja + da-tegevusnimi* ning *peab, peaks + ma-tegevusnimi*. Otsese agendiviite puudumine pehmendab direktiivi (vt Metslang 2004: 250). Neljal juhul on agendiks kuulaja (*sa pead + ma-tegevusnimi, võid + da-tegevusnimi*), seitsmel juhul kõneleja ja kuulaja koos (*peame, me ei peagi + ma-tegevusnimi, saame + da-tegevusnimi*), ühel juhul kõneleja ise (*ma pean + ma-tegevusnimi*).

Modaalverbi sisaldavad direktiivid pigem väljendavad juhiseid või ettepanekuid edasiseks tegevuseks kui nõuavad kohest tegutsemist. Need direktiivid jagunevad kahte rühma.

Osa direktiive esitatakse paralleelselt verbaalselt ja mitteverbaalselt, näiteks osutab isa samal ajal sõrmega või kääriotsaga paberil olevale kujundile, võtab liimi, paneb raamatu lapse ette lauale, et laps vaataks joonist.

Teised direktiivid esitatakse ainult verbaalseid vahendeid kasutades. Neist omakorda on osa üldisemad direktiivid, mis ei ole otseselt seotud ninasarviku tegemisega (nt *sa pead ikka vaatama kas seal teisel pool on vajalikud asjad*). Teise osa direktiivide puhul uurib isa enda käes olevat detaili või joonist ning annab juhiseid hilisemaks tegevuseks.

Isa algatatud direktiivisekvents jätkub järgmiselt.

Ühtedel juhtudel nõustub laps direktiiviga. Kui käsklause puhul oli eelistatud järelliikmeks vastuvõtja tegevus, siis modaallausete puhul direktiivi tavaliselt kohe ei täideta.

Näites (3) lepitakse kokku edasine tegevus. Isa küsib, kumb hakkab liimima. Laps soovib, et isa liimiks. Isa püüab last protsessi kaasata, tehes kõigepealt ettepaneku liimimist vaadata, ning seejärel paneb ette, et laps võiks mõnda detaili liimimisel kinni hoida. Laps nõustub, kuid isa algatab üleküsimisega paranduse: *sobib niimodi*. Laps viib paranduse läbi tõusva intonatsiooniga partikliga *jaa*.

(3) I: kas `sina (.) liimid või `mina liimin.

(0.4)

L: sina.

(1.0)

I: ja sina `vaatad kuidasmoodi ma `liimin.

(0.3)

L: jah.

(0.5)

I: **hhhhh mõne `koha pääle** ((I paneb paberitüki lauale, võtab liimi, hakkab pudelit lahti keerama)) **võid teha sellist asja et oiad oma: `väikse sõrmega teda** ((L tõstab korraks pilgu I käes olevale liimipudelile)) **`kinni.**

(0.5)

L: jah

(0.4)

I: `sobib niimodi. ((I paneb liimipudeli korgi lauale))

(.)

L: jaa?

Teiseks võib laps direktiivi täitmisest keelduda. Keeldumise kohta oli kaks näidet.

Näites (4) teeb laps ettepaneku liimima hakata. Isa nõustub sellega ja esitab lapsele suunatud direktiivi *s=sa pead `liimi tooma*. Pärast 1,3-sekundilist pausi algatab laps vahesekvensi küsimusega *aga ma=i=tea `kust*. Isa vastusele *kapist* järgneb jällegi pikk paus ning lapse parandusealgatus *`kõrgelt=või*. Isa täpsustab liimi asukohta. Järgmises vóorus keeldub laps direktiivi täitmast, esitades omakorda uue direktiivi: *too `sina*. Voor on vormistatud kui mitte-eelistatud: eelneb 2,4-sekundiline paus ning keeldumist/palvet põhjendatakse. Seletusi või põhjendusi kasutatakse ootusvastase, mitte-eelistatud või mittekooperatiivse tegevuse puhul sageli, et muuta see kaasvestleja jaoks aktsepteeritavaks (Heritage 1988), seda nii naabruspaari järelliikmete kui ka esiliikmete puhul (nt palvetes) (Antaki 1994, Schegloff 2007, Taleghani-Nikazm 2006). Mehhiko laste mittenõustuvaid/keelduvaid vastuseid uurinud Rosa Graciela Montes (2009) on leidnud, et lapsed kasutavad ühelt poolt eri tüüpi pehmeid vahendeid (edasilükkamine, seletu-

sed, põhjendused), kuid teiselt poolt tagasilükkamist või keeldumist tugevdavaid strateegiaid (otsene vastuvaidlemine, keelduja soovide või vajaduste rõhutamine). Eri strateegiate kasutamine oleneb aktiitüübist, millele reageeritakse, ning osalejate õigustest/kohustustest ja vanuse- ja staatussuhetest. Meie näites kasutas laps nii otsest keeldumist (*ei*) kui ka pehmendavat põhjendust. Analüüsitud vestluses on see ainus palve, mida põhjendatakse. Teisi käskivas kõneviisis direktiive ei pehmentata (sama on leidnud ka Metslang 2004), pehmentamiseks kasutatakse muid vahendeid nagu modaalsed konstruktsioonid, küsilause vorm jms.

Keeldumine kutsub omakorda esile reaktsiooni (vt Karrebæk 2010). Järgnevad isa algatatud vahesekventsid, isa aktsepteerib keeldumist ja toob ise liimi.

- (4) L: liimime `kohe.
(0.8)
I: liimime `kohe.
(.)
L: mhmh
(0.3)
I: **s=sa pead `liimi tooma.**
(1.3)
L: aga ma=i=tea `kust. ((L pöörab pilgu oma käes olevale tükile))
I: kapist.
(1.9)
L: `kõrgelt=või.
(0.5)
I: no säält=õõõõõ kus need `liimid on seal::: `nurgas ((L vaatab korraks I-le otsa)) on selline
(2.4) ((L paneb käed näo juurde, pöördub kõrvale))
L: ei too `sina see on natuke ((L pöörab pilgu oma tükile)) liiga `kõrgel.
(.)
I: se=on `kõrgel=vä. ((I võtab laualt käärid, hakkab oma tükki lõikama))
(0.5)
L: < mhmh >
(6.4) ((I lõikab, L vaatab hajevil enda ette, sügab nina alt))
I: mida `sina niikaua teed. ((I paneb käärid lauale))
(1.9) ((L vaatab oma tükki))
L: ootan.
(2.3)
I: sul `on ((L pöörab pilgu alla)) `tegevus sis, `ootamise `tegevus.
(0.3)
L: \$ jah. \$ ((L tõstab pilgu oma tükile))
(1.1)
I: teeme nii, (1.0) ma tulen `kohe tagasi. ((I tõuseb püsti, läheb teise tuppa))

Kolmandaks ei pruugi laps direktiivile reageerida. Järellikme said direktiivid, mis esitati nii verbaalseid kui mitteverbaalseid vahendeid kasutades: isa näitas lapsele konkreetselt, mida teha. Kui isa aga annab juhiseid hilisemaks tegevuseks joonist

või enda käes olevat detaili uurides, kusjuures agendile ei viidata või on agendiks kõneleja ja kuulaja mõlemad, ei tõlgenda laps isa voores esiliikmena, millele peaks reageerima. Isa omakorda ei osuta mingil viisil, et ta soovib järelliiget saada. Vähestes näidetes on eksplitsiitselt väljendatud, et tegevus ei toimu kohe (*ja `see tuleb `pärast < `siia saada. >; nii, `seda peab ka `kuidagimoodi `voltima pärast*). Paaris näites algatab laps paranduse, kuid pärast parandussekventsi lõppu direktiivile järelliiget ei anna.

Näites (5) uurib isa üht ninasarviku detaili, võrreldes seda joonisega. Laps lõikab välja teist detaili ning esitab küsimuse selle kohta.

(5) ((I paneb paberitüki käest kaugemale lauale, võtab teise tüki kätte. L lõikab enda käes olevat detaili.))

I: **nii, `seda peab ka `kuidagimoodi `voltima pärast.** ((I uurib käes olevat tükki ja joonist))

(2.3)

L: kas selle `musta sis lõikan `välja. ((L pöördub I poole, näitab käes olevat lehte, osutab kujundile kääriotsaga))

Neljas rühm on piirijuhtum, mis ei mahu traditsioonilisse naabruspaarisüsteemi. Naabruspaari definitsiooni kohaselt on esiliige ja järelliige eri kõnelejate esitatud (Schegloff, Sacks 1973: 295). Analüüsitava vestluses oli ka selliseid isa direktiive, mille ta ise täitis lapse nõusolekut ootamata (6). Iseloomulikud on direktiiviketid: samas voores järgneb üksteisele mitu direktiivi, mis saavad järgnevaid tegevusi.

(6) I: **siin on `niimodi=et** ((I paneb käärid lauale)) **(0.6) sin on vaja hakata teda natukene `voltima.**

(7.9) ((I voldib, L vaatab käsipõsakil tema tegevust))

I: nii, (0.3) ((I võtab kaugemalt laualt videokasseti karbi, paneb enda ette lauale)) võtame `selle, (1.1) voldime siin `peal.

(16.7) ((I voldib, L vaatab käsipõsakil tema tegevust))

Laps kasutas modaalselt konstruktsiooni (*tuleb + da*-tegevusnimi) üks kord, algatades sellega paranduse. Isa viis paranduse läbi ja selgitas edasist tegevust.

3.1.3. Modaalkonstruktsioonita väitlause

Mitmuse 1. isikus väitlauseid väljendavad ettepanekuid, juhatastes sisse uue tegevuse või järjestades järgmised tegevused. Isa vormistas sel viisil 11, laps 5 direktiivi. Niisuguste direktiivide adressaadiks on nii kõneleja kui kuulaja. Järelliikmeks on valdavalt nõustumine (*jah, kordus*), millele võib veel järgneda esiliikme ütleja kinnitus. Nõustumist näitab ka isa direktiivi täiendamine lapse poolt: laps väljendab arusaamist, et sel viisil tegutsemine on vajalik.

Kui ettepaneku ühiseks tegevuseks teeb isa, tegutseb ta pärast lapse nõusoleku saamist ise edasi. Näites (7) teeb isa ettepaneku sarve ja pea vahele liimi panna. Laps nõustub ettepanekuga ning isa kordab tema nõustuvat voores sõna-sõnalt. Laps lõikab samal ajal teist detaili välja, isa liimib ise sarve ja pea kokku.

(7) I: **paneme vahele `liimi `ka.**

(1.2)

L: siis ta `jääb.

(0.3)

I: siis ta `jääb,

(73.2) ((I urgitseb noaga pea kallal, liimib. L lõikab, ei vaata I tegevust))

On aga ka selliseid näiteid, kus direktiivile järelliiget ei anta ning järelliikme puudumisele ei osutata kui ootusvastasusele. Samuti on mõned näited, mille puhul isa täidab direktiivi ise ilma lapse nõustumist ootamata (vt näide (6): *võtame `selle, (1.1) voldime siin `peal.*).

Lapse ettepanekutele kohest tegevust ei järgne. Ühte direktiivi (vt näide (4): *liimime kohe*) pole võimalik kohe täita, enne tuleb teisest toast liim tuua. Teistes näidetes reastab laps edaspidised tegevused. Isa nõustub sellega või ei reageeri direktiivile.

Ainsuse 2. isikus väitlauseid väljendavad käske. Isa esitas niisuguses vormis kaks direktiivi. Üks neist on määratud koheseks täitmiseks. Laps nõustub direktiiviga, algatab paranduse ja pärast parandussekventsi lõppu täidab direktiivi.

Teises direktiivis (8) on eksplitsiitselt väljendatud, et see tuleb täita hiljem. Laps esitab küsimuse *miks mina*, eelkõneleja juttu vaidlustades ja küsija arvates ebaõiglusele selgitust nõudes (vt Laanesoo 2010: 54), sisuliselt keeldudes direktiivi täitmisest. Isa hakkab küsimusele vastama, kuid ei lõpeta voo. Laps esitab nüüd keeldumise põhjuse. Eemal istuv ema teeb ettepaneku kasutada PVA-liimi, toetades lapse arvamust, et tavaline paberiliim ei liimi piisavalt tugevalt. Isa võtab ema ettepaneku vastu ja seejärel on ka laps nõus direktiivi täitma (*liimime `kohe.*).

(8) ((I võtab laualt käärid, lõikab käes olevat tükki. L tõstab enda käes oleva tüki kõrgemale ja vaatab seda, siis paneb käed koos paberitükiga lõua juurde ja pöörab pilgu I tegevusele. I paneb käärid lauale.))

I: **`nii, (0.8) `pärast paned siia liimi `vahele,** ((L pöörab pilgu enda käes olevale tükile, mõlemad vaatavad oma paberitükki ja keerutavad seda käes))

(1.2)

L: `miks `mina. ((mõlemad vaatavad oma paberitükki ja keerutavad seda käes))

(1.0)

I: < sebärast=et > ((mõlemad vaatavad oma paberitükki ja keerutavad seda käes))

L: äkki see tuleb jälle varsti `lahti. ((mõlemad vaatavad oma paberitükki ja keerutavad seda käes))

(0.5)

E: äkki peaks ((L pöörab pilgu E-le)) seda Peve `aa liimiga tegema see paberil `liim plõksatab ((I pöörab pilgu E-le)) varsti `lahti ma arvan. ((L pöörab pilgu enda käes olevale tükile))

(0.3)

I: see. ((I pöörab pilgu oma paberitükile, tõstab selle korraks kõrgemale)) (5.8) ((mõlemad vaatavad oma paberitükki ja keerutavad seda käes))

- I: võime `Peve`aaga ka `teha. (2.5) `nii? ((L tõstab pilgu I tegevusele))
 (3.5) ((I keerutab oma tükki käes, L vaatab tema tegevust))
- L: liimime `kohe.
 (0.8)
- I: liimime `kohe.

3.1.4. Muid vormistusviise

Kolm isa direktiivi on vormistatud **üldküsilausena**. Ka need väljendavad ettepanekuid. Erinevalt käskivas kõneviisis direktiivist on küsimusena vormistatud direktiiv orienteeritud sellele, et vastuvõtja võib mitte tahta direktiivi täita. Kahe küsimuse agendiks on nii kõneleja kui kuulaja (*kas nüüd (.) paneme (0.7) `sarve kokku või.; nii, kas paneme siia liimi `vahele=vä.*), ühe küsimuse agendiks kuulaja (*taad midagi `kleepida.*). Nagu mitmuse 1. isikus väitlauseliste ettepanekutegi puhul nõustub laps ettepanekuga ning isa sooritab tegevuse ise.

Üks isa käsk (9) on vormistatud **verbita fraasina**. Ninasarviku üks osa on kokku liimitud ning isa annab lapsele kleepimiseks järgmise detaili, seda lapse poole sirutades ja öeldes `see ka. Laps alustab uut alateemat enda väljalõigatud kujundit otsides ning isa vastuse ajal täidab direktiivi, vajutab liimitud tüki kokku.

- (9) I: näed, (.) `ilusasti jäi `kinni ka. (0.8) ((keerutab tükki käes)) kohe ma annan sulle sele `teise poole `ka. ((võtab liimi, hakkab liimima)) (6.3) nii. ((pane liimi lauale)) (1.0) `see ka. ((sirutab tüki L poole)) (0.6)
- L: `kuhu=se `kaaren nüüd jäi. (0.4)
- I: ma=i=tea. ((L vajutab isa käes olevat tükki)) (0.8) pärast `otsime.

Vaadeldud vestluses oli kaks isa **pakkumist ainsuse 1. isikus**. Searle'i liigituse järgi ei kuulu need direktiivide hulka, kuid TÜ dialoogiaktide tüpoloogias (Hennoste, Rääbis 2004) on ka endast lähtuvad aktid, mis ootavad partneri nõustumist, arvatud direktiivide hulka. Analüüsitud näidetes ootab isa enne tegutsemist lapse nõusolekut (*tead mina `lähen ja `lõikan ned `noaga: ned `augud siia `sisse, (.) `sobib.*)

3.2. Direktiivid naabruspaari järelliikmena

Naabruspaari järelliikmena väljendavad direktiivid vaadeldavas vestluses nõustumist, lubamist, palvet, keeldu, juhust ja ettepanekut. Niisuguses positsioonis oli 11 isa ja 4 lapse direktiivi. Direktiivid on järelliikmeks küsimusele või ettepanekule. Järelliige on omakorda sageli uue naabruspaari esiliikmeks. Vormistusviisidest on kasutusel käskiv kõneviis (käsklused), modaalseid konstruktsioone sisaldavad väitlused, ainsuse 2. isikus ning mitmuse 1. isikus väitlused (vt tabel 2).

Tabel 2. Naabruspaari järeliikmena esinevate direktiivide vormistus

Vormistusviis	Isa	Laps	Kokku
Käsklause	4	1	5
Modaalset konstruktsiooni sisaldav väitlause	3	1	4
Väitlause mitmuse 1. isikus	2	2	4
Väitlause ainsuse 2. isikus	2	–	2
Kokku	11	4	15

Käskivas kõneviisis oli neli isa direktiivi ja üks lapse direktiiv. Nagu on leidnud Leelo Keevallik (2008), esineb käskiva kõneviisi vorm naabruspaari järeliikmes eelkõige küsimuse vastusena ning tema põhiline funktsioon on eelnevas vóorus ette pandud tegevusega nõustumine. Küsimuse vastusena kasutas isa käskivat kõneviisi ka minu analüüsitud vestluses.

Näites (10) küsib laps väljalõigatava detaili kohta selgitust. Metslang (1981: 30) käsitleb seda tüüpi küsimusi kaalutlusküsimustena, mis annavad vastajale õiguse tegutsemist lubada, soovitada, käskida või keelata. Isa kasutab vastuses käskivas kõneviisis verbi *lõika*, mis annab loa tegutseda, ja lisab täpsema selgituse. Laps üritab selgitust katkestada valjuhääle protestiga, et ta on juba aru saanud.

(10) L: kas selle `musta sis lõikan `välja. ((L pöördub I poole, näitab käes olevat lehte, osutab kujundile kääriotsaga))

(1.4) ((I vaatab L käes olevat lehte, paneb paberitüki käest ära, võtab käärid))

I: .hhhhhhh (0.5) **eee `süia** ((I võtab L käest lehe ära)) (.) **`lõika, vaata `näed, (.) süin on seline** ((I hakkab lõikama)) **(.) must `kriips, sis=se tähendab se[da]**

L: [JAA] MA=`TE[AN MA=`TEAN.]

I: [et et] `lõikad selle kuni (1.0)

((I osutab kääriotsaga kujundile)) kuni `siia, aga seda `välja ei `lõika. ((I veab kääriotsaga mööda kujundit)) (.) sebarast `siit lähed edasi näed siit selle ((I võtab paberi kääride vahele, aga ei lõika))

L: JAA MA `TEAN MA `TEAN MA SAIN `ARU. ((I annab paberi L-le))

I: selge.

(0.4)

L: MA SAIN JUBA `ENNEM aru.

Isa kasutab käskivat kõneviisi ka keelu vormistamisel (*et kas `sinna peab ka sellise `lõike tegema. / mt=.hh ära `praegu `tee.*).

Laps vormistab käskivas kõneviisis isa käsule reageeriva keelduva vóoru (vt näide (4): *ei too `sina see on natuke liiga `kõrgel.*). Vóor algab eitussõnaga, seejärel esitab laps palve, et isa liimi tooks, ja põhjendab oma keeldumist ja palvet. Lapse vóor on omakorda uue naabruspaari esiliige, mille järeliikmeks on isa tegevus (liimi toomine).

Modaalsed konstruksioonid (*tuleb + da-tegevusnimi, ei tohi + da-tegevusnimi, ei või*) on vastuseks tegevusettepanekule või küsimusele. Isa kasutas sel viisil vormistatud direktiive kolm korda, laps üks kord. Isa voor algab kõikides näidetes nõustuva partikliga *jah*, sellele järgneb ettepanekuga nõustumine seda korrates või ümber sõnastades ning seletus või põhjendus. Laps kasutas modaalsel konstruksiooni vastuses isa kaalutlusküsimusele.

Väitlauseid ainsuse 2. isikus on vastuseks küsimusele ja väljendavad juhi-seid (*sebärast `siit lähed edasi näed siit selle*).

Väitlauseid mitmuse 1. isikus on samuti küsimuse vastuseks või väljendavad ettepanekuga nõustumist. Nad on omakorda uue naabruspaari esiliikmed (ettepanekud), millele järgneb nõustumine või tegevuse läbiviimine (vt näide (4): *liimime `kohe. / liimime `kohe. / mhmh*).

4. Kokkuvõte

Direktiivide suur hulk analüüsitavas vestluses on tingitud sellest, et suhtlus on orienteeritud eesmärgi täitmisele (vt Vine 2009, Craven, Potter 2010). Isa uurib tööjuhendit ja juhendab last. Väga selgelt tuleb välja lapsevanema ja lapse ebasüm-meetria: isa esitab direktiive märgatavalt rohkem kui laps.

Direktiivid struktureerivad dialoogi: isa (harvem laps) annab korralduse/juhise järgmiseks tegevuseks või järjestab edasised tegevused. Vaadeldavas vestluses esinevad direktiivid nii naabruspaari esi- kui järelliikmena.

Naabruspaari esiliikmena väljendavad direktiivid käsku, palvet, keeldu, juhust, ettepanekut ja pakkumist. Keskised vormistusvõtted on modaalsed konstruksioonid ja käskiv kõneviis, vähem kasutatakse muid väitlauseid ja küsilauseid.

Erinevalt vormistatud esiliikmed kutsuvad esile eri tüüpi järelliikmeid. Käskivas kõneviisis direktiivid on orienteeritud täitmisele ning täidetakse mitteverbaalselt. Kui laps ei täida direktiivi isa soovitud viisil, kordab isa direktiivi tugevamal kujul, kuni saavutab eesmärgi. Imperatiivi vormis direktiiv näitab, et kõneleja on õigus-tatud kuulaja tegevust suunama.

Modaalverbi sisaldavad direktiivid pigem väljendavad juhiseid või ettepanekuid edasiseks tegevuseks kui nõuavad kohest tegutsemist. Vähemalt formaalselt on nad orienteeritud vastuvõtja võimalusele või soovile direktiivi täita. Järelliikmeks on enamasti nõustumine, direktiiv täidetakse kohe vaid vähestes näidetes. Muude väitlauseliste ning küsilauseliste direktiivide tüüpiliseks järelliikmeks on samuti nõustumine.

Norm, et eelistatud järelliige on tegevus või nõustumine, ei määra aga tegelikku vastust. Vestluses on ka niisuguseid direktiive, mida laps keeldub täitmast. Isa aktsepteerib keeldumist ning täidab direktiivi ise või juhendab last selle täitmisel.

Direktiivid on tihedalt seotud kaasneva tegevusega. Kui modaalse konstruksiooni abil moodustatud või *meie*-vormis direktiiv esitatakse paralleelselt verbaalselt ja mitteverbaalselt (nt isa osutab joonisele või ulatab detaili lapse poole), tõlgendab laps seda kui esiliiget ning annab järelliikme. Kui direktiiviga mitteverbaalselt vest-luskaaslasel suunatud tegevust ei kaasne, siis sellele järelliiget ei anta ning esimene kõneleja ei osuta järelliikme puudumisele kui ootusvastasusele.

Traditsioonilises naabruspaaride käsitluses on esiliige ja järelliige eri kõnelejate esitatud. Analüüsitavas vestluses on ka niisuguseid näiteid, mille puhul isa täidab direktiivi ise lapse nõusolekut ootamata.

Naabruspaari järelliikmena väljendavad direktiivid vaadeldavas vestluses nõustumist, lubamist, palvet, keeldu, juhust ja ettepanekut. Direktiivid on järelliikmeks küsimusele või ettepanekule. Järelliige on omakorda sageli uue naabruspaari esiliikmeks. Vormistusviisidest on kasutusel käskiv kõneviis, modaalsed konstruktsioonid, väitlused ainsuse 2. ning mitmuse 1. isikus.

Analüüs näitab, et direktiive on oluline süstemaatiliselt käsitleda koos kaasneva tegevusega. Edaspidi tuleks uurida direktiivisekventse nii vabas argivestluses kui ka muudes eesmärgile suunatud situatsioonides; nii lapsega suhtlemisel kui ka niisugustes vestlustes, milles ei kajastu lapse ja lapsevanema ebasümmeetria.

Transkriptsioonimärgid ja lühendid

.	langev intonatsioon
,	poollangev intonatsioon
?	tõusev intonatsioon
(.)	mikropaus (0.2 sek või lühem)
(0.5)	pausi pikkus sekundites
`sõna	rõhutamine
[]	pealerääkimine
=	kahe iseseisva üksuse kokkuhääldamine
SÕNA	valjem lõik
< >	aeglustatud lõik
\$ \$	naerev hääl
e::i	hääliku venitus
si-	poolelijäänud sõna
.hh	sissehingamine
mt	matsutamine
(())	transkribeerija kommentaar
I	isa
L	laps
E	ema

Viidatud kirjandus

- Andersen, Elaine 1990. *Speaking with Style. The Sociolinguistic Skills of Children*. London: Routledge.
- Antaki, Charles 1994. *Explaining and Arguing: Social Organization of Account*. London, Thousand Oaks, New Delhi: Sage Publications.
- Aronsson, Karin; Cekaite Asta 2007. Getting things done in family life – Directive trajectories and moral order. – 10th International Pragmatics Conference, Göteborg, 8–13 July 2007, Abstracts, 140.
- Aronsson, Karin; Thorell, Mia 1999. Family politics in children's play directives. – *Journal of Pragmatics*, 31 (1), 25–47. [http://dx.doi.org/10.1016/S0378-2166\(98\)00050-2](http://dx.doi.org/10.1016/S0378-2166(98)00050-2)
- Blum-Kulka, Shoshana 1987. Indirectness and politeness in requests: same or different? – *Journal of Pragmatics*, 11 (2), 134–146. [http://dx.doi.org/10.1016/0378-2166\(87\)90192-5](http://dx.doi.org/10.1016/0378-2166(87)90192-5)
- Brown, Penelope; Levinson, Stephen 1987. *Politeness. Some Universals in Language Usage*. Cambridge: Cambridge University Press.

- Couper-Kuhlen, Elizabeth; Selting, Margret 1996. Towards an interactional perspective on prosody and a prosodic perspective on interaction. – M. Selting, E. Couper-Kuhlen (Eds.). *Prosody in Conversation*. Cambridge: Cambridge University Press, 11–56.
- Couper-Kuhlen, Elizabeth; Selting, Margret 2001. Introducing interactional linguistics. – M. Selting, E. Couper-Kuhlen (Eds.). *Studies in Interactional Linguistics*. Amsterdam, Philadelphia: John Benjamins, 1–22.
- Craven, Alexandra 2010. Children's responses to directives: embodied compliance and verbal resistance. – International Conference on Conversation Analysis ICCA10, Mannheim, 4–8 July 2010, Abstracts, 58–59.
- Craven, Alexandra; Potter, Jonathan 2010. Directives: Entitlement and contingency in action. – *Discourse Studies*, 12 (4), 419–442. <http://dx.doi.org/10.1177/1461445610370126>
- Curl, Traci; Drew, Paul 2008. Contingency and action: A comparison of two forms of requesting. – *Research on Language and Social Interaction*, 41 (2), 129–153. <http://dx.doi.org/10.1080/08351810802028613>
- Drew, Paul 1984. Speakers' reportings in invitation sequences. – J. M. Atkinson, J. Heritage (Eds.). *Structures of Social Action: Studies in Conversation Analysis*. Cambridge: Cambridge University Press, 129–151.
- Ervin-Tripp, Susan M. 1976. Is Sybil there? The structure of some American English directives. – *Language in Society*, 5 (1), 25–66. <http://dx.doi.org/10.1017/S0047404500006849>
- Fukushima, Saeko 1996. Request strategies in British English and Japanese. – *Language Sciences*, 18 (3-4), 671–688. [http://dx.doi.org/10.1016/S0388-0001\(96\)00041-1](http://dx.doi.org/10.1016/S0388-0001(96)00041-1)
- Gerassimenko, Olga; Hennoste, Tiit; Kasterpalu, Riina; Koit, Mare; Rääbis, Andriela; Strandson, Krista; Valdisoo, Maret; Vutt, Evely 2007. Kliendi soovide automaatne tuvastamine eestikeelsetes infodialoogides. – *Eesti Rakenduslingvistika Ühingu aastaraamat*, 3, 135–154.
- Gerassimenko, Olga; Kasterpalu, Riina; Koit, Mare; Rääbis, Andriela; Strandson, Krista 2010. Direktiivsed aktipaarid eestikeelsetes infodialoogides ja nende automaatne tuvastamine. – *Eesti Rakenduslingvistika Ühingu aastaraamat*, 6, 67–86. <http://dx.doi.org/10.5128/ERYa6.05>
- Gleason, Jean B. 1987. Sex differences in parent-child interaction. – S. U. Philips, S. Steele, C. Tanz (Eds.). *Language, Gender, and Sex in Comparative Perspective*. Cambridge: Cambridge University Press, 189–199.
- Goffman, Erving 1981 [1979]. *Footing*. – *Forms of Talk*. Oxford: Basil Blackwell, 124–159.
- Goodwin, Marjorie Harness 1990. *He-Said-She-Said: Talk as Social Organization among Black Children*. Bloomington: Indiana University Press.
- Goodwin, Marjorie Harness 2006. Participation, affect, and trajectory in family directive/response sequences. – *Text and Talk*, 26 (4-5), 515–543. <http://dx.doi.org/10.1515/TEXT.2006.021>
- Hakulinen, Auli; Selting, Margret (Eds.) 2005. *Syntax and Lexis in Conversation: Studies on the Use of Linguistic Resources in Talk-in-Interaction*. Amsterdam: John Benjamins.
- Heinemann, Trine 2006. Will you or can't you?: Displaying entitlement in interrogative requests. – *Journal of Pragmatics*, 38 (7), 1081–1104. <http://dx.doi.org/10.1016/j.pragma.2005.09.013>
- Hennoste, Tiit; Koit, Mare; Rääbis, Andriela; Strandson, Krista; Valdisoo, Maret; Vutt, Evely 2003. Directives in Estonian information dialogues. – V. Matoušek, P. Mautner (Eds.). *Text, Speech and Dialogue. Proceedings of the 6th International Conference, TSD 2003, České Budějovice, Czech Republic, 8.–12.09.2003*. Berlin: Springer, 406–411.
- Hennoste, Tiit; Lindström, Liina; Rääbis, Andriela; Strandson, Krista; Vellerind, Riina 2001. vä ja teised. Ühe vestluse pisianalüüs. – R. Kasik (Toim.). *Keele kannul. Pühendusteos Mati Ereli 60. sünnipäevaks*. Tartu Ülikooli eesti keele õppetooli toimetised, 17. Tartu: Tartu Ülikooli Kirjastus, 90–112.

- Hennoste, Tiit; Rääbis, Andriela 2004. Dialogiaktid eesti infodialoogides: tüpologia ja analüüs. Tartu: Tartu Ülikooli Kirjastus.
- Heritage, John 1988. Explanations as accounts: a conversation analytic perspective. – C. Antaki (Ed.). *Analyzing Everyday Explanations: a Casebook of Methods*. London: Sage, 127–144.
- Huls, Erica; van de Peppel, Anne-Lies; van Wijk, Carel 2007. The development of a directive repertoire. – 10th International Pragmatics Conference, Göteborg, 8–13 July 2007, Abstracts, 21–22.
- Karrebæk, Martha 2010. Refusal sequences in kindergarten interaction. – International Conference on Conversation Analysis ICCA10, Mannheim, 4–8 July 2010, Abstracts, 246–247.
- Kasterpalu, Riina; Gerassimenko, Olga 2006. Vestlusanalüüs. – I. Tragel, H. Õim (Toim.). *Teoreetiline keeleteadus Eestis 2*. Tartu Ülikooli üldkeeleteaduse õppetooli toimetised, 7. Tartu: Tartu Ülikooli Kirjastus, 112–126.
- Keevallik, Leelo 2002. Grammatika suhtluses. – R. Pajusalu, I. Tragel, T. Hennoste, H. Õim (Toim.). *Teoreetiline keeleteadus Eestis*. Tartu Ülikooli üldkeeleteaduse õppetooli toimetised, 4. Tartu: Tartu Ülikooli Kirjastus, 89–104.
- Keevallik, Leelo 2003. From Interaction to Grammar: Estonian Finite Verb Forms in Conversation. *Acta Universitatis Upsaliensis. Studia Uralica Upsaliensia*, 34. Uppsala.
- Keevallik, Leelo 2005. Suhtluskeeel uurimine partiklistunud verbivormide näitel. – *Keel ja Kirjandus*, 7, 8, 535–548, 630–642.
- Keevallik, Leelo 2008. Käskiv kõneviis nõustavas voores ja vastuste tüpologia. – *Emakeele Seltsi aastaraamat*, 54 (2007), 94–106.
- Laanesoo, Kirsi 2010. Ümberpööratud polaarsusega küsimused eesti argivestluses. *Magistritöö*. Käsikiri TÜ eesti ja üldkeeleteaduse instituudis.
- Law, Chung Wa 2009. The use of request strategies in Cantonese school aged children. – 11th International Pragmatics Conference, Melbourne, 12–17 July 2009, Abstracts, 270.
- Leung, Cheung-Shing Sam; Wong, Lornita 2009. Use of request strategies in Cantonese-speaking preschool children. – 11th International Pragmatics Conference, Melbourne, 12–17 July 2009, Abstracts, 272.
- Liddicoat, Anthony J. 2007. *An Introduction to Conversation Analysis*. New York: Continuum.
- Malahhov, Mariliis 2011. Rollide võtmine ja pakkumine 3–4-aastaste eesti laste rollimängus. *Magistritöö*. Käsikiri TÜ eesti ja üldkeeleteaduse instituudis.
- Márquez Reiter, Rosina 2000. *Linguistic Politeness in Britain and Uruguay: a Contrastive Study of Requests and Apologies*. Amsterdam, Philadelphia: John Benjamins.
- Metslang, Helle 1981. *Küsilause eesti keeles*. Eesti NSV Teaduste Akadeemia Keele ja Kirjanduse Instituut. Tallinn: Valgus.
- Metslang, Helle 2004. Imperative and related matters in everyday Estonian. – *Linguistica Uralica*, 4, 243–256.
- Mihkels, Krista 2011. Õpetaja algatatud parandussekventsids eesti algkoolitundides: kõne ja mittevokaalse suhtlusmodaalsuse kasutamine. Käsikiri autori valduses.
- Montes, Rosa Graciela 2009. Rejections and refusals: An analysis of children's dispreferred responses. – 11th International Pragmatics Conference, Melbourne, 12–17 July 2009, Abstracts, 127–128.
- Ochs, Elinor; Schegloff, Emanuel A.; Thompson, Sandra A. (Eds.) 1996. *Interaction and Grammar*. Cambridge: Cambridge University Press.
- Pajusalu, Karl; Pajusalu, Renate 2004. The conditional in everyday Estonian: It's form and functions. – *Linguistica Uralica*, 4, 257–269.
- Pomerantz, Anita 1984. Agreeing and disagreeing with assessments: some features of preferred/dispreferred turn shapes. – J. M. Atkinson, J. Heritage (Eds.). *Structures of Social Action: Studies in Conversation Analysis*. Cambridge: Cambridge University Press, 57–101.

- Sacks, Harvey 1987 [1973]. On the preferences for agreement and contiguity in sequences in conversation. – G. Button, J. R. E. Lee (Eds.). *Talk and Social Organisation*. Clevedon: Multilingual Matters, 34–69.
- Schegloff, Emanuel A. 1987. Between micro and macro: contexts and other connections. – J. C. Alexander, B. Giesen, R. Münch, N. J. Smelser (Eds.). *The Micro-Macro Link*. Berkeley, Los Angeles, London: University of California Press, 207–234.
- Schegloff, Emanuel A. 1990. On the organization of sequences as a source of “coherence” in talk-in-interaction. – B. Dorval (Ed.). *Conversational Organization and its Development*. Norwood, New Jersey: Ablex Publishing Corporation, 51–77.
- Schegloff, Emanuel A. 2007. *Sequence Organization in Interaction. A Primer in Conversation Analysis*. Vol. 1. New York: Cambridge University Press. <http://dx.doi.org/10.1017/CBO9780511791208>
- Schegloff, Emanuel A.; Sacks, Harvey 1973. Opening up closings. – *Semiotica* 8 (4), 289–327. <http://dx.doi.org/10.1515/semi.1973.8.4.289>
- Searle, John R. 1976. A classification of illocutionary acts. – *Language in Society*, 5 (1), 1–23. <http://dx.doi.org/10.1017/S0047404500006837>
- Searle, John R. 1979. *Expression and Meaning: Studies in the Theory of Speech Acts*. Cambridge: Cambridge University Press.
- Sidnell, Jack 2010. *Conversation Analysis. An Introduction*. Sussex: Wiley & Sons.
- Taleghani-Nikazm, Carmen 2006. Request Sequences: The Intersection of Grammar, Interaction, and Social Context. Amsterdam: Benjamins.
- Teemets, Helena 2010. Ettepanekute ja pakkumiste väljendamine eesti argidialoogides. *Bakalaureusetöö. Käsikiri TÜ eesti ja üldkeeleteaduse instituudis*.
- Tulviste, Tiia 2001. Can variation in mother-child interaction be explained by context and collectivistic attitudes? – *Applied Psycholinguistics* 22 (4), 541–553. <http://dx.doi.org/10.1017/S0142716401004039>
- Vine, Bernadette 2009. Directives at work: Exploring the contextual complexity of workplace directives. – *Journal of Pragmatics* 41 (7), 1395–1405. <http://dx.doi.org/10.1016/j.pragma.2009.03.001>
- Wingard, Leah 2006. Parents' inquiries about homework: The first mention. – *Text and Talk* 26 (4-5), 573–596. <http://dx.doi.org/10.1515/TEXT.2006.023>
- Wootton, Anthony J. 1981. The management of grantings and rejections by parents in request sequences. – *Semiotica* 37 (1-2), 59–89. <http://dx.doi.org/10.1515/semi.1981.37.1-2.59>

Võrgumaterjalid

Suulise eesti keele korpus. <http://www.cl.ut.ee/suuline1/suulisekorpus/> (15.12.2011).

Andriela Rääbise (Tartu Ülikool) uurimisvaldkonnad on suuline kõne, telefonisuhtlus, dialoogide struktuur.
andriela.raabis@ut.ee

DIRECTIVE SEQUENCES IN A CONVERSATION OF FATHER AND DAUGHTER: A CASE STUDY

Andriela Rääbis

University of Tartu

This article studies directive sequences in the conversation of a father and his five-year-old daughter, using the methodology of interactional linguistics and concepts of conversation analysis. The analysis is based on a video recording of a conversation (42 minutes), during which paper rhinoceros is made. The father produced 92 directives and the child 15. The study shows that directives are bound up with concurrent action: verbal as well as non-verbal resources are implemented to construct an activity. Directives function both as first and second parts of adjacency pairs. Directives that occur as first pair parts express orders, requests, prohibitions, instructions, proposals and offers. The main devices for the expression of directives as first pair parts are modal constructions and the imperative. Differently formulated directives elicit various types of reactions. Directives expressed by the imperative typically elicit an activity. Directives constructed using other means typically make verbal compliance relevant as the next action. Directives as second pair parts express compliance, permissions, requests, prohibitions, instructions and proposals. They occur as responses to questions or proposals.

Keywords: interactional linguistics, conversation analysis, child language, adjacency pair, directive, Estonian

KAS FORMANDID PEEGELDAVAD EMOTSIOONE?

Kairi Tamuri

Ülevaade. Formante peetakse üheks emotsiooni peegeldavaks akustiliseks parameetriks – kõneleja emotsionaalne seisund võib põhjustada muutusi häälikute formantstruktuuris. Uurimuse eesmärgiks on teada saada, kas ja kuidas mõjutab lausungi emotsioon eestikeelse etteloetud emotsionaalse kõne lühikesi vokaale ning ettelugeja artikulatsiooni täpsust. Selleks mõõdeti etteloetud emotsionaalse ning neutraalse kõne lühikeste rõhuliste vokaalide *a*, *i*, *u* esimest ja teist formanti. Uurimistulemustest selgus, et emotsioonipaariti (k.a neutraalne kõne) on statistiliselt olulised erinevused olemas vokaalide *a* ja *i* esimeste formantide (F1) keskmistes. Teine formant (F2) emotsioonide eristamisel olulist rolli ei mängi. Vokaalide redutseerumist ja artikulatsiooni täpsust uuriti eukleidiilise kauguse arvutamise abil. Tulemused näitasid, et viha- ja rõõmu-lausungites jääb kõneleja artikulatsiooni täpsus võrreldes neutraalse kõnega enam-vähem samaks, kurbuse-lausungites see aga langeb.*

Võtmesõnad: emotsionaalne kõne, akustika, formantanalüüs, artikulatsiooni täpsus, eesti keel

1. Sissejuhatus

Paljud emotsioonisünteesi ja -tuvastusega seotud uurimused on näidanud, et igal emotsioonil on just talle omased akustilised tunnused. Kuulajad suudavad emotsioonid kõnes ära tunda ainuüksi heli põhjal, ilma kõnelejat nägemata (nt telefonivestlustes) (Bachorowski 1999). Kõnelõigu emotsiooni ollakse võimelised määrama isegi siis, kui kõik selle lõigu sõnad on tähenduseta (Goudbeek jt 2009). See, et kuulajad suudavad ainult hääle järgi emotsioone usaldusväärsetl ära tunda, toobki kaasa oletuse, et häälel kannab kogu informatsiooni kõneleja emotsionaalse seisundi kohta¹ ja et emotsioonide häälelised väljendused on erineva akustilise mustri (nt Paeschke jt 1999, Banse, Scherer 1996).

* Artikkel on valminud tänu projektidele EKT1 ja SF0050023s09.

¹ Seda vaid siis, kui kuulaja ei näe kõnelejat, sest vastasel juhul muutub akustiliste tunnuste domineerimine emotsiooni äratundmisel küsitavaks.

1.1. Formandid

Üheks akustiliseks parameetriks, mida emotsionaalse kõne puhul uuritakse, on formandid. Formantide abil saab teha järeldusi selle kohta, kuidas emotsioonid üksteisest akustiliselt erinevad ning kuidas mõjutavad emotsioonid häälikute kvaliteeti ja kõneleja artikulatsiooni täpsust.

Formandid on akustilise energia kontsentratsiooni piirkonnad kõnespektris. Kõikidel helilistel häälikutel on oma formantstruktuur ning nende foneetilist kvaliteeti tajutakse harilikult esimesest kolmest formandist (F1, F2 ja F3).

Formantide asukohad ei ole fikseeritud, vaid sõltuvad kõnetrakti suurusest ja kujust. Kõnetrakt muutub, kui varieerub häälduselundite (nt keel, lõug ja huuled) asend. Samuti muudavad kõnetrakti suurust ja kuju emotsioonid. Kui kõneleja emotsionaalne seisund muutub, siis muutuvad tema näoilme ning häälduselundite asend ja lihaspinge (Ververidis, Kotropoulos 2006, Johnstone, Scherer 2000). Naeratades öelduna kõlab lausung akustiliselt üsnagi erinevalt sellest, kui öelda sama asja kulmu kortsutades (Tartter 1980). See omakorda tekitab kõnesignaalil akustilisi muutusi, mis peegelduvad formantmuutrites.

Uurimistulemused on näidanud, et emotsioonid mõjutavad kõige rohkem F1 ja F2 väärtusi (nt Tolkmitt, Scherer 1986).

1.2. Artikulatsiooni täpsus ja redutseerunud häälikud

Põhjus, miks kuulatakse foneetilisi erinevusi kahe vokaali vahel, peitub selles, et vokaalidel on erinev asukoht kahemõõtmelises vokaaliruumis, mida iseloomustavad vokaali kõrgus ehk vertikaalne dimensioon ja vokaali eespoolsus-tagapoolsus ehk horisontaalne dimensioon. Need kaks mõõdet on seotud keele asukohaga suus vokaali moodustamisel. Jonathan Harrington (2010: 46) kirjeldab vokaale kahemõõtmelise jaotusega F1 x F2. Vokaali kõrguse kasv seostub F1 väärtuse vähenemisega, vokaali tagapoolsuse suurenemine seostub aga F2 kahanemisega. Seega: vokaali *i* F1 < vokaali *e* F1-st ja vokaali *a* F2 < vokaali *ä* F2-st.

Kui vokaaliruum kitseneb, siis kõneleja artikulatsiooni täpsus väheneb. Vokaalid, mida hääldatakse vähem selgelt, kipuvad redutseeruma, s.t nad kaotavad oma kvaliteeti. Kui ilmneb kõrvalekalle vokaali n-ö normaalsest positsioonist akustilises ruumis, siis nimetatakse seda vokaali tsentraliseerumiseks (ingl *centralization*). Tsentraliseerumise puhul moodustub vokaal kõneleja vokaaliruumi keskpunkti lähedal ja sarnaneb švaa-häälikuga, millel pole selget kvaliteeti. (Harrington 2010: 190–193)

Uurimistulemustest (Tolkmitt, Scherer 1986) on selgunud, et näiteks stressis või depressioonis inimene ei artikuleeri helilisi häälikuid sama jõupingutusega nagu neutraalse kõne puhul. Sellisel juhul häälikud redutseeruvad ning artikulatsiooni täpsus langeb. Harrington (2010: 190–193) kirjutab, et selge kõne puhul asetsevad vokaalid teineteisest kaugemal ning kuulajal on erinevate häälikute vahel kergem vahet teha. Vokaaliruum muutub suuremaks ning artikulatsiooni täpsus tõuseb.

2. Uurimisküsimus ja -materjal

Formantanalüüsi eesmärk on leida vastused kahele küsimusele: 1) kas eestikeelse etteloetud emotsionaalse kõne lühikesi vokaale mõjutab lausungi emotsioon ning 2) kas emotsioonid mõjutavad ettelugeja artikulatsiooni täpsust.

Uurimistöö akustiline baas on Eesti Keele Instituudi eesti emotsionaalse kõne korpus ehk EEKK². Korpus toetub seisukohale, et emotsioonid on hääle põhjal piisavalt hästi ära tuntavad ka loomulikus, mittenäideldud kõnes, ja et mittenäideldud kõne on loomuliku kõnesünteesi eeldus (Iida jt 2003). Korpuse loomisel on lähtutud printsüübist, et lõigu sisu kutsub lugejas esile mingi emotsiooni. Seetõttu pole lugejale ette öeldud, mis emotsiooniga teksti lugeda. Korpus sisaldab ühe naishääle etteloetud ajakirjanduslikke tekstilõike, mis on segmenteeritud lauseteks, sõnadeks ja häälikuteks. Korpuses sisalduvate lausete emotsioonid (viha, rõõm ja kurbus) või neutraalsus on määratud tajukatsetega (vt Altrov 2008).

Uurimismaterjalis on EEKK laused, mille emotsiooni või neutraalsust on tajunud üle 50% kuulajaist. Formantanalüüsiks valiti neist lausetest välja kolm lühikest vokaali: *a*, *i* ja *u* (vokaale on kokku 2395, täpsemat jaotust vt tabelist 1). Need on eesti keeles sagedasti esinevad vokaalid, mis moodustavad kolmnurkse vokaaliruumi (*a* on madal tagavokaal, *i* kõrge eesvokaal, *u* kõrge tagavokaal), mille abil saab hästi kirjeldada kõneleja artikulatsiooni täpsust. Vokaalide juures oli oluline ka nende rõhuline positsioon. Selle tingis soov leida vokaale, mis oleksid võimalikult vähe redutseerunud, sest on teada, et mitterõhuliste silpide häälikud kipuvad oma kvaliteeti kaotama (vt nt Moon, Lindblom 1994). Kuna eesti keeles asub rõhk sõnades tavaliselt esimesel silbil, siis pärinevad väljavalitud vokaalid sõnaalgulistest positsioonidest, mille struktuur on VC, CVC või CCVC³.

Tabel 1. Formantanalüüsi materjal

Vokaal	Emotsioon	Vokaalide arv
<i>a</i>	viha	395
	rõõm	279
	kurbus	244
	neutraalne	319
	KOKKU	1237
<i>i</i>	viha	220
	rõõm	146
	kurbus	154
	neutraalne	145
	KOKKU	665
<i>u</i>	viha	138
	rõõm	133
	kurbus	115
	neutraalne	107
	KOKKU	493
<i>a, i, u</i>	KOKKU	2395

² Vt <http://peeter.eki.ee:5000/> (28.09.2011).

³ V – vokaal; C – konsonant.

3. Meetod

Uurimisküsimustele vastuse leidmiseks mõõdeti esmalt ettelõetud emotsionaalse ning neutraalse kõne lühikeste rõhuliste vokaalide a , i ja u esimest ja teist formanti. Harrington (2010: 172–183) kirjutab, et otsus vokaali kvaliteedi kohta tehakse akustiliste andmete põhjal mingil kindlal ajahetkel ja harilikult on selleks koht, mis ühtib vokaali kestuse keskpunktiga (ingl *vowel target*). Suuresti valitakse vokaali keskpunkt formantide mõõtmiseks seepärast, et naaberhäälikute kontekstiline mõju on siis mõõdetavale vokaalile kõige väiksem ning selles punktis on vokaal kõige stabiilsem, s.t formantväärtused muutuvad kõige vähem.

Vokaalide redutseerumist ja artikulatsiooni täpsust mõõdeti eukleidilise kaugusega⁴, et näha, kas vokaalid on võrreldes neutraalse kõnega tsentraliseerunud või mitte ning kas vokaaliruum on võrreldes neutraalse kõnega emotsiooniti suurenenud või vähenenud. (Vt Harrington 2010: 191–195) Et eukleidilist kaugust mõõta, määrati esmalt kõneleja vokaaliruumi keskpunkti (ingl *centroid*). Selleks arvatati välja neutraalne häälik x , võttes arvesse kõigi korpuses leiduvate ja uurimismaterjali tingimustele vastavate sõnaalguliste struktuuride VC, CVC või CCVC lühikeste vokaalide keskmised.

Akustiliste parameetrite mõõtmisel kasutati programmi Praat⁵ ning kõne andmebaaside süsteemi EMU⁶ abi. Formantväärtuste automaatse arvutamise vead korrigeeriti käsitsi ning pärast seda analüüsiti mõõtmistulemusi statistikakeskkonnas R⁷. R-i abil viidi läbi ANOVA-analüüs⁸.

Klaus R. Scherer (1986) on välja töötanud emotsiooniakustika valdkonnas mudeli, mis ennustab emotsiooni mõju häälelisele väljendusele. Schereri *component process model* e CPM võtab arvesse psühholoogilisi ja füsioloogilisi mõjusid, mis kaasnevad emotsiooni väljendamisega, ning näitab, et on olemas emotsioonispetsiifilised akustilised muustrid. Scherer kirjeldab emotsiooni kui kohanemismuutuste (ingl *adaptive changes*) sarja, mis on vastastikusel seoses. Emotsiooni tekkimisest impulsi saanud närvisüsteem mõjutab nii kõneleja hingamist kui ka tema kõneorgani lihaspinget ning see toob kaasa erinevusi kõnesignaali akustikas. Nt midagi väga ebameeldivat avaldub tihti neelu ja kõri pitsituses, mille tagajärjel läheb kõnetrakt pingesse ning väljatulev hääli on kõrgema sagedusega (Thompson, Balkwill 2006).

CPM-mudel ennustab muu hulgas, kuidas muutuvad emotsioonide mõjul vokaalide F1 ja F2 väärtused võrreldes neutraalse kõnega. Võrdlus antud töö uurimistulemuste ja CPM-mudeli ning varasemate uurimuste vahel esitatakse 6. peatükis.

4. Emotsioonidega kaasnevad muutused formantväärtustes

Senised uurimused emotsioonide formantväärtuste kohta pole andnud päris ühesid tulemusi. Põhjus võib peituda erinevas uurimismaterjalis (näideldud vs. esile kutsutud emotsiooniga vs. spontaanne kõne; heli vs. heli + pilt), erinevas uuritavas

⁴ Eukleidilise kauguse valem: $d(x, V_{em}) = \sqrt{(F1x - F1V_{em})^2 + (F2x - F2V_{em})^2}$, kus x on neutraalne vokaal ja V_{em} vaadeldava emotsiooni vokaali keskmine.

⁵ <http://www.praat.org/> (28.09.2011).

⁶ <http://emu.sourceforge.net/> (28.09.2011).

⁷ <http://www.r-project.org/> (28.09.2011).

⁸ Tabeli genereerimiseks kasutati R'i skripti (vt http://peeter.eki.ee:5000/anova_aiu.R), lähteandmeid vt http://peeter.eki.ee:5000/vokaalid_aiu.csv (28.09.2011).

keeles ja kultuuris ning erinevas emotsioonidele antavas sisus (nt viha all võib mõista nii raevu kui ka lihtsalt ärritust, rõõmu all rahulolu või vaimustust, kurbuse all pettumust või muret). (Burkhardt jt 2006, McIntyre, Roland 2006, Wilting jt 2006, Douglas-Cowie jt 2003, Scherer 2003) Samuti erineb viis, kuidas emotsioone kirjeldatakse. Domineerivad kaks lähenemist – kategoriaalne (ingl *category approach*) ja dimensiooniline (ingl *dimensional approach*). Esimene lähenemine määratleb emotsioonid kategooriatena, nt kurbus, viha ja rõõm. Teine lähenemine kasutab emotsioonide kirjeldamiseks dimensioone. Kaks enam kasutatavat dimensiooni on valents (positiivne vs. negatiivne) (ingl *valence*) ja virgumine (madal vs. kõrge) (ingl *arousal*). (McIntyre, Roland 2006, Cowie, Cornelius 2003) Kui panna emotsioonid kahemõõtmelisele teljestikule, siis on rõõm kõrge virgumise ja positiivse valentsiga emotsioon, viha kõrge virgumise ja negatiivse valentsiga emotsioon ning kurbus madala virgumise ja negatiivse valentsiga emotsioon.

Varasemad uurimused prantsus- (Goudbeek jt 2009), inglisis- (Murray, Arnott 2008, Tartter 1980), saksa- (Kienast, Sendlmeier 2000, Paeschke jt 1999) ja jaapanikeelse (Iida jt 2003) emotsionaalse kõne kohta on näidanud, et esinevad mingid üldised tendentsid teatud emotsioonide ja nende akustiliste korrelaatide vahel.

4.1. Viha ja selle võrdlus teiste emotsioonidega

Võrreldes teiste põhiemotsioonide ja neutraalse kõnega on viha puhul täheldatud vokaalide kõrgemat F1 sagedust (Goudbeek jt 2009, Murray, Arnott 2008, Juslin, Laukka 2003) ning vokaalil *a* oluliselt madalamat F2 sagedust (Goudbeek jt 2009). Samuti on viha-emotsiooniga vokaalide juures märgitud nende kõrgemat artikulatsiooni täpsust (Murray, Arnott 2008, Juslin, Laukka 2003, Kienast, Sendlmeier 2000, Paeschke jt 1999).

Viha kohta on saadud ka teistsuguseid tulemusi. Iida jt (2003) uurimusest selgus, et kõrgema helikõrguse ja kiirema tempoga emotsioonidel (nagu viha) on vokaaliruum kitsenenud ning vokaalid rohkem redutseerunud.

4.2. Rõõm ja selle võrdlus teiste emotsioonidega

Rõõmu-lausungite vokaalidel on võrreldes teiste põhiemotsioonide ja neutraalse kõnega täheldatud F1 ja F2 väärtuste tõusu, kusjuures F1 on enamikul vokaalidel rohkem mõjutatud kui F2 (Juslin, Laukka 2003, Kienast, Sendlmeier 2000, Tartter 1980).

Martijn Goudbeek kolleegidega (2009) märkis vokaalide *a*, *i* ja *u* F1 keskmise tõusu kõrge virgumisega emotsioonide (nt rõõm) puhul, kuid võrreldes madala virgumisega (nt kurbus) emotsioonidega oli neil vokaalil *a* madalam F2 keskmine. Positiivse emotsiooniga (nt rõõm) vokaalidel oli F2 keskmine kõrgem kui negatiivse emotsiooniga (nt viha) vokaalidel.

Rõõmu-lausungite vokaalide artikulatsiooni täpsuse kohta on saadud kahe-suguseid andmeid – artikulatsiooni täpsus kas tõuseb (nt Murray, Arnott 2008, Juslin, Laukka 2003, Johnstone, Scherer 2000) või on sarnane neutraalse kõne artikulatsiooni täpsusega (Kienast, Sendlmeier 2000).

4.3. Kurbus ja selle võrdlus teiste emotsioonidega

Kurbuse-lausungite vokaalidele on võrreldes teiste põhiemotsioonide ja neutraalse kõne vokaalidega omane F1 keskmise langus (Juslin, Laukka 2003). Märgitud on ka seda, et negatiivse valentsiga emotsioonide (nt kurbus) vokaalidel on F2 keskmine madalam kui positiivsete emotsioonide (nt rõõm) vokaalidel. (Goudbeek jt 2009)

Kurbuse-lausungite vokaalidele on omane ka madal artikulatsiooni täpsus (Murray, Arnott 2008, Juslin, Laukka 2003, Kienast, Sendlmeier 2000).

Emotsioonide tekitatud muutusi heliliste häälikute formantväärtustes on Juslin ja Laukka (2003) selgitanud sellega, et viha ja rõõm on kõrge virgumisega emotsioonid, mis panevad kõneleja rääkimisel rohkem pingutama (ta püüab selgemalt artikuleerida). Lihaspinge tekitab kõris aga pitsituse. Pingutamine ja pitsitus viivad kõnetrakti lühenemiseni ning toovad kaasa kõrgema F1 ja täpsema artikulatsiooni. Rõõmu puhul muudab kõnetrakti lühemaks ka naeratamine (Tartter 1980). Madala emotsionaalse virgumisega kurbuse puhul pole aga kõris sellist lihaspinget ja kõneleja räägib lõdvestunumalt. See tingib madalama F1 ja ebatäpsema artikulatsiooni.

5. Uurimistulemused: vokaalide *a*, *i*, *u* esimese ja teise formandi väärtused emotsiooniti

Et saada teada, kas eestikeelse ettelõetud emotsionaalse kõne lühikeste vokaalide F1 ja F2 väärtused on mõjutatud oluliselt lausungi emotsioonist võrreldes neutraalse kõnega ning kas emotsionaalse kõne vokaalide formantväärtused erinevad üksteisest oluliselt, mõõdeti vokaalide *a*, *i*, *u* F1 ja F2 väärtused ning arvatati emotsiooniti nende keskmised (vt tabel 2). Siin ja edaspidi on väärtused esitatud nii hertsides kui ka barkides (psühhoakustiline skaala).

Tabel 2. Vokaalide *a*, *i*, *u* esimese ja teise formandi keskmised emotsiooniti

Vokaal	Emotsioon	F1 (Hz)	F1 (Bark)	F2 (Hz)	F2 (Bark)
<i>a</i>	viha	615	5,9	1412	10,7
	rõõm	644	6,1	1440	10,8
	kurbus	558	5,4	1460	10,9
	neutraalne	635	6	1448	10,9
<i>i</i>	viha	386	3,9	2652	14,9
	rõõm	382	3,8	2658	14,9
	kurbus	359	3,6	2642	14,9
	neutraalne	403	4	2660	14,9
<i>u</i>	viha	427	4,3	1298	10,1
	rõõm	418	4,2	1272	10
	kurbus	410	4,1	1334	10,3
	neutraalne	430	4,3	1274	10

Tabelist 2 selgub, et erinevused emotsionaalse kõne ja neutraalse kõne vokaalide F1 keskmiste vahel on väikesed. Kõige enam eristub neutraalsest kõnest kurbuse-emotsioon (vokaal *a* 635 Hz / 6 Bark vs. 558 Hz / 5,4 Bark ja vokaal *i* 403 Hz / 4 Bark vs. 359 Hz / 3,6 Bark). Emotsiooniti erinevad F1 keskmised kõigil kolmel vokaalil, kuid ka siin ei ole erinevused alati suured. Kõige rohkem erineb teistest emotsioonidest kurbus vokaalide *a* ja *i* puhul. Vokaali *a* juures viha vs. kurbus (615 Hz / 5,9 Bark vs. 558 Hz / 5,4 Bark) ja rõõm vs. kurbus (615 Hz / 5,9 Bark vs. 558 Hz / 5,4 Bark) ning vokaal *i* juures viha vs. kurbus (386 Hz / 3,9 Bark vs. 359 Hz / 3,6 Bark).

Erinevused emotsionaalse kõne ja neutraalse kõne vokaalide F2 keskmistes on samuti väikesed. Kõige enam eristub neutraalsest kõnest kurbuse-emotsioon vokaali *u* juures (1274 Hz / 10 Bark vs. 1334 Hz / 10,3 Bark). Emotsiooniti erinevad F2 keskmised vokaalide *a* ja *u* puhul. Vokaali *a* juures erineb kõige rohkem viha kurbusest (1412 Hz / 10,7 Bark vs. 1460 Hz / 10,9 Bark) ning vokaali *u* juures rõõm kurbusest (1272 Hz / 10 Bark vs. 1334 Hz / 10,3 Bark).

Et saada teada, kas emotsioonirühmade erinevused on statistiliselt olulised, tehti formantväärtustele ANOVA-analüüs. Tulemused on esitatud tabelites 3 ja 4.

Tabel 3. ANOVA-analüüsi tulemused emotsionaalsete lausete lühikeste vokaalide *a*, *i*, *u* F1 kohta. (Statistiliselt oluline erinevus on märgitud halli taustaga)

Vokaal	Emotsioonipaar	DF	F value	Pr(>F) Hertzi skaala	Pr(>F) Barki skaala
<i>a</i>	viha vs. rõõm	672	6,43	0,0115	0,03
	viha vs. neutraalne	712	4,01	0,0457	0,04
	viha vs. kurbus	637	20,17	0,0001	0,01
	rõõm vs. neutraalne	596	0,65	0,4211	0,69
	rõõm vs. kurbus	521	33,04	0,0001	0,01
	neutraalne vs. kurbus	561	35,42	0,0001	0,01
<i>i</i>	viha vs. rõõm	364	0,45	0,5010	0,44
	viha vs. neutraalne	363	5,70	0,0174	0,02
	viha vs. kurbus	372	14,77	0,0001	0,01
	rõõm vs. neutraalne	289	6,48	0,0115	0,01
	rõõm vs. kurbus	298	6,81	0,0095	0,01
	neutraalne vs. kurbus	297	27,41	0,0001	0,01
<i>u</i>	viha vs. rõõm	269	0,45	0,5013	0,39
	viha vs. neutraalne	243	0,11	0,7380	0,78
	viha vs. kurbus	251	2,11	0,1478	0,14
	rõõm vs. neutraalne	238	0,78	0,3778	0,31
	rõõm vs. kurbus	246	0,38	0,5384	0,60
	neutraalne vs. kurbus	220	2,59	0,1089	0,11

Tabelist 3 nähtub, et emotsioonipaariti on statistiliselt olulised erinevused ($p < 0,05$) vokaalide F1 keskmistes nii *a* kui ka *i* puhul (v.a vokaali *i* viha vs. rõõm). Vokaali *u* puhul pole erinevused F1 keskmistes emotsiooniti olulised.

Kui võrrelda emotsioone neutraalse kõnega, erinevad sellest oluliselt vokaali *a* puhul viha- ja kurbuse-emotsioon ning vokaali *i* juures kõik kolm emotsiooni.

Tabel 4. ANOVA-analüüsi tulemused emotsionaalsete lausete lühikeste vokaalide *a, i, u* F2 kohta. (Statistiliselt oluline erinevus on märgitud halli taustaga)

Vokaal	Emotsioonipaar	DF	F value	Pr(>F) Hertzi skaala	Pr(>F) Barki skaala
<i>a</i>	viha vs. rõõm	672	1,77	0,1839	0,28
	viha vs. neutraalne	712	3,69	0,0553	0,05
	viha vs. kurbus	637	4,42	0,0358	0,07
	rõõm vs. neutraalne	596	0,13	0,7195	0,49
	rõõm vs. kurbus	521	0,55	0,4600	0,50
	neutraalne vs. kurbus	561	0,25	0,6176	0,93
<i>i</i>	viha vs. rõõm	364	0,07	0,7926	0,84
	viha vs. neutraalne	363	0,19	0,6667	0,63
	viha vs. kurbus	372	0,19	0,6610	0,40
	rõõm vs. neutraalne	289	0,02	0,8948	0,81
	rõõm vs. kurbus	298	0,31	0,5753	0,40
	neutraalne vs. kurbus	297	0,48	0,4896	0,30
<i>u</i>	viha vs. rõõm	269	0,71	0,4015	0,38
	viha vs. neutraalne	243	0,51	0,4758	0,40
	viha vs. kurbus	251	1,14	0,2859	0,32
	rõõm vs. neutraalne	238	0,00	0,9623	0,96
	rõõm vs. kurbus	246	3,27	0,0718	0,07
	neutraalne vs. kurbus	220	2,55	0,1118	0,09

Tabelist 4 selgub, et emotsiooniti on statistiliselt oluline erinevus ($p < 0,05$) F2 keskmistes vaid vokaali *a* hertside puhul, kus F2 eristab viha kurbusest ($p = 0,0358$). Barki skaalal statistiliselt olulist erinevust siin ei ole. Ülejäänud F2 keskmised ei erine oluliselt ühegi uuritud vokaali puhul ei emotsiooniti ega ka võrdluses neutraalse kõnega.

6. Formantanalüüsi tulemuste võrdlus CPM-mudeli ja varasemate uurimustega

CPM-mudel ennustab, et võrreldes neutraalse kõnega tõuseb viha-lausingites heliliste häälikute F1 ja langeb F2. Kui kõrvutada see ennustus antud uurimuse tulemustega, siis selgub, et neutraalse kõne vokaalidest erinesid viha puhul *a* ja *i*, mille F1 keskmised vastupidiselt CPM-i ennustusele langesid. Vokaalide *a, i* ja *u* F2 keskmised neutraalse kõne F2-st ei erinenud. Varasemates uurimustes on viha puhul täheldatud nii vokaalide kõrgemat F1 sagedust (Goudbeek jt 2009, Murray,

Arnott 2008, Juslin, Laukka 2003) kui ka vokaali *a* oluliselt madalamat F2 sagedust (Goudbeek jt 2009). Neid tulemusi käesolev uurimus ei kinnitanud.

Rõõmu puhul ennustab CPM-mudel F1 kahanemist. Antud uurimuse tulemused näitavad, et rõõmu-lausungites erineb neutraalse kõne vokaalide F1-st statistiliselt oluliselt vaid *i* F1, mille väärtus tõesti langeb. Teised uurimused on näidanud, et rõõmu vokaalide F1 ja F2 väärtused tõusevad (Juslin, Laukka 2003, Kienast, Sendlmeier 2000, Tartter 1980) ning neil on madala virgumisega emotsioonidest (nt kurbus) vokaali *a* F2 keskmine madalam. Samuti on rõõmu-lausungite vokaalidel F2 keskmine kõrgem kui negatiivse emotsiooniga (viha ja kurbus) lausungite vokaalidel. (Goudbeek jt 2009) Käesoleva uurimuse tulemustes polnud rõõmu puhul F2 väärtustes emotsiooniti ega võrdluses neutraalse kõnega statistiliselt olulisi erinevusi. Seega ka rõõmu puhul ei ühtinud antud uurimuse tulemused varasematega.

Kurbuse kohta ennustab CPM-mudel, et võrreldes neutraalse kõnega heliliste häälikute F1 keskmine tõuseb ning F2 keskmine langeb. Antud uurimuse tulemused näitavad, et kurbuse-emotsiooni puhul erinevad vokaalide F1 keskmised statistiliselt oluliselt neutraalse kõne vokaalidest *a* ja *i*, mille F1 väärtused langevad. Varasemates uurimustes on kurbuse-lausungite vokaale iseloomustatud neutraalsest kõnest madalama F1 keskmisega (Juslin, Laukka 2003). Negatiivse valentsiga kurbuse puhul on märgitud ka vokaalide madalamat F2 keskmist võrreldes positiivsete emotsioonidega (nt rõõm) (Goudbeek jt 2009). Käesolevast uurimusest selgub, et ühegi vokaali F2 väärtused ei erine statistiliselt oluliselt emotsiooniti (v.a vokaalil *a* viha vs. kurbus Hertzi skaalal, vt tabel 4) ega võrreldes neutraalse kõnega.

Tulemustest võib järeldada, et rõõmu ja viha väljendatakse eesti keeles teisiti kui keeltes, mille kohta on seni uurimusi tehtud. Osaline sarnasus on aga kurbuse väljendamises. Samas peab silmas pidama seda, et erinevuste põhjus võib peituda ka erinevas uurimismaterjalis.

7. Artikulatsiooni täpsus emotsiooniti

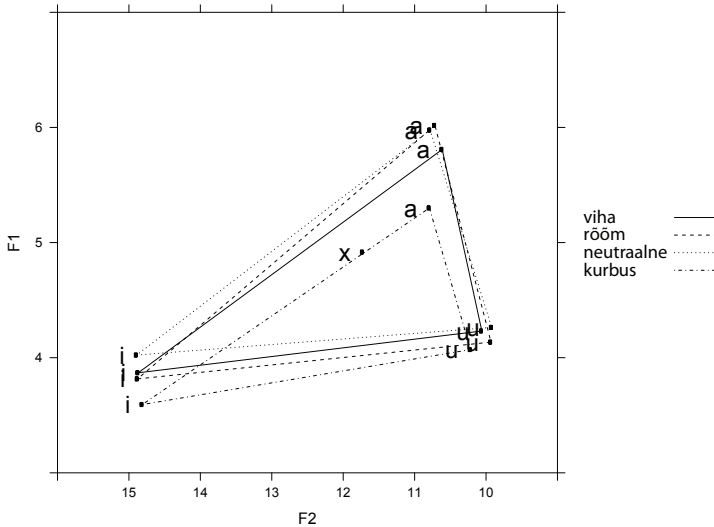
Selleks, et saada vastus küsimusele, kas emotsioonid mõjutavad ettelugeja artikulatsiooni täpsust, arvatati esmalt välja lugeja neutraalne häälik *x* (F1 = 511 Hz / 4,9 Bark ja F2 = 1745 Hz / 11,7 Bark) ning seejärel eukleidiline kaugus emotsionaalse ja neutraalse kõne vokaalide *a*, *i* ja *u* keskmiste ning neutraalse hääliku *x* vahel kasutades valemit $d(x, V_{em}) = \sqrt{(F1x - F1V_{em})^2 + (F2x - F2V_{em})^2}$, kus *x* on neutraalne vokaal ja V_{em} vaadeldava emotsiooni vokaali keskmine.

Eukleidilise kauguse mõõtmistulemused on esitatud tabelis 5.

Tabel 5. Eukleidiline kaugus vokaalide ja neutraalse hääliku *x* vahel (Hertzi skaala / Barki skaala)

Vokaal	<i>x</i> – viha	<i>x</i> – rõõm	<i>x</i> – kurbus	<i>x</i> – neutraalne
<i>a</i>	348,9/1,4	332,7/1,5	288,8/0,9	321,8/1,4
<i>i</i>	915,6/3,3	922,1/3,4	909,8/3,4	921,3/3,3
<i>u</i>	454,8/1,7	482,1/1,8	423,2/1,6	477,9/1,8

Tabelist 5 selgub, et emotsionaalse kõne vokaalide *a*, *i* ja *u* artikulatsiooni täpsus emotsiooniti ja võrreldes neutraalse kõnega eriti ei erine. Erandiks on kurbuse-emotsiooni *a*, mis on võrreldes neutraalse kõne ja teiste emotsioonidega tsentraalsele häälikule *x* märkimisväärselt lähemal. Kurbuse-emotsiooni *a* artikulatsiooni täpsus on seega muutunud ning vokaal on oma kvaliteedi kaotanud. Vt joonis 1.



Joonis 1. Emotsioonid vokaaliruumis Barki skaalal

Kui võrrelda antud uurimuse artikulatsiooni täpsuse tulemusi varasemate uurimustega, saab öelda, et kurbuse puhul need osaliselt kattuvad – paljud uurijad on kurbuse-lausungites märkinud artikulatsiooni täpsuse langemist (Murray, Arnott 2008, Juslin, Laukka 2003, Kienast, Sendlmeier 2000). Viha kohta on saadud erinevaid tulemusi. Osa uurijaid väidab, et viha puhul artikulatsiooni täpsus tõuseb (Murray, Arnott 2008, Juslin, Laukka 2003, Kienast, Sendlmeier 2000, Paeschke jt 1999, Scherer 1986), teised aga märgivad artikulatsiooni täpsuse langemist (Iida jt 2003). Antud uurimuse tulemuste põhjal võib öelda, et viha puhul artikulatsiooni täpsus võrreldes neutraalse kõnega oluliselt ei muutu.

Rõõmu-lausungite vokaalide artikulatsiooni täpsuse kohta on saadud samuti kahesuguseid andmeid – artikulatsiooni täpsus on kas tõusnud (nt Murray, Arnott 2008, Juslin, Laukka 2003, Johnstone, Scherer 2000) või on olnud sarnane neutraalse kõne artikulatsiooni täpsusega (Kienast, Sendlmeier 2000). Käesoleva uurimuse andmed olulist artikulatsiooni täpsuse tõusu rõõmu-lausungites võrreldes neutraalse kõnega ei näita.

Kurbuse artikulatsiooni täpsuse vähenemise tendents kinnitab seda, et kurbust väljendatakse keeleli sarnaselt. Teiste emotsioonide kohta saadud tulemuste vasturääkivusi võib seletada nii uurimismaterjali erinevuse kui ka nende emotsioonide kultuurispetsiifilisema väljenduslaadiga.

8. Kokkuvõte

Uurimuse eesmärgiks oli teada saada, kas ja kuidas mõjutab lausungi emotsioon eestikeelse etteloetud emotsionaalse kõne lühikesi vokaale ning ettelugeja artikulatsiooni täpsust.

Formantanalüüs näitas, et eesti emotsioonid viha, rõõm ja kurbus mõjutasid vaadeldud lühikestest vokaalidest *a*, *i* ja *u* vokaale *a* ja *i*. Emotsioonide mõju neile vokaalidele väljendus nende F1 väärtuste muutumises. Kõige selgemini eristus teistest emotsioonidest kurbus. Vokaalide *a*, *i* ja *u* F2 väärtusi emotsioonid statistiliselt oluliselt ei mõjutanud.

Artikulatsiooni täpsuse mõõtmised näitasid, et emotsioonidel on sellele väikene mõju olemas. Kuigi viha- ja rõõmu-lausungites jäi artikulatsiooni täpsus võrreldes neutraalse kõnega enam-vähem samaks, siis kurbuse-lausungites see langes.

Viidatud kirjandus

- Altrov, Rene 2008. Eesti emotsionaalse kõne korpus: teoreetilised toetuspunktid. – Keel ja Kirjandus, 4, 261–271.
- Bachorowski, Jo-Anne 1999. Vocal expression and perception of emotion. – Current Directions in Psychological Science, 8 (2), 53–57. <http://dx.doi.org/10.1111/1467-8721.00013>
- Banse, Rainer; Scherer, Klaus R. 1996. Acoustic profiles in vocal emotion expression. – Journal of Personality and Social Psychology, 70 (3), 614–636. <http://dx.doi.org/10.1037/0022-3514.70.3.614>
- Burkhardt, Felix; Audibert, Nicolas; Malatesta, Lori; Türk, Oytun; Arslan, Levent M.; Auberge, Veronique 2006. Emotional prosody – does culture make a difference? – Proceedings of Speech Prosody. Dresden, Germany. May 2–5, 2006.
- Douglas-Cowie, Ellen; Campbell, Nick; Cowie, Roddy; Roach, Peter 2003. Emotional speech: Towards a new generation of databases. – Speech Communication, 40 (1-2), 33–60. [http://dx.doi.org/10.1016/S0167-6393\(02\)00070-5](http://dx.doi.org/10.1016/S0167-6393(02)00070-5)
- Goudbeek, Martijn; Goldman, Jean Philippe; Scherer, Klaus R. 2009. Emotion, dimensions and formant position. – INTERSPEECH-2009, 1575–1578.
- Harrington, Jonathan 2010. – Phonetic Analysis of Speech Corpora. Chichester: Wiley-Blackwell.
- Iida, Akemi; Campbell, Nick; Higuchi, Fumito; Yasumura, Michiaki 2003. A corpus-based speech synthesis system with emotion. – Speech Communication, 40 (1-2), 161–187. [http://dx.doi.org/10.1016/S0167-6393\(02\)00081-X](http://dx.doi.org/10.1016/S0167-6393(02)00081-X)
- Johnstone, Tom; Scherer, Klaus R. 2000. Vocal communication of emotion. –M. Lewis, J. Haviland (Eds.). Handbook of Emotion, 2nd ed. New York: Guilford, 220–235.
- Juslin, Patrik N.; Laukka, Petri 2003. Communication of emotions in vocal expression and music performance: Different channels, same code? – Psychological Bulletin, 129 (5), 770–814. <http://dx.doi.org/10.1037/0033-2909.129.5.770>
- Kienast, Miriam; Sendlmeier, Walter F. 2000. Acoustical analysis of spectral and temporal changes in emotional speech. – R. Cowie, E. Douglas-Cowie, M. Schroeder (Eds.). Speech and Emotion: Proceedings of the ISCA workshop. Newcastle, Co. Down, 92–97.
- McIntyre, Gordon; Roland, Göcke 2006. Researching emotions in speech. – Proceedings of the 11th Australian International Conference on Speech Sciences & Technology, 264–269.
- Moon, Seung-Jae; Lindblom, Björn 1994. Interaction between duration, context, and speaking style in English stressed vowels. – Journal of the Acoustical Society of America, 96 (1), 40–55. <http://dx.doi.org/10.1121/1.410492>

- Murray, Iain R.; Arnott, John L. 2008. Applying an analysis of acted vocal emotions to improve the simulation of synthetic speech. – *Computer Speech and Language*, 22 (2), 107–129. <http://dx.doi.org/10.1016/j.csl.2007.06.001>
- Paeschke, Astrid; Kienast, Miriam; Sendlmeier, Walter F. 1999. Fo-contours in emotional speech. – *Proc. ICPHS 99*, San Francisco, Vol. 2, 929–933.
- Scherer, Klaus R. 1986. Vocal affect expression: A review and a model for future research. – *Psychological Bulletin*, 99 (2), 143–165. <http://dx.doi.org/10.1037/0033-2909.99.2.143>
- Scherer, Klaus R. 2003. Vocal communication of emotion: A review of research paradigms. – *Speech Communication*, 40 (1-2), 227–256. [http://dx.doi.org/10.1016/S0167-6393\(02\)00084-5](http://dx.doi.org/10.1016/S0167-6393(02)00084-5)
- Tartter, V. C. 1980. Happy talk: Perceptual and acoustic effects of smiling on speech. – *Perception and Psychophysics*, 27 (1), 24–27. <http://dx.doi.org/10.3758/BF03199901>
- Thompson, William F.; Balkwill, L.-L. 2006. Decoding speech prosody in five languages. – *Semiotica*, 158 (1/4), 407–424. <http://dx.doi.org/10.1515/SEM.2006.017>
- Tolkmitt, Frank J.; Scherer, Klaus R. 1986. Effect of experimentally induced stress on vocal parameters. – *Journal of Experimental Psychology: Human Perception and Performance*, 12 (3), 302–313. <http://dx.doi.org/10.1037/0096-1523.12.3.302>
- Ververidis, Dimitrios; Kotropoulos, Constantine 2006. Emotional speech recognition: Resources, features, and methods. – *Speech Communication*, 48 (9), 1162–1181. <http://dx.doi.org/10.1016/j.specom.2006.04.003>
- Wilting, Janneke; Krahmer, Emiel; Swerts, Marc 2006. Real vs. acted emotional speech. – *Proceedings of Interspeech 2006 ICSLP*, Pittsburgh, PA, USA, 805–808.

Võrgumaterjalid

- ANOVA-analüüsi lähteandmed. http://peeter.eki.ee:5000/vokaalid_aiu.csv (28.09.2011).
- ANOVA-analüüsi tabeli genereerimiseks kasutatud statistikakeskkonna R skript. http://peeter.eki.ee:5000/anova_aiu.R (28.09.2011).
- Boersma, Paul; Weenink, David 2011. Praat: doing phonetics by computer, Version 5.2.35 [Computer program]. <http://www.praat.org/> (28.09.2011).
- Eesti emotsionaalse kõne korpus. <http://peeter.eki.ee:5000/> (28.09.2011).
- Kõne andmebaaside süsteem EMU. <http://emu.sourceforge.net/> (28.09.2011).
- Statistikakeskkond R. <http://www.r-project.org/> (28.09.2011).

Kairi Tamuri (Eesti Keele Instituut) uurimisvaldkond on emotsionaalse kõne akustika.
kairi.tamuri@eki.ee

DO FORMANTS SPEAK OF EMOTIONS?

Kairi Tamuri

Institute of the Estonian Language

Formants belong to the acoustic parameters addressed by the research of emotional speech. The aim of the present study was to find out how, if at all, the utterance emotion of Estonian read emotional speech may affect short vowels and articulatory precision.

The acoustic basis of the study was the Estonian Emotional Speech Corpus (EESC) of the Institute of the Estonian Language. The Corpus contains non-acted speech (journalistic passages) read by a female voice, which have been segmented into sentences, words and speech sounds. The emotional colouring (anger, joy, sadness) or neutrality of the Corpus sentences has been ascertained by perception tests. The research material consisted of such Corpus sentences whose emotion or neutrality had been confirmed by more than half of the listeners. Three short vowels, *a*, *i* and *u*, were picked from those sentences and subjected to formant analysis. Those three are frequent Estonian vowels defining a triangular vowel space (*a* being a low back vowel, *i* a high front vowel and *u* a high back vowel) enabling a fine description of a speaker's articulatory precision.

To solve the research question, measurements were conducted on the first and second formants (F1 and F2) of the short stressed vowels *a*, *i*, *u* as segmented from read Estonian emotional and neutral speech. The results showed that when measured pairwise (partners including neutral sentences) significant differences are revealed between the F1 means of *a* and *i*. F2, however, plays no significant role in emotion differentiation. Vowel reduction and articulatory precision were investigated by using Euclidian distance. According to the results, in the case of angry and joyous utterances articulatory precision was more or less the same as in neutral ones, while a fall signalled sad utterances.

Keywords: emotional speech, acoustics, formant analysis, articulatory precision, Estonian

ALGUPÄRASTE JA TÕLGITUD EESTIKEELSETE TEKSTIDE ERISTATAVUSEST NING NEILE OMISTATUD OMADUSTEST

Marju Taukar

Ülevaade. Tõlkeuurimuses on üheks läbivaks uurimisteemaks selgelt äratuntav tõlgitud keel. Artiklis kirjeldan taju-uuringut, mille eesmärgiks oli kindlaks teha, kas lugeja teeb vahet algupärasel tekstil ja tõlgitud tekstil ning mille alusel ta ühe või teise kasuks otsustab. Vastustest selgus, et 1) algupärased ja tõlgitud tekstid ei olnud selgelt eristatavad, 2) oletataval tõlgitud tekstil arvati olevat keelelisi puudusi ning 3) algupärase teksti sõnavara oli vastajate arvates loovam ja eestipärasem. Uurimuse tulemused näitasid selgelt vajadust jätkata korpusuuringuid algupärase ja tõlgitud tekstide sõnavaraliste erinevuste kohta.

Võtmesõnad: tõlgitud keel, kirjeldav tõlkeuurimus, tõlkeuniversaalid, eesti keel

1. Tõlgitud keel ja tõlkeuniversaalid

Tõlkeuurimuses on viimastel aastakümnetel uuritud tõlgitud keelt algupärase keele kontekstis. Sellise lähenemise eelduseks on, et tõlgitud keel võib erineda algupärasest. Lähtutakse arusaamast, et tõlgitud keel ei ole ilmtingimata kõrvalekalle normist, tõlge lihtsalt on kuidagi teistsugune kui algupärane tekst (Baker 1993). On uuritud tõlgete ühisjooni (nt Mauranen 2004); leitud, et tõlgitud keelt sihtkeele kontekstis uurides on oluline teha vahet ebapädevast tõlkimisest tulenevatel vigadel ja algupärasest keelekasutusest erineval keelekasutusel (Tirkkonen-Condit 2002), kuid jääb täpsustamata, milliste kriteeriumide alusel seda vahet teha. Üldjoontes arvatakse tõlketekst olevat konservatiivsem, keelekasutuselt standardsem ja ametlikum, sest tõlkides silutakse mittestandardset kirjaipilti ja regionaalset kõnepruuki ning tekst on ühtlasem (Baker 2004: 172).

Tõlketeksti lihtsustamisest ja eksplitseerimisest rääkisid oma uurimuste põhjal 1980. aastatel Ria Vanderauwera ning Shoshana Blum-Kulka ja Edward A. Levenston (Laviosa 2002). Tõlkimisprotsessis toimuva on kokku võtnud Mona

Baker (1993), esitades hüpoteesi tõlkeuniversaalidest, mille kohaselt leiavad tõlkimisel aset muutused, mis ei sõltu keeltest ega tõlkimise suunast. Nende muutuste tagajärjel on tõlgitud tekstid sama keele algupärastest tekstidest tuntavalt erinevad. Lisaks lihtsustamisele ja eksplitseerimisele nimetas ta võimaliku tõlkeuniversaalina ka normaliseerimist – levinumate sihtkeele konstruktsioonide ülekasutamist. Hilisemastes uurimustes on lisandunud teisigi nähtusi, nt sanitiseerimine, mis on lähteteksti kohandamine sihtkeele lugejale vastuvõetavamaks (Kenny 1998: 1). Sonja Tirkkonen-Condit (2002: 209) esitas unikaalsete üksuste hüpoteesi, mille kohaselt esineb tõlgitud tekstis algupärasega võrreldes vähem selliseid elemente, millel puudub lähtekeelne leksikaalne vaste. Gideon Toury (1995) esitatud kasvava standardiseerimise seadus ja interferentsi seadus on tõenäosuslikud ning võtavad arvesse erinevaid keeleväliseid muutujaid.

Bakeri väljatoodud tõlkeuniversaalidest said alguse paljud uurimused ning teda võib pidada ka korpuspõhise tõlkeuurimuse alusepanijaks. Korpusmeetod võimaldas näha palju enam kui varasemad meetodid. Seepärast on ka algne uurimuste rohkus ja hilisem kriitika igati arusaadav.

Kuigi uurimusi oli palju, jäid tõlkeuniversaalid ise täpselt määratlemata ning nende piirid ja alaliigid on ähmased. Samuti võeti uutes uurimustes kasutusele uusi termineid, mille puhul ei olnud alati aru saada, kas tegemist oli uue termini või uue mõistega. Ei ole päris selge, mis vahe on lihtsustamisel ja normaliseerimisel või sanitiseerimisel ning millises seoses nad on eksplitseerimise või standardiseerimisega. Nii eksplitseerimine kui lihtsustamine muudavad mõlemad teksti jälgimise lihtsamaks, seega on piir nende vahel üsna ähmane (Pym 2008). Mõistete hägusust tunnistatakse viimase aja uurimustes ikka ja jälle, kuid nende defineerimisele ei ole lähedale jõutud.

Ka Baker (2004: 172) kirjutab tõlkeuniversaalidest ja korpusuurimusest kui meetodist kriitiliselt ja möönab, et senini ei ole osatud päris selgelt välja öelda, millisel moel tõlked algupärasest keelest erinevad, kuid arvab siiski, et erinevused peaksid korpusuurimuse abil esile kerkima.

Toury on juhtinud võimalikest universaalidest, seadustest või normidest rääkides tähelepanu nende määratlematusele ja võimaliku defineerimise üle ka arutlenud, kuid jõudnud samuti vaid probleemi olemasolu teadvustamiseni. Näiteks kui öelda, et “tõlkimisel toimub eksplitseerimine”, jäävad välja juhud, kui eksplitseerimist ei toimu või toimub implitseerimine, ning meie väide on vale, ning kui öelda jällegi, et tõlkes on eksplitseerimine kõrvuti implitseerimisega, siis pole me öelnud midagi uut (Toury 2001: 18). Linn Øveråsi inglise ja norra keele tõlgete uurimusest on näha, et lisaks eksplitseerimisele, mida tema otsis, toimus ka implitseerimine (Øverås 1998). On ilmne, et tõlkeuniversaalid või tõlkes toimuvad muutused tuleks täpsemalt defineerida, sest mainitud universaalid on sõnastatud nii üldiselt, et pole aru saada, kus algab üks ja lõpeb teine, või kas üks on teise alaliik. Juliane House (2008) kritiseerib tõlkeuniversaale üldisemalt. Ta seab kahtluse alla nii nende otsimise vajaduse kui nende õigustamise pelgalt sellega, et korpusmeetod sellist uurimist võimaldab.

Tõlkeuniversaale puudutav kriitika on põhjendatud, kuid et tõlkeuniversaalid on olnud kirjeldava tõlkeuurimuse üks olulisi pidepunkte, ei saa sel alal tehtud uurimustest mööda vaadata. Tõlgete eriline keelekasutus on nähtav ka vaatamata sellele, et universaale sellisel kujul, nagu neist algselt räägiti, ei ole. Tõlke uurimisel

sihtkeele kontekstis on mõtet, eriti kui antud kultuuris ei ole tõlgitud keelt selliselt uuritud. Kõigis uurimustes ilmneb vähemalt mõni nüanss tõlgete süstemaatilise erinevuse kohta algupärastest. Minu uurimuse eesmärgiks on saada esialgne pilt tõlgitud eesti keelest.

2. Uuringu eesmärk, meetod, materjal

Tõlgitud teksti ja algupärase teksti eristamist uurides leiab Tirkkonen-Condit (2002), et kõigepealt peaks välja selgitama, kas lugeja teeb vahet tõlkel ja algupärasel tekstil – kas tõlgitud keel on tõepoolest nii selgelt äratuntav. Hiljem on tõlgete äratundmisel põhineva taju-uuringu läbi viinud ka Jarmo Harri Jantunen (2004), kelle tulemused langevad Tirkkonen-Conditiga kokku nii õigete vastuste protsendi kui põhjenduste poolest. Mõneti sarnase uuringu on masintõlke uurimise raames teostanud Marco Baroni ja Silvia Bernardini (2006) itaalia keele kohta. Ühest küljest sarnanevad ka nende tulemused siinse ja varasemate uuringutega – vastajad ei suutnud tõlgitud tekstidel ja algupärastel tekstidel selgelt vahet teha. Teisest küljest sai äratundmisega paremini hakkama arvutiprogramm, näidates, et vähemalt itaalia keeles on tõlgitud tekstidel ühiseid jooni, mida inimene ei märka, kuid mida arvuti suudab tuvastada.

Siin kirjeldatud taju-uuringu eesmärgiks oli teada saada, kas tõlgitud teksti ja algupärase teksti on võimalik eristada, s.t kas tõlgitud keel on tõepoolest nii spetsiifiline, et lugeja seda märkab. Samuti huvitasid hinnangud ja arvamused tõlgete keele ja algupärase keele kohta. Uuring on kvalitatiivset laadi.

Uurimus seisnes 27 eestikeelse tekstilõigu määramisel tõlkeks või algupäraseks tekstiks. Tekstidest 12 olid algupärase ja 15 tõlgitud. 11 olid ilukirjandustekstid, 3 juhendid ja 3 ajalehe-/ajakirjaartiklid. Tekstide pikkus oli 150–250 tähemärki. Kõik olid toimetatud tekstid täpselt sellisel kujul, nagu need lugejani jõuavad. Tekstides ei asendatud isiku- ega kohanimed; tõlkele või algupärasele tekstile viitavaid vihjeid sisaldavad lõigud jäid uurimusest välja. Kuigi võiks arvata, et tekstides leiduvad pärisnimed mõjutavad vastajaid, esines tegelikult näiteks võõrapäraseid nimesid nii algupärastes kui tõlgitud tekstides. Mõne teksti puhul toodigi esile võõrapäraseid nimesid, kuid need ei olnud kunagi ainsaks, mille põhjal otsus langetati. Näiteks nenditi, et nimed on imelikud, kuid tekst arvati olevat igati sujuv ning eestipärane.

Vastajaid oli taju-uuringus 23. Neist 6 olid tõlkijad, 14 toimetaja või tõlkija eriala üliõpilased ning 3 vastajat ei olnud tegevad keelega seotud alal.

Vastajatel paluti märkida kõigepealt iga teksti juurde, kas nende arvates on tegemist tõlkega või algupärase eestikeelse tekstiga, ning lisaks kirjeldada, miks nad nii arvavad, või vähemalt märkida tekstis need kohad, mis nende arvamust ühele või teisele poole kallutasid. Enamasti mingi selgitus ka anti. Kõige vähem lisasid kommentaare vastajad, kes ei olnud tegevad ühelgi keelega seotud alal. Muid märkimisväärseid erinevusi ei olnud.

3. Tulemused

Kui tõlketekste oli katses 56% ja algupäraseid 44%, siis tõlkeks märgiti 60,9% ja algupäraseks 35,8% ning 2,8% juhtudel ei osatud kummagi kasuks otsustada. Õigeid vastuseid oli 59,7%.

Kõige rohkem (õigesti) tõlkeks märgitud teksti puhul oli otsus üksmeelne 95%, kõige rohkem (samuti õigesti) algupäraseks märgitud teksti puhul nõustus 78% vastajatest. Joonisel 1 on esitatud vastuste jagunemine tekstiti. Tekstid on reastatud tõlkeks märkimiste arvu järgi ning siingi võib näha, et algupäraseid tekste ja tõlketekste on tõlkeks märgitud üldjoontes sarnaselt, tõlketekste pisut sagedamini.

3.1. Algupäraseks peetud tekstide puhul väljatoodud jooned

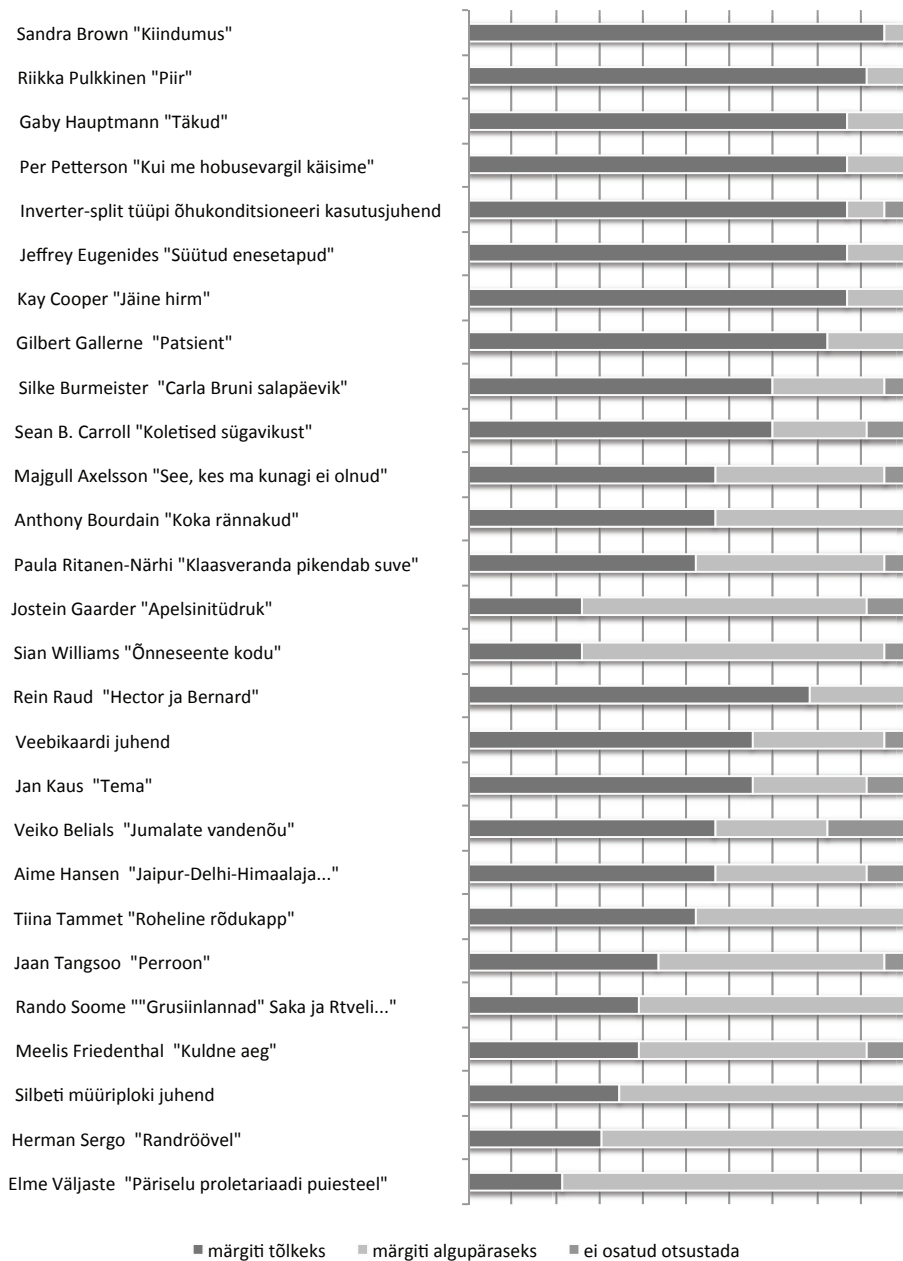
Tekst arvati olevat algupärane, kui sellel ei täheldatud mingeid puudusi: miski ei häirinud, tekstis puudusid võõrapärasused ja vastuolud. Märgiti veel, et tekst oli sorav, sujuv, ladus või loomulik ning isegi pikad laused olid arusaadavad ja voolasid kenasti.

Sageli öeldi nii lauseehituse kui sõnavara kohta, et see on eestipärane, huvitav või leidlik. Mõni tekst võis olla tõlgitud teksti kohta keelekasutuselt liiga vaba või eriline. Algupärasele tekstile viitavad leidlikud sõnad olid näiteks *jõmlik*, *muserdab*, *puunott*, *meikar*, *johtuv*, *ajad pada*. Algupärastes tekstides ja tõlgetes äramärgitud sõnad on koondatud lissasse 1.

3.2. Tõlkeks märgitud teksti puhul väljatoodud jooned

Tõlkeks märgitud tekstide puhul põhjendati valikut sagedamini ja täpsemalt kui algupäraseks märkimisel. Põhjendused puudutasid peamiselt tekstide süntaksit ja leksikat. Allpool on esitatud katseisikute hinnangud, millega põhjendati teksti märkimist tõlkeks; lisatud on hinnangut illustreerinud näited, mille algupära katseisik ei teadnud.

1. Kõige rohkem pöörati tähelepanu sõnajärjele ja lausestusele, mis võis olla vale, kohmakas, konarlik.
 - (1) Asjaolu, et teda üldse ballile kutsuti, üllatas mõnevõrra. (Friedenthal 2005: 41)
2. Mõnes tekstis hinnati lauseid pikkadeks ja lohisevateks. Lause pikkuse kohta mainiti mitmel korral, et algupärase teksti korral teeks autor mõne pika lause mitmeks lauseks, kuid tõlkes on originaali struktuuri järgimine põhjendatud.
 - (2) Siis ilmnas kuidagi sujuvalt, et tegelikult oskab meie prantsuse külaline ka päris talutavat inglise keelt, minu kolleegid aga saavad vajaduste piires hakkama vene keelega, eriti üks neist, kellele tõlgineiu oli korraga hakanud pakkuma külalisest suurematki huvi, ehkki tema poolt daamile meie kultuuriarealile nii tüüpilise nihkega avaldatud tähelepanu ei pälvinud märkimisväärset vastulahkust. (Raud 2004: 111)



Joonis 1. Tekstide tõlkeks ja algupäraseks märkimise jaotus

3. Lauseehituse piires räägiti veel võõrastest, kohmakatest ja lähtekeelele viitavatest konstruktsioonidest.
 - (3) Nad maiustasid jäätisepallidega, kui Rennie küsis: “Kas teie tunnete hirmuhoogusid?” (Brown 2003: 145)
4. Leiti rektsioonivigu (4) ja ebaühtlust aja- (5) ja isikuvormide (6) kasutamisel.
 - (4) Mõnikord panevad **tugevate ärevushoogude tõttu vaevlevad** inimesed sääare kummilindi käe ümber. (Brown 2003: 145)
 - (5) Ta käed **on** mu mälus palju selgemad kui tema nägu – nahk **oli** punane ja küüned kollased. (Axelsson 2006: 44)
 - (6) Kui nad **tunnevad** hirmu enda üle võimust võtvat, siis **lastakse** kumm laksatades lahti. (Brown 2003: 145)
5. Tõlkimise arvele pandi ka mitmesugused kordused: mitmes tekstis leiti olevat liiga palju sidesõnu (7), mõnes asesõnu (8) ja ühes kordus nimisõna **tüdruk** (9).
 - (7) Punased **ja** kollased limonaadikastid **ja** tormine meri nende taga **ja** tusane taevas nende kohal. (Kaus 2006: 100)
 - (8) **Keegi** ei mõistnud, **mis meil sel** aastal hakkas või miks **me** nii meeletult vihkasime surnud putukate koorikuid **meie** elus. (Eugenides 2005: 52)
 - (9) Ukse tagant kiirgas eredamat valgust ja **tüdruk** seisis valgusvihus, uksepiidad raamisid **tüdruku** piirjooned selgeks ja õhtuks saali paigutatud ringlevad punktvalgustid silitasid **tüdruku** siluetti kiiresti, kuni pöördusid jälle eemale. See oli just nii, nagu ta oli kujutlenud: **tüdrukul** olid lumehelbed juustes. (Pulkinen 2009: 168)

Viimasel juhul on tõepoolest tegemist tõlkega, kuid mitte eesti sooneutraalse asesõna vältimisega, nagu pakuti. Ka soomekeelses originaalis esines sõna *tyttö* lühikese lõigu jooksul 14 korda. Näite (7) korduste kohta ütles üks vastaja: “häirisid sõnakordused, kuigi hiljem mõeldes tundus loogiline, et sõnakordused esineksid pigem algupärasest tekstis.” Mõned kordused esinesidki algupärasest tekstis, mõned ka tõlkes. Ühe algupärase teksti puhul leiti üksmeelselt, et pidev sidesõna kasutamine lause alguses viitab tõlkele.
6. Sõnavara osas toodi välja üldist võõrapärasust, samuti mõningaid võõrapäraseid väljendeid. Vähemal määral mainiti ära eesti keeles vähem kasutatavaid või konteksti ebasobivaid sõnu; mõnda sõna või väljendit peeti ka otseselt valeks. Kritiseeritud väljendid ja sõnad olid näiteks *jouetu kiskuma puruks, krapasakas valge sein, kas polnud, heilutas, plombiir, oivaliselt*. Loend võõrapärasena tundunud sõnadest on lisas 1.

Suurimad erinevused vastajate hinnangutes olidki sõnavara osas. Ühe sõna kohta võidi arvata, et see on tore ja eestipärane, ning järgmises vastuses leiti, et seesama

sõna on väga imelik või võõrapärane (nt *lossida* (Sergo 2009: 289), *ülemriukameister* (Hauptmann 2005: 117)). Kõige enam erinesid sõnavara puudutavad arvamused ainsa varasema (Sergo 2009 [1988]) teksti puhul. Nii toodi ühelt poolt omapäraste ja eestipärastena ning teiselt poolt ebaloomulikuks ja otsetõlkeks peetavatena välja sõnad ja fraasid nagu *ilm juba praegu nii külm, et kõrvalestades kinni, jääks tinutada, kuidagi ikka madala pealt lahti tõmbavad, lasti lossida, siis pool tundi aitajat tuult*. Arvamused läksid väga selgelt lahku ka kõigi kolme ajakirjandusteksti puhul: *vollepiigad, odadeks propaganda malelaulal, mustad nõud ja muu jama, jalamaid, imeõrna varjundiga valge, krapasakas valge sein* olid mõne vastaja jaoks eestipärased, mõne jaoks kõlasid eesti algupäraga teksti kohta liiga veidralt.

4. Tõlgitud keele olemus lugeja arvates

Siinse uurimuse materjalis ei olnud tõlgitud ja algupärased tekstid vastajate jaoks väga kergesti eristatavad. Uurimuse eesmärgiks oli esiteks teada saada, kas tõlgitud keel on tõepoolest nii spetsiifiline, et lugeja tõlgitud teksti ära tunneb, ning teiseks, mis tekstis tõlgitud keelele viitab. Vastustest nähtus, et 1) selget vahet tõlgitud tekstil ja algupärasel tekstil ei tehtud, 2) oletataval tõlkel arvati midagi viga olevat ja tõlkest räägiti pigem negatiivses toonis ning 3) oletatava algupärase teksti sõnavara arvati olevat eestipärasem ja leidlikum kui tõlgitud teksti sõnavara.

Nii Tirkkonen-Conditi kui Jantuneni tulemused sarnanevad käesoleva tajuuuringu tulemustele. Õigete vastuste protsendid on vastavalt 61,5 (Tirkkonen-Conditi 2002: 210), 62,6 (Jantunen 2004: 561) ning 59,7.

Eestikeelsete tõlgitud tekstide ja algupäraste tekstide määramisel tulid esile vastajate isiklikud eelistused ja negatiivne eelhoiak tõlgitud teksti suhtes. Oletatava tõlketeksti kommenteerimisel kasutati pigem negatiivseid väljendeid, ja sageli väljendati ka isiklikke keelelisi eelistusi. Lisaks tõlkele või algupärasele tekstile viitavatele kommentaaridele tehti teksti kohta parandusettepanekuid, arvati näiteks, et mõni sõna sobiks paremini kui mõni teine. Kui arvutiprogramm on suutnud tõlgitud ja algupärasel tekstil paremini vahet teha kui inimene (Baroni, Bernardini 2006), siis on võimalik, et põhjused peituvad isiklikes eelistustes. Tõlgitud tekstid on ehk küll erinevad, kuid vastajad lähtusid sageli oma eelistustest ning tõlget arvati pigem õigeks peetud normist kõrvalekalduvaks.

Sõnavara osas kerkisid vastustes esile kaks peamist joont. Esiteks kasutatakse vastajate arvates algupärastes tekstides rikkalikumat sõnavara ning teiseks leiti tõlgitud tekstides võõrana mõjuvaid sõnakombinatsioone. Kõige üksmeelsemalt tõlkeks arvatud teksti puhul toodi tõlkele viitavana esile võõrapäraseid või kohmakaid sõnu ning sõnaühendeid, nagu *tunnete hirmuhoogusid, kaitseb aju väärsignaali eest, linnast eemal, ratsaklubi liige, kummilint, maiustasid jäätisepallidega*.

Kuigi ei öeldud, et tõlgitud teksti sõnavara on otseselt konservatiivsem, mainiti korduvalt, et oletatava algupärase teksti sõnavara on hea või leidlik või et tõlkija mõne vähem kasutatava või kõnekeelse sõna peale ei tuleks. Teiselt poolt toodi aga oletatava tõlgitud teksti puhul esile võõrapärast sõnavara. Leidlikku ja võõrapärast sõnavara eristada ei olnud võimalik ning piir nende vahel on kas õrn või subjektiivne. Teisalt võib sõnavara leidlikkus või võõrapärusus olla seotud kollokatsioonide või

korduvate muustritega ning erilised sõnad peavad lugeja jaoks asuma tuttavas must-
ris, sest eraldiseisvana või vales kohas esinevad erilised sõnad tunduvad kohmakad.

Korduvaid mustreid on tõlgitud ja algupärase keele puhul uuritud mitmetes keeltes ning enamasti on leitud, et need on tõlgitud ja algupärase keeles tõepoolest erinevad (nt Balaskó 2008, Jantunen, Eskola 2002). Neist võib sõltuda, kas keel mõjub loomuliku või võõrana ning leitud on, et kinnistunud konstruktsioonidel, nn poolfabrikaatidel on eriti oluline roll just tehnilises tõlkes (Heltai 2004) ja tehnilises tekstis üldse. Siinse uuringu tulemustest ühtegi otsust kommentaari tehnilise tõlke ja poolfabrikaatide seose kohta ei esinenud, kuid mitmel korral avaldati arvamust, et tehniliste tekstide puhul on keerulisem vahet teha algupärase ja tõlkelise päritoluga tekstidel. Üheks põhjuseks on ilmselt sõnavara, kuid oma roll võib olla rohkel valmis konstruktsioonide kasutamisel.

Seniste tõlgitud ja algupärasest keelt võrdlevate uurimuste ja siinse taju-uuringu tulemused erinevad kohati. Vastajad näisid olevat arvamusel, et kui tekstis midagi ei häiri, siis on tegemist algupärase tekstiga. Mainitud varasemate uurimuste põhjal võiks öelda, et kui tõlkimisel toimub normaliseerimine, siis peaks tõlgitud tekst olema sihtkeele vormi sobituv ning konventsionaalne. Seda aga tõlgitud tekstilt ei oodatud. Teisalt, kui tekstis esinev sõnavara oli vastajate arvates leidlik, siis märgiti tekst algupäraseks. Sõnavara lihtsustamisest ja üldistamisest tehtud uurimustes on leitud, et tõlgetes kasutatakse rohkem sagedamini esinevat sõnavara ning selle põhjuseks on eesmärk kirjutada kergesti loetav tekst. Tõlkes toimuvate muutuste põhjuseid ei ole enamasti käsitletud, piirdutud on võimalike muutuste otsimisega ja tõlgitud keele olemuse kirjeldamisega.

5. Kokkuvõtteks

Artikkel käsitles tõlgitud keele uurimist algupärase keele kontekstis. Kirjeldavas tõlkeuurimuses on selliste uurimuste lähtekohaks olnud sageli tõlkeuniversaalid, seadused või normid. Teisalt ei ole veel jõutud nende täpse defineerimiseni ning puhuti on kahtluse alla seatud nende olemasolu ja põhjendatus. Siiski tõdetakse ikka ja jälle, et tõlgitud tekst erineb algupärasest tekstist. Siinse uuringu kontrollis seda seisukohta eestikeelsete tekstide materjalil. Tõlgete äratundmisel põhineva taju-uuringu tulemusena selgus, et väga selget vahet tõlkel ja algupärasel tekstil ei olnud võimalik teha. Tõlgitud keelt peeti enamasti lugeja poolt õigeks peetavast normist kõrvalekalduvaks. Kui tekstis leidis leidlikku või põnevamat sõnavara, siis arvati tegemist olevat algupärase tekstiga. Algupärasena toodi välja üksikuid sõnu, samas kui tõlketekstist arvati märku andvat pigem ebaõnnestunud, eesti keelele võõrad sõnaühendid.

Varasemate uurimuste ja siinse taju-uuringu tulemuste valguses tuleb selgelt välja vajadus jätkata korpusmeetodil põhineva uurimusega, mis annaks aimu, milline on eestikeelne tõlgitud tekst tegelikult. Taju-uuringu vastuste seast kerkis edasist uurimist vajavana esile seisukoht, et algupärane sõnavara on eestipärane ja leidlik ning tõlketeksti fraasid võõrapärased.

Materjali allikad

Tõlgitud tekstid

- Axelsson, Majgull 2006. Need, kes ma kunagi ei olnud. Tõlkinud Raili Väin. Tallinn: Pegasus.
- Bourdain, Anthony 2003. Koka rännakud. Tõlkinud Juhan Habicht. Tallinn: Varrak.
- Brown, Sandra 2003. Kiindumus. Tõlkinud Mati Piirimaa. Tallinn: Ersen.
- Burmeister, Silke 2009. Carla Bruni salapäevik. Tõlkinud Maris Jakobson. Tallinn: Olion.
- Carroll, Sean B. 2010. Koletised sügavusest. – Tarkade Klubi, 5 (41). Tallinn: Presshouse.
- Cooper, Key 2008. Jäine hirm. Tõlkinud Leena Suits. Tallinn: Ersen.
- Eugenides, Jeffrey 2005. Süütud enesetapud. Tõlkinud Liia Piirmets. Tallinn: Pegasus.
- Gaarder, Jostein 2005. Apelsinitüdruk. Tõlkinud Karel Zova. Tallinn: Koolibri.
- Gallerne, Gilbert 2008. Patsient. Tõlkinud Margot Endjärv. Tallinn: Eesti Raamat.
- Hauptmann, Gaby 2005. Tükud. Tõlkinud Tiina Aro. Tallinn: Tormikiri.
- Inverter-split tüüpi õhukonditsioneeride kasutusjuhend. http://www.bvtpartners.ee/content/files/AWI_O26_35HPDC1_user_manual_Estonian.pdf (06.02.2011).
- Petterson, Per 2010. Kui me hobusevargil käisime. Tõlkinud Kai Mai Aja. Tallinn: Pegasus.
- Pulkkinen, Riikka 2009. Piir. Tõlkinud Monika Pais. Tallinn: Pegasus.
- Ritanen-Närhi, Paula 2009. Klaasveranda pikendab suve. – Tehnikamaailma Kodu ja Ehitus, 8. Tallinn: Ühinenud Ajakirjad AS.
- Williams, Sian 2011. Önneseente kodu. – Kodu ja Aed, dets. Tallinn: OÜ Formamedia.

Algupäraseid tekstid

- Belias, Veiko 2006. Jumalate vandenõu. Tallinn: Varrak.
- Friedenthal, Meelis 2005. Kuldne aeg. Tallinn: Tuum.
- Hansen, Aime 2009. Jaipur-Delhi-Himaalaja: reisikohvrast leitud lood. Tallinn: Varrak.
- Kaus, Jan 2006. Tema. Tallinn: Tuum.
- PRIA veebikaardirakenduse kasutusjuhend. https://kls.eesti.ee/pria_avalik_kaart/doc/veebikaardi_kasutusjuhend.pdf (06.02.2011).
- Raud, Rein 2004. Hector ja Bernard. Tallinn: Tuum.
- Sergo, Hermann 2009 [1988]. Randröövel. Tallinn: Pegasus.
- Silbeti müüriplokkide kasutusjuhend. <http://www.silbet.ee/kasutusjuhend.pdf> (06.02.2011).
- Tammet, Tiina 2010. Roheline rõdukapp. – Kodu ja Aed, okt. Tallinn: OÜ Formamedia.
- Soome, Rando 2009. “Grusiinlannad” Saka ja Rtveli, kangelased tiigrinahas. – Sporditäht, 14. jaanuar. Tallinn: Kalev Meedia.
- Tangsoo, Jaan 2010. Perroon. Tallinn: Eesti Raamat.
- Väljaste, Elme 2007. Päriselu proletariaadi puiesteel. Tallinn: Varrak.

Viidatud kirjandus

- Balaskó, Maria 2008. What does the figure show? Patterns of translationese in a Hungarian comparable corpus. – *trans-kom*, 1 (1), 58–73.
- Baker, Mona 1993. Corpus linguistics and translation studies: Implications and applications. – Mona Baker, Gill Francis, Elena Tognini Bonelli (Eds.). *Text and Technology*. Amsterdam: John Benjamins, 233–250.
- Baker, Mona 2004. A corpus-based view of similarity and difference in translation. – *International Journal of Corpus Linguistics*, 9 (2), 167–193. <http://dx.doi.org/10.1075/ijcl.9.2.02bak>
- Baroni, Marco; Bernardini, Silvia 2006. A new approach to the study of translationese: Machine-learning the difference between original and translated text. – *Literary and Linguistic Computing*, 21 (3), 259–274. <http://dx.doi.org/10.1093/lc/fqio39>

- Heltai, Pál 2004. Ready-Made language and translation. – Gyde Hansen, Kirsten Malmkjær, Daniel Gile (Eds.). *Claims, Changes and Challenges in Translation Studies: Selected Contributions from the EST Congress, Copenhagen 2001*. Benjamins Translation Library, 50. Amsterdam, Philadelphia: John Benjamins Publishing Company, 51–71.
- House, Juliane 2008. Beyond intervention: Universals in translation. – *trans-kom*, 1 (1), 6–19. http://www.trans-kom.eu/bdo1nr01/transkom_01_01_02_House_Beyond_Intervention.20080707.pdf (5.05.2011).
- Jantunen, Jarmo Harri 2004. Suomennosten kieliasu puntarissa. – *Virittäjä*, 4, 559–572. http://www.kotikielenseura.fi/virittaja/hakemistot/jutut/2004_559.pdf (5.05.2011).
- Jantunen, Jarmo Harri; Eskola, Sari 2002. Käännössuomi kielivarianttina: syntaktisia ja leksikaalisia erityspiirteitä. – *Virittäjä*, 2, 184–207. http://www.kotikielenseura.fi/virittaja/hakemistot/jutut/2002_184.pdf (24.03.2011).
- Kenny, Dorothy 1998. Creatures of habit? What translators usually do with words. – *Meta: Journal des Traducteurs / Meta: Translators' Journal*, 43 (4), 515–523.
- Laviosa, Sara 2002. *Corpus-based Translation Studies: Theory, Findings, Applications. Approaches to Translation Studies*, 17. Amsterdam: Rodopi.
- Mauranen, Anna 2004. Corpora, universals, and interference. – Anna Mauranen, Pekka Kujamäki (Eds.). *Translation Universals. Do They Exist?* Benjamins Translation Library, 48. Amsterdam, Philadelphia: John Benjamins: 65–82.
- Øverås, Linn 1998. In search of the third code. – *Meta: Journal des Traducteurs / Meta: Translators' Journal*, 43 (4), 557–570. <http://id.erudit.org/revue/meta/1998/v43/n4/003775ar.pdf> (27.04.2011).
- Pym, Anthony 2008. On Toury's laws of how translators translate. – <http://www.scribd.com/doc/3735779/Toury-Laws> (5.05.2011).
- Tirkkonen-Condit, Sonja 2002. Translationese, a myth or an empirical fact? A study into the linguistic identifiability of translated language. – *Target*, 14 (2), 207–220. <http://dx.doi.org/10.1075/target.14.2.02tir>
- Toury, Gideon 1995. *Descriptive Translation Studies and Beyond*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Toury, Gideon 2001. Probabilistic explanations in Translation Studies: Universals – or a challenge to the very concept? – Gyde Hansen, Kirsten Malmkjær, Daniel Gile (Eds.). *Claims, Changes and Challenges in Translation Studies: Selected Contributions from the EST Congress, Copenhagen 2001*. Benjamins Translation Library, 50. Amsterdam, Philadelphia: John Benjamins Publishing Company, 15–25.

Marju Taukar (Tallinna Ülikool), uurimisvaldkonnaks on tõlkimine ja tõlgitud keel.
marjutaukar@gmail.com

Lisa 1.

Algupäraseks peetud sõnad ja fraasid

kurk kuivas ropult
johtuv
jõmlik
jäässe tõmmata
kõhedaks
käed-jalad tööd täis
lainelokse
latrajail
meikar
muserdab
nišši
osmiku
piskuks ajaks
sõnavahet
täielik gei
vennaarmust pajatav
ropult
puunott
säherdune
võhma

Tõlkeks peetud sõnad ja fraasid

andestust?
desserttaldrikutele
haletsusväärne mälukaotanud
heilutas
ei tulegi usaldada
jõuetu kiskuma puruks
kastide armee
keegi kolmas
krapsakas valge sein
kummilint
laev madalal
maiustaksid jäätisepallidega
läheksid kuulajatel toidupalad kurku
plombiir
saab mu kogemus kinnitust
tuleb tagasi nagu kummitus
B. lükkus kaugele ära
oma enamuses mitte
lossida
ratsaklubi

ON THE IDENTIFIABILITY AND CHARACTERISTICS OF TRANSLATIONS AND NON-TRANSLATIONS IN ESTONIAN

Marju Taukar

Tallinn University

The focus of descriptive translation studies has for a few decades been on translated language in the context of non-translated language. The impulse for research has been an attitude that translations in a language differ linguistically from non-translations in the same language. Even though the difference is usually stressed as not deriving from straightforward mistakes, the common attitude towards translations and also the prevailing view of translations in the study discussed in this article is that there is something wrong with the translation. The aim of the study was to investigate whether readers can distinguish between translations and non-translations in Estonian and what linguistic features are described as creating the impression of a translation or a non-translation. The two broad features that, according to respondents, indicate non-translationality are a flowing, fluent text and creative vocabulary. Features suggesting that a text is a translation are deviations from the linguistic norm (or what the respondent regards as the norm) and unnatural lexical strings. The article places the study in the context where translation universals have been sought for to show what lies behind the impression of strangeness claimed present in translations. This context involves also a discussion of the critique of translation universals.

Keywords: translational language, descriptive translation studies, translation universals, Estonian

KUIDAS MÄÄRATA EESTI KEELE SÕNAVARA TUNDETOONE?

Ene Vainik

Ülevaade. Artiklis antakse ülevaade sõnavarauksuste tundetoonide määramise teoreetilistest alustest ja praktilistest võimalustest. Esmalt kirjeldatakse ülesande olulisust, siis teoreetilisi lähte-eeldusi ja vaetakse neist tulenevaid meetodilisi võimalusi. Juhtumianalüüsina kirjeldatakse eesti keele põhisõnavara sõnastiku märksõnade tundetoonide hindamise meetodikat, katse läbiviimist ja tulemusi ning võrreldakse neid varasemate eesti keele emotsioonisõnavara uurimustega. Artikkel näitab, kuidas eesti lingvistika kontekstis uudset üleannet lahendada meetodiliselt võimalikult korrektselt, kuid arutleb ka küsitavuste ja perspektiivide üle.*

Võtmesõnad: emotsionaalne valents, meetodika, sõnavara, afektikalduvus, põhisõnavara sõnastik, eesti keel

1. Sissejuhatus

Asjaolu, et on tekkinud vajadus pöörata tähelepanu eestikeelsete sõnadega tüüpiliselt kaasnevatele emotsionaalsetele lisatähendustele – nn tundetoonidele – tuleneb sellest, et on ülesandeid, mille lahendamiseks ei piisa keele tähistusfunktsiooni uurimisest, tarvis läheb teadmisi ka keele ekspressiivsest funktsioonist ja emotsioonide väljendumisest (Foolen 1997, Hübler 1998).

Üks selliseid ülesandeid on eestikeelse kõnesünteesi loomulikustamine (Altrov, Pajupuu 2008). Selleks, et ette loetavatele lausungitele oleks võimalik lisada inimlikku emotsionaalsust, on ühest küljest tarvis uurida ja modelleerida emotsioonide prosoodilist väljendumist (Iida jt 2003, Ververidis, Kotropoulos 2006), teisalt aga õpetada süsteem kirjaliku teksti emotsionaalsust või selle puudumist määrama (Vainik 2010, Pajupuu jt, ilmumas). Katsetes määrata teksti, lõigu, lause ja isegi osalause emotsiooni on üksmeelselt leitud, et taolise süsteemi osaks peab olema sõnade tüüpilisi emotsionaalseid lisatähendusi esitav sõnastik (ülevaadet vt Vainik 2010). Eestikeelsete sõnade tundetoonide ehk emotsionaalsete valentside määra-

* Artikkel on valminud sihtfinantseeritava teadusteema SF0050023s09 ja Eesti Keele Instituudi baasrahastusprojekti "Eesti keele sõnavara afektikalduvuse määramise raamtööd aastal 2010" toetusel.

misest on vaja kõnelda just nüüd, kuna Eesti Keele Instituudis on koostamisel uut tüüpi eesti keele põhisonavara sõnastik (nn põhisonastik), mis koondab ligikaudu 4000 kõige sagedasemat ning aktiivsele keeleõppijale kõige tarvilikumat eesti keele sõna koos mitmesuguse informatsiooniga selle tüüpilise kasutamise kohta (vt Kallas, Tuulik 2011). Uurimused on näidanud, et sellisest sõnavara hulgast piisab haritud kõrgtasemel keeleoskaja vajadusteks (Pajupuu jt 2009). Kuna süntesaatori poolt ette loetav tekst ei ületa keerukuse poolest haritud eesti keele oskaja taset, siis on loomulik mõelda, et põhisonastik võiks olla põhjaks ka kõnesünteesi loomulikustamise eeltingimuseks olevale emotsiooni määramisele tekstis (Pajupuu jt, ilmumas).

Teisest küljest on nähtud vajadust teavitada sõnavara tüüpilistest tundetoonidest ka teise keele õppijat. Kakskeelsete isikutega tehtud uurimused on näidanud, et kuigi üldine teise keele oskus on neil kõrgtasemel, võib just emotsionaalse väljenduse olukordades tekkida äkki sobivatest sõnadest puudus ja võidakse kergesti emakeelele ümber lülituda (Pavlenko 2008b). Eesti keele põhisonavara sõnastik on mõeldud pakkuma keeleõppijast kasutajale aktiivset tuge ka sellistes olukordades.

Nii lähtubki ülesanne määrata põhisonastiku sõnavara eriomased tundetoonid kahest praktilisest vajadusest ning kavatsusest need vajadused ühitada. Ühitamine tundub realistlikuna, kuna mõlemad ülesandepüstitused – nii kõnesünteesiks vajalik sõnavara märgendus kui aktiivse keeleõppija teavitamine tundetoonidest – lähtuvad ühisest eeldusest, et vähemalt teatud osa keele sõnavarast ei kannu mitte üksnes tähendusi ja/või funktsioone, vaid edastab ka emotsionaalset informatsiooni. Lisaks eeldatakse, et too emotiivne varjund on teataval määral konventsionaalne ning seega peaks seda põhimõtteliselt saama kindlaks teha ja ka sõnastikus fikseerida. Ülesandepüstituse praktilise poole pealt näeb asi välja nii, et põhisonastiku koostajad on kokku pannud märksõnastiku (tulevane maht üle 4500, artikli kirjutamise hetkel 3015) ja määrata tuleb neid iseloomustavad tundetoonid. Küsimus on selles, kuidas teha seda teoreetiliselt läbimõeldud alusel ning meetodiliselt võimalikult korrektset.

Artikli esimene pool on pühendatud teoreetilistes lähteolukordades orienteerumisele ning neist tulenevate meetodite sobivuse vaagimisele. Juhtumianalüüsi osas kirjeldatakse eesti keele põhisonavara sõnastikule rakendatud konkreetset meetodit, esitatakse väljavõtte hindamiste tulemustest ning võrreldakse neid varasemate uurimustega. Lõpuosas arutletakse rakendatud meetodi tugevate ja nõrkade külgede üle ning pakutakse välja võimalusi, kuidas tundetoonide määramist veelgi eesmärgipärasemaks muuta.

2. Teoreetilised ja meetodilised küsimused

2.1. Millisest teoreetilisest taustast lähtuda?

Nähtuse kohta, et loomuliku keele sõnadel on kalduvus edasi anda mitte üksnes ideesid, vaid ka emotsionaalset või hinnangulist suhtumist, on kasutatud erinevaid nimetusi – konnotatsioon, tundetoon, emotsionaalne valents. Käesolevas artiklis kasutatakse peamiselt eestikeelset nimetust tundetoon, mingil määral aga ka

terminit emotsionaalne valents. Valents kui metafoorne (keemiast pärit) termin võeti psühholoogias kasutusele 1930-ndatel aastatel sündmuste ja nähtuste kas külgetõmbavuse (positiivne valents) või eemaletõukavuse (negatiivne valents) tähistamiseks. Peamiselt kasutatakse seda terminit tunnete, afektide, käitumiste (nt lignemine vs. vältimine), eesmärgi (saavutamine vs. mitte-saavutamine jmt) hedoonilisest toonist rääkimiseks, ehkki välistatud pole ka hinnangud muudelt alustelt ja nende mitmekihilisus (Colombetti 2005). Enne tundetoonide praktilist määramist ja sobiva meetodi valimist on hea end teoreetiliselt positsioneerida küsimuse suhtes, kas ja kuidas stabiilselt saavad tundetoonid suvalise sõnavaraüksuse juurde üldse kuuluda.

Traditsiooniline lingvistiline lähenemine on sõnaga kaasneva tundetooni suhtes kõige skeptilisem: nende olemasolu küll mõõndakse, kuid seda peetakse sõna mõistelise tuuma “sentimentaalseks juurdekasvuks”, mida on suure variatiivsuse tõttu pea võimatu määratleda ja millest teadusele kasu pole (Sapir 2009: 50–51). Leksikograafias lähtutakse objektiivsetest tõenditest keele tegeliku kasutamise kohta kas kirjalikus või suulises suhtluses ja mitte selle “potentsiaalist”. Öeldakse, et sõnaraamatus tuleb kirjeldada nähtusi, mis on sagedased ja laialt levinud (esinevad korpuses eri tekstitüüpides ja kuuluvad keele “tuuma”) (Atkins jt 2008). Traditsioonilisest puhtlingvistilisest ja leksikograafilisest tähenduse kirjeldusest on tundetoonide kirjeldamine ja määramine jäänudki välja arvatavasti just neile omistatud subjektiivsuse ning muutlikkuse tõttu. Üksnes sõnastikes kasutatavad stiilmärgendid (nt *halvustav*, *hellitlev*) viitavad mõningal määral sõnade emotsionaalsele lisaväärtusele ja tüüpilisele kontekstile (Voll 2002).

Psühholoogias seevastu ei kahelda, et ka üksiksõna nagu iga muugi stiimul võib esile kutsuda emotsionaalseid reaktsioone ja et neil reaktsioonidel võib olla mingil määral stabiilne ja kinnistunud iseloom; arutletakse pigem selle üle, kuidas üksiksõnade iseloomulike valentside põhjal määrata lause ja kogu teksti valentsi (Bestgen 1994). Selle, lähtekohalt biheivioristliku mõtteviisi klassikaks on C. E. Osgoodi, G. J. Suci ja P. H. Tannenbaumi ([1957] 1975) uurimus, milles esmakordselt kirjeldati nn semantilise diferentsiaali meetodit ja selgitati välja semantilise ruumi kolm universaalset dimensiooni: *hinnang*, *tugevus* ja *aktiivsus*, mille suhtes paigutuvad väidetavalt niihästi keele kõik sõnad kui ka mittesõnalised mõisted – nimed, pildid, narratiivid – ja üldse kõik, mida küsimustiku ja polaarsetest adjektiividest moodustatud skaalade abil mõõdetakse.

Psühholingvistilise lähenemise järgi seevastu ei eeldata, et emotsionaalsus peaks olema potentsiaalselt keele iga sõna iseloomustav parameeter. Küll aga leidub iga inimese jaoks siiski hulk sõnu, mille kasutamisega emotsionaalne konnotatsioon võib kaasneda. Väidetavalt kujutavad emotsionaalsed keelendid endast isiku mentaalse leksikoni eraldi klassi, erinedes nii abstraktsetest kui konkreetsetest sõnadest abstraktsuse, ettekujutatavuse ja konteksti kättesaadavuse osas (Pavlenko 2008a). Kuulumine samasse emotsionaalsesse kategooriasse võib olla viisiks, kuidas on mentaalse leksikoni üksused omavahel seotud. See oletus sobib kokku ka kognitsiooni ja emotsioonide seoseid selgitava nn assotsiatiivse võrgustiku teooriaga (Bower, Forgas 2000: 94). Selle kohaselt praimivad sõnad, kui nad on antud isiku jaoks emotsionaalselt tähendusrikkad, selle emotsiooniga seostuvate assotsiatsioonide ja mälestuste meenumise.

Üldjoontes sarnane on ka kognitiivses lingvistikas omaks võetud nn entsüklopeedilise semantika kontseptsioon (vt nt Evans, Green 2006, Langacker 1987). Viide entsüklopeediale ei tähenda siin mitte autoriteeti ja üldkehtivust, vaid semantiliste tegurite hõlmavust. Üksiksõna vaadeldakse sisendsõlmene (ingl *access node*) tervesse teadmiste ja kogemuste võrgustikku, milles tippudena esinevad entiteedid ja kaartena ühendavad neid omavahelised suhted (vt Langacker 1987). Entsüklopeediline semantikakäsitlus ühendab teadmisi ja kogemusi, mille subjekt on omandanud, sõltumata omandamise viisist, ajast ja "õigsusest" ning hõlmab kindlasti mõningat varieerumist isikuti. Veelgi lähemal mõistete subjektiivsetele representatsioonidele seisab uusempiristlik simulatsiooniteooria. Väidetakse, et sõnavaraüksus kutsub kognitiivsete representatsioonide tasandil esile tähistatava situatsiooni simulatsiooni, selle osalise "läbimängu" aju multimodaalse kogemuse representatsioonide "teatris". Mingi sõna "töötlemisel" aktiveeritakse samade tajumodaalsustega seotud keskused, mis on kasutusel sama sõnaga tähistatavas situatsioonis osalemisega või selle vaatlemisega. (Barsalou 2003, Evans, Green 2006, Feldman Barrett 2006, Gallese, Lakoff 2005.) Paljude, kui mitte kõigi loomuliku keele mõistete olemuslikuks osaks on situatsioonile "läbimängus" antav hinnang selle kohta, kas seda nähtust tuleks pigem vältida või on see huvipakkvalt kasutootav. Need hinnangud võivad ilmnedada nt teadvustamata liikumistendentsina kas objekti suunas või sellest eemale.

Kokkuvõttes võib öelda, et lähenemised erinevad selle poolest, kas sõnade emotsionaalseid omadusi peetakse pigem osaks keele (juba realiseerunud) struktuurist või osaks inimese mentaalsest struktuurist, kus see esineb alles potentsiaalina. Vastusest sellele küsimusele olenevad meetodid, mida sõna tähenduse ja kaastähenduste määramisel soovitatakse kasutada. Esimesel juhul (traditsiooniline ja leksikograafiline lähenemine) arvatakse sõnavaraüksuse tähendus ja sellega kaasnev tundetoon sugenevat kasutuskontekstist ja seega tuleb tähenduste määramisel võtta arvesse eelkõige andmeid keele(kasutuse) seest. Teisel juhul (psühholoogiline, psühholingvistiline, kognitiivne ja simulatsiooniteooria) usutakse pigem, et omadused kuuluvad sõnavaraüksuse juurde ka ilma lingvistilise kontekstita, seosed esinevad tähenduste ja kaastähenduste representatsioonide vahel pigem assotsiatsioonidena ja seega on nende tuvastamiseks õigustatud kaasata ekstralingvistilist informatsiooni (Valitutti jt 2004). Käesolevas artiklis kirjeldatud uurimuse aluseeldused sobivad kokku kõigi lähenemistega peale traditsioonilise ja leksikograafilise.

2.2. Millist meetodit valida?

Nagu öeldud, lähtutakse käesolevas uuringus teoreetilisest seisukohast, et sõnad on abstraktsed sümbolid, mille omadused on tuvastatavad ka ilma konteksti ja kasutusituatsioonita. Sõnavara tundetoonide määramise ülesanne kujutab endast ülesannet eksplitseerida seos, mis on sõnal kui stiimulil ja selle emotsionaalsel tähendusel kas siis semantilisel, kogemuslikul või tajude simulatsiooni tasandil. Järgnevas tutvustatavad meetodid tuginevad valdavalt semantilistele seostele ja inimeste kas eksplitsiitsetele või implitsiitsetele teadmistele, ainult kaks viimasena tutvustatavat meetodit kasutavad emotsioonidega kaasnevaid kogemuslikke ja aju tasandil esinevaid multimodaalsuse efekte.

Introspektsioon on meetod, mille puhul uurija tugineb enda teadmistele, kogemustele ja keelepädevusele. Introspektsioon on aktsepteeritav uurimismeetod näiteks semantikas (vt Huumo 2008). Samas, kuivõrd tundetoonid on osaliselt ekstrasemantiline nähtus, millest on haaratud mitte üksnes mõisteliste seoste võrgustik, vaid ka individuaalsed emotsionaalsed kogemused ning tajusimulatsioonid, ei pruugi ühe uurija sisevaatlus anda piisavalt usaldatavat tulemust. Lisaks eri uurijate hinnangute võimalikule mittekokkulangemisele ei pruugi anda korratavat tulemust ka sama uurija introspektsioonilised hinnangud eri aegadel. Nende karide ületamiseks kasutatakse nt psühholoogias kahe sõltumatu eksperdi meetodikat ja usaldatakse seda osa tulemustest, kus hinnangud kattuvad (vt nt Aavik, Allik 2002). Võidakse kasutada ka nn domineeriva arvamuse meetodit, mispuhul eksperte on kolm ja arvesse läheb see arvamus, mida on kaks (vt Pennebaker jt 2001). Introspektsiooni usaldatavust saab tõsta ka korduvate hindamiste seeriat kasutades, kus vahepeale jäävad piisavalt pikad ajavahemikud ning arvesse lähevad sarnaselt langetatud otsused.

Introspektsiooni implitsiitsete aspektide esile toomiseks on lingvistikas käibel tehnikad, mida on nimetatud mõnevõrra erinevalt: propositsiooniline meetod (Pavlenko 2008a), lausete kontrasteerimise meetod (Cruse 2000), süntaktilis-grammatilised testid (Orav 2006). Nende olemus seisneb selles, et uuritav sõna pistetakse mingisse etteantud lüngaga lausesse, mille vastuvõetavust uurija siis oma introspektsioonile tuginedes hindab. Nii võiks üritada testida näiteks positiivset tundetooni lausega “X on hea ja ilus” ja negatiivset lausega “X on kole ja paha”. Sõna *sõda* näiteks ei sobi esimesse lausesse, aga sobib teise. Küsimus on selles, et mida võimaldaks taoline test teada saada sellist, mida me sõja kohta niigi ei teaks? Sõna *vaidlus* näiteks taolise testiga ära määrata ei õnnestu – kumbki lause poleks eriti mõttekas, kuigi põhimõtteliselt mitte võimatu. Taolised tehnikad põhinevad asjaolul, et sõna eri tähendused on keelesüsteemis sõna teiste tähendustega erinevates suhetes. Kui tundetoon (positiivsus vs. negatiivsus) on osa tähendusest, võib loota seda taoliste testidega määrata. Kui tegu on aga pragmaatilise konnotatsiooniga, mis tähendusele n-ö hinnangulise suhtumisena “külge kleepub”, pole taolistest testidest abi.

Küsitlusmeetodi puhul välistatakse hinnangute subjektiivsus, kaasates hindamisprotsessi rohkem inimesi; enamasti kasutatakse mitte-eksperte, et saada teada n-ö argitasandi teadmist. Küsitlused võivad olla avatud või suletud küsimustega, kirjalikud või suulised. Näiteks nn välimeetod näeb ette vähemalt 25 kuni 30 inimese kaasamist ja suulist intervjuerimist (Sutrop 2001). Korralikult koostatud ja läbiviidud küsitluste puhul on tulemuste kehtivus ja usaldatavus väljaspool kahtlust, ainsaks puuduseks on selle meetodi ressursimahukus. Olemuselt küsitlus on ka semantilise diferentsiaali meetod (Osgood jt 1975 [1957]). Katseisikutel (vähemalt 30, parim 100) lastakse sõnu/mõisteid/nähtusi hinnata polaarsetest adjektiividest koostatud 7-palliste skaalade suhtes. Ühe üksuse hindamiseks läheb vaja minimaalselt 12 skaalat, kuivõrd iga väidetavalt universaalse dimensiooni (hinnang, aktiivsus, tugevus) mõõtmise jaoks läheb tarvis 4 skaalat, mis peavad olema kohandatud vastavalt uuritavale objektile. Semantilise diferentsiaali kriitikana on välja toodud, et selle usaldatavust mõjutab sõnade polüseemia ja sotsiaalne soovitavus – isikud kalduvad andma hinnanguid, mida nad peavad soovitatavateks (Heise 1970). Teise puudusena võib mainida selle ressursinõudlikkust, mistõttu selle rakendamine

nii suuremahulisele ainestikule nagu seda on põhisonastik võtaks tohutult aega ja tööjõudu. Ainuüksi minimaalse skaalade ja katseisikute arvuga tuleks langetada ligi 1,5 miljonit hinnangut. Veelgi suurejoonelisem on patenteeritud süsteem¹, mille on välja töötanud W. Chase ja Connotative Intelligence Corporation². Selles süsteemis kasutatakse 100–200 mõningase ettevalmistusega isikust koosnevat hindajate “tiiki”, millest siis mõne konkreetsema sõnakogumi jaoks valitakse juhuslikult, aga esindatuse printsiipi järgides (mehed-naised, vanuseastmed) 24 hindajat. Hindamised toimuvad mitmes astmes ja väga diferentseeritud kategooriate süsteemi järgi. Tulemused esitatakse visuaalsete profiilidena. Hindajatel on hindamiseks spetsiaalne tarkvara, mille vahendusel nad hindamist sooritavad. Positiivsed ja negatiivsed emotsioonid ei välista üksteist, vaid profiil kajastab kogu arvamuste spektrit.

Eraldi meetodite grupi moodustavad füsioloogilised inimkatsed, näiteks tuginemine naha galvaanilise reaktsiooni muutuste tuvastamisele. Kui sõna on emotsionaalse laenguga, siis nahareaktsioon muutub (vt Pavlenko 2008a ja tema viiteid). Jällegi peab usaldatava tulemuse jaoks olema katseisikuid hulgi. Moodne meetod on ka ajulainete mõõtmine (EEG). Selle rakendamine sõnade tundetoonide määramiseks tugineks mõistete multimodaalsuse ja simulatsiooniteooriatele. Sõnade emotsionaalsete tähendustega kaasnevaid ajulaineid on mingil määral uuritud Hiinas (Skrandies, Chiu 2003). Enne eestikeelsete sõnade mõõtma asumist oleks vaja aga ulatuslikke katseid ja võrdlusmaterjali selle kohta, kas ja mil määral ajulainete muster sõnaliste stiimulite peale muutub. Ka selle meetodi jaoks on vaja palju katseisikuid ja lisaks on katseseadmed kallid.

Eesti keele põhisonavara sõnastiku märksõnade (üle 4500 üksuse) tundetoonide määramisel jäävad esialgu paratamatult kõrvale füsioloogilised inimkatsed, nimekirja teisest otsast jääb vähese usaldatavuse tõttu vaatluse alt välja lihtne introspektsioon, ressursimahukuse tõttu jääb esialgu kõrvale ka semantiline diferentsiaal ja sellel põhinevad tehnikad. Juhtumiuuringus lähtusime meetodi ressursinõudlikkuse ja usaldatavuse optimaalsest suhtest ja rakendasime põhisonastikule domineeriva arvamuse meetodit pisut modifitseeritud kujul, mida kirjeldatakse täpsemalt allpool.

2.3. Mida ja kui täpselt määrata?

Kui ülesandeks on tundetoonide määramine konkreetse sõnastikuprojekti raames, siis tuleb paratamatult ka küsida, millisel kujul peab olema tulemus. Kas iga märksõna kohta tuleb näidata täpselt, millise emotsiooniga ta tüüpiliselt seostub? Või on piisav, kui näidata ära, kas sõna kaldub olema valdavalt positiivse või negatiivse tundetooniga? Kas tundetoonide esitamisel põhisonastikus lähtuda psühholoogilisest realistlikkusest, keeleõppija vajadusest või keeletehnoloogilise rakenduse (tekstist emotsiooni tuvastamise) huvidest?

Psühholoogilistes uuringutes on näidatud, et diskreetsed emotsioonikogemused (ja -kategooriad nagu hirm, rõõm, uhkus) pole psühholoogiliselt primitiivsed. Baasilisemad on emotsiooni meeldivust ja tugevust iseloomustavad skaalad valents (hinnangulisus) ja aktivatsioon (Feldman Barrett 2004). Inglise keele kohta on sellistel eristustel põhinevaid nn afektisõnastikke ka tehtud. Üsna varsti pärast semantilise diferentsiaali meetodi leitudamist (Osgood jt 1957) olid selle meetodiga

¹ Patenteeritud USA-s “System for identifying connotative meaning, kood: US 6,389,415 B1”.

² <http://www.connotative.com/EmoDict.htm> (25.04.2011).

mõõdetud 1000 inglise keele kõige sagedasema sõna profiilid valentsi, tugevuse ja aktivatsiooni suhtes (Heise 1965). Suuremat hulka märksõnu (4000) sisaldab C. Whisselli (1989) sõnastik, milles on sõnade mõju mõõdetud seitsmepallisel skaalal meeldivushinnangute ja aktivatsiooni tugevuse suhtes.

Keeleõppija jaoks, eriti ühiskondades, milles leidub erineva etnilise ja rassilise või religioosse grupikuuluvuse taustaga isikuid, on tähtis teada võõras kultuuris sõnadega kaasneva võivaid konnotatsioone, mis on seotud pigem kultuuriliste väärtushinnangutega ja stereotüüpidega kui individuaalpsühholoogiliste eristustega³. Kakskeelsetes sõnastikes pole märksõnade tundetoonide läbivat märkimist siiski seni teadaolevalt kasutatud. Tavalisem on stilistiliselt markeeritud keelendite äramärkimine stiilmärgenditega nagu *halvustav*, *hellitlev*, *pidulik* vmt. Kuna valdav osa keele sõnavarast on siiski tähendusvarjundi poolest neutraalne, oleks tundetooni läbiv märgendamine sõnastiku kasutajale pigem koormav lisainformatsioon, mida tal suhteliselt harva vaja läheb. Küll on sellest kasu keeleõppijale, kes oma väljendusoskust soovib lihvida.

Keeletehnoloogilistest vajadustest lähtuvates lähenemistes kirjeldatakse sõnavara afektiivseid lisatähendusi kompleksse kategooriate süsteemi abil. Näiteks nn Harvardi nimekirja⁴ iseloomustavast 182-st karakteristikust on tervelt 20 seotud emotsionaalselt relevantsete eristustega (positiivne ja negatiivne, toetav ja vägi-valdne hoiak, nõrk, tugev, võim, alistumine, valu, mõnu, tunne, emotsionaalne, erutus, voores, pahe, rõhutamine, vähendamine). Teise näitena nn Clairvoyance'i afektisõnastikus koosneb märksõna (arv 3772) kirjeldus viiest väljast: a) lemma; b) lihtsustatud sõnaliik; c) afektiklassi määrang (peensusteni välja arendatud, 86 erinevat); d) sõna kesksuse kaal antud afektiklassi liikmena (vahemikus 0–1); e) sõna intensiivsuse kaal antud afektiklassi liikmena (0–1). Sõnastik on valdavalt koostatud ja märgendatud käsitsi, toetudes koostajate keelepädevusele (Grefens-tette jt 2006: 96).

Kuivõrd põhisõnastiku tundetoonide määramise ülesanne lähtub kahest eri vajadusest, siis peab see olema: a) piisavalt informatsiooniküllane, et rahuldada keeletehnoloogilise rakenduse vajadusi; b) keeleõppijale soovi korral käepärast, ent mitte koormama teda ülearuse informatsiooniga. Nendele tingimustele näib enim vastavat psühholoogilisele realismile iseloomulik kirjelduste granulaarsuse tase, kus ei eristata konkreetseid emotsioonikategooriaid (nagu hirm, viha, rõõm, armastus jne), kuid eristatakse positiivset ja negatiivset tundetooni ning aktivatsiooni taset, s.t neutraalse ja tundetoonilise keelendi eristamist.

Põhisõnastikku koostatakse Eesti Keele Instituudi veebipõhises sõnastikusüsteemis **ΞLex**. Põhisõnastiku kasutajaliideses tuleb keeleõppija jaoks leppemärgiga (nt emotikoniga, tähtlühendiga) näidata märksõna potentsiaalne tundetooni väärtus (positiivne, negatiivne või vastuoluline), neutraalsust kui vaikeväärtust pole vaja näidata. Keeletehnoloogilise rakenduse (tekstist emotsiooni tuvastamise) tarbeks sisaldugu andmebaasis aga rohkem välju. Lisas 1 on esitatud vaade põhisõnastiku toimetamiskeskonnast, kus “semantikagrupile” järgneb “emotsiooni-grupp” ehk tundetooni määrangute tööväli. Nagu näha, on selles lahtreid rohkem kui üksnes neutraalsus vs. tundetoon (ning võimalus määrata viimase iseloomu kas positiivseks, negatiivseks või vastuoluliseks). Vajadust ülejäänud väljade järele selgitatakse artikli viimases osas.

³ <http://www.connotative.com> (25.04.2011).

⁴ <http://www.wjh.harvard.edu/~inquirer/Home.html> (16.04.2010).

3. Eesti keele põhisõnavara sõnastiku tundetoonide määramine

Määramise hetkel koosnes põhisõnastik 3015 märksõnast, puudu olid ase- ja sidesõnad. Peamiselt määrati nimisõnu, verbe, adjektiive, määrsõnu, numeraale – sõnu, mis enam-vähem mahuvad kategooria “täistähenduslikud sõnad” alla. Sõnastiku märksõnade tähendusjaotus polnud määramise faasis veel tehtud, kuid see pole määratavate tundetoonide rakendusi arvestades ka oluline, kuivõrd ei arvuti, mis teksti automaatselt analüüsima hakkab, ega eesti keelt võõrkeelena õppija, kes tekstis tundmatut sõna kohtab, ei tea sissejuhatuses, et tegu võiks olla polüsemse sõnaga ja milline alltähendustest just antud tekstis on relevantne. Lähtuti põhimõttest, et kui sõnal on tundetoon kas või ühes alltähenduses, siis arvuti/keeleõppija jaoks on otstarbekas teda sellest hoiatada. Tundetoon on miski, mis võib, aga ei pruugi sõna kasutamise kaasneda.

3.1. Meetod

Kasutati nn domineeriva arvamuse meetodit (Pennebaker jt 2001), selle erinevusega, et hindajaid oli neli ja domineeriva arvamuse väljaselgitamine selle tõttu pisut komplitseeritum.

Hindajad. Hindajad olid eksperdid: kõrgharidusega lingvistid, kes omavad mõningaid taustateadmisi ka emotsioonidest ja nende avaldumisest mitmesugustel keeletasanditel. Kõik olid eesti keelt emakeelena kõnelejad ja eeldati, et nad jagavad meie kultuurile vaikumisi omast väärtuste ja emotsionaalsete hinnangute süsteemi. Hindamisel soovitati juhendada mitte isiklikest emotsioonidest, vaid kultuuriomasest normatiivsest hinnangust. Ekspertide töö oli tasustatud.

Hinnangud. Kõiki hinnatavaid sõnu tuli hinnata ühel järgnevast neljast viisist: p – positiivne, n – negatiivne, o – neutraalne; p/n – kui ei olnud neutraalne, kuid ka lihtsalt positiivne ja negatiivne ei tundunud sobivat (sisaldas mingit vastuolu, võimalik, et sõna oli näiteks polüsemne vmt).

Läbiviimine. Märksõnastik esitati elektroonse andmetabeli veergudes. Igaüks hindas omaette. Koormus jaotatud ühe nädala peale, keskmiselt tehti 600 hindamist päeva kohta. Kaks hindamiste seeriat, vahemik 2 kuud, rutiini vältimiseks sõnad teisel katsel pöördjärjestuses.

Andmetöötlus. Tabelis teisendati kogutud emotsionaalsushinnangud numbrilisteks väärtusteks järgnevalt: p = 1; n = -1; n/p = 0; neutraalsete hinnangute lahtrid jäeti tühjaks. Tulemusest filtreeriti välja kokkulangevad hinnangud ja märgiti need suhtelise üksmeelsuseindeksiga (SÜ) järgnevalt: 1 – absoluutselt domineeriv arvamus, s.t kõigi nelja hindaja arvamus langes kokku; 0,75 – domineeriv arvamus (kolm ühe vastu); 0,6 – suhteliselt domineeriv arvamus (kaks kokkulangevat arvamust kahe omavahel lahkneva arvamuse vastu); 0,5 – vastuoluline arvamus (arvamused jagunevad pooleks kaks kahe vastu); 0,25 – ülivastuoluline arvamus (kõik neli on erineval arvamusel).

Tulemuste arvessevõtt. Märgendamise suhtes olulise domineeriva arvamusena (positiivne, negatiivne, neutraalne) läksid arvesse need, mille SÜ oli suurem või võrdne 0,6. Vastuoluliseks märgiti juhtumid, mille SÜ = 0,5 või vähem ning need

juhtumid, kus hindajad olid üksmeelsed tundetooni vastuolulisuse suhtes. Domineerivad arvamused selgitati kahes katses eraldi ja siis võrreldi neid omavahel.

Lõplik märgendusettepanek. Lähtuti seisukohast, et neljase valikvastuse puhul on piisav tuvastusprotsent 50% (Vogt jt 2008). Pärast kahe katse võrdlust tehti järgnev ettepanek: stabiilselt positiivne, negatiivne ja neutraalne – märkida need, mille keskmine üksmeelsuseindeks kahes katses jäi vahemikku 0,6–1; vastuoluliseks märkida need, mis a) said stabiilselt üksmeelseid hinnanguid vastuolulisuse kohta (keskmine indeks vahemikus 0,6–1) ja b) said ebastabiilseid hinnanguid (üksmeelsuseindeks väiksem kui 0,6) kahe katse ja isikute lõikes.

3.2. Tulemused ja arutelu

Tabel 1 iseloomustab domineerivalt positiivseid, negatiivseid, neutraalseid ja vastuolulisi hinnanguid saanud arvamuste jaotumist eri katsetes, arvamuste muutumist kahe katse vahel ning üldkokkuvõttes (märgendusettepanek). Numberid iseloomustavad üldisi hinnangute proportsioone ega ütle midagi üksikute sõnade ega üksikute hindajate hinnangute stabiilsuse/muutlikkuse kohta. Tabel 1 näitab põhisõnastiku märksõnade tundetoonide distributsioonide üldist stabiilsust: laias plaanis pisut enam kui 60% sõnadest on täiesti neutraalsed ega vaja seega põhisõnastikus tundetooni märgendit. Ülejäänud 40% sõnadest jagunevad positiivse, negatiivse ja vastuolulise märgendi vahel enam-vähem võrdsetes proportsioonides. Enim muutlikkust näitab üles vastuoluliste hinnangute kategooria. See on põhjustatud asjaolust, et üks hindajatest-ekspertidest loobus teises katses võimalusest märkida tundetooni vastuolulisust. Kuivõrd see kategooria on n-ö “prügikast”, kuhu sattusid lisaks üksmeelsele vastuolu hinnangule ka need sõnad, millede osas kas katseisikute või hindamisseeriade lõikes oli enim varieeruvust, siis lõplikes märgendusettepanekutes selle ühe hindaja strateegiamuutus arvatavasti olulist mõju ei avaldanud. Küll aga osutus õigeks otsus kasutada kolme eksperdi asemel nelja eksperti, kuna praeguse protseduuri korral kõnealune “eksperiment eksperimentis” lõpptulemusi (võrreldes esimese katse proportsioonidega) oluliselt siiski ei muutnud.

Tabel 1. Tundetoonide distributsioon hindamise eri seeriates

Tundetoon	I hindamine		II hindamine		Muutus (n I – n II)		Märgendusettepanek	
	n	%	n	%	n	%	n	%
neutraalne	1978	65,61	2091	69,35	113	5,71	1898	62,95
positiivne	411	13,63	488	16,19	77	18,73	373	12,37
negatiivne	366	12,14	420	13,93	54	14,75	356	11,81
vastuoluline	260	8,62	16	0,53	-244	-93,85	388	12,87

Tabel 2 esitab näiteid sõnade kohta, mis said eelpool kirjeldatud domineeriva arvamuse meetodi alusel määratud kas neutraalseks, positiivseks, negatiivseks või vastuoluliseks. Iga veeru ülemine ots sisaldab sõnu, mille määrangute suhtes valitses kõige suurem üksmeel (SÜ = 1,00 tähendab seda, et kõik neli eksperti olid samal arvamusel). Arvatavasti valitseb nende sõnade suhtes tõesti kultuuriline

hoiakute konsensus, s.t positiivne on nende sõnade tundetoon, mis tähistavad meil kultuuriliselt üldaktsepteeritud positiivseid väärtusi, omadusi, tegevusi, emotsioone ning negatiivne nende sõnade tundetoon, mille denotaatide suhtes on kultuuriliselt tõrjuvad hoiakud. Positiivseks hinnatud sõnade skaala teises otsas ($S\ddot{U} = 0,60$, s.t suhteliselt domineeriv arvamus kaks sarnast kahe erineva vastu) leiduvad sõnad, mis ei paikne ilmselt kultuuriliselt positiivsete nähtuste kategooria tuumosas, vaid pigem perifeerias. Siin on konsensus vaid osaline. Sama kehtib negatiivsete tundetoonide skaala alaotsa kohta. Põhjused ilmselt erinevad – võib-olla sõnade polüseemsus, või hindajate maailmavaade (nt *jumal*), suhtumine kehakaaluküsimustesse (nt *paks*). Samamoodi leiab siit sõnu, mille hinnangulisus on alles välja kujunemas (nt *neeger*).

Igal juhul annab hinnangute keskmise üksmeelsuse määr informatsiooni selle kohta, kui võrd keskne on antud sõna positiivsete ja negatiivsete tundetoonidega sõnade kategooria liikmena. See informatsioon tuleks andmebaasi kaasata keeletehnoloogilise rakenduse tarbeks (eeskujuna Grefenstette jt 2006). Lisas 1 toodud tööväljal võiks lahter “tugevus” olla koht, kus näiteks suhtelise üksmeelsuseindeksi näol iseloomustada antud hinnangute tugevust.

Neutraalsete sõnade veeru ülemine ots ($S\ddot{U} = 1,00$) ei tekita ilmselt kelleski küsimusi, skaala teises otsas on aga sõnad, mille neutraalsus pole sama üksmeelselt hinnatud ($S\ddot{U} = 0,60$). Kahe hindaja jaoks neljast iseloomustas neid sõnu siiski mingisugune tundetoon, puudus aga üksmeel. Selles osas on neutraalse kategooria ääreala sõnad võrreldavad vastuolulisi hinnanguid saanud sõnade kategooriaga. Ratsionaalselt mõeldes ei tundu olevat kuigi suurt erinevust sõnade *ärimees* (neutraalne, $S\ddot{U} = 0,60$) ja *jahimees* (vastuoluline, $S\ddot{U} = 0,80$) võimalike konnotatsioonide vahel. Kummagi puhul on ühiskonna suhtumine mingil määral ambivalentne, nende elukutsete esindajate tegevusse mahub teatud huvigruppide poolt nii moraalseks kui ebamoraalseks peetavaid tegevusi. Vastuoluliste tundetoonidega sõnade veeru alumises otsas on sõnad, mille hinnangud olid kõige ebastabiilsemad, varieerudes suuresti nii hindajate kui hindamisseeuriate lõikes. Nii näiteks leiab sealt sõnad, mille tähenduses sisaldub implitsiitselt nii negatiivseid elemente (*appi*, *hoiatama* ja *hoiduma* – kõik seotud mingi hädaohuga), kuid samas ka positiivseid elemente (ohu vältimine). Vastuolulisus, mis on sees juba sõna semantikas, peegeldub ka tundetoonile antavates hinnangutes. Osa afektiivse suhtumise ambivalentsist võib tuleneda sellest, et emotsioonidel on valents vähemalt kahes suhtes: kas kogeva subjekti Ego suhtes või kohalike kultuurinormide suhtes. Esimene väljendab emotsiooni hedoonilist kvaliteeti (meeldiv vs. ebameeldiv) ja teine pigem kultuurist tingitud moraalseid hinnanguid (õilis vs. taunimisväärne). Need kaks dimensiooni ei ole alati vastavuses. (Le Guen 2009, Colombetti 2005)

3.3. Võrdlus varasemate uurimustega

Põhisõnastiku märksõnade hulgas oli 34 kokkulangevat sõna nendega, mida 100 mõlemast soost ja eriealist eestlastest katseisikut produtseerisid vastavalt kas siis positiivsete või negatiivsete emotsioonide ja isikuomaduste vaba loetlemise katsetes (vt täpsemalt Vainik 2002 ja Orav 2006). See võrreldav osa moodustab küll kõigest 1,1% põhisõnastiku mahust, kuid mingisugust tulemuste kokkulangevuse

Tabel 2. Näiteid tundetoonilistest ja neutraalsetest sõnadest

Neutraalsed	SÜ	Positiivsed	SÜ	Negatiivsed	SÜ	Vastuolulised	SÜ
aastane	1,00	aitama	1,00	alkohoolik	1,00	jahimees	0,88
edastama	1,00	energiline	1,00	avarii	1,00	tervis	0,88
eraisik	1,00	hoolima	1,00	haigus	1,00	töötasu	0,88
ettevõte	1,00	inimlik	1,00	hirm	1,00	armee	0,80
jalgratas	1,00	issi	1,00	hukkuma	1,00	efekt	0,80
keetma	1,00	edu	1,00	hävitama	1,00	kindlustus	0,80
kohver	1,00	julgus	1,00	inetu	1,00	koolilaps	0,80
kujunema	1,00	medal	1,00	karistama	1,00	kunst	0,80
osutama	1,00	lemmik	1,00	korruptsioon	1,00	kunstnik	0,80
porgand	1,00	mõnus	1,00	kurbus	1,00	naljakas	0,80
samm	1,00	pilvitu	1,00	maailmasõda	1,00	odav	0,80
sügis	1,00	puhkus	1,00	mõrv	1,00	pikaajaline	0,80
tellimus	1,00	rõõm	1,00	peavalu	1,00	pikem	0,80
tähendama	1,00	tervislik	1,00	pori	1,00	rebima	0,80
vahemaa	1,00	unistus	1,00	segane	1,00	saun	0,80
võrdlema	1,00	õnnelik	1,00	süü	1,00	spetsiaalne	0,80
ööpäev	1,00	isamaa	1,00	töötus	1,00	suurendama	0,80
sõlmima	1,00	eestimaalane	0,60	vaev	1,00	suvila	0,80
meelitama	0,60	jumal	0,60	varas	1,00	sõrmus	0,80
peremees	0,60	kodune	0,60	vihane	1,00	sünnikoht	0,80
põhjustama	0,60	kohvik	0,60	hall	0,60	tunnistus	0,80
raha	0,60	kuld	0,60	hapu	0,60	töomees	0,80
saatus	0,60	lihtne	0,60	jahe	0,60	valmis	0,80
teater	0,60	mõistus	0,60	külmetama	0,60	kulutama	0,55
tigu	0,60	neitsi	0,60	langus	0,60	perekond	0,55
tsirkus	0,60	parandus	0,60	loosung	0,60	romaan	0,55
tuli	0,60	populaarsus	0,60	neeger	0,60	sigat	0,55
tädi	0,60	sirge	0,60	paks	0,60	viin	0,55
unenägu	0,60	suutma	0,60	süütama	0,60	ülemus	0,55
vari	0,60	tulevane	0,60	tume	0,60	tagajärg	0,55
võistlus	0,60	usk	0,60	vallutama	0,60	sünnitama	0,50
õrn	0,60	vabandama	0,60	võim	0,60	appi	0,43
ärimees	0,60	vähenoudlik	0,60	õhuke	0,60	hoiatama	0,43
viitsima	0,55	üllatama	0,60	õppemaks	0,60	hoiduma	0,38

kontrolli saab nende põhjal siiski teha. Kolmekümnel juhul (88%) langesid loetelukatsetes moodustunud positiivse ja negatiivse hinnangu kategooriad kokku tundetoonihinnangutega käesolevas uurimuses kirjeldatud domineeriva arvamuse meetodil. Seega võib väita, et nelja eksperdi arvamused langesid nende kolmekümne nelja sõna osas valdavalt kokku 200 (kahe uurimuse peale kokku) tavakeelekasutajast katsealuse arvamusega, mida on põhjust pidada representatiivseks. Pole põhjust eeldada, et ekspertide arvamust peaks vähem usaldama ka ülejäänud märksõnastiku sõnade tundetooni hindamisel.

Mingis väga piiratud osas saab põhisõnastiku hinnatud tundetooni võrrelda ka emotsioonisõnade semantika uuringu tulemustega, milles Vainik (2004) lasi 100 (võrdselt eri soost ja eri vanuses) katseisikul sõnu hinnata kokku seitsmel polaarsetest omadustest koosneval skaalal. Meetod on väga lähedane semantilise diferentsiaali meetodiga, mida muidu peetakse kõige usaldusväärsemaks sõnade hinnangulisuse mõõtmise meetodiks (vt eestpoolt). Võrrelda saab nendest tulemustest hinnanguid skaalal meeldiv-ebameeldiv, mida võib samastada emotsioonide negatiivse ja positiivse tundetooniga. Siin oli põhisõnastikuga kokkulangevaid sõnu 14 ja nende hinnangud langesid 100% ulatuses kokku ekspertide hinnangutega käesolevas uurimuses, kaasa arvatud sõnale *uhkus* iseloomulik ambivalents. Kuigi üldistustega peab olema ettevaatlik, kinnitab siiski ka see võrdlus, et käesolevas uurimuses kasutatud meetodikat ja eksperte võib pigem usaldada kui mitte.

4. Kokkuvõte ja üldine arutelu

Kasutatud meetodi positiivse küljena võib mainida selle suhteliselt vähest ressursinõudlikkust – kulus *ca* 4 inimtöönädalat. Teiseks äratub usaldust suur kokkulangevus tulemuste osas, mis olid võrreldavad varasemate uuringutega. Kolmas hea näitaja on see, et antud meetodiga saadud määrangutulemusi on provisoorselt juba rakendatud tekstilõikude emotsionaalse valentsi automaatseks hindamiseks ja sealjuures küllaltki edukalt (Pajupuu jt, ilmumas). Osutub, et määrangud valitud granulaarsuse tasemel – neutraalsus *vs.* tundetoon ning viimase eristused positiivne, negatiivne ning vastuoluline – on piisavad keeletehnoloogilise rakenduse jaoks, mis kalkuleerib pikemate tekstilõikude emotsionaalset valentsi.

Keeletehnoloogilise lahenduse jaoks oleks veelgi selgem, kui sõnade emotsionaalsed valentsid oleks polaarsemad, s.t sõna tundetoon jaguneks selgelt positiivseks ja negatiivseks ega oleks vastuoluline, kuna viimasega on tekstilõigu emotsionaalse valentsi kalkuleerimisel vähem peale hakata. Seega võidakse tunda vajadust sõnavara “ülepolariseerida” – omistada hinnanguid “positiivne” ja “negatiivne” enam, kui see on reaalse elu hinnangute ambivalentsile mõeldes õigustatud. Ülepolariseerimist peaks autori arvates siiski kindlasti vältima, seda eriti keeleõppijast sõnastikukasutaja huve silmas pidades – talle võib eesti keel tunduda liiga komplitseeritud, kui sellega kaasnevad tugevad polaarsed hinnangud sõnade küljes, mille sisu sellist kalduvust ei eelda.

Problemaatiline on ka see, kas kirjeldatud domineeriva arvamuse meetodit oleks mõttekas rakendada märksõnastiku “mittetäistähenduslikele” sõnadele – ase- ja sidesõnadele. Kui näiteks sõna *ja* kohta võib eeldada teatud kalduvust positiivsuse poole (peaaegu homonüüm jaatussõnaga *jah*), siis seda, kas *sest* on pigem posi-

tiivne või negatiivne, oleks ekspertidel arvatavasti väga raske oma kultuurilisele ja keelelisele kompetentsile toetudes öelda. Lahendust tuleks siin otsida teistsugustest meetoditest. Näiteks võiks arvesse võtta selle sõnavara distributsiooni emotsioonide (ja nende valentsi) lõikes, mis esineb eesti keele emotsionaalse kõne korpus⁵ piisava kindlusega tuvastatud lausetes. Senine analüüs on näidanud, et näiteks kurbust väljendavates lausetes esineb tihti just oma “väikeste sõnade”, s.t ase- ja määrsõnade kasutamise muster (Vainik 2010).

Veel oleks arvatavasti keeletehnoloogilisel rakendusel (mitte aga niivõrd keeleõppijal) kasu, kui põhisõnastikus oleks lisaks tundetoonile määratud potentsiaalse afektikalduvuse klass. Afektikalduvuse all mõeldakse siin keeleüksuste võimet aidata kaasa emotsioonide (ja üldisemalt afektiivsuse) kirjeldamisele, väljendamisele, mõjutamisele ja tekstis tuvastamisele. Klass tähendab seda, et eri tüüpi sõnad annavad lause üldisesse afektiivsusesse erineva panuse olenevalt sellest, kas tegu on nn sisusõnadega, mis emotsioone kas nimetavad või väljendavad, või on tegu nn funktsioonisõnadega, mille rolliks on vahendada lause üldist ekspressiivsust. Seda, et emotsioonide väljendumine keeles hõlmab mitmeid keeletasandeid ja et ka sõnavara eri gruppidel on täita eri rollid, on märgitud nt tekstilingvistikas (Biber, Finegan 1989, Ochs, Schieffelin 1989).

Esialgul võiks lähtuda viiest klassist, mis emotsioonide väljendamisest või/ja kirjeldamisest põhimõtteliselt eri viisil oma panuse annavad: a) nn võtmesõnad ehk emotsionaalse denotatsiooniga sõnad (otsene emotsioonisõnavara nagu *viha, rõõm, kurbus* jne, emotsioonide väljendustegevused nagu *nutta, naer, kallistus* jne); b) nn leksikaalne külgetõmme ehk afektiivselt laetud sõnavara, sh nii tänu denotatiivsele kui konnotatiivsele hinnangule (nt omadussõnad *hea, halb, suur, kõhn, paks*, isikuomadused *laisk, tige, kade, töökas*, käitumisaktide nimetused nagu *valetamine, petmine, abistamine* jne), aga ka emotsioone tekitavad nähtused (*päike, loodus, sõda, infarkt*); c) ekspressiivne sõnavara (hüüdsõnad, nt *ah, ahah, ai, oh, ohoo, oi*), slängiväljendid, vormelid (*aitäh, appi*) jm; d) sõnavara, kus põimuvad ülalmainitud dimensioonid (nt piltlikud emotsioonikirjeldused, mis denotatsiooniga osutavad emotsiooni läbielamise protsessile, on aga ka ekspressiivsed (s.t kujundlikud, väljendusjõulised) ja lisaks tihti hinnangulised (nt *vihast vahutama, moka tõllakil*). Sedasorti leksikat põhisõnavara sõnastik praegu aga ei sisalda. Eraldi afektikallaku klass e) on intensiivsust, evidentsiaalsust, tõenäosusmodaalsust, deiktilisust jmt kõneleja ja kõnesündmuse sisu vahekorda esile toovad sõnad (*arvatavasti, väga, nii, võib-olla* jne). Lisas 1 esitatud väljale “afektikalduvuse klass” on mõeldud sisestada just sedasorti informatsiooni.

Afektikalduvuse näol on tegu hajusate piiridega ja koostiselt heterogeense kategooriaga, mida määratlev printsiip on funktsionaalne. See on hõlmavam, kui käesolevas artiklis käsitletud sõnavara tundetoonilisus ehk emotsionaalne valents. Afektikalduvuse arvessevõtt tundub perspektiivikas selles mõttes, et kuigi 75% jooksvast ajalehetekstist koosneb umbes 4000 sagedasemast sõnast (Pajupuu jt 2009), ei ole praegu veel selge, kui suur protsent eesti keele lausete n-ö “jooksvatest sõnadest” on täistähenduslikud ja kui suur on nende seas afektirehkendust võimaldavate mitteneutraalsete sõnade arv. Afektikalduvuse klasside arvessevõtt aitaks tuvastada ka neid lauseid, milles väljendub ainult emotsionaalne ekspressiivsus ja mitte ilmingimata hinnangulisus.

⁵ <http://peeter.eki.ee:5000/> (25.04.2011).

Problemaatiline on afektikalduvuse märkimise puhul aga asjaolu, et informatsiooni selle kohta saadaks sõnade tegelikust kasutamisest eesti emotsionaalse kõne korpusas ja seega on need andmed meetoodiliselt teist moodi kogutud ning nad erinevad ka teoreetilise lähte-eelduse poolest selle kohta, kas keelt tuleb uurida reaalses kasutuses või siis pigem potentsiaalina inimeste peades (vt ptk 1.1). Selleks, et erinevalt kogutud ja määratud informatsioon oleks andmebaasis ka erinevalt talletatud, on põhisonastiku “emotsioonigrupi” toimetamisvaates (vt lisa 1) ette nähtud lahter “määramise alus”. Sinna tuleb siis märkida alus, mille põhjal on otsus langetatud. Seega, kuigi on tehtud valik üht teatud tüüpi informatsiooni eelistamise ees, ei ole välistatud ka võimalus edaspidi toetuda teistsugustele meetoditele (vt ptk 1.2).

Lõpetuseks võiks öelda, et sissejuhatuses mainitud kahe praktilise ülesande lahendamine samaaegselt on kahtlemata võimalik, kuid tundetoonide määramist ja märkimist see ikkagi mõnevõrra komplitseerib. Teha tuleb valikuid, mis ei tundu kummagi ülesande seisukohast just ideaalsed. Lahenduseks on siin see, et sõnastiku andmebaasi väljadel sisaldub rohkem ja detailsemat infot, kui seda kuvatakse keeõppijast kasutajale.

Viidatud kirjandus

- Aavik, Toivo; Allik, Jüri 2002. The structure of Estonian personal values: A lexical approach. – *European Journal of Personality*, 16 (3), 221–235. <http://dx.doi.org/10.1002/per.439>
- Aitchison, Jean 2003. *Words in the Mind: An Introduction to the Mental Lexicon* (3rd ed.). Blackwell Publishing.
- Altrov, Rene; Pajupuu, Hille 2008. The Estonian emotional speech corpus: Release 1. – *The Third Baltic Conference on Human Language Technologies*. Vilnius: Vytauto Didžiojo Universitetas; Lietuvių kalbos institutas, 9–15.
- Atkins, B. T. Sue; Rundell, Michael 2008. *The Oxford Guide to Practical Lexicography*. Oxford: Oxford University Press.
- Barsalou, Laurence W. 2003. Situated simulation in the human conceptual system. – *Language and Cognitive Process*, 18 (5–6), 513–562. <http://dx.doi.org/10.1080/01690960344000026>
- Bestgen, Yves 1994. Can emotional valence in stories be determined from words? – *Cognition and Emotion*, 8 (1), 21–36. <http://dx.doi.org/10.1080/02699939408408926>
- Biber, Douglas; Finegan, Edward 1989. Styles of stance in English: Lexical and grammatical marking of evidentiality and affect. – *Text*, 9 (1), 93–124. <http://dx.doi.org/10.1515/text.1.1989.9.1.93>
- Bower, Gordon H.; Forgas, Joseph P. 2000. Affect, memory and social cognition. – Eric Eich, John F. Kihlström, Gordon Bower, Joseph P. Forgas, Paula M. Niedenthal (Eds.). *Cognition and Emotion*. Oxford: Oxford University Press, 87–168.
- Colombetti, Giovanna 2005. Appraising valence. – *Journal of Consciousness Studies*, 12 (8–10), 103–126.
- Cruse, Alan D. 2000. *Meaning in Language. An Introduction to Semantics and Pragmatics*. Oxford: Oxford University Press.
- Evans, Vyvyan; Green, Melanie 2006. *Cognitive Linguistics: An Introduction*. London: Lawrence Erlbaum.
- Feldman Barrett, Linda 2004. Feelings or words? Understanding the content in self-report ratings of experienced emotion. – *Journal of Personality and Social Psychology*, 87 (2), 266–281. <http://dx.doi.org/10.1037/0022-3514.87.2.266>

- Feldman Barrett, Linda 2006. Solving the emotion paradox: Categorization and the experience of emotion. – *Personality and Social Psychology Review*, 10 (1), 20–46. http://dx.doi.org/10.1207/s15327957pspr1001_2
- Foolen, Ad 1997. The expressive function of language: Towards a cognitive semantic approach. – Susanne Niemeier, Rene Dirven (Eds.). *The Language of Emotions: Conceptualization, Expression and Theoretical Foundation*. Amsterdam, Philadelphia: John Benjamins Publishing Company, 15–32.
- Gallese, Vittorio; Lakoff, George 2005. The brain's concepts: The role of the sensory-motor system in conceptual knowledge. – *Cognitive Neuropsychology*, 22 (3-4), 455–479. <http://dx.doi.org/10.1080/02643290442000310>
- Grefenstette, Gregory; Qu, Yan; Evans, David A.; Shanahan, James G. 2006. Validating the coverage of lexical resources for affect analysis and automatically classifying new words along semantic axes. – James G. Shanahan, Yan Qu, Janyce M. Wiebe (Eds.). *Computing Attitude and Affect in Text: Theory and Applications*. Dordrecht: Springer, 93–107.
- Heise, David R. 1965. Semantic differential profiles for 1000 most frequent English words. *Psychological Monographs*, 79 (601). American Psychological Association.
- Heise, David R. 1970. The semantic differential and attitude research. – Gene F. Summers (Ed.). *Attitude Measurement*. Chicago: Rand McNally, 235–253.
- Huumo, Tuomas 2008. Joko lingvistin nojatuoli joutaisi kaatopaikalle? Introspektio-lingvistiikan asemasta kognitiivisessa kielitieedessä. – *Emakeele Seltsi aastaraamat*, 53 (2007), 163–180.
- Hübler, Axel 1998. *The Expressivity of Grammar. Grammatical Devices Expressing Emotion Across Time*. Berlin, New York: Mouton de Gruyter.
- Iida, Akemi; Campbell, Nick; Higuchi, Fumito; Yasumura, Michiaki 2003. A corpus-based speech synthesis system with emotion. – *Speech Communication*, 40 (1-2), 161–187. [http://dx.doi.org/10.1016/S0167-6393\(02\)00081-X](http://dx.doi.org/10.1016/S0167-6393(02)00081-X)
- Kallas, Jelena; Tuulik, Maria 2011. Eesti keele põhisõnavara sõnastik: ajalooline kontekst ja koostamispõhimõtted. – *Eesti Rakenduslingvistika Ühingu aastaraamat*, 7, 59–76. <http://dx.doi.org/10.5128/ERYa7.04>
- Langacker, Ronald W. 1987. *Foundations of Cognitive Grammar I. Theoretical Prerequisites*. Stanford: Stanford University Press.
- Le Guen, Olivier 2009. The ethnography of emotions: A field worker's guide. – Asifa Majid (Ed.). *Field Manual*, Vol. 12. Nijmegen: Max Planck Institute for Psycholinguistics, 31–34.
- Ochs, Elinor; Schieffelin, Bambi 1989. Language has a heart. – *Text*, 9 (1), 7–25. <http://dx.doi.org/10.1515/text.1.1989.9.1.7>
- Orav, Heili 2006. *Isiksuseomaduste sõnavara semantika eesti keeles*. *Dissertationes Linguisticae Universitatis Tartuensis*, 6. Tartu: Tartu Ülikooli kirjastus.
- Osgood, Charles; Suci, George. J.; Tannenbaum, Percy. H. 1975 [1957]. *The Measurement of Meaning*. Urbana, Chicago: University of Illinois Press.
- Pajupuu, Hille 1994. A nation's autostereotype. – *Oral Memory & National Identity. Papers of the International Conference held in Tallinn, September 18–19, 1993*. Tallinn, 31–39.
- Pajupuu, Hille; Kerge, Krista; Alp, Pilvi 2009. Sõnavara loomulik rikkus haritud keeleoskaja tekstides. – *Eesti Rakenduslingvistika Ühingu aastaraamat*, 5, 187–196. <http://dx.doi.org/10.5128/ERYa5.12>
- Pajupuu, Hille; Kerge, Krista; Altrov, Rene (ilmumas). Detecting emotional valence of text by using a small dictionary. – *Empiricism and Analytical Tools for Applied Linguistics in the 21st Century*. Salamanca: Universidad de Salamanca.
- Pavlenko, Aneta 2008a. Emotion and emotion-laden words in the bilingual lexicon. – *Bilingualism: Language and Cognition*, 11 (2), 147–164. <http://dx.doi.org/10.1017/S1366728908003283>

- Pavlenko, Aneta 2008b. Structural and conceptual equivalence in the acquisition and use of emotion words in a second language. – *The Mental Lexicon*, 3 (1), 91–120. <http://dx.doi.org/10.1075/ml.3.1.07pav>
- Pennebaker, James W.; Francis, Martha; Booth, Roger J. 2001. *Linguistic Inquiry and Word Count*. Mahwah: Erlbaum Publishers
- Sapir, Edward 2009. *Keel. Sissejuhatus kõne uurimisse*. Tallinn: Eesti Keele Sihtasutus.
- Skrandies, Wolfgang; Chiu, Ming-Jang 2003. Dimensions of affective semantic meaning – behavioral and evoked potential correlates in Chinese subject. – *Neuroscience Letters*, 341 (1), 45–48. [http://dx.doi.org/10.1016/S0304-3940\(03\)00137-X](http://dx.doi.org/10.1016/S0304-3940(03)00137-X)
- Sutrop, Urmas 2001. List task and a cognitive salience index. – *Field Methods*, 13 (3), 263–276. <http://dx.doi.org/10.1177/1525822X0101300303>
- Vainik, Ene 2010. Kuidas õpetada kõnesüntesaatorile empaatiat? Emotsiooni automaatse tuvastuse võimalustest eestikeelses kirjalikus lauses sisalduva info põhjal. – *Eesti Rakenduslingvistika Ühingu aastaraamat*, 6, 327–347. <http://dx.doi.org/10.5128/ERYa6.20>
- Vainik, Ene 2004. *Lexical Knowledge of Emotions: The Structure, Variability and Semantics of the Estonian Emotion Vocabulary*. *Dissertationes Linguisticae Universitatis Tartuensis*, 5. Tartu: Tartu Ülikooli kirjastus.
- Vainik, Ene 2002. Millest on tehtud eestlaste emotsioonisõnavara? – *Keel ja Kirjandus*, 8, 537–553.
- Valitutti, Alessandro; Strapparava, Carlo; Stock, Oliviero 2004. Developing Affective Lexical Resources. – *PsychNology Journal*, 2 (1), 61–83.
- Ververidis, Dimitrios; Kotropoulos, Constantine 2006. Emotional speech recognition: Resources, features, and methods. – *Speech Communication*, 48 (9), 1162–1181. <http://dx.doi.org/10.1016/j.specom.2006.04.003>
- Vogt, Thurid; Andre, Elisabeth; Wagner, Johannes 2008. Automatic recognition of emotions from speech: A review of the literature and recommendations for practical realisation. – Christian Peter, Russell Beale (Eds.). *Affect and Emotion in Human-Computer Interaction*. LNCS, 4868. Heidelberg: Springer, 75–91. http://dx.doi.org/10.1007/978-3-540-85099-1_7
- Voll, Piret 2002. Tähenduse konnotatiivne aspekt ja selle kajastamine ükskeelses sõnaraamatus. Käsikirjaline magistritöö Tartu Ülikoolis.
- Whissell, Cynthia M. 1989. *The dictionary of affect in language*. – Robert Plutchik, Henry Kellerman (Eds.). *Emotion: Theory, Research, and Experience*. New York: Academic Press, 113–131.

Võrgumaterjalid

- Eesti emotsionaalse kõne korpus <http://peeter.eki.ee:5000/> (25.04.2011).
- Firma Connotative Intelligence koduleht <http://www.connotative.com> (25.04.2011).
- Kontentanalüüsi hõlbustava veebivara General Inquirer koduleht <http://www.wjh.harvard.edu/~inquirer/Home.html> (16.04. 2010).

Ene Vainiku (Eesti Keele Instituut) uurimisobjekt on eesti keele emotsioonisõnavara, selle struktuur, semantika ja varieerumine. Huvipakkuvateks teemadeks on ka psühholingvistika, kognitiivne semantika ja keele ning emotsioonide interaktsioon laiemalt.
Ene.Vainik@eki.ee

Lisa 1. Vaade EKI eesti keele põhisõnavara sõnastiku toimetamisliidese alale "Emotsioonigrupp"

Kõide: 1. köide A - Y

märksõna

üllatus

Rada [üllatus*]: c:A[1]/c:S[1]/c:tp[1]/c:emg[1]

Toimetamisala XML Tabelina Vaatena

sisu	
eelnev ühisinfo	
tähendusnumbri plokk	1
grammatikagrupp	
semantika grupp	
emotsioonigrupp	
tundetoon	positiivne
määramise alus	4E1
afektikalduvuse klass	
tugevus	0.75
kuupäev	2011-11-26T15:19:29
tähendusgrupp	
semantilised suhted	

Vaade

üllatus

♪ <üllatus, üllatuse, üllatust, üllatusse;
mitmus: üllatused, üllatuste, üllatusi 11 S>

DETECTING EMOTIONAL VALENCIES FOR THE ESTONIAN VOCABULARY

Ene Vainik

Institute of the Estonian Language

The need for detection of the emotional valencies of the Estonian vocabulary originates in two practical applications. First, there is a need for a device capable of automatic detection of emotion based on the information present in written Estonian in order to model the naturalness of speech synthesis. Such a device can rely on a support dictionary that covers the most frequent and central part of the vocabulary. Second, in the course of compiling the Estonian Basic Dictionary (around 4000 entries), there is (among other layers of lexicographic description) also the task of informing users of the dictionary about the potential emotional valence of the entries. Both practical approaches are based on an assumption that emotional meanings and/or connotations are not only subjective phenomena but are at least partly shared and conventional. The idea of the study is to use the lexical resources of the Basic Dictionary as the support dictionary for the purposes of automatic emotion detection.

The present article tackles the theoretical background and accounts for the selection of an optimal method for detecting the emotional valencies of words. The theoretical approaches were found to differ mostly in whether they take the emotional properties of words as part of the structure of language (its usage as a fact) or as a part of the mental structure of the speaker (as a potential). In the present study the second approach is followed. The method suggested for the detection of emotional values – the method of predominant opinion – relies on the speaker's evaluations.

The selected method is tested in the case study, and the results and further prospects are discussed in the final part of the article. The procedure of evaluation was carried out in two series (with an interval of two months) and using four independent experts. An agreement rate stronger than 0.6 was set as a threshold for tagging the items in the database as neutral, positive, negative or ambivalent (average in the two series of evaluation tasks). The level of granularity of the evaluations was based on psychological reliability and the rationale of giving enough information for the process of automatic emotion detection but not too much for the human user of the dictionary. The method used in the case study gave quite reliable results as compared with some previous studies into the vocabulary of emotion and personal traits. However, to account better for the potential of words to contribute to the expression, description or infliction of emotion, the evaluative dimension may not suffice. Therefore, a more detailed classification of lexical affect bias is tentatively suggested at the end of the paper. The practical application of the classification should be proved or disproved by further studies

Keywords: emotional valence, method, vocabulary, affect biasedness, basic dictionary, Estonian

FOCUS ON THE FORM VERSUS FOCUS ON THE MESSAGE IN *LINGUA FRANCA* CONVERSATION

Natalja Zagura

Abstract. The present article looks at the speech of two Estonian twelfth-grade students in an informal conversation in English. The aim of the article is to characterise the speech of these two non-native speakers of English in terms of the communication strategies used and their effectiveness as judged by the further progress of the conversation. It is observed that the student who normally gets good marks at school considers correct linguistic forms to be important, but at the same time she also has difficulties with using effective communication strategies when production problems arise in a conversation. The second student focuses primarily on conveying his message clearly and, when language-related difficulties occur, he skilfully employs effective communication strategies such as restructuring, circumlocution, approximation, word coinage etc. It is concluded that in order to communicate in a foreign language successfully, it is sometimes not enough to have a good command of the language and adequate amount of practice, some learners may also gain from explicit instruction in communication strategy, which can motivate them to be more active and fluent speakers of the target language.*

Keywords: language learning, conversation analysis, communication strategies, English, Estonian

1. Introduction

According to the Estonian National Curriculum for Secondary School (2011: Appendix 2, section 3.1.1), one of the main aims of foreign language learning at school is that the students are able to communicate in the target language, both orally and in writing. Secondary school graduates are expected to speak fluently, with mostly correct grammar, and taking into account the cultural norms of the target language. Most non-native speakers (NNS) inevitably have certain difficulties while commu-

* The research has been partly supported by the Estonian Research Foundation grant No. 8008.

nicating in a foreign language. Nevertheless, it depends on the individual what they consider to be a problem in their foreign language use and how they deal with that.

Salla Kurhila (2005) has investigated interaction between native and non-native speakers of Finnish, pointing to a difference in the participants' orientation towards grammatical correctness. In the data she studied a tendency could be noticed that in conversations in Finnish NNS oriented to grammatical details and correctness, thus "doing being a second language learner" (*ibid.*: 157). Native speakers, on the other hand, mostly did not consider grammatical correctness an important factor in conversation, going beyond the level of linguistic forms and thus avoiding emphasis on linguistic asymmetry (*ibid.*). Kurhila suggests that this discrepancy may impede successful communication and lead to misunderstanding and even negative stereotyping.

Learners of English are more likely to use this language for communication with NNS as English performs the function of *lingua franca* around the world. It seems that talking to other NNS of English, people should be more relaxed, not focus so much on the possible linguistic asymmetry and stop emphasising one's identity as a language learner. Indeed, some learners of English mostly prefer fluency over accuracy, focusing primarily on the message they want to convey rather than linguistic forms. Nevertheless, it can still be observed that some NNS of English treat grammatical correctness as being a highly important factor in oral interaction and make an extra effort to use the correct language. Thus, they talk a diligent language learner's identity into being. It is impossible to say which approach is better because the choice would depend on a number of factors, such as the context, the roles of the conversational partners, their aims, attitudes etc. Both accuracy and fluency are important aspects of communication in a *lingua franca* and the most appropriate proportion of both should be identified by the interactants themselves in a particular situation.

The present article looks at one example of an informal *lingua franca* conversation where NNS of English are involved. The speech of two Estonian twelfth-grade students provides the data for the analysis – one of them mainly focuses on the message he is trying to convey, the other often pays attention to the linguistic forms she uses. The aim of the article is to characterise the speech of these NNS in terms of the communication strategies used and their effectiveness as judged by the further behaviour of other conversational partners. As this analysis is based on qualitative (rather than quantitative) data from a limited corpus, no far-reaching conclusions and implications can be pointed out. Nevertheless, some suggestions can be made in relation to the communication strategies that are likely to be effective in similar contexts. Moreover, some possible implications for foreign language (not exclusively English) instruction are discussed in the final section of the paper.

2. Data, method and key concepts

The data used for the present study consists of slightly more than two hours of naturally occurring informal conversation. This data originates from the corpus MUSU (*TÜ Multimodaalse suhtluse andmebaas*), the sub-corpus ISU (*Interaktiivsete suhtlussituatsioonide allkorpus*). The conversation has been both video-

recorded and audio-taped. The researcher was not present in the room at the time of the conversation, which hopefully contributed to its authenticity. Moreover, as the conversation lasted for quite a long time, the participants were able almost to forget about the recording being made and to talk relatively naturally. There are five 19–23 year-old students involved, three of them being Estonian twelfth-graders who at the time of recording were about to graduate from the secondary school (further coded as E1, E2 and E3), the other two are exchange students from Spain (further coded S1 and S2), who at the time of recording were studying English language and literature at the University of Tartu. The Estonian students have studied English for approximately 10 years and have reached level B₂, one of them probably even C₁. The international university students speak English at level C₁. The present study takes a closer look at the speech of two Estonian students coded as E1 and E2.

The method of Conversation Analysis (CA) (see Hutchby, Wooffitt 1998, ten Have 1999) has been used for transcribing and analysing the conversation. This method has been chosen since it can treat the roles/identities of the participants (e.g. language learner, student, teenager) as a dynamic factor and makes it possible to relate the relevance of a particular category to certain interactional phenomena. Moreover, CA is an ideal method for detailed qualitative analysis of “talk-in-interaction” (Drew, Heritage 1992: 4).

In the course of analysis, the phenomena of interest are discussed within the framework of communication strategies (CS). The term “communication strategy” was coined by Selinker (1972) in his seminal paper on “interlanguage”, where he discussed the “strategies of second language communication” (*ibid.*: 229). CS research attracted many scholars after Canale and Swain (1980) included CS in their influential model of communicative competence, presenting them as primary constituents of one of the subcompetences, strategic competence. From that time on, numerous studies have been addressing the issues related to the description, use and teachability of CS. Despite and also partly due to the great scholarly interest in the topic, there is little agreement about what CS are, how they could be classified and whether they can or should be learnt in the classroom. The various definitions and taxonomies have been summarised and categorised by Dörnyei and Scott (1997).

In broad terms, CS can be defined as the strategies used by NNS when they lack some language structures or skills necessary for solving communication problems (Hennoste, Vihalemm 1999) or, as Faerch and Kasper (1983: 81) put it, “potentially conscious plans for solving what to an individual presents itself as a problem in reaching a particular communicative goal”. In essence, CS represent the same mechanism which in CA is called “repair organisation”. In the present article the notion of CS rather than repair organisation is used in the course of analysis since it involves various detailed taxonomies. Labelling the phenomena analysed in the present study as particular CS makes the discussion clearer and more systematic. For the purpose of this study the taxonomy suggested by Hennoste and Vihalemm (1999) has been selected as it systematises a great variety of strategies in a clear and logical way. Moreover, it uses the same criteria as repair organisation in CA does: at which level of communication the strategy is used (turn level, topic level or phrase/word level); to whose text it is applied; who is the initiator of the use of a strategy and who suggests the solution; what is the direction of the process of strategy use; and finally, who comments on the process of strategy use or on the

solution, if anyone (*ibid.*: 1585). This taxonomy is a supplemented and systematised version of the one suggested by Dörnyei and Scott (1997: 188–194). Due to limits of space, the taxonomy itself is not discussed in the present article. The relevant CS, however, are characterised in the course of analysis. The present study focuses on the word/phrase-level strategies as they are predominantly used by the interactants.

As for the distribution of various CS and the relationship between strategy use and target language proficiency, relatively little attention has been paid to these issues. Moreover, the findings of relevant studies may even contradict each other. According to Nakatani and Goh (2007: 215), this discrepancy can be related to several reasons: scholars use different criteria for identifying the proficiency level of non-native speakers; moreover, they use different taxonomies of CS and thus count CS differently. Speaking of the general trends, some researchers claim that higher-proficiency learners use fewer CS because of the adequate linguistic and sociolinguistic knowledge they have at their disposal. Other scholars have determined that higher-level learners still use some achievement strategies (e.g. topic shift, paraphrase etc.), while lower-level learners tend to prefer reduction strategies (e.g. message abandonment) (*ibid.*). It should be noted that the present study does not aim to provide any quantitative data concerning the frequency and distribution of certain CS. This article presents detailed analysis of some specific instances of CS use and discusses the possible rationale for and outcomes of this use.

3. Communication strategies in *lingua franca* conversation

3.1. Focus on the form

One of the Estonian students participating in the conversation, a 19 year-old female, throughout the conversation tended to pay attention to the linguistic forms she uses – she tried to choose grammatically correct forms (e.g. verb forms, tenses, prepositions) and use the English words which closely correspond semantically to the Estonian words that she had in mind. It can be argued that the issue of using semantically precise words does not fall under the category of orientation towards using the correct form. Nevertheless, it has been included here as it reflects the NNS's desire to use as precise language as possible, to translate the words from the native language (L1) into the target language (TL) directly and succinctly, without resorting to any creativity.

(1) Topic: leisure time in Estonia¹

- 1 E1: well Estonians do go out for ee (.) in the (.) at weekends.
- 2 eeee but eee (0.2) usually (.) in ee (.) at in workdays. there there
- 3 are not so many people at the streets like if (.) if you come out at about
- 4 10 o'clock the the s- streets maybe (0.2) completely empty.
- 5 S1: yes, yes

¹ The transcription conventions used in the examples are explained at the end of the article.

Example (1) illustrates the strategy of self repair, which could be observed in the speech of E1 on several occasions throughout the two-hour conversation. Having started a phrase ‘go out for (weekends)’ (line 1), E1 probably decides that the preposition ‘for’ would not be correct in this context and thus she starts searching for a different suitable preposition, exposing her so-called “mental path” (Kurhila 2005: 146) towards the target word, which she (correctly) decides to be ‘at’ (line 1). Immediately after that she struggles to choose the correct preposition to precede the noun ‘weekdays’. Having started with ‘in’, E1 repairs it to ‘at’ and then again (incorrectly) to ‘in’ (line 2). This demonstrates that E1 considers it important that she uses correct prepositions even in an informal conversation in English despite the fact that in this context incorrect use of a preposition would not radically change the meaning of the utterance. Displaying the effort she takes in formulating the utterance correctly, E1 orients to her role as a NNS/language learner.

This extract also contains strategies that are meant to stall for time, such as pausing to gain time (Est *edasilükkamispaus* (Hennoste, Vihalemm 1999: 1590)), using certain sounds to fill pauses (line 2) (Est *üneem* (*ibid.*)) and self-repetition (lines 2, 3 and 4). However, it seems that these strategies are used for formulating one’s ideas rather than because of the production difficulties related to the foreign language use.

Other participants in the conversation patiently wait for E1 to formulate her ideas in the form she finds appropriate and then the discussion continues. No signs of misunderstanding can be noticed but, again, it seems that even incorrect use of prepositions, with no attempts to correct mistakes, would not have impeded this conversation and it would have proceeded exactly the same way.

(2) Topic: stereotypes about Estonia

- 1 E1: I know that ee in 2008 when the Germans came here. (0.5)
- 2 that ee (0.2) we had a lot of that kind of ee (.) mm (.) *suhtumine*
- 3 ((looks at E2))
- 4 E2: attitude
- 5 E1: attitude (0.2) .hh that eee (0.5) we are still part of Russia=
- 6 E2: =yeah
- 7 E1: Estonia is a modern country (0.5) it’s (0.2) we don’t have
- 8 electricity, we don’t have warmth, we don’t .hh (0.5) we are just aaa
- 9 E2: an extension of Russia=
- 10 E1: =just an extension of Russia. (.) we are so slow we are not developed
at all

In example (2) E1 is looking for the English word for ‘*suhtumine*’. At first she pauses, using the filler ‘kind of’ and the sounds ‘ee’ and ‘mm’. However, having failed to retrieve the necessary word, she takes advantage of the fact that there are also other speakers of Estonian involved in the conversation and she just switches to Estonian, while also looking at E2, as if asking for help (indirect appeal for help). She does get the help in the form of ‘attitude’ (line 4) and repeats this response to indicate acceptance of the suggested word. It is clear that E1, who is in general a diligent student and a high-level speaker of English, does know the word ‘attitude’ but at that point it just slipped her mind. In such a situation she resorted to the strategy

that is far too often used in language classroom – switching to L1 and asking the teacher or classmate for help/translation. In itself, it might be considered an effective strategy as the desired word/phrase is found very quickly. Nevertheless, in a situation where there are no other speakers of E1's L1 present, she would not be likely to get this help. Therefore, another CS, such as circumlocution (paraphrase) or approximation (use of a superordinate or a related term) could be more effective in such a situation. E1 might have chosen code-switching and appeal for help as it seemed the easiest way to get the right word; however, it is also possible that she has too little experience with circumlocution or approximation and therefore fails to use them in this situation.

Further in the conversation (lines 7-8) E1 successfully uses the strategy of restructuring – she leaves the problematic utterances unfinished and formulates new ones, with the structure which she feels more comfortable with. At the end of line 8 she again has trouble with finding the right word and reacting to the pause filled with 'aaa' (an indirect appeal for help) E2 suggests the possible expression, which E1 happily accepts by repeating it (line 10). As the discussion about popular stereotypes is continuing and other participants develop the points suggested by E1, it can be claimed that here the use of CS had a positive influence on the progress of the conversation.

(3) Topic: Estonian Song Festival

- 1 E1: it's amazing to be here (0.2) under this ee ((gesture showing an
- 2 arch, looks at E2 and E3)) (.) *kaar* ((gesture showing an arch))
- 3 ((E1 smiles, E3 laughs))
- 4 (0.8)
- 5 S2: also it was different (.) it is (0.5) I was surprised that you do that (.)
- 6 because in Spain you (0.2) ok you can find a lot of people together
- 7 but you can't find the people being in that way (.) all together,
- 8 singing,

Example (3) demonstrates that sometimes E1 is more creative. In lines 1-2, failing to remember or maybe not knowing the word 'arch', she uses the strategy of miming – describes the phenomenon non-verbally, with a gesture. However, at the same time she looks at E2 and E3, thus indirectly appealing for help and then also switching to the Estonian '*kaar*'. Since E2 and E3 cannot help (E3 laughs in a friendly way to mitigate the inability to help), E1 also smiles in a slightly embarrassed way and just abandons the message; after a pause of 0.8 sec (line 4) S2 takes the turn. He continues with the same general topic, although it is unclear whether he has actually understood the message in lines 1-2. It seems possible to suggest that E1's CS of miming was actually successful; what could have confused S2 is the fact that the utterance was never finished and thus the message remained unclear.

It can be seen from the examples above that E1, who is actually a diligent student and a successful language learner at school, orients to grammatical correctness in her speech and also occasionally needs assistance in finding the suitable word in English. She thus clearly positions herself as a NNS and language learner in this informal *lingua franca* conversation. It can be suggested that in this context, excessive attention to linguistic forms is not vital and it might seem a bit strange

or unnecessary to some conversational partners. As for E1's difficulty with finding the right English words, it could be a result of her slight nervousness. Being a good language learner and having used English mostly in the classroom context, she might not be used to having such difficulties and therefore cannot use CS automatically. Because of her somewhat limited experience of real-life interaction in English, E1 might be slightly confused and therefore instinctively resorts to the CS often used in the classroom – code switching and appeal for help, which would not necessarily be the most productive ones in a conversation with people who do not speak her L1.

In the conversation analysed, E1 did not use any other types of CS, such as circumlocution or approximation, for example.² This can be due to her slight nervousness or limited experience, as suggested before. However, it should also be taken into account that the corpus analysed is relatively small – the conversation lasted for slightly more than two hours and due to E1's good command of English, she did not need to use a lot of CS. The typical examples of all the strategy types found in E1's speech have been discussed above.

3.2. Focus on the message

Another Estonian student, a 19 year-old male, although (according to his marks at school) slightly less proficient in English than E1, orients to his NNS status considerably less. He pays less attention to the search for grammatically correct forms and skilfully uses CS in order to overcome occasional production difficulties.

(4) Topic: Lower density of population in Estonia

- 1 E2: it's it's a northern country (0.2) it's (0.5) ((shows upwards)) north
- 2 from the the (0.2) globe (0.2) it's (0.5) there's less population here

In example (4) E2 is talking about the relatively low density of population in Estonia. Several instances of self repetition, pausing and also restructuring (leaving an utterance unfinished and using an alternative wording for the intended message; line 2) can be seen here. However, it seems that this hesitation results from E2's attempt to formulate his idea better rather than from the production difficulties related to the use of a foreign language. It is worth attention, though, how E2 uses mime (a gesture showing upwards) and approximation (using a superordinate 'globe') in order to explain his idea that the density of population in Estonia is smaller than in most countries due to the country's position relatively far to the north of the equator. For some reason having forgotten the word 'equator', he uses a general term 'globe' (equator being a component of it). The resulting explanation is rather clumsy, but conversational partners have no difficulty with understanding the idea and further continue the conversation on this topic. Thus, it is a positive fact that E2 has not abandoned the message because of the lack of a suitable word, but used approximation to convey his idea.

² I would like to thank the anonymous reviewer for pointing out that this essential information was missing in the first version of the article.

(5) Topic: Tourist attractions in Pärnu

- 1 E2: and in Pärnu. (1.2) there are the (0.5) bicycles called tandems
2 a two-man bicycles (0.5) I: suggest you to rent them (0.5)
3 ((E3 laughs)) and ride (.) around the town.
4 E3: *omad kogemused*
5 S1: hehee
6 E3: from his experience
7 S1: yeah
8 E2: for for about 25 kroons or something you can rent (.) rent
9 the bike and ride around the city for an hour. and ee (0.2) it's it's
10 a lot of fun!

When talking about tourist attractions in Pärnu (example (5)), E2 touches upon an opportunity to rent tandems there. The word 'tandem' is the same in both English and Estonian, it is just pronounced slightly differently. Here, E2 pronounces it in an Estonian way. This could be, of course, a pronunciation error; however, it is also possible that E2 is not sure whether the word is the same in English – he just transfers it from Estonian and for the sake of clarity uses circumlocution, describing it as a 'two-man bicycle'. Still, an alternative interpretation is also possible – E2 may be unsure whether his conversational partners (S1 and S2) know this term and defines it for them. S1 and S2 smile and nod, which could be considered as a sign of them getting the point.

(6) Topic: Summer employment for young people

- 1 E2: in Estonia, here in Tartu we have ee (.) some kind of a (0.2) *malev*.
2 *maleva* (0.5) *kuidas sa tõlgid seda* ((looks at E3)) (0.2)
3 we have some kinds of events. (.)
4 when young children come together .hh (0.2) and they do
5 some kind of ee (0.5)
6 E3: yeah, public work=
7 E2: =public works (0.5) so they paint fences, clean eee (0.5) graveyards
8 and so on .hh (0.2) and ee for a week pay (0.2) eee (.) two emm
9 two-week pay was about (.)eee (.)twenty euros I guess (0.2)
10 two weeks in summer
11 S2: mmm

Circumlocution is also needed when the L1 terms that do not exist or are rarely used in the TL are discussed. In the example (6) above E2 talks about summer employment opportunities for young people. Quite naturally, he does not know the English for '*malev*'. At first, he pauses, then says the Estonian word (strategy of code-switching) and addressing E3 presents a direct appeal for help (line 2). Within the pause of 0.2 sec having understood that E3 does not know the translation either, he decides to describe what '*malev*' is (circumlocution). This is a good decision because even if E2 found the precise term, there is no guarantee that S1 and S2 would know this term and understand it. The description suggested by E2, on the other hand, makes it clear for the international students what the word '*malev*' stands for.

(7) Topic: The history of Estonia

- 1 E2: when you came to Estonia. (0.2) did you know anything (0.5) about
2 Estonia being (0.2) aaa (0.5) sub-country for Russia (0.5) maybe (0.5)
3 S1: yeea (0.2) we learnt something (0.2)
4 E2: that we were the Russian extension (0.5) have you heard
5 anything like that?
6 (1.0)
7 S2: no
8 S1: do you mean (0.2) that Estonia was part of Russia.
9 E2: we were a part of Russia but as a separate state

In the final example (above), when talking about the history of Estonia, E2 does not know how to name the status that Estonia had within the Soviet Union. So, he coins the word “sub-country”, knowing that the prefix sub- often implies an object/subject being a part of something larger and thus getting a word with quite a transparent meaning. This coinage, however, seems somewhat ambiguous to the Spanish students. S1 carefully confirms that they have studied “something”, probably meaning that they are aware of the history of the Soviet Union and its member states, but he is not quite sure what this status of “sub-country” could mean. Having noticed that S1 is puzzled, E2 makes another attempt, this time suggesting another coined term – “the Russian extension”. Unfortunately, it does not make the situation much clearer for the Spanish students and after a pause of 1.0 sec S2 has to confess “No”. S1 uses the strategy of an interpreted summary, suggesting his understanding and checking whether this interpretation is correct (line 8). It can be seen that the strategy of word coinage was effective and in general terms S1 has understood what E2 was trying to say (i.e. Estonia being a part of the Soviet Union). However, it is important for E2 to emphasise the view that while being a part of the Soviet Union, Estonia was still a separate entity. Such a description seems confusing to the Spanish students, they do not manage to understand E2’s point completely and eventually the topic is abandoned. Despite this misunderstanding, it is still possible to claim that the strategy of word coinage can be a productive one or at least it can be used when other resources are not available.

The examples above characterise the strategy types used by E2 and they demonstrate that although E2 has certain difficulties when speaking English, he uses CS skilfully and is able to communicate freely, mostly being able to convey his message successfully. He does make grammatical mistakes but (in the conversation analysed) does not focus on them, thus (completely appropriately) he is “doing being” a young man, a student, an Estonian etc. rather than a language learner. It is difficult to say whether E2 is just naturally a skilful communicator or has extensive experience of interacting in English, or maybe both. What is certain is that, when having some difficulty in communication in English, he does not get stuck because of that but immediately attempts to overcome the problem. As discussed above, the strategies of circumlocution, approximation and word coinage (in some situations also code-switching and appeal for help) can be effective aids in overcoming perceived problems in *lingua franca* interaction. There are certainly also other effective CS, but in this particular dataset it can be observed that the abovementioned types of CS have been used.

4. Conclusions

The present article has looked at a relatively small corpus of *lingua franca* communication with the aim to study in detail and characterise the speech of two young Estonians – both NNS of English. One of the participants in the conversation tended to orient to the form – grammatically correct language and precise words (consequently, succinct language), thus emphasising her identity as a language learner. Moreover, she was relatively helpless when she did not know or remember a certain word³. Another twelfth-grader mostly did not consider the form to be of major importance and used a variety of CS to overcome production difficulties and convey the intended meaning as precisely as possible. In the course of interaction he assumed the roles of a young man, a student, an Estonian etc. rather than that of a language learner.

On the basis of the analysis it is possible to suggest that in similar informal *lingua franca* conversations grammatical correctness is not considered vital as long as errors do not interfere with the meaning conveyed. NNS might be tempted to switch to L1 and ask for help, as they are used to doing in the language classroom. If there are other people speaking this L1 involved in the conversation, the abovementioned strategies are likely to be effective. However, code-switching would not be an effective strategy if the conversational partners do not share any other language in addition to the *lingua franca* currently used. If the interactants do not despair in the case of a communication problem and use some CS, like restructuring, approximation, circumlocution or even word coinage, they are more likely to succeed in conveying their message and achieving the communicative goal.

Some people just are good communicators and encountering a problem they automatically use a suitable CS. Some NNS of a language might acquire this ability naturally, in the course of practice – the more experience they have, the more automatic the CS use becomes, yet there could also be some language learners who would gain from explicit instruction in CS use. It is a fact that some NNS are afraid of making mistakes when speaking a foreign language. For this reason they prefer to abandon current topics and messages when they encounter some difficulty or problem. If their negative attitude towards errors and risk-taking is changed, and potentially productive CS are introduced and practised, NNS might be more motivated to use them and, in general, also be more motivated to communicate in the foreign language. In turn, the more practice they have, the more active and competent speakers they can become.

Indeed, to date, a number of studies on CS instruction have been conducted⁴. Although scholars admit that it is somewhat difficult to evaluate the effect of strategy instruction, they recognise that there is evidence that such instruction can produce positive results. Having seen in the present study that even the language learner who is considered diligent and successful might have difficulties in real-life communication in the TL, it seems especially important to identify the most effective methods of strategy instruction and to develop suitable didactic materials.

³ A similar phenomenon could be observed in another study (Zagura 2006), where the performance of twelfth-graders in mock gate-keeping interviews was analysed.

⁴ For an overview, see e.g. Nakatani, Goh 2007.

Abbreviations and transcription conventions

- CA Conversation Analysis
CS communication strategies
E1 Estonian student 1 (female)
E2 Estonian student 2 (male)
E3 Estonian student 3 (female)
L1 first/native language
NNS non-native speaker
S1 Spanish (university) student 1 (male)
S2 Spanish (university) student 2 (male)
TL target language
(0.5) The number in brackets indicates a time gap in tenths of a second.
(.) A dot enclosed in a bracket indicates a pause in the talk of less than two-tenths of a second.
= The 'equals' sign indicates 'latching' between utterances.
[] Square brackets between adjacent lines of concurrent speech indicate the onset and end of a spate of overlapping talk.
.hh A dot before an 'h' indicates speaker in-breath. The more h's, the longer in-breath.
hh An 'h' indicates an outbreath. The more h's, the longer breath.
(()) A description enclosed in a double bracket indicates a non-verbal activity or comments.
: Colons indicate that the speaker has stretched the preceding sound or letter. The more colons the greater the extent of the stretching.
. A full stop indicates a stopping fall of tone.
, A comma indicates 'continuing' intonation.
? A question mark indicates a rising inflection. It does not necessarily indicate a question.
und Underlined fragments indicate speaker emphasis.

References

- Canale, Michael; Swain, Merrill 1980. Theoretical bases of communicative approaches to language teaching and testing. – *Applied Linguistics*, 1 (1), 1–47. <http://dx.doi.org/10.1093/applin/1.1.1>
- Drew, Paul; Heritage, John 1992. Analysing talk at work: An introduction. – Paul Drew, John Heritage (Eds.). *Talk at Work. Interaction in Institutional Settings*. Cambridge: Cambridge University Press, 3–65.
- Dörnyei, Zoltán; Scott, Mary Lee 1997. Review article. Communication strategies in a second language: Definitions and taxonomies. – *Language Learning*, 47 (1), 173–210. <http://dx.doi.org/0.1111/0023-8333.51997005>
- Estonian National Curriculum for Secondary School = Gümnaasiumi Riiklik Õppekava. <https://www.riigiteataja.ee/akt/114012011002> (28.09.2011).
- Faerch, Claus; Kasper, Gabriele (Eds.) 1983. *Strategies in Interlanguage Communication*. London: Longman.
- Hennoste, Tiit; Vihalemm, Triin 1999. Võõrkeele suhtlusstrateegiad. – *Akadeemia*, 125 (8), 1571–1608.
- Hutchby, Ian; Wooffitt, Robin 1998. *Conversation Analysis. Principles, Practices and Applications*. Cambridge: Polity Press.
- Kurhila, Salla 2005. Different orientations to grammatical correctness. – Keith Richards, Paul Seedhouse (Eds.). *Applying Conversation Analysis*. Basingstoke, New York: Palgrave Macmillan, 143–173.

- Nakatani, Yasuo; Goh, Christine 2007. A review of oral communication strategies: Focus on interactionist and psycholinguistic perspectives. – Andrew D. Cohen, Ernesto Macaro (Eds.). *Language Learner Strategies*. Oxford, New York: Oxford University Press, 207–227.
- Selinker, Larry 1972. Interlanguage. – *International Review of Applied Linguistics*, 10 (1–4), 209–231. <http://dx.doi.org/10.1515/iral.1972.10.1-4.209>
- ten Have, Paul 1999. *Doing Conversation Analysis. A Practical Guide*. London: SAGE Publications.
- Zagura, Natalja 2006. The communicative language ability of secondary school leavers in an English-language gate-keeping interview. – Unpublished graduation thesis of the Teacher Training Programme. University of Tartu, Department of English language and literature.

Natalja Zagura (University of Tartu) has used conversation analysis for studying the following issues: strategic language use and influence of gender roles on communication, the use of English and Estonian interlanguages in an institutional context as well as the use of communication strategies in the speech of Estonian learners of English.
natalja.zagura@ut.ee

SÕNUMILE VÕI KEELELISELE VORMILE KESKENDUMINE VÕORKEELES VESTLUSES

Natalja Zagura

Tartu Ülikool

Artikkel käsitleb kahe eesti 12. klassi õpilase kõnet mitteametlikus ingliskeelses vestluses. Eesmärgiks on iseloomustada nende kahe inglise keelt võorkeelena kõneleja kõnet suhtlusstrateegiate kasutuse seisukohalt ning vestluse edasise kulu põhjal analüüsida nende strateegiate kasutamise edukust.

Analüüsist ilmneb, et see õpilane, kes on koolis võorkeeleõppes edukas, peab grammatiliselt korrektset keelt tähtsaks ka kõnes. Samas tekib tal raskusi tõhusate suhtlusstrateegiate kasutamisega, kui mingi suhtlusprobleem vajab lahendamist. Suure tõenäosusega valib ta koodivahetuse ja abiküsimise strateegiate seast. Teine õpilane peab aga esmatähtsaks hoopis sõnumi edastamist, ning seda võimalikult arusaadaval kujul. Kui tema keelevahendite või -oskuste komplekt on ebapiisav, kasutab see õpilane edukalt selliseid tõhusaid suhtlusstrateegiaid nagu näiteks restruktureerimine, mõiste/sõna/väljendi väljajäätlemine kirjelduse või näite abil, mõtte väljajäätlemine üldisema tähendusega sõna kaudu ning isegi uute, keeles olematute sõnade tuletust.

Jõutakse järeldusele, et kui inimesel on soov võorkeeles edukalt suhelda, ei piisa mõnikord vaid heast keeleoskusest ja suhteliselt suurest keelepraktikast. Mõne keeleõppija puhul on eriti kasulik tõhusate suhtlusstrateegiate tutvustus ja nende harjutamine, mis omakorda võib motiveerida inimest olla aktiivsem ja julgem võorkeele kasutaja.

Võtmesõnad: keeleõpe, vestlusanalüüs, suhtlusstrateegiad, inglise keel, eesti keel

Trükitud 150 eksemplari