

Eesti Rakenduslingvistika Ühing

Eesti Rakenduslingvistika Ühingu aastaraamat 11

Estonian Papers in Applied Linguistics 11

Toimetanud
Helle Metslang, Margit Langemets
ja Maria-Maren Sepper

Tallinn 2015

EESTI RAKENDUSLINGVISTIKA ÜHINGU AASTARAAMAT 11 **Estonian Papers in Applied Linguistics 11**

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

Toimetuskolleegium / Advisory board:

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

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

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)
ERIH PLUS (European Science Foundation)
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)
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/ajakirjad

Eesti Rakenduslingvistika Ühingu aastaraamatu 11. numbri väljaandmist on toetanud
Haridus- ja Teadusministeeriumi riiklik programm "Eesti keel ja kultuurimälu II".

Autoriõigus / Copyright: autorid ja Eesti Rakenduslingvistika Ühing

Kujundus / Design: Piia Ruber, Sirje Ratso

ISSN 1736-2563 (trükis)

ISSN 2228-0677 (võrguväljaanne)

[doi:10.5128/ERYa.1736-2563](https://doi.org/10.5128/ERYa.1736-2563)

SISUKORD

Kes vastutab? Liiklusõnnetuses osalejate keeleline konstrueerimine <i>Who's responsible? The linguistic construction of participants in traffic accidents</i> Katrín Aava	7
External factors and the interference of L1 Estonian on L2 English pronunciation: An apparent-time study <i>Välised tegurid keeleomandamisel ja eesti keele kui emakeele interferents inglise keele kui teise keele hääldeuses</i> Kristiina Ader, Merilin Miljan	21
Statistilised meetodid ühendverbide tuvastamisel tekstikorpusest <i>Statistical methods for Estonian particle verb extraction from text corpora</i> Eleri Aedmaa	37
Eesti ja soome keele vastastikune mõistmine üksiksõna- ja tekstitasandil: lingvistilised tegurid, mõistmisprotsess ja sümmeetria <i>Mutual intelligibility of Estonian and Finnish context-free words and texts: Linguistic determinants, comprehension process and symmetry</i> Annekatriin Kaivapalu	55
Korpusleksikograafia uued võimalused eesti keele kollokatsiooni-sõnastiku näitel <i>New possibilities in corpus lexicography based on the example of the Estonian Collocations Dictionary</i> Jelena Kallas, Kristina Koppel, Maria Tuulik	75
Morphonotactics in L1 acquisition of Lithuanian: TD vs. SLI <i>Morfonotaktika leedu keele omandamisel esimese keelena: tüüpilise arenguga vs. spetsiifilise kõnearengupuudega lapsed</i> Laura Kamandulytė-Merfeldienė	95
Milleks mulle eesti keel? Riigikeele oskuse vajalikkusest vene koolinoorte pilgu läbi <i>Why do I need Estonian? The importance of official language competence in the eyes of Russian teenagers</i> Birute Klaas-Lang, Kristiina Praakli	111

Läbirääkimiste modelleerimine arvutil <i>Modelling negotiations on the computer</i> Mare Koit	127
Narrative production weakness in Russian dyslexics: Linguistic or procedural limitations? <i>Narratiiviloome nõrkus vene düslektikutel: keeleline või protseduuriline piiratus?</i>	141 157
Aleksandr N. Kornev, Ingrida Balčiūnienė	
Direktiivsed kõneaktid ning nende kasutamise dünaamika isa ja ema kõnes: juhtumiuuring <i>Directive speech acts and the dynamics of their usage in father's and mother's speech: A case study</i>	159
Helen Kõrgesaar	
Kui lapsega ei räägi üksnes ema: valik termineid eesti laste- ja hoidjakeele kohta <i>When the mother is not the only one talking to a child: Selection of terms in Estonian child language and caregivers' speech</i>	177
Helen Kõrgesaar, Airi Kapanen	
Event coreference detection in Estonian news articles: Preliminary experiments <i>Sündmust väljendavate keelendite samaviitelisuse tuvastamine eesti keele uudistekstides: esialgsed katsetused</i>	189
Siim Orasmaa	
Using micro-contexts to describe a writing process in Estonian as a second language across proficiency levels <i>Mikrokontekstid eesti keele kui teise keele kirjutamisprotsessi kirjeldamisel keeleoskustasemeti</i>	205
Olga Pastuhhova	
The linguistic construction of the giftedness discourse in the media texts of historical and digital times <i>Andekuse diskursuse keeleline konstrueerimine ajaloolistes ja digiajastu meediatekstides</i>	223
Halliki Põlda	
Factorial structure of SILL revisited: adaptation of SILL for Estonian EFL learners <i>Keeleõppestrateegiate mõõtevahendi (SILL) faktorstruktuuri uuring: SILL-i adapteerimine Eesti inglise keele õppijate jaoks, keeleõppestrateegiate kasutamise kultuurilised iseärasused</i>	241
Katrin Saks, Äli Leijen, Kandela Õun, Karin Täht	

Evidentsiaalsuse ja episteamilise modaalsuse suhetest eesti lastekeeles käitumiskatsete põhjal <i>Evidentiality and epistemic modality combine in the Estonian morpheme -vat but are different categories in acquisition: Experimental evidence</i> Anne Tamm, Reili Argus, Airi Kapanen, Andra Kütt, Kadri Suurmäe	263
Keeleteaduslike andmete ruumilisi visualiseerimisvõimalusi <i>Techniques for spatial data visualization in linguistics</i> Kristel Uihoaed, Aki-Juhani Kyröläinen	281
The treatment of lexical collocations in EFL coursebooks in the Estonian secondary school context <i>Leksikaalsete kollokatsioonide käsitlemine inglise keele kui võõrkeele õpikutes Eesti keskkooli kontekstis</i> Liina Vassiljev, Liljana Skopinskaja, Suliko Liiv	297

KES VASTUTAB? LIIKLUSÕNNETUSES OSALEJATE KEELELINE KONSTRUEERIMINE

Katrin Aava

Ülevaade. Liiklusõnnetuste kajastused sisaldavad ühiskondlikke hinnanguid selle kohta, milline tähendus on inimesel, roll ja vastutus osapooltel ning kelle süül asjad juhtuvad. Uuringus kasutatud kriitiline diskursuseanalüüs on meetod, mis võimaldab tuvastada selliseid tekstides varjatud ühiskondlikke kokkuleppeid ning tekstidesse kodeeritud eelistatavaid tõlgendusi. Mida näitab lingvistiline analüüs osasündmuste agentsuse, asjaolude esituse ja teksti vaatepunkti kohta ning kas need tulemused on kooskõlas lugejaarvamuse uurimisega? Selleks analüüsisin 2013. aasta oktoobris-novembris politsei edastatud 26 liiklusteadete õnnetuste kohta ülekäigurajal. Selgitamaks kujutatava sündmuse hoiakulise tõlgenduse sissekirjutamist teksti, küsitlesin 104 vastajat.

Liiklusteade pealaused on passiivlaused ning sõnastatud lause-semantiliste tunnuste agentiivne/mitteagentiivne alusel juhtumitena (õnnetused toimuvad), mitte tegevuste, protsesside ega tegudena. Liiklusteade esimese astme kõrvallauses on aktiivsed osalised jalakäija otsa sõitvad autod ning sõidukilt löögi saanud jalakäijad. Autojuhti nimetatakse teise astme kõrvallauses. Uudised on raamistatud juhi vaatepunktist. Selline uudiste raamistus on kooskõlas lugejaarvamuse uuringuga: lugejad interpreteerivad liiklusteateid autojuhile soodsamal viisil. Mida rohkem sõnastati asjaolusid, mille vastu eksis jalakäija (helkuri puudumine, tumedad riided) või mis segasid autojuhti ülekäigurajal (halb nähtavus), seda rohkem tajuti süüdlasena jalakäijat.

Võtmesõnad: kriitiline diskursuseanalüüs, raamistamine, vaatepunkt, liiklusteade

1. Lähtekohad ja eesmärk

Liiklusuudised on olulised tekstid, mis sisaldavad ühiskonnale omast suhtluskultuuri ja väljakujunenud sotsiaalseid kokkuleppeid. Liiklusuudised annavad edasi sõnumeid selle kohta, milline tähendus on inimesel, milline roll ja vastutus on jalakäijal ja autojuhil. Nii väljenduvad liiklusuudistes ühiskonnas aktsepteeritud suhted. Teed on avaliku ruumi osa, kus kõigil kodanikel peaksid olema võrdsed õigused. Samas toimub erinevate huvigruppide võitlus selle eest, kes on tee kasutajana paremas positsioonis. Neid suhteid väljendatakse keeleliselt: nt kas jalakäijat ning ratturit nähakse liikluse segajana või loomuliku osana liiklusest.

Seniste uuringute põhjal domineerib liiklusõnnetuste diskursuses autokeskne lähenemine, kus rõhutatakse eelkõige suurte autode turvalisust (Freund, Martin 1997). Selline olukord jätab autota avalikus ruumis liiklejad halvemasse positsiooni. Autojuhikesksusel on ühiskonnas ka pikem kultuuriline ning ajalooline taust: kultuuriliselt on konstrueeritud head elu, mehelikkust ja ka võimu diskursust võimsate autode kaudu (Whatley 1988, Sheller, Urry 2000: 738). Lucie Vallières (2006) on uurinud Kanada näitel, kuidas neoliberaalsed poliitilismajanduslikud diskursused mõjutavad liiklusohutuse kampaaniaid, pannes vastutuse liiklusõnnetuse eest üksikindiviididele. Liikluskasvatus on suunatud sellele, et jalakäija ei tohi segada autosid, kujundades sedasi autokeskset liikluskultuuri.

Autotootjad on enda majandushuvide kaitseks mõjutanud seadusandlust, linnaplaneerijad on pikka aega tegutsenud selle nimel, et luua linnaruumis eelkõige autojuhtidele paremad tingimused. See väljendub ka keelekasutuses. Maailmapanga ja autotootjate koostöös valminud globaalse teedeohutuse partnerlusprogrammi sõnasagedusanalüüs (Roberts jt 2006) näitas, et dokumentides kasutatakse vähem sõnu nagu *kiirus, kiiruspiirang, jalakäija, avalik transport, kõndimine, jalgrattasõit*, eelistatumad väljendid on *kool, kampaania, juhikoolitus, teavitus-, reklaamitahvel*. Tähelepanu on juhitud ka ühendriikides viimastel aastatel toimunud keelelisele pöördele: jalakäija surmaga lõppenud õnnetustes saab õnnetuse põhjustaja süüdistuse *hooletus juhtimises*, mitte *surma põhjustamises* (Oatman-Stanford 2014).

Artikli fookuses olevate liiklusõnnetuste toimumispaigaks on ülekäigurada, mis on see osa teest, kus jalakäija peab saama turvaliselt teed ületada. Kehtiva liikluseaduse (2010) § 2 defineerib liiklusõnnetuse legaalterminina: "liiklusõnnetus on juhtum, kus vähemalt ühe sõiduki teel liikumise või teelt väljasõidu tagajärjel saab inimene vigastada, surma või tekib varaline kahju". § 35 kohaselt "[r]eguleerimata ülekäigurajale lähenedes peab juht sõitma sellise kiirusega, mis on piisavalt väike, et mitte ohustada jalakäijat, kes on astunud või astumas ülekäigurajale". Liikluseaduse kohaselt pole mingil põhjusel õigustatud ülekäigurajal sõiduteed ületajale otsasõit. Kas see väljendub ka tegelikult liikluseõnnetustes? Eesmärk on selgitada, millised suhted loomulikustatakse liikluseadetes ning kelle vaatepunkti need esindavad?

Kriitilise tekstianalüüsi kohaselt pole keelekasutus neutraalne, kuna teksti koostaja esitab sõnumi kellegi vaatepunktist, mis teeb sündmuste objektiivse kirjeldamise võimatuks (Fairclough 1989). Diskursuse kord määrab, milline on ühes või teises kontekstis loomulik keelekasutus. Diskursus on sotsiaalne praktika, kommunikatsiooni viis teatud sotsiaalses, kultuurilises, ajaloolises, poliitilises situatsioonis (Dijk 2005: 3). Kriitilise tekstianalüüsi metodoloogiaga on tegelelud Norman Fairclough (1991, 1992, 1995, 2001, 2005), Ruth Wodak (1996), Fairclough

ja Wodak 1997, Teun van Dijk (2005) jt. Kui diskursuseanalüüsi tehes kirjeldatakse keelekasutust, siis kriitilist diskursuseanalüüsi tehes selgitatakse ka keelekasutuse põhjuseid (Fairclough 1995: 43).

Kriitiline diskursuseanalüüs seob sotsiaalsed teooriad tekstianalüüsiga, analüüsis tuntakse huvi, kuidas tekstid toimivad suhtluses. Fairclough' (1992) järgi on diskursusel kolm mõõdet: kõneldud, kirjutatud keeletekst; diskursuslik käitumine, mis sisaldab teksti tootmist ja selle tõlgendamist, ning sotsiokultuuriline käitumine, mis hõlmab sotsiaalseid ja poliitilisi suhteid. Fairclough tegeles sotsiolingvistilise uurimusega, uurides teksti sõnavara, grammatikat, semantikat. Samuti uuris ta sotsiaalseid struktuure ning praktikaid, mis tekstidesse kirjutatakse.

Meediatekstide kriitilises diskursuseanalüüsis kasutatakse ka *raamistamise* mõistet (nt Serafini 2003 jt). Raamistamine annab võimaluse lisada oma tõlgendus hajuksatele faktidele (McQuail 2000: 307). Uudiste diskursust käsitletakse sotsiokognitiivse protsessina, kus osalevad allikad, ajakirjanikud ning auditoorium, kes loovad ühiselt jagatud kultuuri, mis baseerub ühistele sotsiaalselt defineeritud rollidele (Pan, Kosicki 2010). Raamistamine on uudiste edastamise formaat, mille kohaselt lugu esitatakse ja ka tõlgendatakse ning see sisaldab hoiakuid sündmuste osalejate suhtes (Huckin 1997: 91–95). Diskursuseanalüüsis pööratakse kõigepealt tähelepanu esitatud info raamistamisele – valitud vaatenurgale, mis sisaldab teatud valikuid ja väärtustamist. Samuti võimaldab raamistamine selgitada kujutatava sündmuse hoiakulise tõlgenduse sissekirjutamist teksti. Raamistamiseks võib kasutada võtmesõnu, mis rõhutavad teatud seisukohti; mingite teemade, seisukohtade vm mainimata jätmist; teatud alternatiive sõnavalikul (Laherand 2008).

Peamised etteheited kriitilisele diskursuseanalüüsile on olnud järgmised: poliitilised ja sotsiaalsed ideoloogiad “loetakse” andmetesse sisse; sotsiaalsed ja keeleteooriad on meetodis tasakaalustamata; kriitilist diskursuseanalüüsi on tihti lahutatud sotsiaalsest kontekstist (Flowerdew 1999, Widdowson 1998). Et kontrollida kvalitatiivselt saadud tekstianalüüsi tulemuste usaldusväärsust kvantitatiivselt, kasutan artiklis trianguleerimismeetodit (Harvey, Macdonald 1993). Kombineerin kvalitatiivuurimuse kujundustüüpe kvantitatiivsete argumentidega, nii on lingvistiline analüüs ühendatud lugejaarvamuse uurimusega.

Analüüsin, kuidas tekstides kirjeldatakse sündmuste osapooli, konflikti ohvreid ja süüdlasi, kuidas tähistatakse sündmuste tegelasi ja millise tähenduse selline raamistamine teksti lugeja jaoks loob. Otsin vastust uurimisküsimustele: mida näitab lingvistiline analüüs sündmuste agentsuse, asjaolude esituse ja teksti vaatepunkti kohta, ning kas need tulemused on kooskõlas lugejaarvamuse uurimusega?

2. Materjal ja meetod

Uuring koosneb kahest etapist. Selgitamiseks, kuidas konstrueeritakse liiklusteadetes osalejate rollid, uurin, kuidas nimetatakse tegevuses osalejaid, väljendatakse agentsust ning milliste tunnuste kaudu neid protsessis määratletakse. Protsessi osalejate rollide kaudu selgitan, milliste tegevuste ja asjaoludega neid seostatakse, kuidas keeleliselt väljendatakse rolle, mida nad on tekstis tegema pandud. Nimelt kirjutatakse faktide paigutuse ning vaatepunkti valikuga teksti sisse ka teksti koostaja suhtumisi: lausete keelelisel valikul osutavad, kelle poolt või vastu kirjutaja

selles tekstis on, millisesse rühma ta tunneb ennast kuuluvat, millisesse rühma ta eeldab lugejat kuuluvat (Kasik 2008: 18–20).

Tegevused eeldavad tegijaks inimest ehk agenti, kes võib olla tagaplaanil või jääda ka varjatuks (Leeuwen 1996). Lausesemantiline analüüs näitab, millal kajastab lause tegevust või tegu, mis on agentsed, ning millal protsessi või juhtumit, mis on automaatsed. Agentiivsus tähendab “Eesti keele grammatika II” järgi agendi ehk situatsiooni suhtes aktiivse, seda kontrolliva osalise olemasolu situatsioonis. Tunnuste duratiivne/momentaanne ja agentiivne/mitteagentiivne alusel liigituvad dünaamilised situatsioonid tegevusteks (agentiivsed duratiivsed situatsioonid), protsessideks (mitteagentiivsed duratiivsed situatsioonid), tegudeks (agentiivsed mitteduratiivsed situatsioonid) ning juhtumiteks (mitteagentiivsed mitteduratiivsed situatsioonid) (EKG II: 24). Seega liigitub lause *Auto sõitis jalakäijale otsa* tegevuse, lause *Jalakäija jäi auto alla* teo ning lause *Ülekäigurajal juhtus õnnetus* juhtumi alla.

Lauseid, milles tegevussubjekt esineb grammatilise subjektina, nimetatakse aktiivlauseteks, lauseid, kus selline vastavus puudub, aga passiivlauseteks. Subjektita konstruktsiooni käsitlemine passiivi alaliigina rajaneb eeldusel, et tegumoe vahetus põhineb eelkõige püüdel vältida mingitel kaalutlustel tegija esiletõstmist ning tegevuse vaatlemist tegevusobjekti aspektist on selle vahetuse võimalikke kaasnähtusi. (EKG II: 30) Kui liiklusteadetes väidetakse, et *jalakäija jääb auto alla*, siis välditakse lauses otsasõitja esiletõstmist ning sündmust kirjeldatakse autojuhi vaatepunktist. Kui aga *auto sõidab jalakäijale otsa*, siis protsessi aktiivne osaline on auto (kuigi peaks olema autot juhtiv inimene), jalakäija roll on jäetud passiivseks.

Uuringus vaatlen, kelle vaatepunktist sündmusi raamistatakse, kes saab rääkida, kellel on võimalus sõnastada, mis toimus ja mis põhjustel toimus, kes on vaikiv hää. Järeldusi teen järgmiste põhimõtete järgi: kui kasutatakse nt väljendeid *ootamatult teele jooksnud, helkur polnud nähtav*, siis on olukorda kirjeldatud juhi vaatepunktist. Nimelt tuli jalakäija ülekäigurajale juhile ootamatult ning juht ei näinud helkurit, mida jalakäija tegelikult kandis. Analüüsin, kuidas pea- ja kõrvalausetega tähendusi edasi antakse, kuidas väljendatakse agentsust, kelle eksimustele tekstis osundatakse ning milliste keeleliste valikutega seda tehakse.

Uurimismaterjaliks on siseministeeriumilt maanteeametile esitatud 2013. aasta oktoobri-novembri 26 liiklusteadet, mis kajastasid õnnetusi olukorras, kus jalakäijal oli eesõigus ehk õnnetused toimusid reguleeritud või reguleerimata ülekäigurajal. Igahommikused liiklusteadet paneb prefektuuridest tulnud lühikirjelduse põhjal kokku Politsei- ja Piirivalveameti korrapidaja (Härson 2014). Oktoobri ja novembri teated valisin põhjusel, et sel perioodil suurenes õnnetuste arv ülekäiguradadel ning oli erinevaid juhtumeid: nendes näidetes esines õnnetusi nii laste kui ka vanemate inimestega, samuti oli surmaga lõppenud õnnetusi ning viiteid helkuri puudumisele. Oluline oli ka asjaolu, et sarnaseid sündmusi esitati erinevas sõnastuses.

Et selgitada, kas need osatulemused on kooskõlas lugejaarvamuse uurimusega, viisin uuringu teises etapis läbi küsitluse selgitamaks, kuidas liiklusteadete lugejad tajuvad süüdlast. Palusin 104 uuringus osalejalt hinnata nende 26 liiklusteate sõnastusi. Liiklusteadete kirjapilt oli muutmata. Kui ühel päeval oli kaks õnnetust, siis nummerdasin need teate esinemise järjekorras: nt 28. okt (2). Vastajad olid vanuses 17–50, nendest kolmandik oli põhi-, kolmandik kesk- ning kolmandik kõrgharidusega. Vastajatest oli mehi 45, naisi oli 59. Vastajate hulgas oli nii

autojuhte kui ka juhiloata inimesi. Liiklusteate sõnastuse põhjal palusin määrata, kes on õnnetuses süüdlane (tõmmata liiklusõnnetuse osalejale joon alla). Kui vastajad märkisid süüdlaseks nii autojuhi kui ka jalakäija, läksid mõlemad vastused arvesse.

Järgnevas peatükis esitan tekstianalüüsi koos näidetega, mis pärinevad sise-ministeeriumi liiklusteadetest. Analüüsiosa iga peatükk sisaldab kõrvuti mõlema uurimisetapi tulemusi: nii kriitilise diskursuseanalüüsi tekstinäited ja järeldused kui ka küsitluste tulemused selle kohta, kuidas lugejad tajuvad liiklusõnnetuste süüdlasi. Küsitluste tulemused on esitatud eraldi kolmel graafikul.

Esmalt on peatükis 3.1. sündmuste agentsuse lingvistiline analüüs. Järgneb sündmuste asjaolude esitus ning vaatepunkti analüüs: materjal on liigendatud esile kerkinud aladiskursuste kaudu, milleks on jalakäija eksimuste diskursus, surmaga lõppenud liiklusteated ning õnnetused lastega.

3. 2013. aasta oktoobri- ja novembrikuu liiklusteade analüüs

3.1. Kas juht sõitis otsa või jalakäija sai autolt löögi?

Lausetasandi analüüsis selgitan, millised tähendused esitatakse pealauses ning millised kõrvallause(te)s, kuidas väljendatakse agentsust. Kõikides oktoobri- ja novembrikuu liiklusteades nimetatakse pealauses, millisel ajahetkel ning kus *õnnetus toimus* ehk sündmus on esitatud juhtumina, kus tegutsejad, seega ka vastutajad puuduvad.

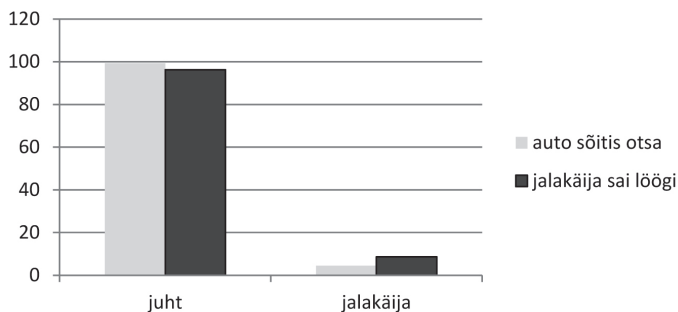
Esimese astme kõrvallaused olid oktoobri- ja novembrikuu uudistes sõnastatud peamiselt kahel viisil. Esimese raamistuse kohaselt (1. okt näide) saame esimese astme kõrvallauses teada, et *auto sõitis jalakäijale otsa*. Selliseid liiklusteateid oli oktoobris 9 ning novembris 4. Teise astme kõrvallauses lisatakse sündmuse agent: öeldakse juhi vanus ja eesnimi. Uudise viimane lause on passiivlause: *jalakäija toimetatakse ära*. Seega on liiklusteade sõnastatud viisil, kus sündmuses osalevad auto ning jalakäija: täpsemalt *sõidab auto jalakäijale otsa*.

- (1) 1. oktoobril kell 16.00 toimus liiklusõnnetus Tallinnas Pärnu maantee 238 juures, kus maastur Mercedes-Benz ML 270, mida juhtis 38-aastane Vadim, sõitis reguleerimata ülekäigurajal otsa teed ületanud 51-aastasele Enele. Jalakäija toimetati Põhja-Eesti Regionaalhaiglasse.

Teise raamistuse kohaselt saab *jalakäija autolt löögi* (7. okt näide). Selliseid teadete sõnastusi oli oktoobris 2 ning novembris 11, seega on liiklusteade sõnastusviis novembris võrreldes oktoobriga muutunud.

- (2) 7. oktoobril kella 07.59. ajal toimus liiklusõnnetus Harjumaal Maardus Keemikute 25 juures, kus reguleerimata jalakäijate ülekäigurajal sõiduteed ületanud 40-aastane Natalja sai löögi seni kindlakstegemata punast värvi kaubikult. Peale liiklusõnnetust aitas kaubiku juht jalakäija bussipeatusesse istuma ning lahkus seejärel sündmuskohalt. Jalakäija toimetati Ida-Tallinna Keskhaiglasse.

Järgnevalt kontrollisin küsitluse vastuseid, et hinnata, kas need kaks erinevat raamistust, kus üks on väljendatud tegevusena (*auto sõidab otsa*) ning teine teona (*jalakäija saab autolt löögi*), on mõjutanud ka vastajate hoiakuid. Selleks et tulemused oleksid võrreldavad, valisin ainult need uudised, kus pole lisainfot helkuri, alkoholi, ilmaolude, surma või muu kohta. Mõlemat raamistust esines kuuel korral. Nende 12 näite puhul võrdlesin, kas vastajad tajusid eri sõnastuste puhul liiklusõnnetuse süülisi erinevalt (joonis 1).



Joonis 1. Vastajate poolt märgitud süüdlane (juht või jalakäija) liiklusteadetes kasutatud väljendite *auto sõitis jalakäijale otsa* ning *jalakäija sai autolt löögi* korral

Nendel juhtumitel, kus sõnastuse kohaselt *auto sõitis ülekäigurajal jalakäijale otsa*, märgiti juhti süüdlaseks keskmiselt 99,5 korda ning jalakäijat 4,5 korral. Kui sõnastuses esines väljend, mille kohaselt *jalakäija sai liiklusvahendilt löögi*, märgiti juhti süüdlaseks keskmiselt 96 korral ning jalakäijat 8 korral. Seega kui jalakäijat esitatakse situatsiooni suhtes aktiivses rollis, suureneb nende vastajate osakaal, kes näevad jalakäija süüd suuremana ning autojuhi süüd väiksemana.

3.2. Jalakäija eksimuste diskursus

Liiklusteadetes eristuvad omaette diskursusena sõnumid, kus juhitakse tähelepanu asjaoludele, mille vastu oli eksinud jalakäija, ning nendele asjaoludele, mis takistas juhti jalakäijat ülekäigurajal märkamast. Oktoobrikuu uudistes on kahel korral lisatud helkuri puudumine (28. okt 1. uudis ja 30. okt näide).

- (3) 28. oktoobril (1) kella 17.05 ajal toimus liiklusõnnetus Jõhvis Pargi tänaval ajutisel reguleerimata ülekäigurajal, kus sõiduauto Audi 80, mida juhtis 55-aastane Alesha, sõitis otsa sõiduteed ületanud jalakäijale, kellel puudus nõuetekohane helkur. 37-aastane Aljona toimetati Ida-Viru Kesksaiglasse.

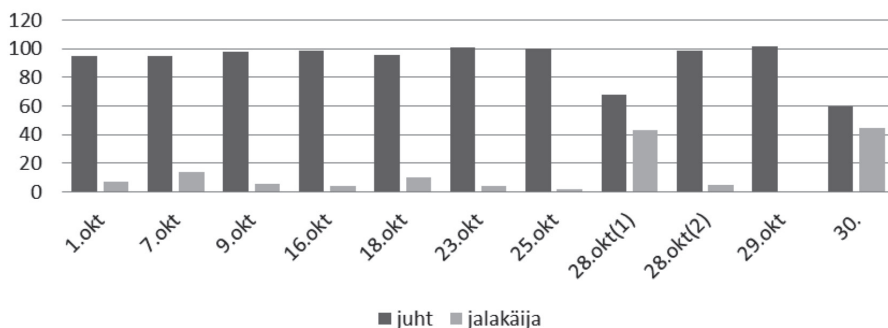
Selliste liiklusteadete sõnastuses ei ole ühtegi viidet neile tunnustele, mille vastu võis eksida juht (nt milline oli kiirus või kas juht rääkis telefoniga vms). Samuti pole arusaadav, kas kõigis ülejäänud õnnetustes kandsid jalakäijad helkurit või heledaid riideid.

Oktoobrikuu uudistes esines ka näide (30. okt), kus helkuri puudumine, jalakäija vanus ning äratoimetamine paigutatakse ühte lausesse. Jalakäija identiteet konstrueeritakse vanuse ning helkuri mitteomamise kaudu.

- (4) 30. oktoobril kella 17.39 ajal toimus liiklusõnnetus Võrus Jüri tänaval reguleerimata ülekäigurajal, kus sõiduauto Opel Astra, mida juhtis 23-aastane Elar, sõitis otsa sõiduteed ületanud jalakäijale. Helkurit mitte kandnud 49-aastane Ave toimetati Võru haiglasse.

Joonisel 2 on 104 vastaja hinnang oktoobrikuu õnnetuste süüdlase kohta. Neis liiklusteadetes, kus jalakäija eksimustele ei viidatud (1.–25. okt, 28 okt), tajusid vastajad süüdlasena valdavalt juhti. Kui uudiste sõnastuses on mainitud, et jalakäija ei kandnud helkurit (28. (1) ja 30. okt, vt näide (3) ja (4)), suurenes oluliselt nende vastajate arv, kelle hinnangul oli õnnetuses süüdlane jalakäija.

7. okt (näide 2) uudisesõnastuse kohaselt *aitas kaubiku juht jalakäija bussipeatusesse istuma ning lahkus seejärel sündmuskohalt*. Selle lisainfo mõjul suurenes nende vastajate arv, kes märkisid süüdlaseks jalakäija. Seda hoolimata asjaolust, et sündmuskohalt lahkunud juhi süü on oluliselt suurem kui sündmuskohale jäänud juhi süü. Vaid ühes uudises (29. okt) ei märgitud süüdlasena jalakäijat: nimelt sõitis selles liiklusteates üks autojuht teisele autojuhile ülekäigurajal otsa.



Joonis 2. 104 küsitleva hinnang oktoobrikuus ülekäigurajal toimunud liiklusõnnetuse süüdlase kohta liiklusteadete põhjal. 28. okt 1. uudises ja 30. okt uudises on viide jalakäija helkuri puudumisele

Liiklusteadetes on ohver vaikiv hää, tema versiooni juhtunust pole kuulda. Jalakäija *toimetatakse ära*. Õnnetuse põhjustaja saab seletuskirjas olukorda kirjeldada enda vaatepunktist. Isegi kui jalakäija kannab helkurit, ei pruugi see olla otsasõitja hinnangul piisav (nt 20. novembri 2. uudis).

- (5) 20. novembril (2) kella 17.02 ajal toimus liiklusõnnetus Haapsalus Tallinna maantee 4 juures, kus reguleerimata ülekäigurajal sõiduteed ületanud 59-aastane Tatjana sai löögi sõidukilt Ford Transit, mida juhtis 40-aastane Arvo. Jalakäija toimetati kiirabi poolt Läänemaa Haiglasse kontrolli. Jalakäijal oli helkur kinnitatud käekoti külge kuid jäi autojuhi jaoks varjatuks, samuti pimestasid vastu tulevate sõidukite tuled autojuhti.

Selline autojuhipoolne sündmuse kirjeldus võimaldab uudisesse lisada rohkesti lisainfot, mis soodustas otsasõitu. Nii on liiklusteated sõnastatud juhi vaatepunktist ning selline liiklusteade raamistus on kooskõlas lugejaarvamuse uuringuga: ka lugejad interpreteerivad liiklusteateid autojuhile soodsamal viisil.

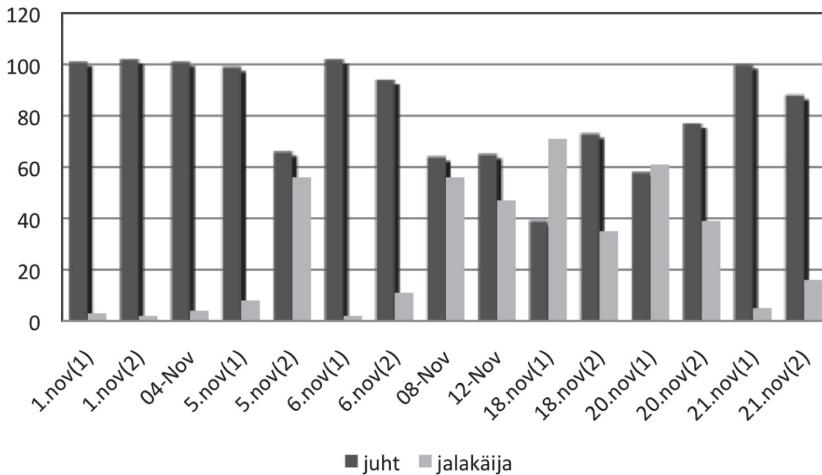
3.3. Surmaga lõppenud õnnetuste liiklusteated

Sama tendents on täheldatav ka surmaga lõppenud õnnetusjuhtumites. Kui liiklusteade (23. okt näide, vt ka 2. joonis) pole lisafakte jalakäija kohta, siis ei märkinud enamik uudise lugejaid süüdlaseks jalakäijat (102 korral märgiti süüdlaseks juht).

- (6) 23. oktoobril kella 15.50 ajal toimus liiklusõnnetus Narva linnas, Kangelaste 29 juures, kus sõiduauto Volkswagen Vento, mida juhtis 22-aastane mees, sõitis otsa reguleerimata ülekäigurajal teed ületavale jalakäijale, 78-aastasele Veerale, kes toimetati Narva Haiglasse, kus ta hiljem suri.

Teistes surmaga lõppenud õnnetuste liiklusteade (3. joonisel 5. nov (2), 20. nov (2)) on lisatud info jalakäija tumedatest riietest või helkuri puudumisest. See teave mõjutab, kuidas teksti lugejad tajuvad süülisust: 5. novembri 2. liiklusõnnetuses märgiti 66 korral süüdlaseks juht, 56 korral jalakäija.

- (7) 5. novembril (2) kella 17.30 ajal toimus liiklusõnnetus Narvas Kadastiku tänaval, kus sõiduauto Peugeot 307, mida juhtis 57-aastane Juri sõitis otsa reguleerimata ülekäigurajal sõiduteed ületanud jalakäijale. Jalakäija 66-aastane Anatoli toimetati Narva Haiglasse. Jalakäija oli riietatud tumedatesse riietesse ja tal puudus helkur. Jalakäija suri haiglas.



Joonis 3. 104 küsitleva hinnang novembrikuus ülekäigurajal toimunud liiklusõnnetuse süüdlase kohta liiklusteade (3. joonisel) põhjal

20. novembri 1. uudises oli neid vastajaid rohkem (vt ka joonis 3), kes märkisid jalakäija süüdlaseks: nimelt 58 vastajat märkisid süüdlaseks juhi ning 61 korral märgiti süüdlaseks jalakäija. Näites on esitatud loetelu selle kohta, kuidas keskkond mõjutab juhti ning lõpuks ka viide helkuri puudumisele.

- (8) 20. novembril (1) kella 07.00 ajal toimus liiklusõnnetus Tallinnas Paldiski maantee 227 juures, kus esialgsetel andmetel reguleerimata ülekäigurajal sõiduteed ületanud 70-aastane Galina sai löögi sõiduautolt Peugeot 206, mida juhtis 35-aastane Jüri. Jalakäija hukkus sündmuskohal. Õnnetuse hetkel oli pime, sadas vihma, teepind oli märg. Jalakäijal helkurit ei olnud.

Jalakäija kohta esitatud lisafaktid liiklusteadete sõnastustes kujundavad lugejate hoiakuid, ka jalakäija surm ei mõjuta olulisel määral vastajate hoiakuid. Mida rohkem on sõnastuses asjaolusid, mis selgitavad juhi seisukohalt raskendavaid asjaolusid (halvad ilmaolud, jalakäija helkuri puudumine), seda vähem tajuvad uudise lugejad, et juht oli liiklusõnnetuse põhjustaja. Seda hoolimata asjaolust, et seaduse järgi tohib helkurita jalakäija ja ka tumedates riietes jalakäija ülekäigurajal teed ületada.

3.4. Õnnetused lastega

18. novembri uudistes paistab silma kaks ühel päeval lastega toimunud õnnetust, kus selgelt eristub, kuidas sündmuse sõnastus mõjutab vastajaid. Esimese teate sõnastusele *ootamatult jalakäijate ülekäigurajale jooksnud 12-aastane poiss sai löögi* (vt ka 3. joonise 18. nov 1. õnnetus) märgiti 39 korral süüdlaseks juht ning 71 korral märgiti süüdlaseks jalakäija.

- (9) 18. novembril (1) kella 08.50 toimus liiklusõnnetus Tallinnas Paekaare 50 juures, kus ootamatult jalakäijate ülekäigurajale jooksnud 12-aastane poiss sai löögi sõiduautolt Renault Megane, mida juhtis 25-aastane Roman. Jalakäija toimetati Tallinna Lastehaiglasse.

Teises 18. novembri uudises on kasutatud neutraalsemaid väljendeid, mille kohaselt *tõukerattaga sõiduteed ületanud 8-aastane poiss sai löögi*. Sellise sõnastuse puhul märgiti 73 korral süüdlaseks juht ning 35 korral jalakäija (vt ka 3. joonise 18. nov 2. õnnetus).

- (10) 18. novembril (2) kella 13.00 ajal toimus liiklusõnnetus Tallinnas Telliskivi 20 juures, kus reguleerimata jalakäijate ülekäigurajal tõukerattaga sõiduteed ületanud 8-aastane poiss sai löögi sõiduautolt Mercedes-Benz, mida juhtis 44-aastane Marika. Laps toimetati Tallinna Lastehaiglasse.

Väljend *jooksis ootamatult ülekäigurajale* on sõnastatud autojuhi vaatepunktist ning suurendab oluliselt nende vastajate osakaalu, kes näevad õnnetuse põhjustajana jalakäijat. Liiklusseaduse kohaselt peab autojuht liikluses arvestama ka lastega. Sõnastustega *sõiduteed ületanud* või *ootamatult jalakäijate rajale jooksnud* saab teksti koostaja selgelt kujundada lugejate hoiakuid, luues uudiste keelekasutusega ka selle tõlgendust.

4. Järeldused ja diskussioon

Liiklusteadete pealaused on passiivlaused ning sõnastatud lausesemantiliste tunnuste agenttiivne/mitteagenttiivne alusel juhtumina (õnnetused toimuvad), mitte tegevuste, protsesside ega tegudena. Liiklusteadete esimese astme kõrvallauses on aktiivsed osalised jalakäija otsa sõitvad autod, ning sõidukilt löögi saanud jalakäijad. Autojuhti nimetatakse teise astme kõrvallauses. Uudised on raamistatud juhi vaatepunktist, millest tulenevalt on sõnumi fookuses eelkõige need aspektid, miks juht ei näe ülekäigurajal jalakäijat, mis segas ja takistas autojuhti.

Selline uudiste raamistus on kooskõlas lugejaarvamuse uuringuga: lugejad interpreteerivad liiklusteateid autojuhile soodsamal viisil. Mida rohkem sõnastati asjaolusid, mille vastu eksis jalakäija (helkuri puudumine, tumedad riided) või mis segasid autojuhti ülekäigurajal (halb nähtavus), seda rohkem tajuti süüdlasena jalakäijat. Juhi vaatepunkti lisamine liiklusteate sõnastusse (*sõiduteed ületanud asemel ootamatult ülekäigurajale jooksnud*) mõjutab oluliselt lugejate hoiakuid.

Sellisel liiklusteadete diskursusel on mitmeid põhjusi. Ühe selgituse kohaselt on inimesed, kes sündmusi konstrueerivad ning uudiseid loovad, empaatilised nende suhtes, kes olid otsasõitjad, ning vähem kaastundlikud ohvrite suhtes, kuna suheldakse autojuhtidega ning ohvritega pole enam kokkupuudet. Tekstide koostajate jaoks on uudiste koostamine rutiinne tegevus, mida tehakse välja kujunenud diskursuse korra järgi. Selliseid tekste koostatakse usus, et nad teenivad liikluskasvatuse eesmärke. Tuuli Härson (2014: 38) väidab: “Kuupäev, kellaeg, koht ja perfektselt tasakaalus osaliste loetelu – neid teateid oleks lihtne kokku panna, aga mis oleks selle mõte? Kui küsida mis tahes olukorras, kes on süüdi, siis küll see süüdlane leitakse. Kui tahame käitumist muuta, siis peab olema teates midagi, millest õppida.” Politsei lähtub Härsoni (*ibid.*) hinnangul põhimõttest, et “iga infokild, mis aitab õnnetust ära hoida, on äramärkimist väärt. Hukkunutega õnnetuste korral mobiliseerime end võimalikult kiiresti, et pressiteate vormis edastada iga õnnetuse taga olevat lugu, mis inimesi kõnetab.” Seega lähtuvad liiklusteadete kirjutajad usust, et nende eesmärk on eelkõige liiklejaid õpetada.

Nii jääb kõrvaliseks kõige olulisem asjaolu: autojuhid sõidavad ülekäigurajal jalakäijatele otsa ning inimesi saab surma. Samuti annab see avalikkusele selge sõnumi: ülekäigurajal õnnetusse sattunud jalakäijad on ise süüdi, kui autojuht neid ei näe. Sedasi pannakse ebaproportsionaalselt suur vastutus liikluses nõrgemale poolele. Artikli tulemused on kooskõlas rahvusvaheliste uuringutega, mille kohaselt liiklusõnnetuste diskursuses domineerib autokeskne lähenemine ning autota liiklejad on avalikus ruumis halvemas positsioonis (Freund, Martin 1997, Whatley 1988, Sheller, Urry 2000).

Ka on oluline asjaolu, et liiklusõnnetuse mõiste on liikluseaduses (2010) sõnastatud kui juhtum ajas ja ruumis, mitte tegevus ega tegu. Seaduses ei nähta vigastust ega surma personaalsena, küll aga juhtumina. Seetõttu lähtub teate edastanud korrapidaja Eesti legaalsest kontekstist. Kuni pole kindlaks tehtud teisiti, on kellegi surm õnnetusjuhtumi tagajärg, mistõttu ei saaks liiklusteadet ka agentselt esitada. Seega peaks lugejaarvamus viitama asjaolule, et legaalne kontekst suunab kirjutaja kaudu ka lugejat. Nii on tähenduse konstrueerimine keerukam protsess, kui käesolev artikkel suudab kirjeldada. Küll tõestas lingvistiline analüüs selgesti, et liiklusteateid on kaldu autojuhi õigustamise suunas ja unustavad legaalse konteksti: ülekäigurajale lähenemise loomuliku legaalse käitumismudeli.

Kindlasti peavad esmaste liiklusteadete sõnastajad hindama, kas selline meediakajastus on tõesti vajalik ja kas selle tulemusena Eestis liikluskultuur tegelikult ka paraneb. Eestis on seatud eesmärgiks vähendada jalakäijatega juhtunud liiklusõnnetuste arvu 2015. aastaks, sest viiest jalakäijatega juhtunud õnnetusest neli juhtub asulasisesel tänaval, õuealal või parklas (vt Abinõude valik... 2004). Kui liiklusõnnetustes hukkunute arv Eestis on vähenenud, siis jalakäijate puhul on see arv kolme aastaga suurenenud (17,7% asemel juba 25%).

Kas senine liikluskommunikatsioon vajaks ümbervaatomist? Olukorra muutmise võimalusena nähakse kaasata jalakäijaid kultuurilise tähenduse loomise protsessi ja tõsta nende mõjuvõimu ühiskonnas. Selleks peab jalakäijate kogukond senisest oluliselt rohkem kaasa rääkima nii kultuuri, kogukonna kui ka turvalisuse küsimustes (Demerath, Levinger 2003). Ühiskonnas on oluline luua tasakaalustatud dialoog, et liikluskultuuri kujundamisel arvestataks kõikide huvigruppide, ka jalakäijate, jalgratturite ning teiste gruppide, seisukohadega.

Üheks lahenduseks oleks liikluskasvatust ning liiklusõnnetusi kajastavate teadete selge lahushoidmine. Liiklusteadetele kehtiksid meediaeetika reeglid, mille kohaselt ei tohi sündmuste kajastamisel asuda kellegi poolele, kallutada avalikku arvamust ühes või teises suunas (Eetikakoodeks). Kui liiklusteaded on edastatud, on alati võimalus pöörata liiklejate tähelepanu turvalisust suurendavale asjaoludele: helkurile, kiivrile, aga ka õigesti valitud kiirusele, kutsuda liiklejaid üles olema vastastikku tähelepanelikumad, tegema paremat koostööd jne. Õnnetuste ohvrid on liikluses vähem kaitstud osa: eelkõige lapsed ja vanemad inimesed. Jalakäija ja auto kokkupõrkes ülekäigurajal on esimene nõrgemas positsioonis ning sellest on võimalus liikluskommunikatsioonis lähtuda.

Viidatud kirjandus

- Abinõude valik jalakäijate ja jalgratturitega toimunud liiklusõnnetuste ning neis kannatanute arvu vähendamiseks: II etapp. [Improvement of road safety situation of vulnerable road users – pedestrians and bicyclists.] Lõpparuanne. Tallinn: Maanteeamet, 2004. <http://www.mnt.ee/failid/Kergl.pdf> (5.5.2014).
- Demerath, Loren; Levinger, David 2003. The social qualities of being on foot: A theoretical analysis of pedestrian activity, community, and culture. – *City & Community*, 2 (3), 217–237. <http://dx.doi.org/10.1111/1540-6040.00052>
- Dijk, Teun A. van 2005. Ideoloogia. Multidistsiplinaarne käsitus. [Ideology: A Multidisciplinary Approach.] Tartu: Tartu Ülikooli Kirjastus.
- Eetikakoodeks. [Code of Ethics.] Eesti Ajalehtede Liit. <http://www.eall.ee/eetikakoodeks.html> (27.9.2014).
- EKG II = Ereht, Mati; Kasik, Reet; Metslang, Helle; Rajandi, Henno; Ross, Kristiina; Saari, Henn; Tael, Kaja; Vare, Silvi 1993. Eesti keele grammatika II. Süntaks. [Estonian Grammar II. Syntax.] Tallinn: Eesti Teaduste Akadeemia Keele ja Kirjanduse Instituut.
- Fairclough, Norman 1989. *Language and Power*. London: Longman.
- Fairclough, Norman 1991. *Discourse Processes and Social Change*. Cambridge: Polity Press.
- Fairclough, Norman 1992. *Discourse and Social Change*. Cambridge: Polity Press.
- Fairclough, Norman 1995. *Critical Discourse Analysis: The Critical Study of Language*. Harlow: Longman.
- Fairclough, Norman 2001. *Language and Power*. Second edition. Harlow: Longman.
- Fairclough, Norman 2005. *Critical discourse analysis in transdisciplinary research*. – R. Wodak, P. Hilton (Eds.). *A New Agenda in (Critical) Discourse Analysis*. Amsterdam: John Benjamins Publishing Company, 53–70.

- Fairclough, Norman; Wodak, Ruth 1997. Critical discourse analysis. – Teun van Dijk (Ed.). *Discourse as Social Interaction*. London: Sage, 258–284.
- Flowerdew, John 1999. Description and interpretation in critical discourse analysis. – *Journal of Pragmatics*, 31 (8), 1089–1099. [http://dx.doi.org/10.1016/S0378-2166\(99\)00049-1](http://dx.doi.org/10.1016/S0378-2166(99)00049-1)
- Freund, Peter E. S.; Martin, George T. 1997. Speaking about accidents: The ideology of auto safety. – *Health*, 1 (2), 167–182.
- Harvey, Lee; Macdonald, Morag 1993. *Doing Sociology: A Practical Introduction*. Macmillan Distribution.
- Huckin, Thomas 1997. Critical discourse analysis. – Tom Miller (Ed.). *Functional Approach to Written Text: Classroom Application*. Washington. <http://files.eric.ed.gov/fulltext/ED417422.pdf#page=87> (15.7.2014).
- Härson, Tuuli 2014. Liikluskommunikatsioon: sõnaga surma vastu. – *Kaja*, 15, 37–38.
- Kasik, Reet 2008. Meediateksti analüüs: eesmärgid ja meetodika. [Analysis of the media texts: Aims and methods.] – Reet Kasik (Toim.). *Tekstid ja taustad V*. Tartu: Tartu Ülikooli Kirjastus, 10–43.
- Laherand, Marika 2008. Kvalitatiivne uurimisviis. [Quantitative Methods.] Tallinn: OÜ Infotrikk.
- Leeuwen, Theo van 1996. The representation of social actors. – Carmen Rosa Caldas-Coulthard, Malcolm Coulthard (Eds.). *Texts and Practices: Readings in Critical Discourse Analysis*. London: Routledge, 32–70.
- Liiklusseadus 2010. [Traffic Act.] <https://www.riigiteataja.ee/akt/102072013012> (29.6.2014).
- McQuail, Denis 2000. McQuaili massikommunikatsiooni teooria. [McQuail's Mass Communication Theory.] Tartu: Tartu Ülikooli Kirjastus.
- Oatman-Stanford, Hunter 2014. Murder machines: Why cars will kill 30,000 Americans this year. <http://www.collectorsweekly.com/articles/murder-machines/> (28.7.2014).
- Pan, Zhondang; Kosicki, Gerald M. 2010. Framing analysis: An approach to news discourse. – *Political Communication*, 10 (1), 55–75. <http://dx.doi.org/10.1080/10584609.1993.9962963>
- Roberts, Ian; Wentz, Reinhard; Edwards, Phil 2006. Car manufacturer and global road safety: A word frequency analysis of road safety documents. – *Injury Prevention*, 12 (5), 320–322. <http://dx.doi.org/10.1136/ip.2006.012849>
- Sheller, Mimi; Urry, John 2000. The city and car. – *International Journal of Urban and Regional Research*, 24 (4), 737–751. http://www.livingrail.eu/images/pdf/2000_sheller_urry_the_city_and_the_car.pdf (20.8.2014).
- Serafini, Frank 2003. Informing our practice: Modernist, transactional, and critical perspectives on children's literature and reading instruction. – *Reading Online*, 6 (6). http://www.readingonline.org/articles/art_index.asp?HREF=serafini/index.html (12.8.2012).
- Vallières, Lucie 2006. *Disciplining Pedestrians? A Critical Analysis of Traffic Safety Discourses*. MA thesis. Simon Fraser University.
- Whatley, Marianne H. 1988. Raging hormones and powerful cars: The construction of men's sexuality in school sex education and popular adolescent films. – *Journal of Education*, 170 (3), 100–121.
- Widdowson, Henry G. 1998. The theory and practice of critical discourse analysis. – *Applied Linguistics*, 19 (1), 136–151. <http://dx.doi.org/10.1093/applin/19.1.136>
- Wodak, Ruth 1996. *Disorders of Discourse*. London–New York: Longman.

Katrin Aava (Tallinna Ülikool) peamised uurimisvaldkonnad on meediatekstide, haridustekstide kriitiline diskursuseanalüüs.
 Narva mnt 29, 10120 Tallinn, Estonia
aava@tlu.ee

WHO'S RESPONSIBLE? THE LINGUISTIC CONSTRUCTION OF PARTICIPANTS IN TRAFFIC ACCIDENTS

Katrin Aava

Tallinn University

Reports on traffic accidents carry societal meanings and judgements about the value of human life, role and responsibility of all parties and who is to be blamed. Critical discourse analysis used in this research is a method that allows to detect such hidden meanings of societal agreements and interpretations encoded in texts. To exemplify how the roles of different parties are construed in traffic news, were analysed 26 police reports about traffic accidents that happened on crossroads in October and November 2013. To find out how such framing influences readers in interpreting the text, were questioned 104 respondents.

The text analysis showed that traffic news stories are being framed through the driver's viewpoint and attention is drawn to the pedestrian's mistakes (lack of a reflector, dark clothing). Traffic reports are phrased in such a way that accidents *happen*, their active participants are *cars which hit pedestrians* and pedestrians who are hit by cars. The questionnaire revealed that the more emphasis is put on the factors that distracted the driver, the more readers perceive pedestrians as responsible for the accident.

Keywords: critical discourse analysis, framing, viewpoint, traffic reports

EXTERNAL FACTORS AND THE INTERFERENCE OF L1 ESTONIAN ON L2 ENGLISH PRONUNCIATION: AN APPARENT-TIME STUDY

Kristiina Ader, Merilin Miljan

Abstract. This study focusses on external factors of second language learning (L2) and their effect on L2 sound production. The aim was to find out whether young adults whose first language (L1) is Estonian speak L2 English with less accent than older speakers of L1 Estonian. Prior studies have claimed that more exposure to the target language lessens the effect of L1 interference (internal factors) (e.g. Piske et al. 2001, Muñoz, Llanes 2014). An apparent-time study was implemented in order to test whether a change in learner setting (i.e. limited exposure to the target language vs. abundant exposure) shows any evidence which can be explained by such external factors. Data was elicited from the speech samples of 97 speakers of L1 Estonian and L1 interference effects were analysed sound by sound. The results show that the younger generation displays less L1 Estonian interference than the older generation. Both greater exposure to the target language and earlier age of L2 learning onset correlate with the ability to produce more sounds target-like.*

Keywords: second language acquisition, L2 phonology, external and internal factors, English, Estonian

1. Introduction

From very early age on, speakers attune to the phonetic system of their first language (L1). Prior work has established that when another language is learned or acquired, the phonetic system of L1 interferes with the new sound system of a foreign language (L2) resulting in a “foreign” accent (e.g. Flege 1991, Piske et al. 2001). The focus of research in L2 phonology has been on pinning down the exact factors which cause such an interference in L2 pronunciation. Until recently, the explanations were based

* We are very grateful to Madis Paukson, Lemme Kitsing, and teachers of Tallinn Mustamäe College for their help and assistance in collecting data, and to Miriam Anne McIlfratrick-Ksenofontov for helping to analyse the data. We also thank all the informants who took part in the study, and the two anonymous reviewers for their useful comments and suggestions.

entirely either on internal, i.e. neuro-cognitive factors such as the age of L2 learning, attitude and motivation (e.g. Scovel 1988, Piske et al. 2001); or on external, i.e. social, factors such as the amount and type of exposure to the pronunciation of the target language, the length of residence in the country of the target language, among others (e.g. Flege, Liu 2001, Brown 2007). Currently researchers have started to question the strict divide between the internal and external factors, especially Moyer (2014), and are discussing the importance of studying both types of factors together in order to find the intersection of these distinct factors. This paper provides a contribution to this debate by studying the pronunciation of native speakers of Estonian in their L2 English in a wider time span. It reports an apparent-time study which surveys two different generations of a population at one point in time with the aim to find out whether and how external factors, such as the quantity and quality of experience in the target language have affected the L2 English phonology of L1 Estonian speakers and how it interplays with the age of L2 learning onset.

The second aim of the study, but no less important, is to provide empirical evidence to the “problematic” areas in L2 English pronunciation of L1 Estonian speakers, as outlined by few observations (Kostabi 1993, Mutt 1971). These findings have implications for teaching the pronunciation of L2 English to L1 Estonian speakers.

1.1. Internal and external factors in L2 pronunciation

As mentioned above, prior studies have established that children acquire the phonetic system of a foreign language relatively easily compared to adult language learners (e.g. Piske et al. 2001, Flege 1991). Thus age is seen as one of the most important factors and formulated as the Critical Period Hypothesis, which states that “complete mastery of an L2 is no longer possible if learning begins after the end of the putative critical period” (Piske et al. 2001: 195). This hypothesis has generated many studies on internal factors affecting L2 pronunciation. For example, Granena and Long (2013) studied Chinese learners of L2 Spanish (starting at the age of 3–6, 7–15, and 16–29) and found that there are sensitive periods in different language domains: the ability to acquire phonology starts to decline before the ability to acquire lexis and collocations, which in turn declines before morphosyntax; the latest ages of onset which yielded a native-like acquisition were 5, 9, and 12, respectively. Piske et al. (2001) analysed the pronunciation of L2 English by L1 Italian speakers, focusing on the possible influence of age on L2 learning, the extent of their L1 use, and the length of residence in the country of the target language, Canada. They found that although both the age of onset of L2 learning and the extent of L1 use play a significant role in gaining the native-like pronunciation, the age of L2 learning onset appeared still to be more influential. Yet there are studies which question the onset of L2 learning as the major factor. For example, Flege et al. (2006) compared the accent in L2 English of native Korean children and adults and found that children had a detectable foreign accent, although “milder” than the one of adults. This raises a question whether the late learners are predetermined to be significantly less successful at achieving native-like pronunciation than younger learners.

Learner’s L1 interference with the pronunciation of an L2 can be attributed to the articulatory settings of L1. That is, speakers tend to produce the sounds of a

new phonological system which sound similar to the sounds in their L1 in a similar manner: they may substitute certain L2 sounds with the sounds of their L1. This is where the difference between late and early learners lies, because younger learners are less attuned to their L1's phonetic system and therefore are able to establish new phonetic categories more easily (e.g. Flege 1991). Yet, there are late learners who have managed to acquire almost native-like pronunciation. As Marinova-Todd et al. (2000) note, researchers tend to focus on the average scores of adult informants and ignore the most successful ones. One of the studies which has specifically focussed on native-like learners is by Abrahamsson and Hyltenstam (2009). They conducted a large-scale study among late and early L2 learners of Swedish (L1 Spanish) and selected a group of most native-like sounding individuals, as picked by L1 Swedish judges. Only a small minority of late learners and a majority of early learners were perceived as L1 speakers of Swedish. Upon further examination, it was found that none of the late learners performed native-like once their L2 pronunciation was scrutinised in detail; neither did the majority of early learners. The study concluded that "nativelike ultimate attainment of a second language is, in principle, never attained by adult learners and, furthermore, is much less common among child learners than has previously been assumed" (Abrahamsson, Hyltenstam 2009: 249). A more recent study by Moyer (2014) analyses several previous case studies which feature late learners who have achieved native-like pronunciation. She observes that the "exceptional" learners have benefited from both internal (e.g. aptitude and motivation) and external factors (e.g. the amount of use and exposure to the target language). For this reason, Moyer emphasises the need for an integrated approach to studying learner factors, as no isolated factor is prevalent enough to determine the whole outcome of pronunciation learning.

To lessen the interference of learner's L1 with the target language, it has been suggested that continuous practice in L2 is necessary (e.g. Piske et al. 2001, Muñoz, Llanes 2014). For instance, Muñoz and Llanes studied two groups of Catalan-Spanish bilingual children and adults (with respective mean ages 10.4 and 22.36) over the period of three months in different English learning settings: home and abroad. The aim was to find what impact different learning settings have on the informants' L2 pronunciation and whether the effect is similar for both adults and children. Speech samples taken before and after the three month period showed that the groups of children and adults who had studied English in the country of the target language had significantly improved their L2 English pronunciation (according to raters' evaluation) in comparison to those who stayed at home. The children's speech was rated more target-like than the one of the adults, possibly because they had spent more time interacting with native speakers abroad. This observation is often referred to as Single System hypothesis (e.g. Flege et al. 1997, Guion et al. 2000), which states that the more learners speak their L1, the more noticeable the L1 accent in their L2 is.

From a pedagogical perspective, the issue of foreign accent is constantly under discussion. One solution suggested in the literature is the intelligibility principle (e.g. Levis 2005, Derwing, Munro 2005), according to which a foreign accent is acceptable, but only tolerated to the extent that it does not affect the intelligibility of speech. As long as there is a mutual understanding, it is unnecessary to demand a native-speaker-like accent from both the learner and teacher. The learning of L2

pronunciation should thus focus on aspects which cause untelligibility (e.g. the production of L2 sounds and prosody) and be instructed accordingly, as stated by Schaetzel and Low (2009), Derwing and Munro (2005), among others. A related question is which variety of English should be considered for teaching: British English or American English or some other. This is a possible cause of confusion among teachers and students alike, as pointed out by, e.g. Levis (2005), Hamid and Baldauf (2013).

Furthermore, the concept of “native speaker” has been scrutinised recently. It is noted that the concept essentially refers to an abstract entity, an ideal, while in reality native speakers can be affected by their L2 or regional varieties. To overcome the issue, some scholars (e.g. Abrahamsson, Hyltenstam 2009) “operationalize” the native speaker according to some basic criteria, e.g. exposure to one certain language since infancy, the same language is used for instruction at school and as a language of primary communication.

1.2. Prior studies on L1 Estonian interference

How Estonian as L1 affects the pronunciation of L2 English has received little attention thus far. Among the few studies is McAllister et al. (2002) which focusses on how L1 phonetic features affect the acquisition of L2 Swedish quantity distinctions and compares the results of L1 Estonian, Spanish and English speakers. The study found that the group of L1 Estonian speakers was the second most successful in achieving a native-like production and perception¹ of quantity, since the quantity distinction is phonologically relevant in Estonian, but not in English or Spanish. The other studies on how Estonian as L1 influences the pronunciation of L2 English are Mutt (1971) and Kostabi (1993). Mutt focusses on describing the characteristics of Estonian accent in L2 English in his textbook, but there is no mention of how the data was elicited. Kostabi (1993: 2) states that previous teaching materials are “[u]seful but mostly intuitive suggestions about the differences between the sounds and prosody of English and Estonian,” a fact which inspired her to carry out a systematic contrastive analysis of L1 Estonian sounds in L2 English. She examined annual tests of first year English majors at the University of Tartu and, as a result, wrote instructions for L1 Estonian learners how to improve pronunciation in L2 English. Kalk (2001) observed students in elementary and primary schools and her findings confirmed the observations made by Mutt (1971) and Kostabi (1993). Kalk also inquired teachers of English about the pronunciation training in classroom and found that not enough time is spent on it.

In sum, all previous studies on how Estonian interferes with the pronunciation of L2 English are qualitative and focus on providing detailed descriptions of L1 Estonian sounds in a sound system of L2 English. One of the aims of this paper is to measure the “foreign” accent in a target language by quantitative methods in order to provide a more objective point of comparison between different groups of learners. Thus this study is the first attempt to provide empirical and quantitative data about the effect of L1 Estonian on L2 English phonetic system.

¹ For example, when asked whether the word was produced correctly or not, Swedish informants success rate was 99.8%, Estonians' 98.1%, English' 89.2% and Spanish' 68.4%.

2. An apparent-time study: impact of internal and external factors in two different cultural situations

The questions which this study addresses are as follows: a) whether and how external factors, specifically the quality and quantity of exposure to L2 English have affected the learning of L2 pronunciation by L1 Estonian speakers; and b) does the later age of L2 learning onset correlate with a “stronger” L1 Estonian accent in L2 English?

In addition, we test two hypotheses. Hypothesis 1 states that younger adults speak L2 English with less accent than elderly L1 Estonian speakers. The assumption is that since younger generation has studied English in a different cultural situation than older informants, they have been more exposed to the target language from early age on; hence, according to Piske et al. (2001) and Flege (1991), they should show less L1 interference with their L2 sound system. The apparent-time study is implemented in order to compare the impact of a different cultural situation on the informants’ pronunciation. Hypothesis 2 states that a naturalistic setting of language learning and more frequent communication with native speakers of English results in less L1 interference in the informants’ L2 English, as pointed out by, e.g. Muñoz and Llanes (2014).

2.1. Method

The data is drawn from speech samples which were recorded during the period of 16–25 April 2014. The speech samples were analysed together with a native speaker of British English². The target of comparison was the BBC accent of English³ for two reasons: it is perceived as the standard of British English, and it is the variety which is mostly taught at Estonian schools (Kull 2006).

2.2. Informants

The total number of informants was 97 speakers of L1 Estonian (41 males, 56 females). They were divided into two groups: a younger generation, Group 1, and older generation, Group 2. Group 1 comprised 68 students from Tallinn Mustamäe College (TMG)⁴ and 4 students majoring in English at Tallinn University (mean age 20.5). Group 2 consisted of 25 middle-aged adults (mean age 53.74). More information on the background of the informants is given in Section 3.

The speech samples from TMG students were gathered during their English lessons by the first author and an assistant. Data from Group 2 was gathered in various places by the first author which explains the smaller sample size.

2.3. Tasks

The participants were presented with two tasks. Task 1 included 20 sentences, see (1)–(3) below as an example. The sentences were compiled on the basis of the textbooks by Kostabi (1993, 2004) and Mutt (1971) and included potential

² The aim was to determine how the speech samples were perceived by a native speaker who is a qualified IELTS examiner.

³ That is, the accent which used to be referred to as received pronunciation (RP).

⁴ Both majoring and minoring students were included.

pronunciation errors (e.g. *honey*, if produced with the accent, it can sound like Estonian *hani* ‘goose’). The test sentences were kept simple, avoiding complicated words and overcrowding with a single phoneme (e.g. tongue twisters such as ‘She sells seashells by the seashore’). Each sentence featured several examples of different phonemes that were pointed out by Mutt and Kostabi as potential errors, e.g. the phoneme /l/ is pronounced clear word-finally, but as dark in the BBC accent of English.

- (1) **Nelly** is a **very** good **swimmer**.
- (2) I had to apply **something cool** to my **wrist** after **stumbling** in the **dark room**.
- (3) **Sir**, may I suggest that **we** have **our coffee** on the **lawn**.

Task 2 was a short story of 15 sentences, see (4) below for an example. As in Task 1, the aim was to keep the test words and sentences simple. Many words which occurred in Task 1 were also used in Task 2. The aim of Task 2 was to check the pronunciation of certain sounds, such as /ŋ/ in ‘stumbling’ and /ð/ in ‘there’.

- (4) **There** was an old **man** sitting by the fireplace. **He** heard **stumbling** at the **door**. Soon entered a **tall man** who **wore ragged** clothes.

Both tasks included a written instruction in the Estonian language.

2.4. Procedure

The tasks were administered to the informants by the first author and an assistant. The informants were allowed to read the tasks once. Then they were asked to dictate the sentences while being recorded. The recording sessions were conducted individually in a quiet room. All the informants received an individual number in order to maintain their anonymity.

2.5. Analysis

Informants’ L1 accent was measured on the basis of the amount of phonemes they produced in the target language and how many of these sounds were from the L1 Estonian sound system. We counted the relative frequencies of instances where the English sound was replaced by an Estonian one. The number of phonemes in English is 35 (Roach 2000), thus the maximum number of possible sounds was 35. When a target phoneme was produced as an L1 Estonian sound up to two times, it was regarded an inconsistent L1 interference, i.e. an occasional occurrence; when at least three times, it was regarded a consistent occurrence. When no L1 sound instead of a target sound was detected, the target sound was considered to be native-like. No testing for statistical significance was conducted. When any word was substituted with another word or pronounced entirely incorrectly, it was not included in the analysis. Finally the results of Group 1 and 2 were compared.

3. Results

3.1. L1 Estonian interference in L2 English

The study provides empirical data on the characteristics of sounds produced by L1 Estonian speakers in L2 English that were noted by Mutt (1971) and Kostabi (1993). The relative frequencies of producing L2 English sounds as L1 Estonian sounds are summarised in Table 1.

Table 1. L2 English sounds produced with L1 interference (%)

English sound		Group 1	Group 2	Both groups	
1	i:	18	56	28	
2	ɪ	35	60	41	
3	u:	32	88	46	
4	ɔ:	22	68	34	
5	v	54	68	58	
6	θ	43	16	36	
7	ð	71	56	60	
8	r	Rolled	7	80	26
		Post-vocalic	93	20	74
9	k	67	72	68	
10	p	63	92	70	
11	t	13	36	20	
12	g	11	24	14	
13	b	39	64	45	
14	d	24	32	26	

The data in Table 1 shows that in the production of L2 English vowels the L1 Estonian speakers tend to retain their L1 vowel quality, i.e. the vowels are produced more close⁵ than in the BBC accent of English. For example, the long vowel sound /i:/ (sound (1) in Table 1), as in *week* in the target language, was kept identical to the L1 Estonian sound by 28% of all the informants. The short vowel sound /ɪ/ (sound (2) in Table 1) of the target language, as in *silk*, was produced much closer, i.e. identically to L1 sound, by 41% of the informants. This indicates that the informants distinguished quantity instead of quality, as in their L1. Another notable tendency was producing the long vowels /u:/ and /ɔ:/ (sounds (3) and (4), respectively in Table 1) of the target language with much stronger lip-rounding, i.e. articulating L1 sounds instead of the target language sounds that have more neutral lip-position; this happened in 46% and 34% of the informants, respectively. These tendencies, i.e. close front and lip-rounding, were also noted by Mutt (1971) and Kostabi (1993).

The production of consonant sounds by the informants also confirms the observations in Kostabi (1993) and Mutt (1971). There is a clear tendency to substitute the labio-dental fricative /v/ with the bilabial approximant /w/ or close back vowel /ʊ/ in L2 English. That is, /v/ (sound (5) in Table 1) was substituted with /w/ or /ʊ/ by 58% of the informants in the word initial position (e.g. in words such as *valley*,

⁵ The tongue is in higher position, closer to the hard palate.

vain, vest); the lenis dental fricative /ð/ (sound (7) in Table 1) was substituted with /d/ by 60% of the informants. Another significant feature which can be attributed to the L1 interference was rolling the L2 post-alveolar approximant /r/ (sound (8) in Table 1) which happened among 80% of the informants in Group 2 and among 7% in Group 1. Interestingly, /r/ was pronounced in the post-vocalic position almost without an exception by the younger generation, while in Group 2 only 20% of the informants pronounced /r/ in that position; the post-vocalic /r/ is unacceptable in the BBC accent, but common in general (or standard) American English.

A clear instance of L1 Estonian interference is the unaspiration of initial fortis plosives in L2 English. The sounds /k/, /p/, and /t/ (sounds (9), (10) and (11) in Table 1, respectively) were left unaspirated, i.e. produced as L1 Estonian sounds. For example, /p/ was produced in such a manner by 70% of the informants. Note that unaspirated initial plosives tend to be perceived as lenis plosives by native speakers of English (Roach 2000). On the other hand, initial plosives /g/, /b/, and /d/ (the respective sounds (12), (13) and (14) in Table 1) were produced as aspirated or voiced, e.g. /b/ by 45% of the informants. Voicing /g/, /b/, and /d/ deviates from the target language sounds which are simply unaspirated. This voicing of initial lenis plosives is not a characteristic of L1 Estonian, but possibly a result of over production in voicing, i.e. hypercorrection.

In sum, Group 1 produced less sounds with notable L1 interference than Group 2. The overall presence of L1 sounds in L2 English by group is given in Table 2. Note that the term “inconsistent” refers to the sounds which were produced as L1 sound according to manner or place of articulation up to two times, while “consistent” refers to phonemes produced as L1 Estonian sounds more than three times.

Table 2. Comparison of L1 interference by groups (%)

Group	Consistent	Inconsistent	No accent
1	11.8	21.9	66.3
2	21.5	30.5	48

Note that the relative frequency of sounds produced inconsistently with L1 interference is higher than the frequency of sounds consistently pronounced with the detectable L1 accent. This indicates that the informants are aware of the difference between the Estonian and English phonemes, but the Estonian phonemes tend to override the English ones at times, possibly due to habit.

3.2. External factors affecting L2 English pronunciation

The study compared two generations of L1 Estonian speakers in apparent-time, focussing on the quality and quantity of experience in L2 English and its interplay with age. The factors under study were the amount of formal English instruction, the number of English lessons a week, the effect of setting on English learning (i.e. the amount of time spent in the country where the target language is spoken, or having an L1 English teacher; both can improve the quality of exposure), and intensity of communication with native or non-native speakers of English.

The first factor affecting L2 pronunciation which clearly emerged was the amount of formal English instruction (see Table 3 below). The informants were asked where they had received English instruction and whether they had also studied English independently. Note that the relative frequencies in Tables 3–6 refer to the amount of L2 English phonemes produced as L1 Estonian sounds consistently, inconsistently or native-like (see also Section 2).

Table 3. The effect of formal instruction on L2 English pronunciation: the amount of L2 English phonemes produced with L1 interference (%)

Factors	Group	Consistent	Inconsistent	No accent
Kindergarten ⁶	1 (5) [†]	10.9	18.2	70.9
	2	–*	–	–
Elementary, primary, and high school ⁷	1 (33)	12.3	22.8	64.9
	2	–	–	–
Primary and high school	1 (11)	12.5	32.7	54.8
	2 (7)	26.1	31	42.9
University	1 (4)	14.2	18.5	67.3
	2 (16)	19.4	31.1	49.5
Additional independent learning	1 (19)	10.2	15.8	74
	2	–	–	–

* No informants fitted in the category, or the number of informants was too small for comparison (also applies to Tables 4–6).

[†] The number of informants in the category (also applies to Tables 4–6).

Table 3 shows that more time spent on learning the target language correlates with a more target-like sound production. In Group 1, the onset of L2 formal instruction in kindergarten results in less L1 interference, as noted in the case of 29.1% of phonemes. Learning English independently in addition to 11–12 years of formal instruction decreases the features of L1 Estonian sounds in L2 English phonemes to 26%. The same correlation holds for Group 2: the onset of L2 English learning at the age of 14–15 resulted in 42.9% of the English phonemes produced target-like. The continuation of learning English at university reduced the instances of detected L1 interference to 50.5%. Overall, Group 2 produced only 49.5% of the phonemes native-like, even despite instruction at the university level. This result is similar to the informants in Group 1 who had least formal instruction, i.e. only at primary and high school level. The number of English lessons per week was also examined, but it had no effect on the informants pronunciation within each group or between the groups.

One of the hypotheses was that a more frequent communication with L1 English speakers has a positive impact on the pronunciation of L2 English (see Hypothesis 2 above). The data in Table 4 shows that this is indeed the case: a monthly communication with native speakers of English results in less L1 Estonian accent in L2 English, i.e. 30.1% of the sounds for Group 1, and 51.4% of the sounds for Group 2. In contrast, the informants in Group 2 who had never spoken to L1 English speakers display L1 interference in 66.9% of the phonemes. Overall, Group 2 showed more signs of L1 interference than Group 1.

⁶ Up to 7-year-old children.

⁷ Elementary school: age 7–11, Primary school: age 11–16.

Table 4. The effect of frequency of communicating with L1 English speakers on the presence of L1 interference (%)

Frequency	Group	Consistent	Inconsistent	No accent
Daily	1 (9)	10.5	22.8	66.7
	2	–	–	–
Couple of times a week	1 (11)	10.9	17.7	71.4
	2	–	–	–
Couple of times a month	1 (18)	11.4	18.7	69.9
	2 (3)	18.1	33.3	48.6
Couple of times a year	1 (24)	11.7	23.9	64.4
	2 (17)	18.2	29.6	52.2
Never	1 (10)	14.9	26.5	58.6
	2 (5)	34.9	32	33.1

However, a more native-like pronunciation in L2 English was not supported by the results of those informants who reported conversing with L1 English speakers on a daily basis. They should have been the most successful group according to the hypothesis. The likely reason is a possibility that the informants interpreted ‘talking’ as ‘messaging online’, since some informants from the group noted that they had learned English online. Still, there were 4 informants in this category who produced nearly 70% of the phonemes identical to the target sounds.

The frequency of communicating in L2 English with other non-native speakers of English was also examined (see Table 5 below). The aim was to see whether it had a similar impact on the informants’ pronunciation as communication with L1 speakers of English.

Table 5. The effect of frequency of communicating with L2 speakers of English on the presence of L1 interference (%)

Frequency	Group	Consistent	Inconsistent	No accent
Daily	1 (26)	10.9	20	69.1
	2 (5)	11.4	26.9	61.7
Couple of times a week	1 (25)	12	22.9	65.1
	2	–	–	–
Couple of times a month	1 (12)	11.5	24	64.5
	2	–	–	–
Couple of times a year	1 (5)	14.3	22.3	63.4
	2 (13)	22.9	31.6	45.5
Never	1 (4)	14.3	21.4	64.3
	2 (3)	38.1	32.4	29.5

Table 5 shows that the frequency of communicating with other L2 English speakers is not relevant to Group 1, since the differences in the amount of communication do not vary much. In contrast, the results of Group 2 do show that speaking to other L2 English speakers has a positive impact on their L2 pronunciation: the occurrence of L1 Estonian sounds in their L2 English decreased as the frequency of communication increased.

To see whether the setting of L2 English learning has any influence on the informants' pronunciation (see Hypothesis 2 above), the informants were asked whether they had lived in an English-speaking country (natural setting) or had a teacher whose L1 was English. The results are shown in Table 6.

Table 6. The effect of the setting of English learning on the presence of L1 interference (%)

Factors	Group	Consistent	Inconsistent	No accent
Natural setting	1 (5)	9.1	10.3	80.6
	2	-	-	-
L1 English teacher	1 (5)	12	13.7	74.3
	2 (5)	20	31.4	48.6
Neither	1 (62)	11.9	23.3	64.8
	2 (20)	23	31.3	45.7

Table 6 shows that having an L1 English teacher or living in an English-speaking country does have a positive effect on achieving a native-like pronunciation. The informants in Group 1 who had neither L1 English teacher nor lived in a target-language country produced L1 Estonian sounds in 35.2% of the phonemes, whereas living in L1 English country diminished the L1 interference to 19.4% of the sounds in the same group. Group 2 showed a smaller difference between those who had an L1 English teacher and who did not. Overall, Group 2 produced less target-like sounds than Group 1 under similar conditions.

4. Discussion

The results of the study show that internal and external factors act together. The younger informants (Group 1) produced more native-like sounds in L2 English because of the drastic change in the learning setting of L2 English and acquisition: nowadays, L2 English learning begins earlier and there is more exposure to the target language via native speakers and media, among others, compared to the setting in which the older generation (Group 2) studied English. In a way, the two groups represent the divide between language learning and language acquisition. That is, Group 2 has mainly received formal instruction and less natural communication, hence is an example of language learning; whereas Group 1 is an example of language acquisition via natural communication.⁸

This study confirms that the external factors noted by prior studies (e.g. Piske et al. 2001, Muñoz, Llanes 2014) contribute to a more target-like pronunciation. Specifically, 60% of the informants reported watching English TV shows or films almost on a daily basis. In relation to this, it was noticed that general American accent was present in speech samples of the many younger informants. For example, Group 1 showed a tendency to substitute /ɑ:/ with the relevant shorter version of the sound which is found in general American English and which occurs when the /r/ is pronounced after the long vowel. This presence of post-vocalic /r/ (produced by 93% of young informants, 7% rolled /r/) is a clear evidence of the increased exposure to the target language, in this case US English due to the US based film

⁸ Thanks to the anonymous reviewer for pointing this out.

industry (Asu 1999, Kull 2006). The older generation featured only some instances of post-vocalic /r/, as 80% of them rolled /r/ instead. As Kull (2006) notes, they have been more exposed to British English.

The data shows a strong correlation between the onset of L2 English learning and a more native-like pronunciation of the target language. Thus the results tally with the observations in Piske et al. (2001) and Flege (1991) who noted a smaller prevalence of L1 accent in the L2 speech of the early learners. Nowadays students start learning English at the age of 8–9 (as reported by 46% of the informants), while the older age group mostly reported starting learning English at the age of 14–15. The earlier onset, as well as the exposure to natural communication, encourages a more target-like pronunciation among the younger group.

The findings of this study support the hypothesis that people with more exposure to the target language experience less interference with the L1 phonetic system. Furthermore, the older informants were more affected by the overall frequency of communicating whether with native or non-native English speakers: the informants who reported communicating with L2 speakers of English on a daily basis showed L1 interference in 38% of the phonemes, while no communication resulted in 70% of the sounds produced with an L1 Estonian accent. Therefore, even communicating to non-native speakers of L2 seems to have a positive impact on L2 pronunciation. The younger group did not show much of an impact by communicating with non-native speakers. The likely reason is that they reported generally more communication in L2 English with other speakers of L2 English than the older group, hence the amount of L1 accent within the younger group was determined by less or more exposure to L1 English speech.

The data of Group 1 also confirms the Critical Period Hypothesis to some extent. The informants of Group 1 who had received English instruction in kindergarten were the second most successful in producing native-like sounds, i.e. 71% of the phonemes were produced with no detectable L1 Estonian accent. Moyer's (2014) conclusions, which emphasise motivation as one of the influential factors, were firmly confirmed by our study: we found that the group of informants who had studied English independently in addition to formal instruction pronounced 74% of the phonemes target-like, thus having the highest result. Thus motivation is indeed an influential factor in L2 learning.

As for L1 interference, this study provides empirical evidence that the main characteristics and sounds of the L1 Estonian that are produced in L2 English speech have remained the same since Mutt (1971) and Kostabi (1993). Yet the quantitative data of this study shows a difference between the younger and older generations. The younger informants produced 66% of the phonemes identical to the sounds of the target language, which is more than the amount of the target-language sounds produced by the older group, 48%. This difference can be explained by a change in language learning setting, hence also in a completely different learning experience: the acquisition of L2 English rather than learning of L2 English. External factors such as the quality and quantity of L2 exposure thus seem to affect L2 pronunciation towards a more native-like pronunciation.

5. Conclusion

The study aimed to measure L1 Estonian interference and therefore to provide a quantitative aspect to the studies of L1 interference. It shows that the sounds produced by L1 Estonian speakers in L2 English have remained the same over time, thus supporting the observations made by Mutt (1971) and Kostabi (1993).

One of the purposes of this study was to determine how external factors, specifically the setting of L2 English learning and the quality and quantity of exposure to the target language affect the pronunciation of target language sounds. Previous studies have found evidence that exposure to the target language in naturalistic setting or to the communication with native speakers of the target language diminishes the characteristics of one's L1 sounds in their L2 speech and enhances native-like pronunciation. The findings of this study support these observations. It was found that external factors such as the frequency of communicating with the native speakers of the target language, spending time in the country of the target language, and earlier onset of language learning reduce the amount of L1 phonetic system's interference.

The comparison of two age groups clearly shows how exposure to native English influences pronunciation: older generation started their English studies at the later age and in a setting that offered less exposure to the English language. This, coupled with less frequent use of English, has made the older group more prone to the influence of L1 Estonian phonetic system. The younger generation, on the other hand, begun learning English at the earlier age in a setting which offers more exposure to the target language and natural communication, thus decreasing the influence of the L1 sound system on their L2 English. This study also highlights the most "problematic" L2 English sounds for L1 Estonian speakers, a knowledge which can be taken into account in teaching pronunciation. Yet the factors discussed in this paper are by no means exhaustive: there are many other factors that can possibly affect L2 pronunciation, such as aptitude and attitude towards language learning. Thus further studies should aim to include a wider range of factors in order to define more precisely the intersection between internal and external factors.

References

- Abrahamsson, Niclas; Hyltenstam, Kenneth 2009. Age of onset and nativelikeness in a second language: Listener perception versus linguistic scrutiny. – *Language Learning*, 59 (2), 249–306. <http://dx.doi.org/10.1111/j.1467-9922.2009.00507.x>
- Asu, Eva-Liina 1999. British or American English: A study of Estonian students' attitudes to pronunciation. – Pilvi Rajamäe (Ed.). *New Britain: The Heritage of the Past and the Challenge of the Future: Proceedings of the 2nd International Tartu Conference on British Studies Held at the University of Tartu, August 24–25, 1998*. Tartu: Tartu University Press, 22–30.
- Brown, H. Douglas 2007. *Teaching by Principles: An Interactive Approach to Language Pedagogy*. 3rd ed. White Plains, NY: Pearson Education.
- Derwing, Tracey M.; Munro, Murray J. 2005. Second language accent and pronunciation teaching: A research-based approach. – *TESOL Quarterly*, 39 (3), 379–399. <http://dx.doi.org/10.2307/3588486>
- Flege, James Emil 1991. Perception and production: The relevance of phonetic input to L2 phonological learning. – Thom Huebner, Charles A. Ferguson (Eds.). *Cross Currents*

- in Second Language Acquisition and Linguistic Theory. *Language Acquisition and Language Disorders* 2. Philadelphia: John Benjamins, 249–289. <http://dx.doi.org/10.1075/lald.2.15fle>
- Flege, James Emil; Frieda, Elaina M.; Nozava, Takeshi 1997. Amount of native-language (L1) use affects the pronunciation of an L2. – *Journal of Phonetics*, 25 (2), 169–186. http://jimflege.com/files/Flege_Frieda_L1_use_JP_97.pdf (18.2.2015). <http://dx.doi.org/10.1006/jpho.1996.0040>
- Flege, James Emil; Liu, Serena 2001. The effect of experience on adult's acquisition of a second language. – *Studies in Second Language Acquisition*, 23 (4), 527–552. http://jimflege.com/files/Flege_Liu_effect_experience_SSLA_2001.pdf (18.2.2015).
- Flege, James Emil; Birdsong, David; Bialystok, Ellen; Mack, Molly; Sung, Hyekyung; Tsukada, Kimiko 2006. Degree of foreign accent in English sentences produced by Korean children and adults. – *Journal of Phonetics*, 34 (2), 153–175. http://jimflege.com/files/Flege_Birdsong_foreign_accent_JP_20062.pdf (18.2.2015). <http://dx.doi.org/10.1016/j.wocn.2005.05.001>
- Granena, Gisela; Long, Michael H. 2013. Age of onset, length of residence, language aptitude, and ultimate L2 attainment in three linguistic domains. – *Second Language Research*, 29 (3), 311–343. <http://dx.doi.org/10.1177/0267658312461497>
- Guion, Susan G.; Flege, James Emil; Loftin, Jonathan D. 2000. The effect of L1 use on pronunciation in Quichua-Spanish bilinguals. – *Journal of Phonetics*, 28 (1), 27–42. http://jimflege.com/files/Guion_Flege_L1_use_JP_2000.pdf (18.2.2015). <http://dx.doi.org/10.1006/jpho.2000.0104>
- Hamid, M. Obaidul; Baldauf, Richard B. Jr. 2013. Second language errors and features of world Englishes. – *World Englishes*, 32 (4), 476–494. <http://dx.doi.org/10.1111/weng.12056>
- Kalk, Jana 2001. Typical Mistakes in the Pronunciation of English Sounds for the Estonian Learner. Diploma thesis. University of Tartu.
- Kostabi, Leili 1993. A Contrastive Analysis of English and Estonian Monophthongs: Teaching the Pronunciation of English Monophthongs to the Estonian Learner. MA thesis. University of Tartu.
- Kostabi, Leili 2004. Pronunciation Practice for Advanced Students. Tallinn: Koolibri.
- Kull, Rasmus 2006. Students' Pronunciation Preferences Based on an Example of the 1st Year Students of English. BA thesis. University of Tartu.
- Levis, John M. 2005. Changing contexts and shifting paradigms in pronunciation teaching. – *TESOL Quarterly*, 39 (3), 369–377. <http://dx.doi.org/10.2307/3588485>
- Marinova-Todd, Stefka H.; Marshall, Bradford D.; Snow, Catherine E. 2000. Three misconceptions about age and L2 learning. – *TESOL Quarterly*, 34 (1), 9–34. <http://www.jstor.org/stable/3588486>
- McAllister, Robert; Flege, James Emil; Piske, Thorsten 2002. The influence of L1 on the acquisition of Swedish quantity by native speakers of Spanish, English and Estonian. – *Journal of Phonetics*, 30 (2), 229–258. http://www.jimflege.com/files/McAllister_Flege_The_influence_of_JP_2002.pdf (18.2.2015). <http://dx.doi.org/10.1006/jpho.2002.0174>
- Moyer, Alene 2014. Exceptional outcomes in L2 phonology: The critical factors of learner engagement and self-regulation. – *Applied Linguistics*, 35 (4), 418–440. <http://dx.doi.org/10.1093/applin/amu012>
- Muñoz, Carmen; Llanes, Ángels 2014. Study abroad and changes in degree of foreign accent in children and adults. – *The Modern Language Journal*, 98 (1), 432–449. <http://dx.doi.org/10.1111/j.1540-4781.2014.12059.x>
- Mutt, Oleg 1971. An Introduction to English Phonetics for the Estonian Learner. 2nd ed. Tartu: Tartu State University.

- Piske, Thorsten; MacKay, Ian R. A.; Flege, James Emil 2001. Factors affecting degree of foreign accent in an L2: A review. – Journal of Phonetics, 29 (2), 191–215. [http://jimflege.com/files/Piske_MacKay_factors_affecting_JP_2001.pdf\(18.2.2015\)](http://jimflege.com/files/Piske_MacKay_factors_affecting_JP_2001.pdf(18.2.2015)). <http://dx.doi.org/10.1006/jpho.2001.0134>
- Roach, Peter 2000. English Phonetics and Phonology: A Practical Course. 3rd ed. Cambridge: Cambridge University Press.
- Schaetzel, Kirsten; Low, Ee Ling 2009. Teaching Pronunciation to Adult English Language Learners. – CAELA Network Briefs. <http://www.cal.org/caelanetwork/pdfs/TeachingPronunciationWeb.pdf> (18.2.2015).
- Scovel, Thomas 1988. A Time to Speak: A Psycholinguistic Inquiry into the Critical Period for Human Speech. Rowley, MA: Newbury House.

Kristiina Ader, a graduate of Tallinn University, the department of English Studies.
Tallinn University, Narva mnt 25, 10120 Tallinn, Estonia
kristiina.ader@gmail.com

Merilin Miljan (Tallinn University, University of Tartu), lecturer in linguistics at the department of English Studies at Tallinn University, a researcher at the University of Tartu. Her research interests focus on morphosyntax in linguistic theories and psycholinguistics.
Tallinn University, Narva mnt 25, 10120 Tallinn, Estonia
merilin.miljan@ut.ee

VÄLISED TEGURID KEELEOMANDAMISEL JA EESTI KEELE KUI EMAKEELE INTERFERENTS INGLISE KEELE KUI TEISE KEELE HÄÄLDUSES

Kristiina Ader¹, Merilin Miljan^{1,2}

Tallinna Ülikool¹, Tartu Ülikool²

Artikkel keskendub teise keele hääldust mõjutavatele välistele teguritele. Eesmärgiks oli uurida, kuidas eesti keel emakeelena avaldab negatiivset mõju inglise keele kui teise keele hääldusele. Vaatluse all olid sellised varasemates uuringutes välja toodud tegurid nagu sihtkeelega kokkupuute kvaliteet ja kvantiteet ning nende koosmõju keeleõppega alustamise vanusega. Artiklis võrreldi eesti keelt emakeelena rääkivate vanemaealiste täiskasvanute inglise keele hääldust noorte täiskasvanute omaga. Saadud tulemusi võrreldi omakorda eelpool mainitud tegurite vaatepunktist, et rohkem teada saada muutunud keeleõppekeskkonna mõjust inglise keele kui teise keele hääldusele.

Uuringu tulemused tõestasid, et osalejatel, kes suhtlevad tihedamini inglise keelt emakeelena kõnelejatega või on pikemalt viibinud ingliskeelses keskkonnas, on sihtkeelelähedasem hääldus. Tugev seos leiti ka sihtkeelelähedasema häälduse ja nooremas eas inglise keele õpingutega alustamise vahel: noorema põlvkonna esindajad, kes alustasid inglise keele õpingutega enamasti algkoolis, suutsid rohkem sihtkeele häälikuid sihtkeelelähedaselt hääldada võrreldes vanema põlvkonnaga. Üks peamisi põhjuseid, miks vanem põlvkond oli rohkem avatud eesti keele interferentsile, seisneb keeleõppekeskkonnas, kus oli vähem võimalusi loomulikuks suhtluseks (ning seega keeleomandamiseks vastandina formaalsele keeleõppele), ning jätkuvas väiksemas kontaktis sihtkeelega.

Tulemustest lähtuvalt võib väita, et keeleõppes ei saa eraldada sisemisi ja väliseid tegureid, vaid neid tuleks vaadelda koos: varajasem keeleõppe algus (sisemine tegur) on küll eeliseks uute häälikute kategooriate moodustamisel, aga välised tegurid, nagu tihe kokkupuude sihtkeelega, vähendavad sisemiste tegurite, nagu emakeele interferents, mõju.

Töö käigus võrreldi osalejate hääldust ka varasemate eesti keele hääldust inglise keeles käsitlevate tööde tulemustega ning leiti, et eesti keele aktsendi omadused on jäänud samaks. Lisaks mõõtis uuring eesti keele kui emakeele interferentsi inglise keele häälduses, tuues välja empiirilised ja kvantitatiivsed andmed eesti keele mõju kohta.

Võtmesõnad: teise keele omandamine, teise keele fonoloogia, sisemised ja välised tegurid, inglise keel, eesti keel

STATISTILISED MEETODID ÜHENDVERBIDE TUVASTAMISEL TEKSTIKORPUSEST

Eleri Aedmaa

Ülevaade. Artiklis võrdlen sõnadevahelise seose tugevuse mõõtmise statistilisi meetodeid, mida kasutatakse arvutilingvistikas püsiühendite tuvastamiseks. Töö põhieesmärk on rakendada viit sümmeetrilist statistikut – t-skoori, vastastikuse informatsiooni väärtust, hii-ruut-statistikut, log-tõepära funktsiooni ja minimaalset tundlikkust – erineva suurusega korpuste peal ja välja selgitada, milline meetod töötab eesti keele ühendverbide automaatsel tuvastamisel kõige paremini. Teine suurem eesmärk on katsetulemuste põhjal uurida, milline on korpuse suuruse mõju statistikute tööle. Lisaks palju testitud nimetatud sümmeetrilistele statistikutele rakendan psühholoogiliselt paremini põhjendatud asümmeetrilisi statistikuks – tinglikku tõenäosust ja ΔP -d – ning toon välja nende eelised sümmeetriliste statistikute ees.*

Võtmesõnad: arvutilingvistika, korpuslingvistika, püsiühendid, statistika, ühendverbid, eesti keel

1. Sissejuhatus

Lingvistikas ei klassifitseerita sõnu mitte ainult nende tähenduste põhjal, vaid ka sellel alusel, milliste teiste sõnadega need koos esinevad (Church, Hanks 1990: 22). Sõnade sagedase koosesinemise põhjuseks võib olla nende endi suur sagedus tekstis, näiteks *see on, ja ka* jne, aga ka see, et nad moodustavad püsiva tavapärase sõnade ühendi keeles ehk püsiühendi (Kaalep, Muischnek 2009: 157–163). Arvutilingvistikas on *püsiühendi* mõiste kõrval rohkem kasutusel *kollokatsiooni* mõiste (Evert 2008: 3), mida näiteks Kaalep ja Muischnek (2002: 173) defineerivad kui sõnaühendit, millesse kuuluvad sõnad esinevad tekstis koos sagedamini, kui võiks eeldada nende eraldiesinemise sageduse põhjal. Sinclairi (1991: 71) definitsiooni järgi on kollokatsioon kombinatsioon kahest sõnast, mis näitavad tendentsi esineda koos loomuliku keele tekstides.

Keele automaattöötlemisel on kollokatsioonid problemaatiline nähtus, sest süntaktilise ja semantilise analüüsi jaoks on oluline mitmesõnalise üksuse või

* Artikkel põhineb autori magistritööl (Aedmaa 2014a).

minimaalse semantilise üksuse äratundmine, mistõttu ei saa analüüsi aluseks võtta tühikutevahelist stringi, vaid kasutada tuleb teistsuguseid meetodeid (Kaalap, Muischnek 2009: 158). Üks kollokatsioonide tuvastamise võimalus on rakendada statistikuid ehk sõnadevahelise seose tugevuse mõõdikuid (ingl *association measures*), mis on statistilised valemid sõnadevahelise seose statistilise tugevuse arvutamiseks (Evert 2008: 5). Artiklis kasutatakse mõisteid *statistik* ja *mõõdik* sünonüümselt. Statistikud on osutunud tulemuslikuks arvukates kontekstides: näiteks lähisünonüümide eristamisel leksikograafias või leksikaal-semantilistes uurimustes, andmekaeves ja masintõlkega seotud ülesannetes (Wiechmann 2008: 257).

Viimase 50 aasta jooksul on enim testitud sümmeetrilisi statistikud (Gries 2013: 4), mis arvutavad igale korpusest leitud sõnapaarile ühe seose tugevuse väärtuse (ingl *association score*), mis näitab kahe sõna vahelise statistilise seose suurust (Evert 2008: 5). Lisaks sümmeetrilistele mõõdikutele on viimastel aastatel kollokatsioonide tuvastamisel arvestatud ka sõnadevahelise seose asümmeetrilisusega, mille tuvastamiseks rakendatakse asümmeetrilisi sõnadevahelise seose tugevuse mõõdikuid. Nende abil on võimalik arvutada igale sõnapaarile kaks seose tugevuse väärtust, mis osutavad, milline sõna kollokaadis on rohkem mõjutatud teise sõna esinemisest ehk näitavad kollokatiivsuse suunda. (Gries 2013: 5–13)

Sümmeetrilisi sõnadevahelise seose tugevuse mõõdikuid on eesti keele ühendverbide peal varem rakendanud Uiboed (2010), kes tuvastas Eesti murrete korpuse kaheliikmelisi ühendverbe ning katsetas kolme murderühma peal eraldi nelja statistikut. Uurimus kinnitas, et ühtegi mõõdikut ei saa pidada teistest ühemõtteliselt paremaks ning erinevat tüüpi statistikud sobivad erinevat tüüpi ülesannete lahendamiseks. Murdematerjali peal töötas küll kõige paremini log-tõepära funktsioon, ent autor ei soovita valida ühte kindlat mõõdikut kogu materjali jaoks. Lisaks loodi 2010. aastal riikliku programmi projekti “Eesti keele koondkorpuse esituse ja kasutusvõimaluste arendamine” raames automaatne veebis kasutatav kollokatsioonide leidja Tasakaalus korpusest¹, kus on rakendatud kolme statistikut. Sümmeetrilisi mõõdikuid on rakendatud ka teistsuguste ülesannete lahendamiseks, näiteks Jelena Kallas (2013) katsetas neid eesti keele sisusõnade süntagmaatiliste suhete tuvastamiseks.

Oma töös rakendan nii sümmeetrilisi kui ka asümmeetrilisi mõõdikuid, et tuvastada püsiühendite hulka kuuluvaid kahest komponendist – afiksaaladverbist ja verbist – koosnevaid ühendverbe. Artikkel on üles ehitatud järgnevalt. Esmalt teen ülevaate sõnadevahelisest seosest ja selle tugevuse mõõtmisest, millele järgneb materjali ja selle töötlemist ning tulemuste hindamist kirjeldav peatükk. Seejärel tutvustan rakendatud meetodeid ning viimaks analüüsin ja võrdlen sümmeetriliste ja asümmeetriliste statistikute töö tulemusi.

2. Sõnadevaheline seos ja selle tugevuse mõõtmine

Ainult sõnade kordumine ei ole piisav alus, et sõnadevahelist seost tugevaks pidada. Sõnadevahelise seose tugevuse statistikute rakendamine on vajalik, sest need aitavad määrata, kas tegemist on n-ö õige püsiühendiga ja kas sõnade vahel on tugev või nõrk seos. (Evert 2008: 5) Koosesinevate sõnade vahelise seose tuvastamiseks

saab rakendada matemaatilisi seose tugevuse mõõdikuid, mis annavad iga sõna-paari jaoks seose tugevuse skoori. Seose tugevuse skoori abil saab välja valida n-õ õiged püsiühendid rakendades sageduse lävendit või järjestades sõnapaarid seose tugevuse väärtuse järgi kahanevalt, mille tulemusena leiab n-õ õiged püsiühendid nimekirja eesotsast. (Pecina, Schlesinger 2006: 652)

Selleks et sõnadevahelise seose tugevuse mõõtmise statistikuid rakendada, tuleb defineerida, mida tähendab sõnade koosinemine. Ka peab otsustama, kas otsitakse vaid n-õ tõelisi kollokatsioone või tahetakse saada ülevaadet kõikidest sõnapaaridest, asetades need sõnadevahelise tugevuse järgi mingile skaalale, eristamata kollokatsioone ja mitte-kollokatsioone. Esimese lähenemise korral peab uurija ise määrama seose tugevuse piirväärtuse, millest ülespoole jäävad sõnapaarid on n-õ tõelised püsiühendid. Teine otsus puudutab kollokatsioonide grupeerimist: kas otsitakse kõige tugevamini seotud sõnade paare või huvitatakse mingi sõna kindlast kontekstist ehk uuritakse missuguste sõnadega vaadeldav sõna kõige rohkem koos esineb. Kaks otsust on üksteisest sõltumatud, kuid sõna konteksti otsimine kombineeritakse tihti sõnapaaride mingisugusele skaalale asetamisega. Kui on tehtud vajalikud otsused, siis saab erinevate sõnadevahelise seose tugevuse mõõtmise meetodite abil leida sõnadevahelise seose tugevuse väärtused: suuremad väärtused osutavad sõnadevahelisele tugevamale seosele, nõrgemad sellele, et sõnapaari kuuluvad sõnad pigem väldivad koosinemist. (Evert 2008: 6)

Kõige lihtsam meetod kollokatsioonide tuvastamiseks tekstikorpusest on nende kokkulugemine ehk leida sõnade koosinemise arv, kuid see pole sõnadevahelise seose tugevuse väärtusena piisav (Manning, Schütze 1999: 153–157). Näiteks kui moodustada korpuse sõnapaaride sagedusloend, siis ilmselt oleks selles üsna kõrgel kohal sõnapaar *ja ei*, mis tegelikult ei ole püsiühend. Kuna mõlemad sõnad on korpuses sagedased, siis on ka nende koosinemine sage. Järelikult tuleb kasutada keerulisemaid meetodeid ja lisaks sõnapaari sagedusele arvesse võtta mõlema sõnapaari kuuluva sõna sagedused ehk marginaal- ehk ääresagedused, arvestada tuleb ka valimimahtu ehk korpuse suurust, kust püsiühendid leitakse (Evert 2008: 17). Nii on sõnadevahelise seose tugevuse mõõtmiseks kasutatavate statistikute jaoks vajalik andmestik järgmine: O on sõnade koosinemise sagedus valimis, f_1 ja f_2 on vastavalt sõnapaari kuuluva esimese ja teise sõna marginaalsagedus ja N on valimimaht (Evert 2004: 36). Lisaks nõuavad statistilised meetodid ka teoreetilist sagedust (ingl *expected frequency*) E , mis osutab sõnade koosinemise teoreetilisele tõenäosusele. Selle arvutamiseks tuleb sõnade marginaalsageduste korrutis jagada valimi suurusega: $E = f_1 * f_2 / N$. (Evert 2008: 18)

Suur hulk statistilisi meetodeid kasutab kahemõõtmelist sagedustabelit (ingl *contingency table*), mis arvestab marginaalsagedustega. Tabel 1 esitab näite kahemõõtmelisest sagedustabelist koos teheteiga, mis illustreerivad vajalike väärtuste leidmist.

Tabel 1. Kahemõõtmeline sagedustabel (Evert 2008)

O11	O	f1 – O	O12
O21	f2 – O	N – f1 – f2 + O	O22

3. Materjal ja tulemuste hindamine

3.1. Materjal

Selle töö uurimismaterjal on Eesti keele koondkorpuse² ajakirjandustekstid – kokku 170 miljonit sõna. Materjal oli eelnevalt programmiga t3mesta (Kaalep 1998, Kaalep, Vaino 1998) morfoloogiliselt analüüsitud, ühestatud ning (osa)lausestatud (Kaalep, Muischnek 2012).

Püsiühendite tekstist tuvastamist alustasin ühendikandidaatide (kõikvõimalike potentsiaalsete ühendverbide) moodustamisega. Ühendverbide kandidaatpaaride moodustamisel lähtusin tekstuaalsest koosinemisest, mille korral moodustavad kaks sõna potentsiaalse püsiühendi, kui nad esinevad ühes ja samas tekstiüksuses, tüüpiliselt lauses või lausungis (Evert 2008: 13), minu töö puhul osalauses, sest ühendverbi moodustavad sõnad saavad esineda vaid samas osalauses. Teksti sõnajärge kandidaatpaaride moodustamisel ma ei arvestanud ning kuna eesmärk oli tuvastada ühendverbe, siis iga osalause sees genereeriti adverbi ja verbi kõikvõimalikud kombinatsioonid, mis moodustavad kandidaatpaaride loetelu. Selle nimekirja peal rakendasin stopp-sõnade loendit ehk eemaldasid ühendid, milles esinev adverb reeglina ühendverbi koosseisus ei esine (nt *ikka, jälle*). Stopp-sõnade loend põhineb “Eesti keele seletava sõnaraamatu” (EKSS) ühendverbide loendil³, mis ühtlasi on selles töös ka kuldstandardiks ehk õigete ühendverbide loeteluks, millega saadud tulemusi võrdlesin. EKSS-i ühendverbide loendist eemaldasid ühendverbid, mis ei ole selles töös kasutatud korpuse märgendamise seisukohalt võimalikud. Sellisteks ühenditeks on sõnavormiga *kätte, minema* ja *tulema* moodustatud ühendverbid, nt *kätte jõudma, kätte maksuma, minema kihutama/minema/viskama* ja *tulema tulema*. *Kätte, minema* ja *tulema* ei saa selles töös kasutatud korpuse märgenduses mitte kunagi adverbi märgendit ja seega on nendega võimatu moodustada kasulikke kandidaatpaare. N-ö müra vähendamiseks eemaldasid 1757 ühendverbi sisaldavast algsest EKSS-i loendist 20 ühendit ning nii sisaldab selles töös kasutatud loend 1737 ühendverbi. Kandidaatpaaride loendi tekitamisel võtsin arvesse vaid EKSS-i ühendite adverbilise komponendi: kui adverbi EKSS-i nimistus ei esinenud, viskasin adverbi sisaldava ühendi kandidaatpaaride loetelust välja.

3.2. Täpsus ja saagis

Seda, kui tõhus on valitud ühendverbide leidmise meetod, saab hinnata täpsuse ja saagise arvutamiseega. Täpsus kirjeldab leitud õigete ühendverbide suhet kõigi leitud ühendite hulgaga ja näitab, kui suur osa leitud ühenditest on õiged ühendverbid. Täpsus jääb 0% ja 100% vahele: 0% tähendab, et ükski leitud ühenditest pole õige ehk EKSS-i ühendverb, 100% näitab, et kõik leitud ühendid on EKSS-i ühendverbid. Saagis väljendab kõigi meetodiga tuvastatud õigete ühendverbide suhet kõigi võimalike õigete ühendverbidega (siin töös EKSS-i ühendverbidega) ning kirjeldab, kui suurt osa õigetest ühendverbidest õnnestus meetodiga andmestikust leida. Saagis jääb samuti 0% ja 100% vahele ning 0% tähendab, et ei leitud ühtegi õiget ehk EKSS-i ühendverbi, 100% aga seda, et on leitud kõik EKSS-is olevad ühendverbid.

² Eesti keele koondkorpus. <http://www.cl.ut.ee/korpused/segakorpus/index.php> (16.12.2014).

³ EKSS-i ühendverbide loetelu ja artiklis esitatud tulemuste põhjalikumad ülevaated on kättesaadavad aadressil kodu.ut.ee/~elieraed/magistrit88_failid/ (16.12.2014).

Tabel 2 esitab siinse töö materjali andmed: materjal koosneb 170 miljonist sõnast ja 25 912 251 osalausest, materjalist genereeriti 67 558 kandidaatpaari ning õigeid ühendverbe on nende seas 1676. Kokku oli õigeid ühendverbe võimalik tuvastada 1737.

Tabel 2. Ülevaade töös kasutatud materjalist

sõnu	170 000 000
osalauseid	24 322 394
kandidaatpaare	67 558
tuvastatud õigeid ühendverbe	1676
õigeid ühendverbe kuldstandardis	1737
täpsus	$(1676/67558)*100 \approx 2,5\%$
saagis	$(1676/1737)*100 \approx 96,5\%$

Sageduse täpsus kogu materjalist ühendverbide tuvastamisel näitab, et 2,5% leitud ühenditest on EKSS-i ühendverbid. Saagise väärtus väljendab, et tuvastati 96,5% kõigist EKSS-i ühendverbidest ehk tuvastamata jäi 61 EKSS-i nimistusse kuuluvat ühendverbi. Sagedusloendi kasutamine tagab suure hulga õigete ühendverbide leidmise, kuid tuvastatud ühendite koguhulgast moodustavad õiged ühendverbid väikese osa.

3.3. Täpsuse kõverad

Sümmeetriliste statistikute tulemuslikkuse hindamiseks on Krenn ja Evert (2001) kasutanud täpsuse kõveraid (ingl *precision curves*). Selleks arvutatakse välja valitud mõõdikutega iga sõnapaari jaoks statistilise seose tugevuse väärtused ning saadud väärtuste järgi reastatakse sõnapaarid kahanevalt ümber. Uue loendi põhjal loetakse koosinevaks vaid teatav hulk ühendeid sagedusloendi esimesest osast (selle väärtuse, millest alates ühendeid enam koosinevaks ei loeta, peab uurija määrama ise). Statistikute väärtuste põhjal tehtud pingeridu kõrvutatakse õigete kollokatsioonide loendiga ning igale võimalikule kandidaatpaaride arvule leitakse statistiku põhjal arvatud tulemuste järgi selle täpsus. (Krenn, Evert 2001: 39–41)

Näiteks võetakse 500 esimest sõnapaari, mis on saanud t-skoor statistiku suuremad väärtused. Seejärel arvutatakse välja täpsus ehk mitu protsenti nendest 500 ühendist kuulub õigete kollokatsioonide loendisse (siinses töös EKSS-i ühendverbide loetellu). Sama saab teha kõikide valitud meetoditega ning erineva arvu kandidaatpaaridega. Selleks et mõõta uuritavate meetodite tulemuslikkust püstitatud ülesande lahendamisel, arvutatakse välja algtaseme täpsus (ingl *baseline precision*) ehk teoreetiline täpsus, mis näitab, kui suur osa kõikidest leitud ühenditest on õiged ühendverbid. Algtaseme täpsus arvutatakse korpusest leitud õigete ühendverbide arvu jagamisel genereeritud kandidaatpaaride arvuga. Seejärel kantakse saadud tulemused täpsuse kõveratena joonisele, kus iga kõver märgib erineva mõõdiku täpsust.

Kui statistiku täpsuse kõver asub valdavalt teoreetilist täpsust märkivast joo-
nest kõrgemal, siis võib mõõdikut pidada ülesande lahendamisel tulemuslikuks ja

statistikut tasub kasutada ülesande edukaks sooritamiseks. Kui statistiku täpsuse kõver asub valdavalt allpool teoreetilist täpsust märkivast joonest, siis ei ole mõõdiku rakendamine ülesande lahendamiseks mõttekas. (Krenn, Evert 2001: 39–41)

Lisaks algataseme täpsusele võib statistikuid võrrelda ka teiste näitajatega. Kuna mitmed tööd (Krenn, Evert 2001, Wermter, Hahn 2006) on tõestanud ka koosesinemise sageduse tõhusust püsiühendite tuvastamise katsetes, siis on siinses töös uuritud, millised on koosesinemise sageduse tulemused võrreldes teiste meetoditega.

4. Sõnadevahelise seose tugevuse mõõtmise meetodid

Sõnadevahelise seose tugevuse mõõdikuid on palju ja nende rakendusala lingvistikas erinevad, kuid kõige rohkem on neid kasutatud kollokatiiivsust käsitlevates uurimustes (Wiechmann 2008: 254–257). Eelnevad tööd on tõestanud, et ühe mõõdiku eelistamine teisele on keeruline ja ei saa üheselt öelda, et ühte gruppi kuuluvad statistikud on paremad kui teised (Evert 2008: 32).

4.1. Sümmetriselised mõõdikud

Sümmetriselised sõnadevahelise seose tugevuse mõõtmise statistikud võib jagada arvutuslike ja matemaatiliste põhimõtete alusel väiksematesse gruppidesse. Näiteks Evert (2008) on jaganud statistikud lihtsateks (nt t-skoor, vastastikuse informatsiooni väärtus) ja statistilisteks (nt hii-ruut-statistik, log-tõepära funktsioon). Lihtsad statistikud mõõdavad sõnadevahelist seost sõnaühendi sageduse võrdlemisel teoreetilise sagedusega, statistilised mõõdikud põhinevad kahemõõtmelistel sagedustabelitel.

Selles töös olen sümmetriselistest statistikutest uurimiseks välja valinud kõige sagedamini sarnaste ülesannete lahendamiseks kasutatavad mõõdikud.

1. **t-skoor** (ingl *t-score*) (Church, Hanks 1990), mille arvutamisel lahutatakse sagedusest teoreetiline sagedus ning see tulemus jagatakse koosesinemise sageduse ruutjuurega:

$$\text{t-skoor: } \frac{O - E}{\sqrt{O}}$$

t-skoor on osutunud kasulikuks näiteks saksa keele prepositsioonifraasi ja verbiühendite tuvastamisel (Krenn, Evert 2001).

2. **Vastastikuse informatsiooni väärtuse** (ingl *mutual information*, MI) (Church, Hanks 1990) arvutamiseks jagatakse osalausete hulk, kus mõlemad sõna-paari liikmed esinevad, teoreetilise sagedusega, ning sellest võetakse omakorda logaritm alusel kaks:

$$\text{MI} = \log_2 \frac{O}{E}$$

MI väärtus on suurem, kui O on palju suurem kui E, ning see tõstab väärtuste loendis kõrgemale harvaesinevaid ühendeid, mille komponendid on samuti väikese esinemissagedusega (Evert 2008: 19).

3. Hii-ruut-statistiku (ingl *chi-squared measure*) (Manning, Schütze 1999) väärtus arvutatakse kahemõõtmelise sagedustabeli väärtuste abil:

$$\text{hii-ruut} = \sum_{ij} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

Hii-ruut-statistik on kahepoolne mõõdik, mis tähendab, et suure positiivse väärtuse saavad nii tugeva kui ka nõrga seosega sõnaühendid ehk suure statistiku väärtuse saavad nii sõnapaarid, mis kindlasti moodustavad püsiühendi, kui ka paarid, mis väldivad koosinemist. Selleks et kahepoolsest mõõdikust saaks n-õ ühepoolne mõõdik, mis eristab nii positiivseid kui ka negatiivseid seoseid, korrutatakse tulemus läbi (-1)-ga kui $O < E$. (Evert 2008: 21)

4. Log-tõepära funktsioon (ingl *log-likelihood measure*) (Dunning 1993) on kõige laialdasemat kasutust leidnud statistik arvutilingvistikas (Evert 2008: 31). Sarnaselt hii-ruut-statistikule arvutatakse ka log-tõepära funktsiooni väärtused kahemõõtmelise sagedustabeli väärtuste põhjal:

$$\text{log-tõepära} = 2 \sum_{ij} O_{ij} \log \frac{O_{ij}}{E_{ij}}$$

Ka log-tõepära funktsioon on kahepoolne mõõdik (Evert 2008: 22) ning vajab ühepoolseks teisendamist (vt eelmine punkt).

5. Minimaalne tundlikkus (ingl *minimum sensitivity*, MS) (Pedersen 1998) on lihtne ja efektiivne mõõdik bigrammide tuvastamiseks. MS arvutatakse sõna bigrammis esinemise sageduse ja sama sõna üldise esinemissageduse võrdlemisel valemi põhjal:

$$\text{MS} = \text{miinimum} \left(\frac{O_{11}}{O_{11} + O_{12}}, \frac{O_{11}}{O_{11} + O_{21}} \right)$$

Esimene väärtus tähistab esimese sõna tundlikkust teise suhtes ehk juhul kui teine sõna esineb, siis kui suure tõenäosusega esineb esimene sõna. Teine väärtus tähistab vastupidist ehk kui tundlik on teine sõna esimese suhtes ehk kui suure tõenäosusega esineb teine sõna tingimusel, et esineb esimene sõna. Kui kahe väärtuse miinimum on 0, siis kaks vaadeldavat sõna ei esine kunagi koos, kui miinimum on 1, siis on sõnade vahel tugev seos ja iga kord kui üks sõna esineb, esineb ka teine. (Pedersen, Bruce 1996: 12)

4.2. Asümmeetrilised mõõdikud

Eelnevalt kirjeldatud statistikud kajastavad kahe sõna vastastikust seost ja selline lähenemine on korpuslingvistikas kollokatsioonide tuvastamise uurimustes domineerinud viimased viiskümmend aastat (Gries 2013: 4). Samas on teada, et inimese teadvuses ei ole seos kahe sõna vahel alati sümmeetriline (Michelbacher jt 2007: 1), ning sümmeetrilised mõõdikud ei tuvasta, kas esimene sõna on abiks teise

püsiühendi komponendi ennustamisel või vastupidi (Gries 2013: 4). Näiteks ühendi *lahku minema* korral, kui osaluses on *lahku*, siis on seal väga suure tõenäosusega ka *minema* ehk seos *lahku* ja *minema* vahel on tugev ja *minema* esinemine on ennustatav *lahku* esinemise järgi. Kuid vastupidi see ei toimi: kui osaluses esineb *minema*, siis *lahku* seal sama suure tõenäosusega ei esine ehk seos *minema* ja *lahku* vahel nii tugev ei ole ning *lahku* esinemine ei ole hästi ennustatav *minema* esinemise järgi. Erinevalt sümmeetrilistest statistikutest ei ühenda asümmeetrilised statistikud kahte väga erinevat tõenäosust, vaid arvutavad kaks väärtust: $p(sõna_1|sõna_2)$ ja $p(sõna_2|sõna_1)$ (Gries 2013: 4). Asümmeetrilistest statistikutest testin selles töös tinglikku tõenäosust ja ΔP -d.

1. Tinglik tõenäosus (ingl *conditional probability*) (Bell jt 2009, Michelbacher jt 2007) on MS-ist tuletatud sõnadevahelise seose tugevuse mõõdik. Tinglik tõenäosus arvutatakse kahe valemi abil:

$$p(sõna_2|sõna_1) = \frac{O_{11}}{O_{11} + O_{12}}$$

$$p(sõna_1|sõna_2) = \frac{O_{11}}{O_{11} + O_{21}}$$

Tinglikku tõenäosust on seose tugevuse mõõdikuna rakendatud vähestes töödes (Gries 2013: 4). Erandiks on Michelbacheri jt (2007, 2011) uurimused, mille tulemusena ilmneb, et tinglik tõenäosus on sobiv asümmeetriliste seoste tuvastamiseks, kuid sümmeetriliste seoste tuvastamisel on statistiku tulemuslikkus madal.

2. ΔP (Ellis 2006, Ellis, Ferreira-Junior 2009) väärtus arvutatakse kahe tõenäosuse põhjal. Esimese tõenäosuse arvutamisel arvestatakse ennustava sõna sagedust. See tähendab, et võetakse arvesse selle sõna sagedus, mille esinemise abil teise sõna esinemist ennustatakse. Teise tõenäosuse arvutamisel ennustava sõna sagedust ei arvestata. ΔP väärtuse leidmiseks lahutatakse esimesest tõenäosusest teine. (Ellis 2006: 11) Kahemõõtmelise sagedustabeli abil arvutatakse ΔP väärtused järgmiselt:

$$\Delta P_{2|1} = p(sõna_2|sõna_1=esineb) - p(sõna_2|sõna_1=puudub) = \frac{O_{11}}{O_{11} + O_{12}} - \frac{O_{21}}{O_{21} + O_{22}}$$

$$\Delta P_{1|2} = p(sõna_1|sõna_2=esineb) - p(sõna_1|sõna_2=puudub) = \frac{O_{11}}{O_{11} + O_{21}} - \frac{O_{12}}{O_{12} + O_{22}}$$

Kui kahe sündmuse tõenäosused on võrdsed, siis nende sündmuste vahel seos puudub ja $\Delta P = 0$. Kui $\Delta P = 1$, siis teise sõna esinemine suurendab vaadeldava sõna esinemise tõenäosust ja kui $\Delta P = -1$, siis teise sõna esinemine vähendab vaadeldava sõna esinemise tõenäosust ja tegemist on negatiivse seosega. (Ellis 2006: 11) Gries (2013: 6–13) toob välja ΔP eelised võrreldes sümmeetriliste statistikutega: ΔP on traditsiooniliste mõõdikutega kõrvutades tundlikum, sest vastupidiselt nendele näitab ΔP , missugune sõna kollokatsioonis väljendab tugevamat või nõrgemat seost teiste sõnadega kollokatsioonis; ΔP on leidnud kasutust psühholoogilistes uurimustes ja osutunud mõõdikuks, mis iseloomustab paremini inimese kognitiivseid võimeid.

5. Meetodite täpsused ja saagised erineva suurusega korpuste põhjal

Selleks et hinnata valimi suuruse mõju mõõdikute tulemustele, võtsin 170 miljoni sõna suurusest korpusest viis valimit: esimene koosnes 5 miljonist sõnast, teised vastavalt 10, 20, 70 ja 170 miljonist sõnast.

5.1. Sümmeetriliste statistikute ja koosinemise sageduse tulemused

Tabel 3 esitab ülevaate koosinemise sageduse tulemustest erineva suurusega korpustest ühendverbide tuvastamisel.

Tabel 3. Sageduse täpsus ja saagis erineva suurusega ajakirjanduskorpustes

Sõnu	Osalauseid	Kandidaatpaare	Õigeid ühendverbe	Täpsus	Saagis
5 mln	707 979	13 141	1351	10,3%	77,8%
10 mln	1 410 474	18 545	1459	7,9%	84,0%
20 mln	2 823 255	26 268	1532	5,8%	88,2%
70 mln	9 640 426	46 863	1628	3,5%	93,7%
170 mln	24 322 394	67 558	1676	2,5%	96,5%

Selgub, et näiteks 20 miljonist sõnast koosneva ajakirjandustekstide valimist genereeritud kandidaatpaaride arv on 26 268 ning tuvastatud õigete ühendverbide arv on 1532, mis on 88,2% kõikidest õigetest ühendverbidest. Kõikidest genereeritud kandidaatpaaridest on sellisel juhul 5,8% EKSS-i kuuluvad ühendverbid. Mida suurem on korpus, seda rohkem kandidaatpaare genereeriti ja õigeid ühendverbe tuvastati. Seega on suurima korpuse puhul saagis kõige suurem (96,5%) ja täpsus kõige väiksem (2,5%).

Tabelis 4 on esitatud iga statistiku 25 suurima väärtuse saanud ühendit ning seal on näha, et MI kõrgemad väärtused on saanud harva esinevad ühendid, millesse kuuluv verb esineb samuti harva. Seetõttu on MI hea harvaesinevate (õigete) ühendverbide tuvastamiseks. Koosinemise sageduse eripärana saab välja tuua asjaolu, et kõrge väärtusega on *olema*-verbi sisaldavad ühendid, sest *olema* on korpuses kõige sagedasem verb. Ka selgub, et suured statistikute väärtused on saanud ühendid, mis EKSS-i ühendverbide loendisse ei kuulu, kuid millesse kuuluvate sõnade vaheline seos on tugev ehk neid võib pidada ühendverbideks, näiteks *kaasa võtma*, *kallale tungima*, *eemale peletama*, *edasi lükkuma*, *esile kerkima*. Kokku oli neid ühendeid umbes 30, kuid täpne arv sõltub sellest, millist seose tugevuse väärtust pidada tugevaks. Selliste ühendite esinemine kinnitab, et püsiühendite tuvastamine tekstikorpusest on kasulik leksikograafide töös.

Tabel 4. Statistikutete 25 suurima väärtusega ühendit ajakirjandustekstides

t-skoor	MI	hii-ruut	log-tõepära	MS	sagedus
vastu võtma	sekka jõratama	kallale tungima	vastu võtma	kallale tungima	välja olema
ette nägema	sekka mahakooruma	eemale peletama	ette nägema	ette nägema	üle olema
välja tulema	sekka taas-taas-taasavaldama	ette kujutama	kaasa tooma	eemale peletama	kokku olema
kaasa tooma	ühte kanseldama	kokku leppima	kokku leppima	ümber lükkama	ära olema
kinni pidama	järel mitte-kunagi-armastama	vastu võtma	ette kujutama	kaasa aitama	vastu võtma
välja kuulutama	tagant turgima	ette nägema	välja kuulutama	maha müüma	välja tulema
kokku leppima	kallale õppimatma	ette valmistama	alla kirjutama	kaasa tooma	ette nägema
läbi viima	kallale trompama	kaasa tooma	ette valmistama	läbi viima	ette olema
alla kirjutama	valla portesteerima	esile tõstma	läbi viima	ette kujutama	läbi olema
välja andma	sekka needistama	alt vedama	kaasa aitama	alla kirjutama	valmis olema
ette kujutama	sekka tšekkama	edasi lükkama	kinni pidama	vastu võtma	välja andma
ette võtma	pärale muganema	alla kirjutama	maha müüma	ette valmistama	kinni pidama
ette valmistama	alt delegeeruma	edasi lükkuma	edasi lükkama	üles kutsuma	kaasa tooma
tagasi tulema	alt tsirukleerima	välja kuulutama	välja tulema	esile tõstma	tagasi olema
alla jääma	üleväl walitsema	kokku pörkama	vahela jääma	edasi lükkama	kokku saama
üle andma	taha praeguma	ümber lükkama	kokku puutuma	kokku leppima	välja kuulutama
kaasa aitama	tasa posisema	külge pookima	ilma jääma	esile kerkima	vastu olema
maha müüma	taga garaaxima	läbi viima	ette võtma	üles astuma	ette võtma
välja töötama	taga pusletama	kokku puutuma	alla jääma	üles ehitama	tagasi tulema
kaasa võtma	taga nõduma	kaasa aitama	kinni maksuma	kinni hoidma	läbi viima
ära kasutama	alt näpsima	maha müüma	esile tõstma	maha laskma	ära tegema
maha võtma	eemale seminaritsema	kõrvale hiilima	tagasi lükkama	alt vedama	ligi olema
kinni maksuma	eemale otsas/muigama	ette heitma	ette heitma	alla kukkuma	kokku leppima
ilma jääma	eemale jääminema	taga ajama	kokku pörkama	kaasa lööma	alla kirjutama
välja tooma	eemale hoidsima	taga kiusama	üles kutsuma	tagasi astuma	üles olema

Ülevaate korpuse mahu suurenemise mõjust mõõdikute tulemustele ja paremusjärjestusele saab tabelist 5, kus on esitatud sümmeetriliste statistikute ja koosinemise sageduse täpsused ja saagised ühendverbide tuvastamisel erineva suurusega korpustest 100, 1000 ja 2000 statistikute suurima väärtuse saanud sõnapaari seas.

Tabel 5. Sümmeetriliste statistikute ja koosinemise sageduse täpsused ja saagised erineva suurusega ajakirjanduskorpustes

stat*	Hindamise-meetod	5 000 000			20 000 000			170 000 000		
		n=100	n=1000	n=2000	n=100	n=1000	n=2000	n=100	n=1000	n=2000
t	täpsus	95,0%	62,6%	42,2%	96,0%	64,5%	46,0%	95,0%	64,2%	46,2%
	saagis	5,5%	36,0%	48,6%	5,5%	37,1%	53,0%	5,5%	37,0%	53,1%
MI	täpsus	9,0%	11,9%	14,1%	8,0%	9,2%	10,2%	4,0%	4,1%	5,6%
	saagis	0,5%	6,9%	16,2%	0,5%	5,3%	11,7%	0,2%	2,4%	6,4%
hii	täpsus	71,0%	40,5%	31,6%	73,0%	48,4%	35,5%	79,0%	54,9%	39,6%
	saagis	4,1%	23,3%	36,4%	4,2%	27,9%	40,9%	4,5%	31,6%	45,6%
log	täpsus	88,0%	60,4%	38,8%	90,0%	63,0%	44,4%	88,0%	62,6%	45,1%
	saagis	5,1%	34,8%	44,7%	5,2%	36,3%	51,1%	5,1%	36,0%	52,0%
MS	täpsus	80,0%	51,0%	35,9%	86,0%	54,7%	38,1%	83,0%	56,2%	40,2%
	saagis	4,6%	29,4%	41,3%	5,0%	31,5%	43,9%	4,8%	32,4%	46,2%
sag	täpsus	73,0%	56,9%	41,6%	73,0%	59,0%	41,9%	73,0%	57,9%	41,4%
	saagis	4,2%	32,8%	47,9%	4,2%	34,0%	48,2%	4,2%	33,3%	47,7%

* Lühendid: stat – statistik, t – t-skoor, hii – hii-ruut-statistik, log – log-tõepära funktsioon, sag – koosinemise sagedus.

Tabelis 5 esitatud tulemused kinnitavad, et t-skooril on kõige paremad tulemused ühendverbide tuvastamisel erineva suurusega korpustest. t-skoori tulemused paranevad korpuse kasvades: kui 5 miljoni sõna suuruse korpuse puhul on t-skoori täpsus 42,2%, siis 170 miljoni sõna suuruse korpuse korral on see 46,2%. Ka saagis paraneb 48,6%-lt 53,1%-ni. Seega, mida suurem on korpus, seda paremad on t-skoori tulemused.

MI tulemused korpuse kasvades halvenevad: kui 5 miljonist sõnast koosnevast korpusest ühendverbide tuvastamisel on MI täpsus 2000 kandidaatpaari lõikes 14,1%, siis 170 miljoni sõna suuruse korpuse korral on täpsus 5,6%. Saagise väärtus väheneb 16,2%-lt 6,4%-le. Seega on korpuse suuruse kasvul tugev negatiivne mõju MI tulemustele.

Hii-ruut-statistiku, log-tõepära funktsiooni ja MS-i tulemused korpuse mahu suurenedes paranevad. Kui 5 miljonist sõnast koosneva korpuse 100 kandidaatpaari hulgas on hii-ruut-statistiku täpsus 71,0% ja 2000 kandidaatpaari juures 31,6%, siis 170 miljoni sõna suuruse valimi korral on need näitajad vastavalt 79,0% ja 39,6%. Ka hii-ruut-statistiku saagis paraneb 45,6%-ni. Log-tõepära funktsiooni tulemused paranevad korpuse kasvades 100 kandidaatpaari seas vähem, 2000 kandidaatpaari hulgas rohkem. 5 miljonist sõnast ja 170 miljonist sõnast koosnevatest korpustest ühendverbide tuvastamisel on 100 kandidaatpaari lõikes log-tõepära funktsiooni täpsused ja saagised võrdsed, 2000 kandidaatpaari seas on väiksema korpuse korral log-tõepära funktsiooni täpsus 38,8% ja saagis 44,7%, 170 miljonist sõnast koosneva korpuse puhul aga vastavalt 45,1% ja 52,0%. 5 ja 170 miljoni sõna suuruse korpuse

võrdlemisel näeb, et MS-i täpsus suureneb 100 kandidaatpaari seas 80,0%-lt 83,0%-ni ja 2000 kandidaatpaari hulgas 35,9%-lt 40,2%-ni, saagis vastavalt 4,6%-lt 4,8%-ni ja 41,3%-lt 46,2%-ni. Järelikult, mida suurem on korpus, seda paremad on nii hii-ruut-statistiku, log-tõepära funktsiooni kui ka MS-i tulemused.

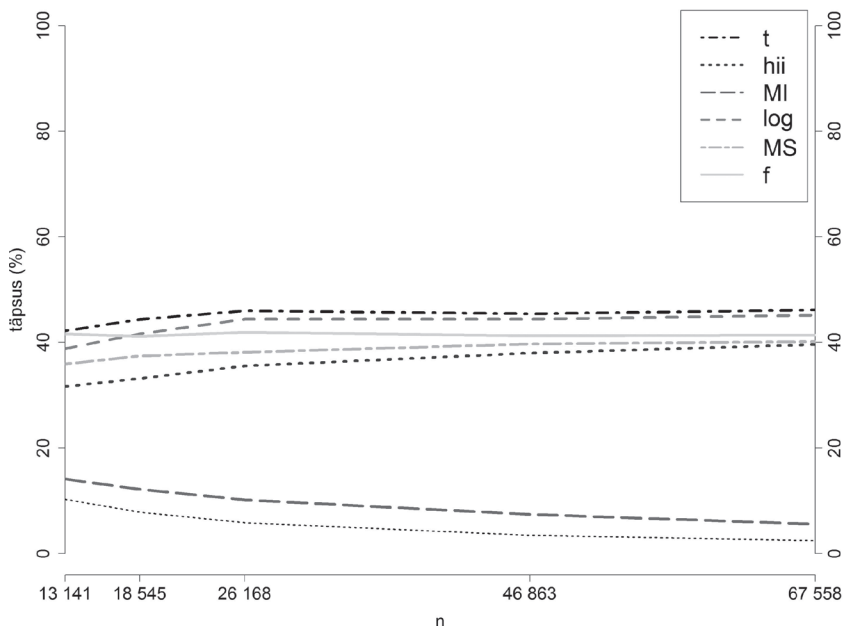
Koosesinemise sageduse tulemused ei muutu oluliselt: 100 kandidaatpaari seas on meetodi täpsus ja saagis korpuse suuruselt olenemata sama. 2000 kandidaatpaari hulgas on täpsus ja saagis kõige suurem 20 miljoni sõnast koosneva valimi korral (vastavalt 41,9% ja 48,2%), 170 miljoni sõna suuruselt korpusest ühendverbide tuvastamisel on koosesinemise sageduse täpsus 41,4%, mis erineb 5 miljoni sõna suurusel valimile vastavast täpsusest (41,6%) vaid 0,2%. Niisamuti on koosesinemise sageduse saagis 170 miljoni sõna suurusel korpuse korral (47,7%) võrreldes 5 miljoni sõna suurusel korpusega (47,9%) kahanenud 0,2%. Järelikult ei ole korpuse suurusel mõju koosesinemise sageduse tulemustele.

Korpuse suurusel ja ka vaadeldaval kandidaatpaaride arvul on mõju mõõdikute paremusjärjestusele. Parimaks mõõdikuks on olenemata korpuse suuruselt või vaadeldavate kandidaatpaaride arvust t-skoor ja halvimal võrreldud statistikuks MI. Kui vaadelda 100 kandidaatpaari, siis on üldiselt MS parem koosesinemise sagedusest, mille tulemused on omakorda paremad kui hii-ruut-statistikul. 20 miljoni sõna suurusel korpuse puhul on koosesinemise sageduse ja hii-ruut-statistiku tulemused võrdsed ja 170 miljoni sõna suurusel korpuses on hii-ruut-statistiku tulemused koosesinemise sageduse omadest paremad. 2000 kandidaatpaari hulgas on üldjuhul log-tõepära efektiivsem kui koosesinemise sagedus, mille tulemused on omakorda paremad kui MS-i ja hii-ruut-statistiku omad. Vaid kõige väiksema korpuse korral on koosesinemise sagedus 2000 kandidaatpaari seas parem kui log-tõepära funktsioon.

Seda, kuidas mõõdikute tulemused muutuvad korpuse suurenedes, illustreerib joonis 1, mis esitab sümmeetriliste mõõdikute ja koosesinemise sageduse täpsuste muutused ühendverbide tuvastamisel erineva suurusel korpustest. Joonise x-telg märgib korpuse suurusel korpusest genereeritud kandidaatpaaride arvu näol, y-telg märgib mõõdikute täpsust, kui arvesse võtta esimesed 2000 statistiku suurima väärtuse saanud kandidaatpaari. Peenike punktiirjoon joonisel tähistab algtaseme täpsust erineva suurusel korpustest.

Joonis 1 näitab, et kõige enam mõjutab 2000 kandidaatpaari hulgas korpuse kasv hii-ruut-statistiku tulemust, sest selle täpsus kasvab korpuse suurenemisega umbes 30%-lt 40%-ni. Samuti paranevad mõnevõrra t-skoori, log-tõepära funktsiooni ja MS-i tulemused. Koosesinemise sageduse tulemused korpuse mahu suurenemisega oluliselt ei muutu. MI tulemused halvenevad korpuse mahu suurendamisega.

Meetodite kõrvutamisel algtaseme täpsusega selgub, et kõikide mõõdikute täpsused on korpuse mahu kasvamisel suuremad kui algtaseme täpsus. Kui vaadelda esimest 2000 suurima statistiku väärtustega kandidaatpaari, siis üldjoontes on kõik mõõdikud eesti keele ühendverbide tuvastamisel tulemuslikud. Paremusjärjestus aga mõnevõrra muutub: kui 5 miljoni sõna suurusel korpuse korral on koosesinemise sagedus log-tõepära funktsioonist efektiivsem, siis suuremate korpuste korral on log-tõepära funktsiooni täpsus koosesinemise sageduse omast suurem. Ka võib öelda, et mida suurem on korpus, seda sarnasemad on koosesinemise sageduse, MS-i ja hii-ruut-statistiku tulemused. t-skoori täpsus on kõikide korpuste korral suurim, kuid suurema korpuse puhul on log-tõepära tulemused t-skoori tulemustega sarnasemad kui väiksema korpuse korral.



Joonis 1. Sümmetriliste mõõdikute ja koosinemise sageduse täpsused 2000 kandidaatpaari seas erineva suurusega ajakirjanduskorpustes

Tabeli 5 ja joonise 1 põhjal saab kokkuvõtlikult öelda, et kui korpuse maht suureneb, paranevad t-skoori, log-tõepära funktsiooni, hii-ruut-statistiku ja MS-i tulemused. Vastupidine mõju on korpuse mahu kasvamisel MI tulemustele. Koosinemise sageduse tulemused jäävad korpuse kasvades peaaegu muutumatuks.

5.2. Asümmeetriliste mõõdikute tulemused

Selleks et uurida, kuidas korpuse mahu suurenemine mõjutab asümmeetriliste statistikute tulemusi, saab vaadelda nende tuvastatud õigete ühendverbide arvu. Täpsema ülevaate korpuse mahu suurenemise mõjust asümmeetriliste statistikute tulemustele saab tabelist 6, kus on esitatud asümmeetriliste mõõdikute tuvastatud õigete ühendverbide arvud erineva suurusega ajakirjandustekstide korpustest.

Tabel 6. Asümmeetriliste statistikute tuvastatud õigete ühendverbide hulk statistiku 50 suurima väärtuse saanud ühendi seas erineva suurusega korpustest

Sõnade arv korpuses	CP(verb adverb)	CP(adverb verb)	$\Delta P(\text{verb} \text{adverb})$	$\Delta P(\text{adverb} \text{verb})$
5 mln	25	1	36	1
10 mln	25	1	39	1
20 mln	26	0	38	0
70 mln	26	0	40	0
170 mln	28	1	41	0

Kui vaadelda seost, kuidas verbi esinemine sõltub adverbi esinemisest samas osalauses, siis nii tingliku tõenäosuse kui ka ΔP tulemused korpuse kasvades paranevad: kui 5 miljoni sõna suurusest korpusest tuvastas tinglik tõenäosus 25 ja ΔP 36 õiget ühendverbi, siis 170 miljoni sõna suuruse korpuse põhjal on 50 suurima väärtuse saanud ühendi seas need numbrid vastavalt 28 ja 41. Vastupidist seost uurides aga tulemused korpuse mahu suurendes ei muutu või halvenevad. Seega oleneb korpuse mahu kasvu mõju asümmeetriliste statistikute puhul sellest, missugust seost sõnade vahel vaadelda. CP(adverb|verb) ja ΔP (adverb|verb) ei tuvasta palju õigeid ühendverbe, kuid sobivad harvaesinevate ühendite tuvastamiseks ja nad sisaldavad informatsiooni ühendite komponentide seoste kohta.

5.3. Sümmeetriliste ja asümmeetriliste mõõdikute tulemuste võrdlus

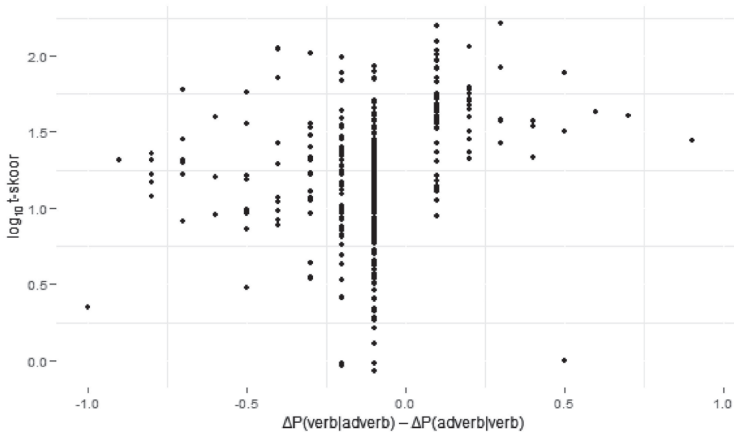
Sümmeetriliste ja asümmeetriliste mõõdikute tulemuste võrdluse eesmärgiks on välja selgitada, kas ja mida täpsemalt asümmeetriliste statistikute rakendamine lõpptulemusele juurde annab. Selleks võrdlen parimat sümmeetrilist statistikut – t-skoori – parima asümmeetrilise statistiku ΔP -ga (sümmeetriliste ja asümmeetriliste mõõdikute võrdlemisest saab põhjalikuma ülevaate autori artiklist Aedmaa 2014b).

Võtan arvesse need ühendid, mille ΔP -de väärtuste vahe (ΔP (verb|adverb) – ΔP (adverb|verb)) on suurem või väiksem kui 0 ehk ühendid, mida võib pidada asümmeetrilisteks. Selliseid ühendeid on ajakirjanduskorpuses 3314, millest 364 on õiged ühendverbid. ΔP võrdlemiseks t-skooriga arvestan mõlema statistiku 3314 suurima väärtuse saanud ühendit, mille seast eraldan õiged ühendverbid. ΔP hindamiseks võrdlen t-skoori tuvastatud õigete ühendverbide loendit ΔP tuvastatud õigete ühendverbide nimekirjaga.

t-skoori 3314 suurima väärtuse saanud ühendi seas on 1082 õiget ühendverbi. Nende ühendverbide võrdlemisel ΔP tuvastatud ühendverbidega selgub, et ΔP õigete ühendverbide hulgas on 34 sellist ühendverbi, mida t-skoor tuvastada ei suuda. t-skoor tuvastab küll rohkem õigeid ühendverbe kui ΔP ja on edukam eesti keele ühendverbide tuvastamisel, ent ΔP tulemused lisavad infot selle kohta, kumb sõna kollokatsioonis teisest rohkem sõltub.

Suurima positiivse ΔP kahe väärtuse vahe saanud ühendverb on *pärale jõudma*. Selles ühendverbis on verb *jõudma* palju rohkem ennustatav afiksaaladverbi *pärale* abil kui vastupidi ehk kui osalauses esineb afiksaaladverb *pärale*, siis esineb seal suure tõenäosusega ka verb *jõudma*, aga kui osalauses esineb verb *jõudma*, siis afiksaaladverbi *pärale* esinemine nii tõenäoline ei ole. Põhjus peitub selles, et *pärale* on afiksaaladverb, mis üldjuhul esinebki ühendverbi *pärale jõudma* koosseisus. Väikseim ΔP väärtuste vahe on ühendverbil *ümber rahvustuma*, milles on afiksaaladverb *ümber* palju rohkem ennustatav verbi *rahvustuma* abil kui vastupidi ehk kui osalauses esineb verb *rahvustuma*, siis esineb seal suure tõenäosusega afiksaaladverb *ümber*, aga kui osalauses esineb afiksaaladverb *ümber*, siis verbi *rahvustuma* esinemine nii tõenäoline ei ole.

Seda, et sümmeetrilised mõõdikud ei sisalda infot ühendite asümmeetrilisuse kohta, kinnitab joonis 2, kus on esitatud ΔP tuvastatud õigete ühendverbide jaotus nende t-skoori ja ΔP (verb|adverb) – ΔP (adverb|verb) väärtuste järgi. Joonise x-teljel on ΔP kahe väärtuse vahe ning y-teljel t-skoori väärtused. t-skoori väärtused on logaritmitud alusel 10.



Joonis 2. ΔP tuvastatud õigete ühendverbide jaotus $\Delta P(\text{verb}|\text{adverb}) - \Delta P(\text{adverb}|\text{verb})$ ja t-skoori väärtuste järgi

Jooniselt 2 selgub, et suure t-skoori väärtusega ühendeid leidub igasuguse ΔP väärtusega ühendite seas, mis tähendab, et t-skoor ja üldistatult ka teised sümmeetrilised statistikud omistavad ühenditele suuri kahesuunalisi väärtuseid arvestamata seda, kumb sõna teisest sõltub. Asümmeetrilise ΔP abil saab aga tuvastada ühendeid, mille komponentide vahel on asümmeetriline seos. Jooniselt 2 on näha ka see, et rohkem on negatiivse ΔP vahega ühendeid ja järelikult on andmestikus rohkem ühendeid, kus verb on paremini ennustatav adverbi järgi kui vastupidi. Samas t-skoor omistab sagedamini suurema väärtuse just nendele ühenditele, kus adverb on paremini ennustatav verbi esinemise järgi kui vastupidi, ning seega tuvastab paremini üht tüüpi ühendverbe.

Kokkuvõttes võib öelda, et ΔP -d ja üldistatult teisi asümmeetrilisi mõõdikuid on mõistlik sümmeetriliste mõõdikute kõrval kollokatsioonide tuvastamisel rakendada, sest nad tuvastavad ühendeid, mida sümmeetrilised mõõdikud ei tuvasta, ning sisaldavad lisainfot ühendite asümmeetrilisuse kohta. Samas teistest oluliselt paremat mõõdikut ei ole ja mõistlik on erinevat liiki statistikuid omavahel kombineerida.

6. Kokkuvõte

Artikkel käsitles sõnadevahelise seose tugevuse mõõtmise meetodeid eesti keele ühendverbide automaatsel tuvastamisel tekstikorpusest. Lisaks sümmeetrilistele statistikutele rakendasin ka kahte asümmeetrilist statistikut eesmärgiga uurida, kuidas erinevad sümmeetriliste ja asümmeetriliste statistikute tulemused.

Selgus, et t-skoori võib pidada kõige efektiivsemaks sümmeetriliseks statistikuks ühendverbide tuvastamisel ajakirjanduskorpusest, kuid kõiki rakendatud statistikuid tasub kasutada ühendverbide tuvastamiseks tekstikorpusest, sest korpuse suurus ja vaadeldavate kandidaatpaaride arv mõjutavad statistikute tööd.

Katsed tõestasid, et ka asümmeetriliste mõõdikute rakendamine on põhjendatud ühendverbide tuvastamisel tekstikorpusest, sest need tuvastavad arvestatava

arvu õigeid ühendverbe ning lisaks sisaldavad informatsiooni ühendisse kuuluvate sõnade seose suuna kohta. Parima lõpptulemuse saamiseks on mõistlik sümmeetrilisi ja asümmeetrilisi statistikuid kombineerida, sest sümmeetrilised mõõdikud ei tuvasta sõnadevahelisi asümmeetrilisi seoseid, mis on iseloomulikud inimese kognitiivsetele võimetele. Erinevate meetodite kombineerimise võimaluste ja nende kombinatsioonide tulemuslikkuse hindamine kollokatsioonide tuvastamisel tekstikorpusest jääb edasise uurimistöö teemaks.

Viidatud kirjandus

- Aedmaa, Eleri 2014a. Sõnadevahelise seose tugevuse mõõtmise statistilised meetodid ühendverbide tuvastamisel. [Statistical Methods for Particle Verb Extraction.] Magistritöö. Käsikiri Tartu ülikooli üldkeeleteaduse osakonnas. <http://hdl.handle.net/10062/44260>
- Aedmaa, Eleri 2014b. Statistical methods for Estonian particle verb extraction from text corpus. – Proceedings of the ESSLLI 2014 Workshop: Computational, Cognitive, and Linguistic Approaches to the Analysis of Complex Words and Collocations, 17–22.
- Bell, Alan; Brenier, Jason M.; Gregory, Michelle; Girand, Cynthia; Jurafsky, Dan 2009. Predictability effects on durations of content and function words in conversational English. – Journal of Memory and Language, 60 (1), 92–111. <http://dx.doi.org/10.1016/j.jml.2008.06.003>
- Church, Kenneth Ward; Hanks, Patrick 1990. Word association norms, mutual information, and lexicography. – Computational Linguistics, 16, 22–29.
- Dunning, Ted 1993. Accurate methods for the statistics of surprise and coincidence. – Computational Linguistics, 19, 61–74.
- Eesti keele seletav sõnaraamat. <http://www.eki.ee/dict/ekss/> (16.12.2014).
- Ellis, Nick C 2006. Language acquisition as rational contingency learning. – Applied Linguistics, 27 (1), 1–24. <http://dx.doi.org/10.1093/applin/ami038>
- Ellis, Nick C.; Ferreira-Junior, Fernando 2009. Constructions and their acquisition: Islands and the distinctiveness of their occupancy. – Annual Review of Cognitive Linguistics, 7 (1), 188–221. <http://dx.doi.org/10.1075/arcl.7.08ell>
- Evert, Stefan 2004. The Statistics of Word Cooccurrences. Dissertation. Stuttgart: Stuttgart University.
- Evert, Stefan 2008. Corpora and collocations. – Anke Lüdeling, Merja Kytö (Eds.). Corpus Linguistics. An International Handbook 2. De Gruyter Mouton, 223–233. <http://dx.doi.org/10.1515/9783110213881.2.1212>
- Gries, Stefan Th. 2013. 50-something years of work on collocations: What is or should be next. – International Journal of Corpus Linguistics, 18 (1), 137–166. <http://dx.doi.org/10.1075/ijcl.18.1.09gri>
- Kaalep, Heiki-Jaan 1998. Tekstikorpuse abil loodud eesti keele morfoloogiaanalüsaator. [An Estonian morphological analyser and using a corpus on its development.] – Keel ja Kirjandus, 1, 22–29.
- Kaalep, Heiki-Jaan; Muischnek, Kadri 2002. Püsiühendite leidmine teksti abil. [Extraction of multiword expressions using text corpus.] – Renate Pajusalu, Tiit Hennoste (Toim.). Tähtsusepüüdja: pühendusteos professor Haldur Õimu 60. sünnipäevaks 22. jaanuaril 2002. Catcher of the Meaning: Festschrift for Professor Haldur Õim on the occasion of his 60th birthday. TÜ üldkeeleteaduse õppetooli toimetised 3. Tartu: Tartu Ülikool, 172–184.
- Kaalep, Heiki-Jaan; Muischnek, Kadri 2009. Eesti keele püsiühendid arvutilingvistikas: miks ja kuidas. [Estonian multiword expressions in computational linguistics.] – Eesti

- Rakenduslingvistika Ühingu aastaraamat, 5, 157–172. <http://dx.doi.org/10.5128/ERYa5.10>
- Kaalep, Heiki-Jaan; Muischnek, Kadri 2012. Osalause tüvastamine eestikeelses tekstis kui iseseisev ülesanne. [Clause splitting as a separate task (in the analysis of Estonian texts).] – Eesti Rakenduslingvistika Ühingu aastaraamat, 8, 55–68. <http://dx.doi.org/10.5128/ERYa8.04>
- Kaalep, Heiki-Jaan; Vaino, Tarmo 1998. Kas vale meetodiga õiged tulemused? Statistikaline tuginenud eesti keele morfoloogiline ühestamine. [Getting correct results with an incorrect method? Morphological disambiguation of Estonian using statistics.] – Keel ja Kirjandus, 1, 30–38.
- Kallas, Jelena 2013. Eesti keele sisusõnade süntagmaatilised suhted korpus- ja õppeleksikograafias. [Syntagmatic Relationships of Estonian Content Words in Corpus and Pedagogical Lexicography.] Tallinna Ülikooli humanitaarteaduste dissertatsioonid 32. Tallinn: Tallinna Ülikool. <http://e-ait.tlulib.ee/id/eprint/303>
- Krenn, Brigitte; Evert, Stefan 2001. Can we do better than frequency? A case study on extracting PP-verb collocations. – Proceedings of the ACL Workshop on Collocations, 39–46.
- Manning, Christopher D.; Schütze, Hinrich 1999. Foundations of Statistical Natural Language Processing. Cambridge (Mass.)–London: MIT press.
- Michelbacher, Lukas; Evert, Stefan; Schütze, Hinrich 2007. Asymmetric association measures. – Proceedings of the Recent Advances in Natural Language Processing (RANLP 2007).
- Michelbacher, Lukas; Evert, Stefan; Schütze, Hinrich 2011. Asymmetry in corpus-derived and human word associations. – Corpus Linguistics and Linguistic Theory, 7 (2), 245–276. <http://dx.doi.org/10.1515/cllt.2011.012>
- Pecina, Pavel; Sclesinger, Pavel 2006. Combining association measures for collocation extraction. – Proceedings of the COLING/ACL on Main conference poster sessions, 651–658.
- Pedersen, Ted 1998. Dependent bigram identification. – AAAI/IAAI, 1197.
- Pedersen, Ted; Bruce, Rebecca 1996. What to infer from a description. Technical Report 96-CSE-04. Southern Methodist University. Dallas, TX.
- Sinclair, John 1991. Corpus, Concordance, Collocation. Oxford: Oxford University Press.
- Uiboed, Kristel 2010. Statistilised meetodid murdekorpuse ühendverbide tüvastamisel. – Eesti Rakenduslingvistika Ühingu aastaraamat, 6, 307–326. <http://dx.doi.org/10.5128/ERYa6.19>
- Wermter, Joachim; Hahn, Udo 2006. You can't beat frequency (unless you use linguistic knowledge): A qualitative evaluation of association measures for collocation and term extraction. – Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics, 785–792.
- Wiechmann, Daniel 2008. On the computation of collocation strength: Testing measures of association as expressions of lexical bias. – Corpus Linguistics and Linguistic Theory, 4 (2), 253–290. <http://dx.doi.org/10.1515/CLLT.2008.011>

Eleri Aedmaa (Tartu Ülikool) uurimisvaldkonnad on korpuslingvistika, püsiühendid, statistilised meetodid keeleteaduses.
Tartu Ülikool, eesti ja üldkeeleteaduse instituut, Jakobi 2, 51014 Tartu, Estonia
eleriaed@ut.ee

STATISTICAL METHODS FOR ESTONIAN PARTICLE VERB EXTRACTION FROM TEXT CORPORA

Eleri Aedmaa

University of Tartu

The present article compares lexical association measures (AMs) for automatic extraction of Estonian particle verbs from the newspaper part of the Estonian Reference Corpus. The main purpose of this study is to ascertain the best symmetrical AM for Estonian particle verb extraction. The central focus lies on the impact of the corpus size on the performance of the compared symmetrical association measures. In addition, asymmetrical AMs have been included in the study to observe their suitability for Estonian particle verb extraction. Five symmetrical association measures have been used, namely the t-test, log-likelihood, χ^2 , mutual information, and minimum sensitivity, as well as two asymmetrical association measures, namely conditional probability and ΔP . The association measures were compared against the co-occurrence frequency of verb and verbal particle.

The analysis of the comparison reveals that the t-test achieved the best precision values and the corpus size has an impact on the performance of the AMs. As the corpus size increased, the performances of the t-test, log-likelihood, χ^2 and minimum sensitivity increased and the precision of mutual information decreased. The performance of (simple) frequency did not change significantly as the size of the corpus increased.

The comparison of symmetrical and asymmetrical AMs revealed that asymmetrical association measures are suitable for the task of Estonian particle verb extraction and provide slightly different and more detailed information about the extracted particle verbs. The results presented in this article confirm that further study of asymmetrical AMs is necessary and more experiments are needed to broaden our knowledge of the performance of asymmetrical AMs.

Keywords: computational linguistics, corpus linguistics, multi-word expressions, particle verbs, statistics, Estonian

EESTI JA SOOME KEELE VASTASTIKUNE MÕISTMINE ÜKSIKSÕNA- JA TEKSTITASANDIL: LINGVISTILISED TEGURID, MÕISTMIS-PROTSESS JA SÜMMEETRIA

Annekatriin Kaivapalu

Ülevaade. Artiklis käsitletakse kirjaliku eesti ja soome keele vastastikust mõistmist üksiksõna- ja tekstitasandil. Eesmärk on selgitada, kas ja kuidas eesti- ja soomekeelsed keelekasutajad mõistavad oma emakeele põhjal lähedast sugulaskeelt ilma sihtkeele õppimis- ja kasutuskogemusest ning mille poolest erineb üksiksõnade mõistmine tervikteksti mõistmisest. Mõistmistestide vastuste ja osalejate eneserefleksiooni analüüsi keskmes on kahe sihtkeele mõistmise võimalik sümmeetria ning mõistmisprotsessi mõjutavad tegurid.

Tulemuste põhjal osutusid eestikeelsed keelekasutajad edukamaks üksiksõnatestis, eriti eesti ja soome keeles samatähenduslike kirjapildilt sarnaste sõnade ning laen- ja võõrsõnade mõistmisel, soomekeelsed keelekasutajad tervikteksti mõistmisel. Mõlema sihtkeele mõistmisel on otsustav keeltevahelise sarnasuse tunnetamine: toetutakse esmajoones eesti ja soome, aga ka muude varem õpitud keelte ja sihtkeele sarnasusele, seostades seda metalingvistiliste ja üldteadmistega ning kontrollides oletusi lause-, lõigu- ja tekstikontekstis. Kontekst toimib omalaadse filtrina, mis selekteerib keelekasutaja oletusi neid aktsepteerides või tagasi lükates. Mõistmisprotsessi komponentide jaotus ja suund varieeruvad indiviiditi ja sõltuvad paljuski osalejate varasematest teadmistest ja kogemustest.*

Võtmesõnad: retseptiivne mitmekeelsus, tunnetatud keeltevaheline sarnasus, eesti keel, soome keel

* Tänan häid kolleege soome keele lektorit Maria-Magdalena Jürvetsoni ja soome-ugri keelte lektorit Ivar Sinimetsa abi eest materjali kogumisel.

1. Retseptiivsest mitmekeelsusest ja selle uurimisest

Euroopa Nõukogu keelepoliitika põhiväärtusi on keeleline mitmekesisus, eesmärkide hulgas Euroopa Liidu kodanike vähemalt kahe keele oskamine lisaks emakeelele, üks neist võiks eelistatavalt olla n-ö väike keel. Keeleoskuse all mõeldakse keeleteaduse vajadustest lähtuvat toimetulekut, seega ei tarvitse keeleoskus olla osaoskusesti võrdsel tasemel, vaid vastavalt vajadusele võib olla parem retseptiivne või produktiivne, suuline või kirjalik keeleoskus. (White paper 1995, EKR) Uut keelt õppides tuginetakse teadustades või teadustamata emakeele ja/või teiste varem õpitud või omandatud keelte oskusele, kasutades teadmisi nende keelte kohta ning omandatud konstruktsioone ja väljendusvahendeid. Sihtkeele õppimine ja kasutamine algab arusaamisest: enne kui keelt on võimalik produtseerida, tuleb sellest kasvõi väikest osa mõista. (Ringbom, Jarvis 2009) Seega on sihtkeele mõistmine, mis teise keele omandamise uurimises seni teenimatult vähe tähelepanu on leidnud, keeleteaduse eelduseks. Sihtkeelest arusaamisel toetub keeleõppija esmajoonel selle sarnasusele lähtekeelena toimiva emakeelega või mõne varem õpitud keelega (Ringbom 2007). Lähedaste sugulaskeelte puhul ulatub keelesüsteemide sarnasus määrani, mis võimaldab sihtkeelt mõista ilma keeleõppeta või märkimisväärse keeleteaduskogemusega.

Eelkirjeldatud põhineb suhtlemine mitmes Euroopa piirkonnas (näiteks rootsi, taani, norra ja islandi keel Skandinaavias, saksa, hollandi ja friisi, tšehhi, slovaki ja sorbi keel, romaani keeled), kus on üldiseks tavaks **retseptiivne mitmekeelsus**: suhtlusviis, mille korral vestluses osalejad kõnelevad oma emakeeles ja mõistavad kaasvestleja keelt (Zeevaert, ten Thije 2007). Viimastel aastatel on vastastikuse mõistmise uurimises hakatud pöörama tähelepanu ka õpitud mittesugulaskeelte, näiteks eesti ja vene keele kasutusele samasuguses funktsioonis, kasutades selle tähistamiseks laiemat mõistet *lingua receptiva* (LaRa). *Lingua receptiva* all mõeldakse nende lingvistiliste, mentaalsete, interaktiivsete ja kultuuridevaheliste pädevuste kogumit, mis aktiveeruvad, kui kuulaja tunnetab kõneleja keelelist tegevust või kõneleja püüab kontrollida arusaamisprotsessi (ten Thije jt 2012). Kuigi retseptiivse mitmekeelsuse uurimises on keskendunud peaaesjalikult suulisele suhtlusele, ilmneb see ka kirjalike tekstide mõistmises. Nimelt kirjalikust arusaamisest lähtub üleeuroopaline projekt EuroComDidact¹ ja selle allprojektid EuroComRom, EuroComGerm ja EuroComSlav, milles samaaegselt õpetatakse ja õpitakse mõistma mitut lähedast sugulaskeelt.

Ka eesti ja soome keele sarnasus on aluseks sihtkeele kohta käivatele oletustele, mida toetavad või ei toeta õppijate arusaamad ja kogemused keeltevahelisest sarnasusest ning millel põhineb sihtkeele mõistmine. Vaatamata arvukatele eesti ja soome keelt võrdlevatele ja vastastikust keeleõpet käsitlevatele uurimustele, millest enamik on ilmunud sarjas “Lähivõrdlusi. Lähivertailuja”², ning eestlaste ja soomlaste retseptiivse mitmekeelsuse põhimõtteid järgiva suhtlemistraditsiooni pikaajalisusele ja tavapärasusele, on eesti ja soome keele vastastikuse mõistmise uurimine alles algusjärgus. Ida-Soome ülikooli koordineeritav võrgustik REMU³ (“Retseptiivne mitmekeelsus: lähedaste sugulaskeelte vastastikune mõistmine”) Tallinna, Helsingi ja Jyväskylä ülikooli ning Åbo Akademi osalusel on seadnud eesmärgiks seda lünka täita ja nii mitmekesisust uurimist seni valdavalt indo-euroopa keeltele keskendunud valdkonnas.

¹ <http://www.eurocomresearch.net/> (2.2.2015).

² <http://www.rakenduslingvistika.ee/ajakirjad/index.php/lahivordlusi> (2.2.2015).

³ www.uef.fi/en/remu2015 (2.2.2015).

Käesolev artikkel käsitleb eesti emakeelega keeletekasutajate soome keele mõistmist ja soome emakeelega keeletekasutajate eesti keele mõistmist kahel tasandil: üksiksõna- ja tekstitasandil. Uurimused viidi läbi Eestis ja Soomes eesmärgiga selgitada, kuidas eesti- ja soomekeelsed keeletekasutajad mõistavad oma emakeele põhjal lähedast sugulaskeelt juhul, kui uurimuses osalejatel puudub sihtkeele õppimis- ja kasutuskogemus. Teine eesmärk on keskenduda eesti ja soome keele mõistmise võimalikule sümmeetriale, analüüsides, kas eestlased ja soomlased mõistavad lähedast sugulaskeelt samal määral; kolmas eesmärk on uurida osalejate mõistmisprotsessi ja seda kujundavaid tegureid ning töötada välja mõistmisprotsessi kirjeldav mudel. Tegemist on projekti REMU raames teostatud kahe eraldi uurimusega. Üksiksõnatesti kavandamisel on siiski mõnevõrra lähtunud terviktekstist, kasutades selles üksikuid tekstis esinenud sõnu. Seetõttu esitatakse artiklis kahe mõistmistasandi tulemused eraldi neid omavahel märkimisväärselt seostamata, rõhuasetusega tekstitasandil. Analüüsi keskmes on mõistmist mõjutavad keelelised tegurid, mida järgnevalt kirjeldatakse.

2. Arusaamist mõjutavad tegurid

Valdavalt germaani, sh Skandinaavia keelele keskendunud retseptiivse mitmekeelsuse uurimises⁴ on analüüsitud nii keelelisi kui ka keeleväliseid tegureid, mis mõjutavad lähedasest sugulaskeelest arusaamist ja selle sümmeetriat (nt Bezooijen, Gooskens 2007, Kaivapalu, Muikku-Werner 2010). Keelevälisest teguritest on enam tähelepanu pööratud kokkupuudetele sihtkeelelega (Gooskens, Kürschner 2009, Gooskens, Heeringa 2014) ja sellesse suhtumisele, aga ka sotsiaalsetele teguritele (Gooskens 2007, Hilton, Gooskens 2013, Gooskens, Hilton 2010). Kui tihedam kokkupuude sihtkeelelega aitab reeglina kaasa selle paremale mõistmisele ka ilma formaalse keeleõppeta, siis sihtkeelde suhtumise roll on osutunud küll keeliti erinevaks, kuid üldiselt arvatust märksa marginaalsemaks (Bezooijen, Gooskens 2007).

Keelelistest teguritest on enim käsitletud lähte- ja sihtkeele vahelist kaugust (ingl *linguistic distance*): ühise päritoluga sõnade hulka, leksikaalse ja grammatilise tähenduse läbipaistvust jne (nt Bezooijen, Gooskens 2007), mida on nimetatud ka tüpoloogiliseks läheduseks (*typological proximity*) või pindmiseks sarnasuseks (*surface similarity*) (vt Kaivapalu, Martin 2014). Tegemist on keelesüsteemide lähedust peegeldava objektiivse ehk tegeliku sarnasusega (Ringbom 2007), mis on sümmeetriline, st keel A on keelega B sama sarnane kui keel B keelega A, ja ajas muutumatu (Ringbom, Jarvis 2009: 107). Keeltevahelise objektiivse sarnasuse defineerimine ja mõõtmine on osutunud tõsiseks väljakutseks: ka enim kasutatud mõõtmismeetodid nagu Levenshteni distant⁵ ja projekti EuroComDidact nn seitsme sõela meetod (Klein, Stegmann 2000) ei ole osutunud keeltevahelise tegeliku sarnasuse praktikas määratlemisel piisavaks (keeltevahelise sarnasuse mõõtmisest vt lähemalt Kaivapalu, Martin 2014). Keeltevahelised tegelikud sarnasused ja erinevused omavad sihtkeele mõistmisel kesksel rolli ja seletavad mõneski osas, miks keele A kõnelejad on keele B mõistmisel edukamad kui keele B kõnelejad keele A mõistmisel (Bezooijen, Gooskens 2007, van Ommen jt 2013, Gooskens, Heeringa 2006).

⁴ Vt <http://www.let.rug.nl/gooskens/?p=project> (2.2.2015).

⁵ <http://xlinux.nist.gov/dads/HTML/Levenshtein.html> (2.2.2015).

Keeltevahelisele tunnetatud ja oletatavale sarnasusele (Ringbom 2007), mida on nimetatud ka subjektiivseks sarnasuseks, psühhotüpoloogiaks (*psychotypology*), tunnetatud läheduseks (*perceived proximity*) või sarnasuse psühholingvistiliseks reaalsuseks (*psycholinguistic reality of similarity*), on sihtkeele mõistmise uurimisel seevastu hoopis vähem tähelepanu pööratud (nt murdeteadlikkuse osast keeltevahelise sarnasuse tunnetamisel Heeringa jt 2008; vt Kaivapalu, Martin 2014). Tegemist on aga sihtkeele mõistmise seisukohalt palju olulisema teguriga kui keeltevaheline tegelik sarnasus. Selleks, et tundmatust sihtkeelest lähtekeele põhjal aru saada, võrdleb keelekasutaja lähte- ja sihtkeele keelendeid, tunnetamaks sarnasust keelte sõnavara ja struktuuri, mallide ja reeglite vahel (Jarvis, Odlin 2000: 538). Tunnetatud sarnasus toimib lähte- ja sihtkeele vahelise ühendusena, mille kaudu ühes keeles olemasolevad teadmised ja oskused siirduvad teise keelde ja viivad selle mõistmiseni (Martin 2006).

Hüpoteetiliselt võivad tegelik ja tunnetatud sarnasus ka kokku langeda, kui keelekasutaja tunnetab täielikult tegelikku sarnasust, kuid seda juhtub siiski väga harva. Keeltevaheline tegelik sarnasus määrab küll suuresti keeleomandamise käigu ja saavutatud keeleoskustaseme (nt Ringbom 1987: 66, Odlin 1989, Schepens jt 2013), kuid selle mõju on võrreldes tunnetatud sarnasusega kaudsem ja piiratum peamiselt keelekasutajate individuaalse eripära tõttu (Ringbom, Jarvis 2009: 106–107). Sarnasuse tunnetamine on seega psühholingvistiline ilming, mis, nagu ka sellele lähedane vastastikune mõistmine (*mutual intelligibility*, sümmeetrilisusest vt Gooskens jt 2010), ei tarvitse tingimata olla sümmeetriline: keele A kasutajal võib keele A ja B sarnasuse tunnetamine olla kergem kui keele B kasutajal. Samuti võib sarnasuse tunnetamine ajas muutuda, kui sihtkeele oskus ja kasutuskogemus kasvavad (Kellerman 1979).

Sihtkeele mõistmist mõjutavate keeleliste tegurite hulka kuuluvad ka siiani vähe uuritud kommunikatiivse keeleoskuse (nt Bachman 1990: 85–87) komponendid: üldteadmised, strateegiapädevus, tekstipädevus (eriti konteksti kasutamine), aga ka metalingvistiline teadlikkus (Kaivapalu, Muikku-Werner 2010), st teadmised inimkeelest kui sellisest ja selle toimimisest ning teadmised lähtekeel(t)est (eriti emakeelest) ja selle/nende varieerumisest. Siinses artiklis keskendutakse üksnes sihtkeele mõistmist mõjutavatele keelelistele teguritele. Keelevälised tegurid (kokkupuude sihtkeelega, suhtumine sellesse, sotsiaalne taust) on nii üksiksõna- kui ka tekstitasandi uurimustes elimineeritud.

3. Lähedase sugulaskeele mõistmine ja selle sümmeetria

Üks projekti REMU eesmärke on võrrelda lähedase sugulaskeele mõistmist üksiksõna-, lause- ja tekstitasandil. Käesolevas osas tutvustatakse lähemalt üksiksõna- ja tekstitasandi mõistmist käsitlevaid uurimusi, mille sihtrühmad on taustategurite osas hästi võrreldavad. Nii üksiksõnade mõistmise testi sooritanud gümnaasiumiõpilased kui ülikooliõpingute alguses olevad üliõpilased, kes osalesid tekstimõistmistestis, paigutuvad kognitiivse arengu mõttes samasse rühma, n-ö noorte täiskasvanute hulka. Samuti on sarnane uurimustes osalejate haridus-, kultuuri- ja sotsiaalne taust ning varasem keeleoskus, ka testi sooritamise tingimused olid samad.

3.1. Üksiksõnatasand

Üksiksõnade mõistmist käsitletakse kahes võrreldavuse huvides samadel alustel kavandatud ja teostatud uurimuses: Maria Heinose (2011) bakalaureusetööl põhinevas artiklis (Muikku-Werner, Heinonen 2012) ja sellele järgnevas magistritöös (Heinonen 2014) keskendutakse soome gümnaasiumiõpilaste eesti üksiksõnade ning Marta Margi (2014) bakalaureusetöös eesti gümnaasiumiõpilaste soome üksiksõnade mõistmisele. Mõlema uurimuse eesmärk oli selgitada, kas ja kuidas aitab eesti ja soome keele sarnasus nende keelte sõnavara mõistmisele kaasa. Samuti sooviti teada saada, kas ja kuidas on mõlema keele sõnavarasse kuuluvate võõr- ja laensõnade mõistmisel kasu muudest varem õpitud võõrkeeltest ning milliseid strateegiaid on vastajad mõistmisprotsessis kasutanud. Leidmaks vastuseid nendele küsimustele, koostati üksiksõnade mõistmise testid nelja tüüpi sõnadest (soomekeelne test vt Mark 2014 ja eestikeelne Muikku-Werner, Heinonen 2012: 184–185):

- 1) homogeensed samatähenduslikud sõnad, mille kirjpilt ja tähendus on eesti ja soome keeles sama või lähedane (nt *kurss* – *kurssi*, *tädi* – *täti*, *naaber* – *naapuri*);
- 2) homogeensed eritähenduslikud sõnad ehk nn valesõbrad, mille kirjpilt on sama või lähedane, kuid tähendus erinev (nt *kalju* – *kalju* 'kiilaspäine', *hallitus* – *hallitus* 'valitsus', *viiner* – *viineri* 'saiake');
- 3) sünonüümidenä või jagatud semantilise välja tõttu lähtekeelega kaudses tähenduseseoses olevad sõnad (nt eesti *vihmavari* 'sateenvarjo' võiks olla soomlasele mõistetav soome keeles lähedase tähendusega sõna *vihma* 'seenevihm' kaudu, sest mõlemad kuuluvad samasse semantilisse välja, st mõlemal juhul on tegemist vihmaga);
- 4) mõne muu keele abil äratuntavad laen- ja võõrsõnad (nt sm *historia*, ee *kostüüm*).

Soomekeelses testis oli 35 eestikeelset, eestikeelses testis 40 soomekeelset sõna, mille hulgas esines nii käändsõnu (sh liitsõnu) kui ka verbe. Testitavad käändsõnad ja liitsõnad esitati uurimuses osalejatele ainsuse nominatiivis, verbid A-infinitiivis⁶. Sõnade arvu erinevus eesti- ja soomekeelses testis on tingitud soomekeelse testi kõrvaleesmärgist: uurida lõunaeesti murdekeele mõju soome keele mõistmisele. Uurimuses osalejate ülesandeks oli tõlkida testisõnad võimalikult täpselt oma emakeelde ja kirjeldada strateegiaid, mille abil nad lahenduseni jõudsid. Soomekeelseid õpilasi ühest Lõuna-Savo maakonna gümnaasiumist osales uurimuses 32, nendest neide 17, noormehi 15. Eestikeelseid õpilasi Keila gümnaasiumist ja Põlva ühisgümnaasiumist oli kokku 47, nendest 27 neidu ja 20 noormeest. Kõik õpilased olid võõrkeelena õppinud inglise keelt, lisaks soome õpilased rootsi ja eesti õpilased vene (43 õpilast), saksa (16), hispaania (2) ja/või prantsuse (8) keelt. Seega olid soome õpilased varem õppinud kaht, eesti õpilased kaht-kolme võõrkeelt. Sihtkeeli eesti ja soome keelt ei olnud uurimuses osalejad varem õppinud, kokkupuude naabrite keelega puudus täiesti või seisnes väheses juhuslikus kogemuses sugulaste ja pereliikmete, turismi, televisiooni või filmide kaudu.

Tabelis 1 on esitatud testisõnade tõlgete õigete vastuste protsent testisõnärühmiti. Vastuste õigsuse hindamisel sai iga täiesti õige vastus kaks ja peaaegu õige vastus ühe punkti, seejärel summeeriti iga sõnärühma punktid ja leiti õigete

⁶ Terminit *A-infinitiiv* on kasutatud eesti *da-* ja soome *A-infinitiivi* ühisnimetusena.

vastuste protsent maksimaalsest võimalikust punktide arvust (Mark 2014: 20, Muikku-Werner, Heinonen 2012: 176).

Tabel 1. Lähedase sugulaskeele üksiksõnade mõistmine (õigete vastuste osakaal protsentides)

Osalejate rühm	Testisõnarühm				
	Kirjapildilt ja tähenduselt sarnased sõnad	Kirjapildilt sarnased, kuid tähenduselt erinevad sõnad	Kaudses tähendusseoses olevad sõnad	Laen- ja võõrsõnad	Keskmine
Eesti õpilased (n = 47)	68,4	2,4	8,0	56,2	33,8
Soome õpilased (n = 32)	43,6	4,7	24,2	35,3	27,0

Tabelist 1 ilmneb, et sõnarühmade mõistmise järjestus on mõlema uurimuse puhul sama: nii eesti kui ka soome õpilaste jaoks olid ootuspäraselt kõige arusaadavamad eesti ja soome keeles kirjapildilt sarnased samatähenduslikud sõnad, seejärel laen- ja võõrsõnad, enim raskusi valmistasid kirjapildilt sarnased, kuid tähenduselt erinevad valesõbrad. Selline tulemus rõhutab keeltevahelise sarnasuse primaarsust mõistmisprotsessis (Ringbom 2007): emakeele ja sihtkeele vahelise sarnasuse kõrval on äärmiselt oluline ka sarnasuse tunnetamine varem õpitud keelte ja sihtkeele vahel, kuigi viimane ei vii emakeele ja sihtkeele sarnasusega võrreldavate tulemusteni. Ka tunnetatud sarnasust otseselt käsitlevad vähesed uurimused (Kaivapalu, Martin 2014) osutavad, et emakeele ja sihtkeele vahelist sarnasust tunnetatakse paremini kui emakeele ja varem õpitud keele sarnasust eelkõige seetõttu, et emakeele puhul on keelekasutajate käsutuses lai geograafilise, sotsiaalse ja ajaloolise varieerumise repertuaar, mis reeglina puudub isegi kõrgetasemelisel teise keele valdajal. Seetõttu piirdub sarnasuse tajumine varem õpitud keele ja sihtkeele kirjaliku kasutuse vahel enamasti ortograafiatasandiga, erinevate tähtede arvu märkamisega.

Sarnasuste kõrval on eesti ja soome õpilaste testitulemustes huvitavaid erinevusi. Kui eesti õpilased on oma soome eakaaslastest paremini mõistnud nii emakeeles kui ka varem õpitud võõrkeeles ja sihtkeeles sarnaseid sõnu, siis soome õpilased on oluliselt edukamad kaudses tähendusseoses olevate sõnade mõistmisel.

3.2. Tekstitasand

Tervikteksti mõistmist käsitlevas uurimuses paluti osalejatel (39 soome üliõpilast Jyväskylä ja Ida-Soome ülikoolidest ning 39 eesti üliõpilast Tallinna ülikoolist) lugeda Eesti ja Soome suhteid käsitlevat sihtkeelset teksti (“Silta meren yli”, “Sild üle mere”), vastata küsimustele teksti kohta, tõlkida teksti nii palju, kui teadmised ja oskused võimaldavad, ning kirjeldada oma mõistmisprotsessi ja selles kasutatavaid strateegiaid (vt allpool 4. osa). Tekst oli tõlgitud Eesti ja Soome kontakte käsitlevast ingliskeelsest tekstist “Bridge over the sea” paralleelselt eesti ja soome keelde nii, et kahe keele konstruktsioonid oleksid võimalikult sarnased: näiteks kui ühes keeles väljendati tõlkides mõtet lauselühendi abil, siis toimiti samuti ka teises keeles, või kui ühes keeles kasutati kõrvallauset, siis tehti seda ka teises (Muikku-Werner 2013: 236). Oma tegevuse peegeldamisel olid osalejatele abiks küsimused:

1. Kuidas proovisid lahendada tekstist arusaamise probleeme, milliseid võtteid selleks kasutasid?
2. Mida oli tekstis kerge mõista?
3. Mida oli tekstis raske mõista?
4. Kas ja kuidas aitasid sind loetletud eesti ja soome sarnasused või erinevused?

Analüüsitava materjali hulgast jäeti välja kõik muu emakeelega kui soome ja eesti ning kakskeelsete üliõpilaste vastused. Kolmandikus testidest oli vastajatele abiks ülesandele lisatud eesti ja soome keele olulisemate sarnasuste loend, teises kolmandikus erinevuste loend, kolmandas kolmandikus testidest abimaterjal puudus. Tekstist arusaamist kontrolliti küsimuste abil (sulgudes on lisatud vastuse eest saadav maksimaalne punktide arv):

- K1. Millisel eesmärgil asutati tekstis mainitud instituut? (3)
- K2. Kui kaugel asuvad teineteisest Eesti ja Soome pealinnad? (2)
- K3. Mille poolest sarnaneb Ungari Eesti ja Soomega? (2)
- K4. Milliseid sarnaseid jooni leiab Eesti ja Soome ajaloost? (3)
- K5. Kuidas arenesid Eesti ja Soome suhted Eesti taasiseseisvumise järel? (3)
- K6. Millest tekstis veel juttu oli? (3)

Kuuest küsimusest viis on seega struktureeritud, tekstist leitavaid kindlaid vastuseid eeldavad, kuues küsimus võimaldab vastajale vabamat valikut otsustamiseks, mida mõistetust ta esile tuua tahab. Küsimuste eest saadud punktide summad on esitatud tabelis 2.

Tabel 2. Tekstimõistmisküsimuste eest saadud summaarne punktide arv

Osalejate rühm	Küsimus						Kokku
	K1	K2	K3	K4	K5	K6	
Eesti üliõpilased	43	51	37,5	20	30	33	214,5
Soome üliõpilased	80	67	43	27,5	57	28,5	305
Võimalik maksimaalne punktide arv	117	78	73	117	117	117	619

Tabelis 2 esitatud punktisummad väljendavad osalejate tekstimõistmise summaarset määra. Niisiis osutavad testi tulemused, et soome üliõpilased on tekstist paremini aru saanud nii kokkuvõttes kui ka viie üksikküsimuse lõikes. Eriti suur on eesti ja soome üliõpilaste punktisummade erinevus esimese ja viienda küsimuse osas. Eesti üliõpilased seevastu on olnud edukamad testi ainsale avatud küsimusele vastates.

Seega andsid üksiksõnade ja tekstimõistmise uurimused lähedase sugulaskeele mõistmise sümmeetria osas erinevad tulemused vastuse taustategurite sarnasusest hoolimata. Järgmises osas käsitletakse osalejate eneserefleksiooni alusel nende mõistmisprotsesse ja -strateegiaid ning püütakse leida tulemustele seletusi.

4. Osalejate mõistmisprotsess ja -strateegiad: keelelised tegurid

Strateegiate all mõistetakse tavaliselt teadvustatud toimimisviisi mõne ülesande lahendamiseks (Oxford 1990). Toimija ei tarvitse ilmtingimata olla toimimisprotsessi enda jaoks eksplitsiitselt sõnastanud, kuid vastuseks küsimusele selle kohta oskab ta kirjeldada, mida ta lahenduse leidmiseks ette võttis ja kuidas toimis. Sellisel juhul on tegemist potentsiaalselt teadvustatud strateegiaga (Martin 1995: 267). Käsitletavate uurimuste sihtrühmad osutusid mõistmisprotsessi ja -strateegiate sõnastamisel märkimisväärselt heterogeenseteks: nii eesti- kui ka soomekeelsete osalejate hulgas oli nii neid, kes kirjeldasid ja põhjendasid oma tegevust üksikasjaliselt, kui ka neid, kes üldse mingeid kommentaare ei esitanud. Osalejad on kirjeldanud tegureid, mis sihtkeele mõistmisele kaasa aitasid: keeltevaheline sarnasus ja selle tunnetamine, metalingvistilised ja üldteadmised, aga ka varasemad juhuslikud teadmised sihtkeelest, oletamine ja kontekst. Järgnevalt käsitletakse mõistmisprotsessi refleksiooni üksikasjalikumalt nii üksiksõnatesti kui ka tervikteksti põhjal. Kuna tervikteksti mõistmist oli peegeldatud ulatuslikumalt, on sellel analüüsis põhirõhk.

4.1. Keeltevaheline sarnasus

4.1.1. Emakeel ja selle variandid ning sihtkeel

Sihtkeele loomes, eriti aga selle mõistmisel tugineb keeleõppija või -kasutaja olemasolevatele teadmistele ja oskustele. Håkan Ringbomi (2007) sõnul tajub keeleõppija mõistmisprotsessis sarnasusi siis, kui vastuvõetav info ja tema varasemad teadmised omavahel seostuvad. Kui teadmised sihtkeelest puuduvad või on äärmiselt piiratud, on peamine allikas sarnasuste otsimisel emakeel või mõni muu varem õpitud keel (Ringbom, Jarvis 2009), mis mõlemad võivad toimida sihtkeele suhtes lähtekeelena. Mõni muu varem omandatud keel on lähtekeeleks eriti juhul, kui see on sihtkeelele lähedane sugulaskeel, keelekasutaja emakeel aga ei ole (Ringbom 2007: 2, Nissilä 2011). Lähte- ja sihtkeele sarnasuse tajumine on sihtkeele mõistmise peamiseks eelduseks, tänu millele suudetaksegi mingit osa sihtkeelest mõista ka ilma eelneva õppeta, eriti kui keelte geneetiline ja/või tüpoloogiline taust on sarnane. Vähesed varasemad uurimusedki (Kaivapalu, Muikku-Werner 2010, Paajanen, Muikku-Werner 2012) osutavad, et peamise strateegiana on sihtkeelest tõlkimisel kasutatud sarnasuste otsimist ennekõike emakeelest, seejärel muudest omandatud keeltest.

Soome ja eesti üksiksõnade mõistmise testides osutuski valdavaks strateegiaks sarnasuste otsimine oma emakeelest. Nii mainiti Margi (2014) uurimuses põhjendusena soome sõnade sarnasust eesti keele omadega kokku 43 testisõna puhul, millest 14 olid sama tähendusega ja viis semantiliselt lähedased. Ülejäänud 24 olid küll sarnase kirjapildiga, kuid erineva tähendusega ja seetõttu juhatasid informante eksiteele. Soome õpilased on maininud eesti testisõna sarnasust soome keelega eriti ulatuslikult sõnade *tädi* 'täti', *suvi* 'kesä' ja *naaber* 'naapuri' puhul. Vastajad olid ka märganud süstemaatilisi suhteid eesti ja soome sõnade vahel ning mainisid seetõttu strateegiana tähtede lisamist ja ärajätmist, näiteks *d* ärajätmist

sõnas *paljastada* ‘paljastaa’. Sama strateegia kasutamine võis samas viia ka väärade lahenduseni, näiteks sõnas *vili* ‘vilja’ veel ühe *l*-i lisamine andis soome keeles tulemuseks *villi* ‘metsik’ (Muikku-Werner, Heinonen 2012: 174). Mõistmise tulemuslikkuse mõttes on niisiis emakeele ja lähedase sugulaskeele sarnasusele toetumise võtmeküsimuseks teadmine (või teadmatuse), millal sarnase sõnavormiga kaasneb ka lähte- ja sihtkeeles sama tähendus ning millal mitte. Emakeele ja sihtkeele sarnasusele toetumise kasutamist mõistmisstrateegiana näitab kaudselt ka üksikute testisõnade mõistmise analüüs. Nii näiteks on kirjapildilt sarnane samatähenduslik sõna *kurssi* ‘kurss’ olnud mõistetav kõikidele soome õpilastele ja kõikidele peale ühe eesti õpilastest. Seevastu riskisõna *lahja*, mis soome keeles tähendab kingitust, eesti keeles vähest rasvasisaldust, kõhna, kondist, nappi või puudulikku, ei ole keegi õigesti mõistnud. Alati ei ole aga riskisõna mõistmine sümmeetriline: sõna *kalju* ei ole õigesti mõistnud ükski eesti õpilane, küll aga 25% soome õpilastest. Siin mängib rolli keeltevaheline tegeliku sarnasuse määr: sõnapaari *kalju – kallio* ühisosa on suurem kui sõnapaaril *kalju – kiilas(päine)* (vt Levenshtein Distance). Ka sihtkeele sarnasust emakeele slängi- ja murdesõnadega olid vastajad kasutanud tähenduse mõistmisel: näiteks eesti sõnale *öökida* ‘oksenta’ lähedane sõna on kasutusel soome slängis, eesti sõna *juua* ‘juoda’ kohta väitsid soome vastajad, et kasutavad seda ise kõnekeeles (Heinonen 2014: 41), *itkeä* ‘nutta’ oli tuttav Võru murde kaudu (Mark 2014: 36).

Terviktekstist arusaamisel lähtusid uurimuses osalejad veelgi enam kui üksik-sõnatestis eesti ja soome keele sarnasusest, nagu on näha näidetest (1)–(6).

- (1) Kerge on mõista sarnaseid sõnu (mida esines võrdlemisi palju, nt “kaarsilta”, “kulttuurisuhteet” jne). Lauseehitus on sarnane, isegi kiillauseid pole väga raske eraldada, samuti on arusaadavad paljud käänded.
- (2) Suomen kielen sanoja muistuttavat sanat, joissa tavurakenne ja taivutukset olivat suunnilleen samat kuin suomen kielessä tai ainakin lähentelivät suomen kielen rakenteita.

Nii nagu mõnes varasemas uurimuses (Kaivapalu, Muikku-Werner 2010), ei toetu osalejad mitte ainult sõnavara, vaid ka keelte struktuuri sarnasusele: emakeelega sarnane silbistruktuur, sõnamuutmine, ajavorm ja lauseehitus hõlbustavad teksti mõistmist. Võõr- ja laensõnu tunnetatakse sageli emakeele sõnadena ega mainita laenuandjat võõrkeelt. Eriti märgatav on see suundumus eesti informantidel, kes mainivad teisi võõrkeeli soomlastest vähem:

- (3) [Lihtne oli mõista] sarnaseid sõnu eesti keelest nagu kultuur jne
- (4) Kerge oli ka mõista võõrsõnu, mis on kasutusel eesti keeles.

Soome vastajate olulisim erinevus võrreldes eesti vastajatega oli märkimisväärselt ulatuslikum soome geograafiliste ja sotsiaalsete murrete, kõnekeele ja arhailise keele kasutamine eesti keele mõistmisel.

- (5) [Yritin ymmärtää] sitomalla sanoja vanhaan suomeen tai sanantoihin (esim. *Soome lahe mõlemal kaldal* → kaldal → kaidal)
- (6) [Auttoivat] Suomen puhekieli ja Turun murre

Kui üksiksõnatestis mainisid murdeid ja kõnekeelt arusaamise allikana vaid mõned soome vastajad ning murdeid mõned eesti vastajad, siis tekstimõistmise

kommentaaries ei maininud eesti vastajad emakeele varieerumist üldse, soome vastajatest seevastu aga peaaegu kõik. Seega avaldab soome keele eesti keelest märksa mitmekesisem varieerumine (vt nt Metslang 1994) mõju lähedase sugulaskeele mõistmise tulemustele: soomekeelne testi sooritaja on harjunud nägema seoseid eri keelevariantide vahel ja tema käsutuses on seetõttu enam vahendeid seoste loomiseks sihtkeelega. Emakeele ulatuslikum varieerumine seletab soome õpilaste edukust üksiksõnatesti kauses tähendusseoses olevate sõnade mõistmisel. Samuti on antud tulemus kooskõlas varasemate uurimustega saksa ja šveitsisaksa keelekasutajate rootsi keele mõistmisest (Berthele 2008) ning rootsi, taani ja norra keele omavahelisest mõistmisest (Gooskens 2004): šveitsisaksa ja norra keelekasutajad osutasid edukamateks lähedase sugulaskeele mõistjateks just emakeele ulatuslikuma geograafilise varieerumise tõttu.

4.1.2. Varem õpitud keel ja sihtkeel

Kui vasteid sihtkeelsetele konstruktsioonidele emakeele ja sihtkeele sarnasuse alusel ei leitud, otsisid vastajad abi teistest varem õpitud keeltest, eriti juhul, kui oli tegemist laen- või võõrsõnadega. Üksiksõnatestis toetusid soome õpilased esmajoones inglise (nt *privileeg* 'etuoikeus', *biifsteek* 'pihvi', *poksida* 'nyrkkeillä') ja rootsi keelele (*röövida* 'ryöstää, ryövätä, rosvota', rootsi *röva*) (Heinonen 2014: 33–35). Seejuures võis sarnasuse tunnetamine põhineda mitte ainult täpsel tõlkevastel varem õpitud keeles, vaid ka mõnel muul assotsiatsioone tekitanud sõnal. Näiteks üks soome vastajatest märkis sõna *röövida* tõlkevasteks *punastua* 'punastada', lähtudes rootsi sõnast *röd* 'punane', teine *piirtää* 'joonistada' rootsi sõna *rita* 'joonistada' põhjal. Viis vastajat olid toetunud rootsi sõnale *röka* või rootsi algupära slängisõnale *röökata* 'suitsetada'. Õige lahendus saadi tõenäolisemalt siis, kui mõne võõrkeelse sõnaga sarnane testisõna esines laensõnana ka soome (kõne)keeles või slängis, nagu eesti sõna *lähifoto* puhul, vrd inglise *photo*, rootsi *ett foto*, saksa *das Foto*, soome kõnekeelne *foto* (Muikku-Werner, Heinonen 2012: 174). Niisiis on sihtkeelse sõna mõistmise võimalus seda suurem, mida mitmekesisemad on seoste tekkimise võimalused.

Eesti õpilased (Mark 2014) otsisid sarnasusi valdavalt enim õpitud võõrkeelest, inglise keelest, mida tehti 11 vaste puhul. Viiel korral jõuti ka õige tähenduseni. Sarnasuse tunnetamine ja sobiva mõistmisstrateegia rakendamine ei tarvitse aga alati õige tõlkevasteni viia, eriti juhul, kui lähtekeele (inglise keele) sõna on mitmetähenduslik. Nii on sõna *halpahintainen* puhul vastajatel tekkinud seosed mitme ingliskeelse sõnaga: vasteteks pakuti näiteks *poolaeg* 'half-time' ja *tähestik* 'alphabet' ning sõna *hint* 'vihje' põhjal *halb eelaimdus*, *halvaendeline* ja *halba aimama*. Mõne sõna puhul, näiteks *historia*, mis oli mõistetav kõikidele testis osalenutele, osutab sarnasusele toetumist õigesti vastanute suur arv, kuigi selleni jõudmise strateegiat on kirjeldanud vaid üksikud. Eesti õpilased otsisid abi ka vene keelest. Õige tähenduseni jõuti vene keelt kasutades vastete *musiikki* 'muusika', *historia* 'ajalugu' ja *komedial* 'komöödia' puhul, eksiti aga vastele *symboloida* 'sümboliseerida' pakutud vastega *sümbol*. Sarnasust kreeka ja ladina keelega leiti kolme vaste puhul, millest kaks (*komedial* 'komöödia' ja *politiikka* 'poliitika') olid õiged. Vaste *komedial* 'komöödia' õige tähenduseni jõuti isegi hispaania keele kaudu. Kuigi

osalejad polnud varem õppinud ei kreeka ega ladina keelt, on nad siiski nende keeltega emakeele võõrsõnade kaudu piisavalt kokku puutunud, et neid kasutada mõne teise võõrkeele mõistmisel. Eesti õpilaste tulemuslikum toetumine keeltevahelisele sarnasusele üksiksõnatestis võibki olla muuhulgas seletatav varem õpitud võõrkeelte suurema mitmekesisusega.

Nii eesti kui ka soome üksiksõnatestis esines võõrkeeltest abi otsimist sellistegi sõnade puhul, kus oleks võinud toetuda emakeelele. Nõnda on pakutud soome sõna *halpahintainen* vasteteks *halb eelaimdus*, *halvaendeline* ja *halba aimama* ingliskeelse sõna *hint* 'vihje' põhjal, eesti sõna *tädi* 'täti' vasteks aga *isä* 'isa' ingliskeelse sõna *daddy* põhjal (Muikku-Werner, Heinonen 2012: 163). Kui ühes sihtkeeles on tegemist vastajatele läbinähtamatu omasõnaga (ee *ajalugu*), millele ei ole sarnaseid sõnu õpitud võõrkeeltes, ja teises sihtkeeles varem õpitud võõrkeelte põhjal mõistetava sõnaga (sm *historia*), kajastub see ka testitulemustes: *ajalugu* polnud mõistetav ühelegi soome õpilasele, *historia* seevastu oli arusaadav kõikidele eesti õpilastele.

Ka tervikteksti mõistmisel toetuti sihtkeele sarnasusele varem õpitud keeltega.

- (7) Proovisin leida sarnasusi õpitud keeltest.
- (8) Myös muiden kielten osaamisesta oli hyötyä, esim. sana “direktor” on tuttu muista kielistä.

Erinevalt üksiksõnatestist nimetati seejuures vähem konkreetseid keeli, pigem mainiti üldisemalt teisi keeli, võõrsõnu, rahvusvahelisi sõnu (*logo*, *direktor*), riikide nimesid. Eesti vastajad ei maininud konkreetseid keeli üldse, soome vastajad toetusid rootsi, inglise ja vene keelele:

- (9) Yritin hyödyntää suomen murteiden tuntemusta, ruotsin kieltä ja venäjää. Yritin nähdä kielen enemmän “vieraana kielenä” kuin “suomen kaltaisena kielenä”.
- (10) Indoeurooppalaiset lainat, kuten *informatsiooni*, *symboliseerida*, *instituut* ja *organiseeriti* [olivat ymmärrettäviä]

Antud erinevus eestlaste ja soomlaste tekstimõistmisstrateegiates võib olla seletatav võõrsõnade suurema osakaaluga eesti keeles; seevastu soomekeelsest tekstist arusaamine ei tarvitse nõudagi ulatuslikumat sarnasuste otsimist varem õpitud võõrkeelest seetõttu, et soome keeles on omasõnu suhteliselt rohkem. Soomekeelses üksiksõnatestis, milles esines ka võõr- ja laensõnu, toetusid eesti õpilasedki testisõnade ja tuttavate võõrkeelsete sõnade sarnasusele.

4.2. Metalingvistilised teadmised

Keeltevaheliste sarnasuste tunnetamine on seotud uurimuses osalejate metalingvistilise teadlikkusega: teadmistega inimkeelest kui sellisest ja selle toimimisest ning teadmistega eriti emakeelest, aga ka teistest varem omandatud keeltest ja sihtkeelest. Tervikteksti mõistmise uurimuses küsiti osalejatelt, kas neid aitasid ülesandele lisatud eesti ja soome keele sarnasuste ja erinevuste loendid. Vastused kinnitasid väiteid sarnasuse primaarsusest sihtkeele mõistmisel ja keeleloomes (Ringbom 2007): kui teadmised keeltevahelise sarnasuse kohta aitasid reeglina teksti mõistmisele kaasa (11), (12), siis erinevuste olulisust ei kommenteeritud.

- (11) [Sarnasuste loetelu aitas] väga palju, kuna olen juba natuke süsteemist aru saanud, et mismoodi need soome ja eesti sõnad sarnanevad (nt *sild – silta*)
- (12) Loetletud sarnasused aitasid mind palju. Nende kaudu seostasin soome keelt palju eesti keelega.

Eesti ja soome keele sarnasuste loendist ei olnud abi vastajatele, kes olid neid seaduspärasusi mingil põhjusel juba varem ise märganud või siis ei suutnud oma tähelepanu ajapuudusel kahe tegevuse vahel jagada:

- (13) Loetelu ei aidanud mind lihtsalt seetõttu, et olen samu seoseid märganud iseseisvalt (nii varasemalt kui ka praegust teksti lugedes).
- (14) Teksti lugedes nendele sarnasustele lihtsalt ei suuda mõelda, kuna suurem tähelepanu koondub teksti üldisele tõlkimisele.

Kahe keele vaheliste süstemaatiliste seaduspärasuste otsimisele viidatakse rohkelt ka mõistmisstrategiate kirjeldamisel:

- (15) Yritin myös ymmärtää uusia sanoja tarkkailemalla äännevastavuuksia.
- (16) [Yritin ymmärtää tekstiä] äänneiden muutoksilla” → kokeilemalla saisi-ko vaikeasti avatuista sanoista ymmärrettävä esim. muuttamalla pitkiä vokaaleja diftongeiksi tms.
- (17) Muuttamalla joitain soinnillisia klusiileja soinnittomiksi lukiessa, esim. *d*-kirjainten lukeminen *t* : nä

Kui keeltevahelist sarnasust, näiteks klusiilide puhul, ei teatud või tunnetati seda erinevusena, osutus sihtkeelse teksti mõistmine keerulisemaks:

- (18) [Hankala oli] kirjoitusasun erilaisuus *g*, *b* ja *d*.

Sama selgus ka üksiksõnatestist. Näiteks oli enamik eesti vastajatest märkinud sõna *särki* ‘särg’ tõlkevasteks *särk* (Mark 2014), sõna *särk* ‘paita’ sai aga suuremalt osalt soome vastajatelt tähenduseks *särki* ‘särg’ (Muikku-Werner, Heinonen 2012: 168).

Selline tulemus pakub rohkesti mõtlemisainet eesti ja soome keele õpetamise osas. Traditsiooniliselt on lähedaste sugulaskeelte (eriti võõrkeeltena) õpetamisel tavaks rõhutada vigade vähendamise eesmärgil keeltevahelisi erinevusi ja oletada, et õppija märkab keeltevahelisi sarnasusi ise, halvemal juhul aga ei peeta sarnasusi lihtsalt oluliseks. Eesti ja soome keele mõistmist käsitlevad uurimused seevastu näitavad taas kord, et tulemusliku sihtkeele töötluse lähtekoht on keeltevaheline sarnasus: sama vastavuse puhul võib otsustavaks saada asjaolu, kas seda tunnetatakse sarnasuse või erinevusena.

4.3. Üldteadmised

Keeleoskus, sealhulgas keelest arusaamine on tihedalt seotud üldteadmistega ümbritseva maailma kohta; ilma selleta ei ole keelekasutus võimalik. Üldteadmistele tuginetakse ka sihtkeele mõistmisel: tuttava valdkonna korral on tekstist lihtsam aru saada. Tervikteksti mõistmise testis osalejad toetusid peamiselt ajaloo- ja geograafiateadmistele, samuti teati tekstis kirjeldatud keelte sugulussuhteid (19), (20).

- (19) Lauseid, mille sisu kohta oli olemas juba mingisugune eelnev teave – näiteks lause ühte keelenterühma kuulumise kohta.
- (20) Tekstin alku oli suhteelisen ymmärrettävää, etenkin sukukieli-osio, sillä se on itselle tuttu aihe.

Mitmetele vastajatele mõlemal pool Soome lahte osutus siiski keeruliseks Nõukogude Liidu ajalooa seonduv ja see asjaolu takistas teksti mõistmist:

- (21) [Raske oli mõista] juttu Nõukogude Liidust.
- (22) Viimeinen kappale oli vaikea ymmärtää ja lauseet, joista kerrotaan Neuvostoliitosta.

Üksiksõnatesti spetsiifika ja kommentaaride vähesuse tõttu ei leia üldteadmistele toetumine selles nii ulatuslikku kajastust kui tervikteksti mõistmist käsitlevas uurimuses, kuid üksikuid viiteid siiski esineb. Nii pakkus spordihuvilisest vastaja testisõna *kuppi* 'tass, kruus, karikas' vasteks *turniir*, sest inglise *cup* esineb mitmete spordialade karikaturniiride (nt *FIFA world cup*) nimetustes (Mark 2014).

4.4. Varasemad juhuslikud teadmised ja oletamine

Kuigi uurimuses osalejad ei olnud eelnevalt õppinud vastavalt eesti või soome keelt ja neil puudus sellega püsivam kokkupuude, on osal vastajatest olnud siiski naabermaa keelest juhuslikke varasemaid teadmisi. Varasemate teadmiste allikatena nimetati näiteks televisiooni, turismireise ja muid lühemaid Soomes või Eestis viibimisi, Soomes või Eestis töötavaid lähedasi ja sõpru jne, aga ka seda, et "on kuskilt meelde jäänud". Nii mainisid eesti vastajad varasemaid juhuslikke teadmisi sõnade *käteinen* 'sularaha', *väkiluku* 'rahvaarv', *itkeä* 'nutma', *sateenkaari* 'vikerkaar', *kuolla* 'surra', *kasvis* 'köögivili', *kuppi* 'tass', *ystävä* 'sõber' ja *lukea* 'lugema' tõlkimisel, kusjuures viimase puhul oli informant olnud teadlik, et eesti *g*-tähele vastab soome keeles *k*-täht (Mark 2014). Üks soome vastajatest mäletas, et oli näinud sõna *joogivesi* veepudelil ja tõlkis selle seetõttu õigesti (Muikku-Werner, Heinonen 2012: 175). Ka mõned tervikteksti mõistmise uurimuses osalejad tõid välja juhuslikud varasemad teadmised sihtkeelest:

- (23) [Kerge oli mõista] sarnaseid sõnu eesti keelega ja neid sõnu, mida tava-keeles ka soomlased palju kasutavad ja mida seetõttu tean.

Kui tunnetatud sarnasus ja teadmised mõistmisele kaasa ei aidanud, püüdsid vastajad oletada, millega võiks tegemist olla:

- (24) Oletasin, mida tahetakse öelda, kui sarnasuste ja teadmiste põhjal ei mõistnud.

Üksiksõnatestis kasutati oletamist näiteks sõna *joogivesi* puhul ja tõlgiti see soome keelde sõnana *juomavesi*. Sellele, et tegemist on oletamisega, viitavad ka küsimärgid ja emotikonid vastuse järel, nagu sõna *sularaha* 'käteinen' puhul: *kulta ...?, joku raha?* (Heinonen, Muikku-Werner 2012: 157). Eesti informandid kasutasid oletamist (nimetades seda loogilisuseks) näiteks sõnade *leijona* 'lövi', *käteinen* 'sularaha' ja *johtaja* 'juhataja' puhul. Esimesel juhul osutus oletus õigeks, teisel

juhul pakuti tähenduseks ekslikult 'tervitama' ja kolmandal juhul lähedast tähendust 'juht' (Mark 2014). Viimasel juhul põhineb oletus keeltevahelisel tunnetatud sarnasusel.

4.5. Kontekst ja sisuseosed

Nagu varasemates uurimustes (nt Kaivapalu, Muikku-Werner 2010, Muikku-Werner jt 2012), osutus ka nüüd tervikteksti mõistmisel äärmiselt oluliseks lause-, lõigu- ja tekstikontekst ning tekstisisesed seosed:

- (25) Püüdsin leida tundmatute sõnade kordusi erinevas kontekstis ja mõista seeläbi nende tähendust.
- (26) Yritin tulkida teksti kontekstissaan, ja hyödynsinkin asiayhteyttä monessa kohdassa.
- (27) Yritin etsiä ymmärtämättömien sanojen läheltä sanoja joita ymmärrän ja miettiä sitten miten sanat voisivat limittyä yhteen järkeväksi kokonaisuudeksi.

Kontekst toimib niisiis mõistmisprotsessi omalaadse filtrina, monitorina, mille vahendusel kontrollitakse keeltevahelise sarnasuse, metalingvistiliste teadmiste ja üldteadmiste paikapidavust. Mõistmisprotsessi komponente ja dünaamikat käsitletakse lähemalt järgmises osas.

5. Mõistmisprotsess

Uurimuses osalejate kommentaarid osutavad, et sihtkeele mõistmise komponendid, mida eespool analüüsisin, on osalejate individuaalses mõistmisprotsessis läbi põimunud ja toimivad interaktsioonis. Osa vastajatest lähtus esmajoones eesti ja soome keele sarnasusest ja püüdis sellele tuginedes ümbritsevast kontekstist aru saada:

- (28) Püüdsin võimalikult paljudele sõnadele leida eestikeelset tähendust. Nendele sõnadele, millele ei leidnud, püüdsin tähenduse leida lause mõttest arusaamisega. Seega tuletasin paljude sõnade tähenduse sellisel viisil.
- (29) Luin uudelleen, hain vihjeitä suomen kielestä, ehitin hankalat sanat ja "täytin" niiden jättämät aukot arvauksella, hain sanaa helpommasta lauseympäristöstä, hyödynsin ennakkotietouttani esim. kielisuhteista.

Mõned liikusid mõistmisprotsessis keelendite sarnasuselt konteksti mõistmisele, toetusid sarnasusele laiemalt, sh sihtkeele ja varem omandatud keele sarnasusele, ja võisid jätta sarnasuse allika selgelt määratlemata:

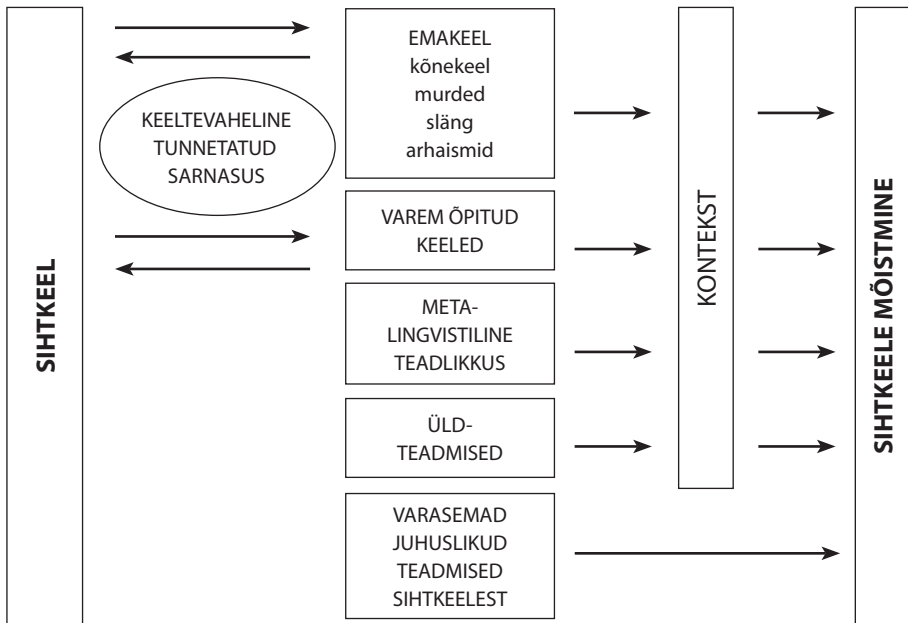
- (30) Keskendusin neile sõnadele, millest aru sain. Kuna osa sõnu on sarnased, siis on võimalik teksti konteksti n-ö ette aimata. Samuti on mulle lihtsamad sõnad teada.
- (31) Vertaamalla ensin sanaa mieleen tulevin synonymeihin, vaihtelemalla vokaaleja, palaamalla tekstissä toiseen kohtaan, jossa sana esiintyi, rinnastamalla joitain sanoja ja lauserakenteita ruotsiin. Osan arvasin pohtimalla vain lauseyhteyttä.

Vastustest nähtub seega, et tunnetatud sarnasust rakendatakse kombinatsioonis oletamise, üldteadmiste, metalingvistiliste ja varasemate teadmistega sihtkeelest.

Teise osa informantide mõistmisprotsessi suund oli eelkirjeldatule vastupidine: lähtuti kontekstist ja püüti sellest leida emakeelega sarnaseid sõnu:

- (32) Lugesin lauset mitu korda, üritades lause tervikust tähendusest aru saada ja kasutades tuttavaid sõnu, mida juba teadsin. Vaatasin lõike tervikuna, et tekstist aru saada. Võrdlesin sõnu eestikeelsetega, et äkki näkkab.
- (33) Päättelin sana lauseyhteyden perusteella ja vertasin niitä suomen samankaltaisiin sanoihin ja mietin, oliko yhtäläisyys järkevää kussakin yhteydessä.

Seega avalduvad sihtkeele mõistmisprotsessis keelekasutajate individuaalsed iseärasused kahel kujul: ühelt poolt lähenemissuunas (üksikult üldisele või üldiselt üksikule), teisalt kombineeruvad sarnasuse tunnetamine ja erinevat tüüpi varasemad teadmised-kogemused vastavalt olemasolevale pagasile, mis osalejati varieerub. Kokkuvõtlikult väljendab lähedase sugulaskeele mõistmisprotsessi järgmine mudel (joonis 1).



Joonis 1. Lähedase sugulaskeele mõistmise mudel

6. Kokkuvõtteks

Nii üksiksõna- kui ka tekstitasandi mõistmise uurimused osutasid, et eesti- ja soomekeelsed keelekasutajad mõistavad lähedast sugulaskeelt arvestataval määral – sõltuvalt keelevahelise sarnasuse astmest kogumahust kolmandiku kuni kahe kolmandiku ulatuses – ka ilma sihtkeele õppimise ja kasutamise kogemusega. Mõistmisprotsessis toetutakse sihtkeele tunnetatud sarnasusele peamiselt emakeelega, aga ka muude varem õpitud keeltega; seostatakse seda metalingvistiliste ja üldteadmistega ning tervikteksti puhul kontrollitakse oma oletusi lause-, lõigu- ja tekstikontekstis. Kontekst toimib omalaadse filtrina, mis selekteerib keelekasutaja oletusi kas aktsepteerides või tagasi lükates. Mõistmisprotsessi komponentide jaotus ja suund on indiviiditi küllalt varieeruv ja sõltub paljuski osalejate varasematest teadmistest ja kogemustest.

Eestlaste soome keele ja soomlaste eesti keele mõistmine ei ole siiski sümmeetriline, vaid erineb üksiksõnade ja tervikteksti osas. Kui eesti õpilased olid soome õpilastest üksiksõnatestis edukamad tervikuna, eesti ja soome keeles kirjapildilt ja tähenduselt sarnaste sõnade ning laen- ja võõrsõnade mõistmisel, siis soome õpilased mõistsid paremini kaudses tähenduseseoses esinevaid sõnu, samuti mõistsid soomekeelsed uurimused osalejad paremini tervikteksti.

Nende tulemuste tõlgendamine vajab kindlasti edasist uurimist. On võimalik, et eesti õpilaste paremale laen- ja võõrsõnade mõistmisele üksiksõnatestis aitab kaasa nende sõnade suurem osakaal eesti keeles soome keelega võrreldes, samuti eesti õpilaste varem õpitud võõrkeelte suurem mitmekesisus. Soome õpilaste edu kaudses tähenduseseoses olevate sõnade mõistmisel võib muu hulgas tingida soome keele laiem geograafiline, sotsiaalne ja ajalooline varieerumine, mis aitab kaasa leksikaalsete seoste nägemisele sõnade vahel. Kuna nii eesti kui ka soome üksiksõnatestis osalejad kommenteerisid oma mõistmisprotsessi üsna tagasihoidlikult, kinnitab osalejate refleksioon neid selgitusi ainult osaliselt. Tervikteksti mõistmise testis osalejad seevastu tõid väga selgelt välja toetumise mitte ainult soome üldkeelele, vaid ka kõnekeelele, geograafilistele ja sotsiaalsetele murretele, vanale soome keelele.

Valdav osa retseptiivse mitmekeelsuse uurimisest on keskendunud üksiksõnade või siis tervikteksti mõistmisele neid omavahel seostamata. Käesolevas artiklis kirjeldatud uurimistulemusi on edaspidi kavas kasutada lähtekohana mõistmisprotsessi holistilisel uurimisel. Võtmeküsimuseks seejuures on, miks ja kuidas keelevahelise sarnasuse tunnetamine realiseerub või ei realiseeru mõistmisena ja millised tegurid sellele kaasa aitavad.

Viidatud kirjandus

- Bachman, Lyle F. 1990. *Fundamental Considerations in Language Testing*. Oxford: OUP.
- Berthele, Raphael 2008. *Dialekt-Standard Situationen als embryonale Mehrsprachigkeit. Erkenntnisse zum interlingualen Potenzial des Provinzlerdaseins*. – *Sociolinguistica*, 22, 87–107.
- Bezooijen, Renée van; Gooskens, Charlotte 2007. *Interlingual text comprehension: Linguistic and extralinguistic determinants*. – Jan D. ten Thije, Ludger Zeevaert (Eds.). *Receptive Multilingualism: Linguistic Analyses, Language Policies, and Didactic Concepts*. Amsterdam: John Benjamins, 249–264.

- EKR = Euroopa keeleõppe raamdokument: õppimine, õpetamine, hindamine. [Common European Framework of Reference for Languages: Learning, teaching and assessment.] Tartu: Haridus- ja Teadusministeerium.
- Gooskens, Charlotte 2004. Norwegian dialect distances geographically explained. – Britt-Louise Gunnarson et al. (Eds.). *Language Variation in Europe. Papers from the Second International Conference on Language Variation in Europe, ICLAVE 2*, Uppsala University, Sweden, June 12–14, 2003. Uppsala: Uppsala University, 195–206.
- Gooskens, Charlotte 2007. Contact, attitude and phonetic distance as predictors of inter-Scandinavian communication. – Jean-Michel Eloy, Tadhg ÓhLfeárnáin (Eds.). *Near Languages – Collateral Languages. Actes du colloque international réuni à Limerick, du 16 au 18 juin 2005*, 99–109.
- Gooskens, Charlotte; Heeringa, Wilbert 2006. The relative contribution of pronunciation, lexical and prosodic differences to the perceived distances between Norwegian dialects. – *Literary and Linguistic Computing*, 21 (4), 477–492. <http://dx.doi.org/10.1093/lc/fql038>
- Gooskens, Charlotte; Heeringa, Wilbert 2014. The role of dialect exposure in receptive multilingualism. – *Applied Linguistics Review*, 5 (1), 247–271. <http://dx.doi.org/10.1515/applirev-2014-0011>
- Gooskens, Charlotte; Heuven, Vincent van; Bezooijen, Renée van; Pacilly, Jos 2010. Is spoken Danish less intelligible than Swedish? – *Speech Communication*, 52 (11–12), 1022–1037. <http://dx.doi.org/10.1016/j.specom.2010.06.005>
- Gooskens Charlotte; Hilton, Nanna Haug 2010. The effect of social factors on the comprehension of a closely related language. – Jani-Matti Tirkkone, Esa Anttikoski (Eds.). *Proceedings of the 24th Scandinavian Conference of Linguistics. Reports and Studies in Education, Humanities, and Theology 5*. Joensuu: University of Eastern Finland, 201–210.
- Gooskens, Charlotte; Kürschner, Sebastian 2009. Cross border intelligibility – on the intelligibility of Low German among speakers of Danish and Dutch. – Alexandra N. Lenz, Charlotte Gooskens, Siemon Reker (Eds.). *Low Saxon Dialects Across Borders – Niedersächsische Dialecte über Grenzen hinweg. Zeitschrift für Dialektologie und Linguistik, Beihefte 138*. Franz Steiner Verlag, 273–297.
- Heeringa, Wilbert; Gooskens, Charlotte; Smedt, Koenraad de 2008. What role does dialect knowledge play in the perception of linguistic distances? – *International Journal of Humanities and Arts Computing*, 1–2, 243–259.
- Heinonen, Maria 2011. “Viineri tuli mieleen, mutta ihan kuin oisin kuullu että nakki?”. Kyselytutkimus eräiden viron sanojen ymmärtämisestä suomen kielen sanojen pohjalta. Kandidaatintutkielma. Itä-Suomen yliopisto.
- Heinonen, Maria 2014. “Koska Virossa kaikki on nurinkurin”. Kyselytutkimus eräiden luetujen ja kuultujen vironkielisten sanojen ymmärtämisestä suomen kielen pohjalta. Suomen kielen pro gradu -tutkielma. Itä-Suomen yliopisto.
- Hilton, Nanna Haug; Gooskens, Charlotte 2013. Language policies and attitudes towards Frisian in the Netherlands. – Charlotte Gooskens, Renée van Bezooijen (Eds.). *Phonetics in Europe: Perception and Production*. Frankfurt a.M.: Peter Lang, 139–157.
- Jarvis, Scott; Odlin, Terrence 2000. Morphological type, spatial reference, and language transfer. – *Studies in Second Language Acquisition*, 22 (4), 535–556. <http://dx.doi.org/10.1017/S0272263100004034>
- Kaivapalu, Annekatrin; Martin, Maisa 2014. Measuring perceptions of cross-linguistic similarity between closely related languages: Finnish and Estonian noun morphology as a testing ground. – Heli Paulasto, Lea Meriläinen, Helka Riionheimo, Maria Kok (Eds.). *Language Contacts at the Crossroads of Disciplines*. Cambridge Scholars Publishing, 283–318.

- Kaivapalu, Annekatrin; Muikku-Werner, Pirkko 2010. Reseptiivinen monikielisyys: miten suomenkielinen oppija ymmärtää viroa äidinkiensä pohjalta? [Receptive multilingualism: How Finnish as a first language helps learners to understand Estonian?] – *Lähivertailuja. Lähivördlusi*, 20, 68–97. <http://dx.doi.org/10.5128/LV20.03>
- Kellerman, Eric 1979. Transfer and non-transfer: Where we are now. – *Studies in Second Language Acquisition*, 2 (1), 37–57. <http://dx.doi.org/10.1017/S0272263100000942>
- Klein, Horst G.; Stegmann, Tilbert D. 2000. *EuroComRom – Die sieben Siebe: Romanische Sprachen sofort lesen können*. Aachen: Shaker Verlag.
- Mark, Marta 2014. Eesti õpilaste soome üksiksõnade mõistmine. [Understanding Finnish words by the Estonian students.] *Bakalaureusetöö*. Tallinna Ülikooli eesti keele ja kultuuri instituut.
- Martin, Maisa 1995. The Map and the Rope: Finnish Nominal Inflection as a Learning Target. *Studia Philologica Jyväskyläensis* 38. Jyväskylä: Jyväskylän yliopisto.
- Martin, Maisa 2006. Suomi ja viro oppijan mielessä. Näkökulmia taivutusmuotojen prosessointiin. – Annekatrin Kaivapalu, Külvi Pruuli (Toim.). *Lähivördlusi. Lähivertailuja*, 17. Jyväskylä *Studies in Humanities* 53. Jyväskylä: University of Jyväskylä, 43–60.
- Metslang, Helle 1994. Temporal Relations in the Predicate and the Grammatical System of Estonian and Finnish. *Oulun yliopiston suomen ja saamen kielen tutkimusraportteja* 39. Oulu: University of Oulu.
- Muikku-Werner, Pirkko; Heinonen, Maria 2012. Lumesadu – ‘tarina’ vai ‘lumikasa’ vai ei kumpikaan? Suomalaiset lukiolaiset viro sanoja tunnistamassa. [“Lumesadu” - ‘tarina’ or ‘lumikasa’ or something completely different? How Finnish senior high school students try to recognise Estonian words.] – *Lähivertailuja. Lähivördlusi*, 22, 157–187. <http://dx.doi.org/10.5128/LV22.06>
- Muikku-Werner, Pirkko 2013. Vironkielisen tekstin ymmärtäminen suomen kielen pohjalta. [Understanding Estonian texts on a Finnish language base.] – *Lähivertailuja. Lähivördlusi*, 23, 210–237. <http://dx.doi.org/10.5128/LV23.09>
- Muikku-Werner, Pirkko; Kaivapalu, Annekatrin; Martin, Maisa 2012. Mutual intelligibility in language learning context: Linguistic and strategic determinants of comprehension of Estonian and Finnish. – Matthias Hüning, Uli Reich (Eds.). *Language and the City. Sociolinguistic Symposium 19*. Berlin, August 21–24, 2012. Berlin: Freie Universität Berlin, 266.
- Nissilä, Leena 2011. Viron kielen vaikutus suomen kielen verbien ja niiden rektioiden oppimiseen. [Impact of the Estonian language on the learning of Finnish verbs and their rections.] *Acta Universitatis Ouluensis B. Humaniora* 99. Oulu: Oulun yliopisto.
- Odlin, Terrence 1989. *Language Transfer. Cross-Linguistic Influence in Language Learning*. Cambridge, USA: Cambridge University Press. <http://dx.doi.org/10.1017/CBO9781139524537>
- Oxford, Rebecca 1990. *Language Learning Strategies. What Every Teacher Should Know*. New York: Newbury House.
- Ommen, Sandrien van; Hendriks, Petra; Gilbers, Dicky; Heuven, Vincent van; Gooskens, Charlotte 2013. Is diachronic lenition a factor in the asymmetry in intelligibility between Danish and Swedish? – *Lingua*, 137, 193–213. <http://dx.doi.org/10.1016/j.lingua.2013.09.009>
- Paajanen, Ilona; Muikku-Werner, Pirkko 2012. Tee on kitsas – onko ‘tee kitkerää’ vai oletteko ‘te saita’? Suomalaiset opiskelijat viroa ymmärtämässä. [“Tee on kitsas” – is ‘tea bitter’ or are you ‘penny-pinching’? Finnish students comprehending Estonian.] – *Lähivördlusi. Lähivertailuja*, 22, 219–258. <http://dx.doi.org/10.5128/LV22.08>
- Ringbom, Håkan 1987. *The Role of the First Language in Foreign Language Learning*. Clevedon: Multilingual Matters.

- Ringbom, Håkan 2007. Cross-Linguistic Similarity in Foreign Language Learning. Clevedon: Multilingual Matters LTD.
- Ringbom, Håkan; Jarvis, Scott 2009. The importance of cross-linguistic similarity in foreign language learning. – Michel H. Long, Catherine J. Doughty (Eds.). Handbook of Language Teaching. Oxford: Blackwell, 106–118.
- Schepens, Job; Slik, Frans van der; Hout, Roeland van 2013. The effect of linguistic distance across Indo-European mother tongues on learning Dutch as a second language. – Lars Borin, Anju Saxena (Eds.). Approaches to Measuring Linguistic Differences. Berlin: Mouton de Gruyter, 199–230.
- Zeevaert, Ludger; Thijs, Jan D. ten 2007. Introduction. – Jan D. ten Thijs, Ludger Zeevaert (Eds.). Receptive Multilingualism. Linguistic Analyses, Language Policies and Dialectic Concepts. Amsterdam: John Benjamins, 1–21.
- Thijs, Jan D. ten; Rehbein, Jochen; Verschik, Anna (Eds.) 2012. International Journal of Bilingualism, 16 (3). Special Issue: Receptive Multilingualism. <http://dx.doi.org/10.1177/1367006911426468>

Võrgumaterjalid

- Charlotte Gooskens. <http://www.let.rug.nl/gooskens/?p=project> (2.2.2015).
- EuroComDidact. <http://www.eurocomresearch.net/> (2.2.2015).
- Levenshtein Distance. <http://xlinux.nist.gov/dads//HTML/Levenshtein.html> (3.2.2015).
- Lähivõrdlusi. Lähivertailuja. <http://www.rakenduslingvistika.ee/ajakirjad/index.php/lahivordlusi> (2.2.2015).
- REMU. www.uef.fi/en/remu2015 (2.2.2015).
- White paper 1995 = White paper on education and training 1995. http://europa.eu/documents/comm/white_papers/pdf/com95_590_en.pdf (2.2.2015).

Annekatriin Kaivapalu (Tallinna Ülikool) uurimisvaldkonnad on sihtkeele õppimine ja omandamine, keeltevaheline mõju, keeltevahelise sarnasuse tunnetamine ja selle mõõtmine, lähisugulaskeele vastastikune mõistmine.

Tallinna Ülikooli eesti keele ja kultuuri instituut, Narva mnt 29, 10120 Tallinn, Estonia
kaivapa@tlu.ee

MUTUAL INTELLIGIBILITY OF ESTONIAN AND FINNISH CONTEXT-FREE WORDS AND TEXTS: LINGUISTIC DETERMINANTS, COMPREHENSION PROCESS AND SYMMETRY

Annekatrin Kaivapalu

Tallinn University

The paper deals with the mutual intelligibility of written Estonian and Finnish on two levels: on the level of individual context-free words and on the level of a text as a whole. The study aims to explain whether and how Estonian speakers understand Finnish and Finnish speakers Estonian with no previous formal learning or exposure to the closely related target language, and how the comprehension of individual context-free words differs from the comprehension of a text. The results of the comprehension tests and the participants' comments on their process of understanding are discussed to shed light on the possible symmetry of mutual intelligibility and on the linguistic determinants having an impact on the comprehension process. The discussion is followed by proposing a model of intelligibility.

According to the test results, the Estonian-speaking participants were more successful in comprehending context-free words, particularly the words with similar meaning and similar spelling, as well as loanwords. Finnish-speaking participants succeeded better in comprehending the whole text. Perceived cross-linguistic similarity has a crucial role in comprehending both of the target languages: the base of understanding is finding similarities between Estonian and Finnish, as well as other previously learned languages and the target language. Those similarities are combined with metalinguistic and general knowledge and the assumptions made are checked in the context of a sentence, a paragraph and a text. Context operates as a kind of filter, either confirming or disproving the assumptions made. The distribution and order of use of the components of the understanding process vary from person to person and largely depend on the participants' previous knowledge and experience.

Keywords: receptive multilingualism, perceived cross-linguistic similarity, Estonian, Finnish

KORPUSLEKSIKOGRAAFIA UUED VÕIMALUSED EESTI KEELE KOLLOKATSIOONISÕNASTIKU NÄITEL

Jelena Kallas, Kristina Koppel,
Maria Tuulik

Ülevaade. Artiklis tutvustame korpusleksikograafia üldisi arengutendentse ja uusi meetodeid. Käsitleme korpuse kui leksikograafilise info allika potentsiaali ning analüüsime, kuidas saab leksikograafilisi andmebaase pool- ja täisautomaatselt genereerida. Vaatleme, mil määral on uusi tehnoloogilisi lahendusi võimalik rakendada Eesti õppeleksikograafias, täpsemalt eesti keele kollokatsioonisõnastiku (KOLS) koostamisel. KOLS on esimene eestikeelne sõnastik, kus rakendatakse andmebaasi automaatset genereerimist nii märksõnastiku kui ka sõnaartikli sisu (kollokatiivse info ja näitelause) tasandil. Tutvustame sõnastiku koostamise üldisi põhimõtteid ja esitame näidisartikli.*

Võtmesõnad: korpusleksikograafia, kollokatsioonisõnastik, korpus-päringüsteem, sõnastikusüsteem, eesti keel

1. Sõnastike korpuspõhine koostamine internetiajastul: vahendid ja meetodid

Korpusleksikograafia eesmärk on luua meetodid, mis võimaldaksid leksikograafilisi üksusi automaatselt tuvastada ja valida. Korpusanalüüsi tulemusi rakendatakse tänapäeval märksõnastiku loomisel, tähendusjaotuste uurimisel, leksikaalsete üksuste erinevate omaduste tuvastamisel (süntaktiline käitumine, kollokatsioonid, semantiline prosoodia eri tüüpi tekstides, leksikaal-semantilised suhted), näitelause valikul ja tõlkevastete leidmisel (Kilgarriff 2013: 77).

Korpuspõhine koostamine eeldab suure hulga erineva sisuga korpuste olemasolu. Näiteks ei pruugi korpus, mis on piisavalt representatiivne õppesõnastiku koostamiseks, olla piisav suure seletava sõnaraamatu jaoks. Teisel pool on oluline, et korpuste sisu oleks võimalik pidevalt täiendada ja uuendada. Adam Kilgarriff (2001: 345) tõdes juba 2001. aastal, et korpuste loomiseks saab edukalt kasutada

* Eesti keele kollokatsioonisõnastiku koostamist toetab Haridus- ja Teadusministeeriumi riiklik programm "Eesti keel ja kultuurimälu II (2014–2018)". Sõnastik valmib 2018. aastal.

veebimaterjali. Veebi põhieelis on see, et tekstid on oma olemuselt autentseid ja juba masinloetavas formaadis. Geoffrey Leech (2007: 145) on osutanud, et korpus on esinduslik ainult siis, kui see on koostatud ja kujundatud hoolikalt valitud materjalist. Ilmselt ei saa seda öelda veebi kohta, mis koosneb laiaulatuslikest igapäevasituatsioonide puudutavatest tekstidest ja mille sisu ei ole keeleteadlaste valitud. Maristella Gatto (2014: 43–45, 67) arvates see just veebi esinduslikuks teebki. Veeb pakub juurdepääsu mitmele žanrile, millest mõni, näiteks akadeemilised tekstid, on korralikud kirjakeelsed tekstid, kuid teised, näiteks blogid, on lähemad kõnekeelele.

Veebi dünaamilise iseloomu tõttu ei saa aga päringu kordamisel olla kindel, et tulemus on sama. See võib osutada probleemiks veebi kasutamisel leksikograafilise allikmaterjalina. Üks võimalik lahendus on laadida otsingu tulemused alla ning luua neist stabiilne ja kontrollitav andmebaas (Gatto 2014: 68–69). Sellist lähenemist rakendatakse näiteks TenTen-korpuste sarja loomisel (Jakubíček jt 2013). TenTen osutab sellele, et ühes veebikorpusel võib olla kuni 10^{10} sõnet. TenTen-korpused on praegu olemas nii suurtele maailmakeeltele, nagu inglise, arabia, hiina, portugali, prantsuse, saksa, itaalia, jaapani, korea, vene ja hispaania keel, kui ka väiksematele, nagu tšehhi, ungari, poola, slovaki ja eesti keel. Eestikeelsete veebilehtede infost koosneva korpuse etTenTen13¹ koostas firma Lexical Computing Ltd. 2013. aasta sügisel. Korpuse on lemmatiseerinud, märgendanud ja ühestanud OÜ Filosoft. Ajakirjandustekstid moodustavad etTenTen13 korpusest 29%, foorumid ja blogid 23%, teabetekstid 9%, 4% moodustavad usulise ja poliitilise sisuga tekstid ning 35% on liigitamata tekstide osakaal².

TenTen-korpuste loomisprotsess on järgmine: esmalt otsitakse tarkvaraga SpiderLing (Pomikalek, Suchomel 2012) välja vastava keele veebilehed ja tõmmatakse tekstid üheks korpuseks kokku; seejärel kustutatakse mittetekstiline materjal ja korduvad tekstid programmidega JusText ja Onion (Pomikalek 2011). Sel viisil on korpuse loomine väga kiire – näiteks koguti 12 miljardit ingliskeelset sõna sisaldav enTenTen12 ainult 12 päevaga.

Kui korpus on loodud, installeeritakse see korpuspäringusüsteemi. Selliste süsteemide viimasesse põlvkonda kuulub näiteks Sketch Engine (Kilgarriff jt 2004; Kilgarriff, Kosem 2012), mida kasutatakse Eesti Keele Instituudis³. Programmi funktsioonid on konkordantsi koostamine ja selle mitmekülgne töötlemine, statistikapõhine kollokaatide leidmine, korpusest sõnaloendite ja sagedusloendite koostamine, sõna süntaktilist ja kollokatiivset käitumist illustreerivate sõnavisandite (Word Sketch) genereerimine, sõnastikunäidete automaatne valimine, tesauruse koostamine. Seega sisaldab tarkvara erinevaid funktsioone, mida saab konkreetse leksikograafilise projekti puhul rakendada. Osa funktsioonidest on universaalsed ja kergesti kohandatavad kõikidele keeltele (eeldusel, et korpus on lemmatiseeritud, morfoloogiliselt märgendatud ja ühestatud), osa funktsioone nõuab aga keele-spetsiifiliste rakenduste loomist. Nii eeldab näiteks sõnavisandite genereerimine iga keele jaoks spetsiaalse sõnavisandite grammatika (Sketch Grammar) kirjutamist (eesti keele mooduli kohta vt lähemalt Kallas 2013: 31–87), heade näitelauseste tuvastamiseks on samuti vajalik võtta arvesse mitmeid keele tüübist tingitud omadusi.

¹ Korpus on kättesaadav aadressil www.keeleeveeb.ee ning programmi Sketch Engine <https://the.sketchengine.co.uk/auth/corpora/kaudu> (29.9.2014). Korpuse nimetuses sisalduv number 13 osutab, et korpus loodi 2013. aastal.

² Vt lähemalt <http://www2.keeleeveeb.ee/dict/corpus/ettenten/about.html> (29.9.2014).

³ Programmi kasutati "Eesti keele põhisisavara sõnastiku" (PSV) koostamisel, hetkel kasutatakse seda ühekülgilise eesti keele seletava sõnaraamatu (Langemets jt 2010) ja eesti keele kollokatsioonisisavara koostamisel.

Korpuspõhisuse põhimõtte tänapäeva leksikograafias on paratamatult avaldanud mõju kogu sõnastike koostamisprotsessile. Annette Klosa (2013: 520–522) eristab sealjuures kuut etappi: 1) projekti kavandamine (sh pilootuuring); 2) allikmaterjali määramine (tekstikorpused, ilmunud paber- või veebisõnastikud, grammatikaõpikud, leksikaalsed andmebaasid, pildipangad jmt); 3) sõnastiku andmebaasi struktuuri loomine; 4) allikmaterjali (eelkõige tekstikorpuste) installeerimine korpuspäringusüsteemi; 5) andmete töötlus; 6) veebiliidese loomine ja avalikustamine.

Juba enne sõnastiku loomist tuleb täpselt määrata, millisele korpusmaterjalile hakkab leksikograaf toetuma. Mõnikord tuleb vastav korpus alles luua, see märgendada ja installeerida korpuspäringusüsteemi. Koostamist saab oluliselt kiirendada, kui andmebaas genereerida pool- või täisautomaatselt. Seejuures võib tulemus olla kahesugune: a) info, mis enam toimetamist ei vaja, näiteks info märksõna sageduse kohta; b) info, mida tuleb käsitada toorandmestikuna ja mis vajab leksikograafilist järeloimetamist. Sellised infoüksused on näiteks tähendusjaotused, definitsioonid, kollokatsioonid ja näitelauseid.

Sõnastiku andmebaasi täisautomaatset genereerimist on rakendatud näiteks sloveeni keele leksikograafilise andmebaasi Slovene Lexical Database (Kosem jt 2013) ja inglise keele leksikaalse andmebaasi DANTE⁴ genereerimisel. Meetod sobib eriti hästi kollokatsioonisõnastike andmebaaside koostamiseks. Automaatgenereerimine toetub tarkvara Sketch Engine sõnaloendi (Word List), sõnavisandi (Word Sketch) ja heade näitelause (Good Dictionary Example ehk GDEX) funktsioonidele. Sõnaloend võimaldab sageduspõhise märksõnastiku loomist, päringut saab teha nt sõnavormipõhiselt, sõnaliigipõhiselt või grammatiliste tunnuste järgi. Sõnavisandi abil saab tuvastada sagedamaid ja kõrge esilduvusega kollokatsioone. GDEX võimaldab leida sobivad näitelauseid. Kosemi jt (2013: 41–42) uurimus näitas, et andmebaasi täisautomaatne koostamine vähendab leksikograafi ajakulu umbes poole võrra.

Poolautomaatset koostamist võimaldab Sketch Engine'i Tickbox Lexicography (TBL) meetod (Kilgarriff jt 2010), mis seisneb selles, et leksikograaf valib sõnavisandit analüüsides konkreetse lekseemi jaoks sobivad kollokaadid ja näitelauseid ning seejärel toimub nende automaatne ülekanne sõnastikusüsteemi. Nii on koostatud näiteks “Macmillan Collocations Dictionary for Learners of English” (Rundell 2012). Meetodit rakendatakse ka suure hollandi keele sõnaraamatu “Algemeen Nederlands Woordenboek” (Tiberius, Schoonheim (ilmumas)) koostamisel.

Joonis 1 illustreerib TBL-i võimalikku kasutust eesti substantiivi *otsus* näitel. Leksikograaf märgib linnukesega, mis kollokaadid sõnaartiklisse sobivad, siis valib iga kollokatsiooni jaoks sobivad laused (joonisel on näidatud kollokatsiooni *õige otsus* näitelauseid), seejärel salvestab tulemuse ja kopeerib valitud üksused sõnastikusüsteemi.

⁴ Vt <http://www.webdante.com/> (29.9.2014).

otsus (*common noun*)
 EstonianNC freq = 262,389 (465.9 per million)

Adj modifier	59,033	2.4	subject of	10,238	1.1	object of	10,614	3.0
<input checked="" type="checkbox"/> lõplik	4,360	10.25	<input checked="" type="checkbox"/> tulema	1,219	6.58	<input checked="" type="checkbox"/> tegema	4,238	8.16
<input checked="" type="checkbox"/> poliitiline	4,241	8.92	<input checked="" type="checkbox"/> sündima	627	8.99	<input checked="" type="checkbox"/> langetama	1,804	11.56
<input checked="" type="checkbox"/> õige	2,918	7.9	<input checked="" type="checkbox"/> jõustuma	540	9.68	<input type="checkbox"/> võtma	1,077	7.14
<input type="checkbox"/> vastav	1,578	7.51	<input type="checkbox"/> tähendama	434	6.66	<input checked="" type="checkbox"/> põhjendama	369	8.94
<input type="checkbox"/> käesolev	1,473	7.41	<input type="checkbox"/> tegema	396	4.74	<input type="checkbox"/> kohaldama	190	8.32
<input checked="" type="checkbox"/> oluline	1,383	6.48	<input checked="" type="checkbox"/> puudutama	376	8.16	<input checked="" type="checkbox"/> tühistama	186	8.47
<input type="checkbox"/> puudutav	1,286	8.62	<input checked="" type="checkbox"/> mõjutama	352	8.0	<input type="checkbox"/> muutma	163	6.19

Tickbox Lexicography - Select Examples

Lemma: otsus
 Gramrel: Adj_modifier
 Template: vanilla

õige

- Õigete otsuste korral väldite soetatava kinnisvara väärtuse langemist aja jooksul .
- Teatud juhtudel on kodakondsuse vahetamine sportlase jaoks ilmselt õige otsus .
- Eesti on varem suutnud rasketel aegadel teha õigeid otsuseid .

Joonis 1. Kollokaatide ja näitelauseite valik Tickbox Lexicography meetodil

Poolautomaatse koostamise eelis on see, et märksõnu saab valida n-ö jooksvalt, sõnastiku koostamise käigus. Teisalt võimaldab see meetod väljundit kontrollida ja paremini jälgida terminite, harva esinevate sõnade, slängi ja erisuguse müra sattumist kollokaatide hulka ja näidetes.

Siinses artiklis vaatleme, mil määral saab automaatset genereerimist rakendada Eesti õppeleksikograafias, täpsemalt, eesti keele kollokatsioonisõnastiku koostamisel. Selle andmebaasi automaatse genereerimise alus on 2014. aastal loodud eesti keele ühendkorpus (Estonian National Corpus)⁵, mille struktuuri tutvustame lähemalt peatükis 3.

2. Eesti keele kollokatsioonisõnastiku koostamispõhimõtted

Eesti keele kollokatsioonisõnastik (KOLS) on Eesti Keele Instituudi 2014. aastal alanud projekt, mille sihtgrupp on peamiselt eesti keele õppijad (B2-C1 keele-ostkustase), kuid ka emakeelsed kõnelejad.

Igal keelel on eriomaseid sõnade kombinatsioone, mille tundmine on vajalik, et oleks võimalik selles keeles loomulikult ja ladusalt rääkida ning kirjutada. Näiteks on eesti keeles täiesti normaalne öelda *tugev vihm*, kuid inglise keeles see nii ei ole (öeldakse *heavy rain*, aga mitte *strong rain*). Kollokatsioonisõnastike eesmärk ongi aidata keeleõppijal valida õigeid keelendeid, mis väljendaksid nende mõtteid loomulikult moel ja teeksid nende teksti sarnaseks emakeelse kõneleja omale.

⁵ Vt <https://the.sketchengine.co.uk/auth/corpora/> (29.9.2014). Praegu on ühendkorpus kättesaadav ainult korpuspäringu tarkvara Sketch Engine kaudu.

Batia Laufer (2011: 44) tõdeb, et paljud keeleõppijad ei ole kollokatsioonidest teadlikud – nad ei tea, et näiteks teatud substantiivid esinevad vaid koos piiratud arvu verbidega. Õppijad võivad aga valida verbi, mis nende emakeeles konteksti sobitub, kuid sihtkeeles on vale. Üsna tihti ei leia kasutajad infot kollokatsioonide kohta isegi siis, kui see on sõnastikus olemas (nt näitelauseste tasandil). Robert Lew (2004: 23) märgib, et enamik algajaid õppijaid ei otsigi sõnastikest kollokatsiooninfot, vaid see huvitab rohkem edasijõudnuid, kellel on suurem teadmine keelestruktuurist.

Meie käsitleme kollokatsioonidena sisusõnade tähenduslikke ja statistiliselt esilduvaid kombinatsioone teiste leksikaalsete ja grammatiliste üksustega. Kollokatsioonid on nt *ere päike, päike paistab, päikest võtma*.

Eesti keele kollokatsioonisõnastiku põhilised infoüksused on märksõna, definiitsioon (vaid mitmetähenduslike sõnade puhul), kollokatsioonid ja näitelauseid. Tähendusi eristame ainult siis, kui see on kollokatsioonide selgema esituse huvides vajalik. Põhjalik polüseemia kirjeldamine pole kollokatsioonisõnastikus eesmärk omaette.

KOLS-i maht on umbes 10 000 märksõna. Võrdluseks olgu siinkohal toodud “Macmillan Collocations Dictionary for Learners of English” (MCD 2010), kus märksõnu on 4500 ja kollokatsioone rohkem kui 121 000, ning “Oxford Collocations Dictionary for Students of English” (OCDSE 2002), kus märksõnade hulk on 9000, kollokatsioone kokku 150 000 ja näitelauseid rohkem kui 50 000.

Reeglina on kollokatsioonisõnastike märksõnadeks vaid sisusõnad. Substantiivid, adjektiivid ja verbid moodustavad kuni 99% ja määrsõnad vähem kui 1% kogu märksõnastikust (Kilgarriiff jt 2014).

KOLS-i märksõnastikku kuuluvad substantiivid, adjektiivid, verbid ja lisaks valikuliselt ka viisiadverbid (nt *salaja, kiiresti*). Eraldi kontrollime, et märksõnastikus on kindlasti ka akadeemilises kirjutamises vajaminev sõnavara (Metslang, Kibar 2012)⁶. Mitmesõnaliste märksõnadena esitame semantiliselt terviklikke, omaette tähenduse ja argumentstruktuuriga ühendverbe (*kaasa lööma, alla andma*), väljendverbe (*silma paistma, silmas pidama*) ja isolaadi ehk ainukordse komponendi ja verbi ühendeid (*andeks andma, tähele panema, kihla vedama, toime tulema*).

Tänapäeva inglise sõnastikes (nt OCDSE 2002, MCD 2010, “Longman English Dictionary”⁷, “Collins English Dictionary”⁸) esitatakse kollokatsioone mitmel viisil:

- kollokatsioonirühm on esitatud märksõna all sõnaliigipõhise koodi ja/või sümboli taga. Rühmitamise aluseks on kollokatsioonikood, nt Adj+N (adjektiiv + substantiiv);
- kollokatsioonid esitatakse loendina eraldi kastikeses (vt joonis 2);
- kollokatsioonid tuuakse näidete sees esile teistsuguse fondi abil;
- kollokatsioonid esitatakse näidete sees eraldi esile toomata;
- kollokatsioonid esitatakse definiitsiooni osana.

⁶ 400-sõnalise üldakadeemilise sõnavara loend on koostatud ülikoolis õpetatavate erialade tekstide põhjal. Loend ei ole valdkonnaspetsiifiline, vaid selssele on valitud üldisemad sõnad, mis on paljudel erialadel ühised.

⁷ Vt <http://www.ldoceonline.com/> (29.9.2014).

⁸ Vt <http://www.collinsdictionary.com/dictionary/english-cobuild-learners> (29.9.2014).


car *noun* W I S I

Menu

Related topics: [Technology](#), [Motor Vehicles](#)

car [countable]

1 a vehicle with four wheels and an engine, that can carry a small number of passengers







COLLOCATIONS ▲

by car
 get in/into a car
 get out of a car
 drive a car
 park a car
 parked car
 take the car (=drive it somewhere)
 car crash/accident
 car chase
 car crime
 police car
 company car (=a car you are given to use by your company)

Joonis 2. Kollokatsioonide esitus sõnastikus “Longman English Dictionary”

Süntagmaatilise info esitamise varasemad uurimused eesti keele üld- ja õppesõnaraamatutes (Langemets jt 2005, Kallas, Tuulik 2011, Kallas 2013) on näidanud, et eesti leksikograafias ei ole välja kujunenud kollokatsioonisõnaraamatute koostamise traditsiooni. Kollokatsioone pole tavaks esitada eksplitsiitselt, enamasti tehakse seda näitelauseste tasandil.

Üks esimesi katseid kollokatsioone eksplitsiitselt esitada tehti “Eesti keele põhisõnavara sõnastikus” (PSV). Kasutaja jaoks eristati kollokatsioonirühmad muust artiklist mummu abil ja paksu kirjaga (vt joonis 3), andmebaasis aga koodidena (vt lähemalt Kallas 2013: 104–126).

näidend *nimisõna* ⟨näidend , näidendi , näidendit ; *mitmus* näidendid, näidendite, näidendeid 

kirjutatud teos, mis on mõeldud näitlejatele teatris esitamiseks
Ta mängib näidendis peaosa.

- **näidendi autor, näidendi tegelased**
- **näidendit kirjutama, näidendit lavastama** *Kes selle näidendi lavastas?*

Joonis 3. Kollokatsioonide esitus PSV-s

KOLS-i koostamisel järgime PSV tarbeks välja töötatud kollokatsioonide esituspõhimõtteid, üritame hoiduda liigsest lahterdamisest, esitame koos läbipaistvate kollokatsioonidega (nt *päikest nautima*) ka läbipaistmatumaid (nt *päikest võtma*) ja lisame vajadusel seletuse.

Kollokatsioone rühmitame sõnaliikide ning morfoloogiliste ja süntaktiliste kategooriate põhjal järgmistesse kollokatsioonirühmadesse (vt tabel 1).

Table 1. KOLS-i kollokatsioonirühmad (koos näidetega)

Substantiivi mallid	
adjektiiv + substantiiv	<i>hea/ilus/kõlav laul</i>
substantiiv (genitiivis) + substantiiv	<i>ekspertide hinnang koosoleku otsus</i>
substantiiv + substantiiv (partitiivis)	<i>viil leiba/juustu/saia</i>
substantiiv (adverbiaalkäändes) + substantiiv	<i>kullast ehted/kõrvarõngad osavõtt konkursist/võistlustest/valimistest</i>
substantiiv (subjekti funktsioonis) + verb	<i>hobune hirrub palavik tõuseb/langeb</i>
substantiiv (objekti funktsioonis) + verb	<i>arvutit sisse lülitama/välja lülitama</i>
substantiiv (adverbiaali funktsioonis) + verb	<i>aktsiasse investeerima</i>
substantiiv + adpositsioon(ifraas)	<i>lepingu kohaselt/järgi uhkus saavutuste/tehtu üle</i>
adverb + substantiiv	<i>raagus puud omaette tuba</i>
substantiiv + verb <i>ma-</i> või <i>da-</i> infinitiivis	<i>meister valetama soov laulda</i>
rinnastustarind võrdlustarind	<i>päike ja tuul elu kui kabaree</i>
Adjektiivi mallid	
adjektiiv + substantiiv	<i>raske otsus rõõmsates toonides/värvides rõõmsal häälel</i>
adverb + adjektiiv	<i>väga aeglane silmatorkavalt hea</i>
adjektiiv (translatiivis) + verb adjektiiv (essiivis) + verb adjektiiv (nominatiivis) + verb	<i>rikkaks saama kahtlasena paistma ilus välja nägema</i>
adjektiiv + verb <i>ma-</i> või <i>da-</i> infinitiivis	<i>raske mõista pädev otsustama/hindama</i>
adjektiiv + adjektiiv	<i>igavene suur</i>
adjektiiv + adpositsioonifraas	<i>kingituste üle rõõmus hull raha järele</i>
rinnastustarind võrdlustarind	<i>rikas ja ilus valge kui lumi must nagu süsi</i>
Adverbi mallid	
adverb + adverb	<i>aina rohkem väga kiiresti</i>
adverb + adjektiiv	<i>väga aeglane</i>
adverb + verb	<i>kiiresti jooksmas</i>
adverb + substantiiv	<i>palju rahvast</i>
adverb + adpositsioonifraas	<i>selja pealt lõhki puusade ümbert pingul</i>
rinnastustarind võrdlustarind	<i>hästi ja kiiresti kergelt kui õhk</i>

Verbi mallid	
adverb + verb	<i>kiiresti jooksm</i>
substantiiv (subjekti funktsioonis) + verb	<i>päike tõuseb/loojub</i>
substantiiv (objekti funktsioonis) + verb	<i>lilli istutama/kastma/korjama/kinkima</i>
substantiiv (adverbiaali funktsioonis) + verb	<i>lugupidamisega/austusega/eelarvamusega suhtuma</i>
adjektiiv (translatiivis) + verb	<i>paksuks minema</i>
adjektiiv (essiivis) + verb	<i>võimatuna tunduma</i>
adjektiiv (nominatiivis) + verb	<i>kummaline näima</i>
finiitverb + infiniitverb	<i>ajab naerma/nutma/iiveldama/oksendama</i> <i>jätab maksmata/tegemata</i>
verb + adpositsioonifraas	<i>vapruse/ülbuse/jutukuse poolest silma paistma</i>
rinnastustarind	<i>kirjutama ja lugema</i>
võrdlustarind	<i>elada või surra</i>

Andmebaasis jagame kollokatsioonid esitatud rühmade kaupa, kuid sõnastiku meta-keeleks on vaid kollokaatide sõnaliigitähised välja kirjutatud kujul. Kollokatsioonide puhul arvestame kollokaatide vormide sagedust ja esitame neid kõige tüüpilisemal kujul, mitte tingimata algvormis. Seega omandab õppija koos kollokatsiooniga ka grammatilise info viisil, mida on lihtne kohe kasutusse võtta. Järgnevalt illustreerime kavandatavat kollokatsioonide esitust substantiivi *arutus* näitel (1).

(1) ARUTLUS nimisõna

OMADUSSÕNAD

- **teoreetiline, avalik, pikk, filosoofiline, loogiline, tõsine, huvitav, sisuline, põhjalik** arutus

NIMISÕNAD

- arutluse **objekt, tulemus, taust, tase**

TEGUSÕNAD

- arutus **käib, toimub, algab, keskendub milledle, jätkub, kestab, tekib**
- arutlust **jätkama, alustama, korraldama, kuulama, juhtima**
- arutlusele **tulema, võtma**
- arutlusel **olema**

KAASSÕNAD

- arutluse **alla** (tulema, võtma)
- arutluse **all** (olema)

Plaanis on esitada iga kollokatsiooni juures näitelauseid, mis avaneksid kollokaadile klõpsates. Enamasti kasutame autentseid korpusnäiteid või nende mugandatud versioone.

3. KOLS-i andmebaasi allikas: eesti keele ühendkorpus

Eesti keele ühendkorpuse suurus on ca 563 mln sõnet ning hetkel on see suurim ja žanriliselt mitmekesisem eesti keele korpus. Korpus koosneb eesti keele koondkorpusest (250 mln sõnet), sh tasakaalus korpusest (15 mln sõnet), ja eesti veebikorpusest etTenTen13 (ca 330 mln sõnet). Tarkvara Sketch Engine abil saab korpuse sisu analüüsida kas kogu korpuse ulatuses, allkorpuste või domeenide kaupa (vt joonis 4).

Text Types

Subcorpus: [info](#) [create new](#)

<p>DOC.BALANCED</p> <p><input type="checkbox"/> no <input type="checkbox"/> yes</p> <p><input type="button" value="Select All"/></p>	<p>WEB DOMAIN</p> <p><input type="text"/></p>
<p>DOC.TEXTTYPE</p> <p><input type="checkbox"/> EstonianRC / fiction <input type="checkbox"/> EstonianRC / legislation <input type="checkbox"/> EstonianRC / parliament <input type="checkbox"/> EstonianRC / periodicals (until 2008) <input type="checkbox"/> EstonianRC / science <input type="checkbox"/> etTenTen / blog <input type="checkbox"/> etTenTen / forum <input type="checkbox"/> etTenTen / government <input type="checkbox"/> etTenTen / informative <input type="checkbox"/> etTenTen / periodicals (2013) <input type="checkbox"/> etTenTen / religion <input type="checkbox"/> etTenTen / unknown</p> <p><input type="button" value="Select All"/></p>	<p>TOP LEVEL DOMAIN</p> <p><input type="checkbox"/> com <input type="checkbox"/> cz <input type="checkbox"/> de <input type="checkbox"/> ee <input type="checkbox"/> eu <input type="checkbox"/> fi <input type="checkbox"/> gov <input type="checkbox"/> info <input type="checkbox"/> net <input type="checkbox"/> org <input type="checkbox"/> ru <input type="checkbox"/> us</p> <p><input type="button" value="Select All"/></p>

Joonis 4. Eesti keele ühendkorpuse allkorpused ja domeenid

Koondkorpuse allkorpused on ajakirjandustekstid kuni aastani 2008 (*periodicals (until 2008)*), ilukirjandustekstid (*fiction*), riigikogu stenogrammid (*parliament*), Eesti ja Euroopa seadused (*legislation*) ja teadustekstid (*science*). Veebikorpuse etTenTen13 allkorpused on ajakirjandustekstid (*periodicals (2013)*), foorumid (*forum*), blogid (*blog*), teabetekstid (*informative*), usutekstid (*religion*), ametlikud tekstid (*government*) ja varia (*unknown*). Leksikograafi vaatevinklist aga täiendavad koondkorpuse ja veebikorpuse teineteist: kui koondkorpuse tekstivalik võimaldab analüüsida eelkõige kirjakeelt, siis veebikorpuse annab parema ülevaate just kõnekeele ja internetikeele kohta.

KOLS-i andmebaasi aluseks on ühendkorpuse tervikuna, kuid soovi korral on toimetamise faasis võimalik uurida kollokatsioonide esinemist allkorpuste kaupa.

4. KOLS-i andmebaasi korpuspõhine automaatne genereerimine

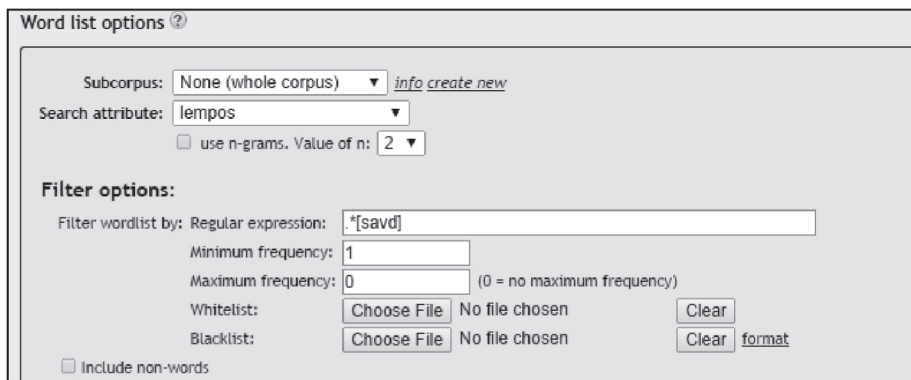
KOLS-i andmebaasi genereerimine toimus korpuspäringutarkvaraga Sketch Engine ja koosnes järgmistest etappidest: 1) märksõnaloendi genereerimine sõnaloendi (Word List) funktsiooni abil, millele järgnes käsitsi kontroll; 2) sõnavisandite grammatika (Sketch Grammar) täiendamine ja andmete ekstraheerimiseks vajalike parameetrite täpsustamine; 3) näitelauseste klassifikaatorite väljatöötamine ja GDEX-i skripti kirjutamine⁹; 4) spetsiaalse rakendusprogrammi kirjutamine ja andmete ekstraheerimine korpuspäringuprogrammist XML-faili kujul; 5) andmete importimine Eesti Keele Instituudi sõnastikusüsteemi EELex¹⁰.

⁹ Skripti süntaksi kohta vt <https://www.sketchengine.co.uk/documentation/wiki/GDEX/Syntax> (5.1.2015).

¹⁰ Vt <http://eelex.dyn.eki.ee/> (29.9.2014).

4.1. Märksõnastik

KOLS-i märksõnastiku koostamisel kasutasime Sketch Engine'i sõnaloendi (Word List) funktsiooni, mis võimaldab regulaaravaldiste abil erisuguse sisuga loendite genereerimist. Joonis 5 näitab funktsiooni päringuakent. Sealt nähtub, et loendit koostades otsib programm läbi kogu korpuse, otsinguatribuut on *lempos* (lemma + sõnaliik), otsitakse vaid substantiiviks, adjektiiviks, verbiks ja adverbiks märgendatud sõnu, lemma minimaalne ettemääratud esinemissagedus korpuses on 1, maksimaalne sagedus ei ole piiratud.



Joonis 5. Programmi Sketch Engine sõnaloendite funktsiooni päringuaken

Märksõnastiku aluseks võtsime 12 000 sagedasemat sõna, mida käsitsi kontrollisime. Kontroll oli vajalik näiteks selleks, et korpuse märgendamise kvaliteedist ja puudulikest ühestamisest tulenevat “müra” kõrvaldada. Tuli eemaldada vale variant märksõnadest, millel esines kaks kirjpilti (*mänedžer* vs. *mänedzher*, *šokk* vs. *shokk*, *režiim* vs. *rezhiim*), lühendid (*EEK*, *EUR*, *TOIM*), pärisnimed (*lõunaleht*, *suurhall*), liitsõna järelosised (*sugune*, *keelne*), vigased liitsõnad (*lapseoode*, *minumeel*, *omavahend*) ja terminid (*süsinikdioksiid*). Lisasime ka “Eesti keele põhisõnavara sõnastiku” (PSV) märksõnad, mis sagedusloendisse ei sattunud (*pott*, *kõhima*, *hakkliha*).

Märksõnade loendi automaatgenereerimise üks puudus on see, et pole võimalik tuvastada mitmesõnalisi märksõnu. Nende lisamine toimub käsitsi sõnastiku koostamise käigus. Kasutame kandidaatide valikul EKSS-i ja PSV perifrastiliste verbide loendit ning analüüsime sõnavisandite abil tuvastatud väljend- ja ühend- verbe. Sõnavisandite grammatika võimaldab tuvastada verbi ja X-iks märgendatud sõnade koosinemisi (nt *pärit olema*, *tähele panema*, *andeks andma*) ning verbi ja afiksaaladverbi funktsioonis esinevate sõnade koosinemisi (nt *kaasa lööma*, *läbi lööma*, *välja lööma*, *kokku lööma*, *lahku lööma*). Samuti toetume eesti keele verbikesksete püsiühendite andmebaasile¹¹ ja tasakaalus korpuse lemmade ja sõnavormide mitmikute (n-grammide) sagedusloenditele¹².

¹¹ Vt <http://www.cl.ut.ee/ressursid/pysiyhendid/index.php?lang=et> (21.5.2014).

¹² Vt <http://www.cl.ut.ee/ressursid/mitmikud/> (21.5.2014).

4.2. Kollokatiivne info

KOLS-i kollokatiivse info allikas on Sketch Engine'i sõnavisandi (Word Sketch) funktsiooni väljund (Kilgarriff jt 2004; eesti mooduli kohta vt lähemalt Kallas 2013: 31–87). Sõnavisand on spetsiaalse sõnavisandite grammatika abil genereeritud kokkuvõtte sõna süntaktilisest ja kollokatiivsest käitumisest. 2013. aastal valminud eesti keele sõnavisandite grammatika versioon 1.5 sisaldas 85 reeglit ning selle väljund oli lemmapõhine, see tähendab, et kõik tuvastatavad kollokaadid olid sõnavisandis viidud lemmadele.

2014. aastal loodud grammatika versioonis 1.6 suudab programm tuvastada kollokaate nii lemma- kui ka sõnavormipõhiselt. On lisatud reeglid, mis võimaldavad tuvastada adverbiali funktsioonis esinevate substantiivide ja verbide kollokatsioone (nt *konkursil/võistlusel/valimistel/võistlustel osalema*), käändumatuteks adjektiivideks märgendatud sõnade ja substantiivide kollokatsioone (nt *vastamata küsimus/kõne*) ning verbi ja nominatiivis adjektiivide kollokatsioone (*uskumatu/imelik/kummaline/võimatu tunduma*). Uues versioonis on kokku 116 reeglit, mille hulgas on 16 *unary*-tüüpi reeglit (võimaldavad analüüsida substantiivide ja adjektiivide morfoloogiliste vormide kasutussagedust), 4 *symmetric*-tüüpi reeglit (tuvastavad rinnastus- ja võrdlusterindeid, nt *päike ja tuul, ilus ja noor, valge nagu lumi*), 16 *dual*-tüüpi reeglit (võimaldavad otsida kahe lemma koosinemisi, nt *päike + paistma/loojuma/tõusma*) ja 80 *colloc*-tüüpi reeglit (võimaldavad tuvastada esiteks kolmest sõnast koosnevaid kollokatsioone, nt *uhke tehtu üle, hoolitsema laste eest*, ja teiseks kollokatsioone, kus otsisõna ja/või kollokaat ei esine lemmana, nt *kari lambaid/hobuseid/lehmi, rääkima aktsendita, suhtuma austusega/lugupidamisega* jne). Eriti kasulik on sõnavormipõhine esitus homonüümide puhul. Joonis 6 illustreerib valikuliselt substantiivi *koor* kollokatsioone.

koor (common noun) EstonianNC freq = 27,820 (49.4 per million)		
osastav modifier 3,372 1.3		
<input type="checkbox"/> koguduse_koor	251	11.05
<input type="checkbox"/> kiriku_koor	85	9.62
<input type="checkbox"/> puu_koor	73	9.41
<input type="checkbox"/> kooli_koor	61	9.16
<input type="checkbox"/> sidruni_koor	55	9.02
>>		
osastav modifies 4,553 1.8		
<input type="checkbox"/> koori_dirigent	125	9.74
<input type="checkbox"/> koori_liige	103	9.47
<input type="checkbox"/> koori_laulja	68	8.89
<input type="checkbox"/> koori_peadirigent	62	8.76
<input type="checkbox"/> koori_repertuaar	57	8.64
>>		
osastav modifies 42 1.3		
<input type="checkbox"/> pakk_koort	12	12.54
<input type="checkbox"/> tonn_koort	6	11.83
<input type="checkbox"/> tükk_koort	5	11.62
<input type="checkbox"/> klaas_koort	4	11.36
<input type="checkbox"/> jagu_koort	3	11.0
>>		
adverbial seesütlev of 341 6.5		
<input type="checkbox"/> laulma_kooris	101	12.87
<input type="checkbox"/> hüüdma_kooris	25	11.13
<input type="checkbox"/> vastama_kooris	20	10.83
<input type="checkbox"/> karjuma_kooris	15	10.43
<input type="checkbox"/> ütleva_kooris	14	10.34
>>		
kaasütlev modifies 675 3.5		
<input type="checkbox"/> keedetud_koorega	54	11.14
<input type="checkbox"/> kartulid_koorega	31	10.43
<input type="checkbox"/> sibul_koorega	13	9.25
<input type="checkbox"/> keedetud_koorega	12	9.14
<input type="checkbox"/> kohv_koorega	12	9.14
>>		
adverbial seestütlev of 34 1.0		
<input type="checkbox"/> lahkuma_koorist	4	11.75
<input type="checkbox"/> puhastama_koorest	3	11.38
<input type="checkbox"/> eralduma_koorest	3	11.38
>>		

Joonis 6. Valik substantiivi *koor* kollokatsioone

Jooniselt 6 on näha, et tuvastatud kollokatsioonid on nt *kooris laulma, kooris vastama/ütleva, koorist lahkuma, kooriga liituma, aga koorega kartulid/sibul, koorest puhastama/eralduma* jne.

Sõnavisandite grammatika versioon 1.6 võimaldab tuvastada kõiki KOLS-is esitatavaid kollokatsioonitüüpe (vt tabel 1).

4.3. Näitelauseid

KOLS-i näitelausete alus on Sketch Engine'i funktsiooni GDEX (Kilgarriff jt 2008) väljund. GDEX on tööriist, mis hindab lausete kvaliteeti ja aitab leksikograafil leida korpusest parimad laused (Kilgarriff jt 2008: 425). Funktsioon töötab justkui sõelana, hinnates lause süntaktilisi ja leksikaalseid tunnuseid ning sortides konkordantse selle järgi, kui hästi need hea lause kriteeriumitele vastavad. Tulemusena pakub tööriist nimekirja näitelausestest, kus eesotsas on head ja lõpupoole halvemad kandidaadid.

Esimeses GDEX-i versioonis inglise keele jaoks kasutati lausete kvaliteedi mõõtmiseks järgmisi klassifikaatoreid (Kilgarriff jt 2008: 426–427):

- lause pikkus on 10–25 sõna;
- lauses esinevad ainult sõnad, mis kuuluvad 17 000 sagedasema sõna hulka;
- lauses ei esine pronoomenid või anafoorid, nagu *this*, *that*, *it* või *one*;
- lause algab suure tähega ja lõppeb kirjavahemärgiga;
- kollokatsioon esineb lause lõpus. Eeldati, et hea näitelause tutvustab esmalt konteksti ja alles siis esineb kollokatsioon, mis sellesse konteksti sobitub. Nii saab kasutaja kollokatsiooni tähenduse konteksti põhjal tuletada, kui ta seda ei tea.

Eelpool nimetatud parameetreid sloveeni keele peal testides selgus, et GDEX-i poolt valitud lausete kvaliteet ei olnud piisavalt hea (Kosem jt 2011: 154–156, Kosem jt 2013: 38–39). Uurimistulemused näitasid, et GDEX nõuab iga keele jaoks eraldi konfiguratsiooni. Nii peeti sloveeni keele puhul kõige olulisemateks parameetriteks lause pikkust (8–30 sõna), väikese sagedusega sõnade läve (minimaalne sagedus korpuses ei tohi olla väiksem kui 3) ning pärisnimede ja pronoomenite puudumist lausetes. Samuti oli oluline märksõna asukoht. Kui märksõna asus lause alguses, ei olnud lause piisavalt informatiivne. Leiti, et on otstarbekas automaatselt eemaldada laused, kus esinevad lemma kordus ja erinevad sümbolid (meiliaadressid, püsilingid). Kvaliteedi parandamiseks oli koostatud ka eraldi loend sõnadest, mis ei tohtinud lauses esineda (nn must nimekiri), näiteks släng ja ropud sõnad. Lisaks määrati Levenshteini distant¹³, mis tagas, et laused erinesid vähemalt 30% ulatuses. Hilisem uurimus tõi esile veel ühe tõsiasja: GDEX-iga saab paremaid tulemusi, kui kavandada iga sõnaliigi jaoks eraldi konfiguratsioon.

Selgitamaks eesti keele GDEX moodulile sobivaid parameetreid, võrdlesime “Eesti keele põhisõnavara sõnastiku” (PSV) ja koostamisel oleva ühekõitelise eesti keele seletava sõnaraamatu¹⁴ (SS1) näitelauseid. Võrdlusmaterjalina kasutasime etTenTen13 korpuse lauseid. Mõõtsime märksõnade (substantiivide, adjektiivide, adverbide ja verbide) parameetreid, nagu sõnade arv lauses, lausete keskmine pikkus, sõna keskmine pikkus ja kõrvallausetega lausete arv. Iga sõnaliigi puhul vaatasime 50 näitelauseid. Seega analüüsisime kokku 600 lauset.

PSV laused on leksikograafide koostatud didaktilised üksused, mille eesmärk on näidata sõna kasutust kontekstis. Ühekõitelise eesti keele seletava sõnaraamatu

¹³ IT-s ja arvutiteaduses tähistab Levenshteini distant valem, mille abil arvutatakse kahe järjendi vahelist erinevust.

¹⁴ Elektrooline käsikiri Eesti Keele Instituudi sõnastikusüsteemis EELex. Sõnastik valmib 2018. aastal.

(Langemets jt 2010) sihtrühm ei ole keeleõppijad, vaid emakeelsed haritud kasutajad. Sõnastiku näitelauseid on suures osas adapteeritud (lauseid on lühendatud, välja on jäänud pärisnimed jmt). Veebikorpuse laused on autentne keelematerjal, mida ei ole leksikograafiliselt töödeldud. Tabelid 2 ja 3 võtavad kokku analüüsi tulemused.

Tabel 2. PSV ja SS1 näitelauseite ning etTenTen13 korpuse lausete parameetrid

	Sõnade arv lauses	Keskmine lause pikkus sõnades	Keskmine sõna pikkus (tm)
Substantiivid			
PSV	3–9	5,08	5,6
SS1	3–12	6,42	6,7
etTenTen13	4–40	15,8	5,2
Adjektiivid			
PSV	3–10	5,08	5,3
SS1	5–11	6,44	6,7
etTenTen13	3–37	15	5,23
Verbid			
PSV	3–7	4,36	6,21
SS1	2–10	4,72	5,66
etTenTen13	6–56	16,9	6
Adverbid			
PSV	3–11	5,44	4,96
SS1	3–13	5,74	6,1
etTenTen13	7–42	16,8	5,64

Parameetrite kvantitatiivne analüüs toob selgelt esile lausete eripära. Õppetstarbelise PSV näitelauseid on tavaliselt üsna lühikesed (sõnade maksimumhulk on kuni 11, lauses on 4,36–5,44 sõna). SS1 laused on samuti üsna lakoonilised: sõnade maksimumhulk ulatub 13-ni ja lause pikkus on 4,72–6,44 sõna. Hoopis teised parameetrid on autentsetel, korpusest pärit lausetel ja vahe on väga suur: sõnade hulk lauses ulatub 56-ni ja lause keskmine pikkus on 15–16,9 sõna.

Näited (2–4) illustreerivad lauseid nimetatud allikates:

- (2) Peremees kütab ahju. (PSV)
- (3) Ta on mees parimais aastais. (SS1)
- (4) Ükski ema ei anna vabatahtlikult oma last ära, selleks peab olema väga mõjuv põhjus, haigus, võimetus toime tulla või mida iganes. (etTenTen13)

Sõna keskmises pikkuses erilist vahet ei ole: see varieerub 4,96 tähemärgist 6,7 tähemärgini. Samas võivad eestikeelsed sõnad olla üsna pikad, nt *kirjanduslikustatamatumatelegi* (30 tähemärki), nii et ilmselt on otstarbekas sõnade maksimaalse pikkuse seadistamine.

Kõrvallauseite analüüs näitas, et osalauseitega lausete osakaal on üsna väike PSV-s ja SS1-s, samas autentses etTenTen13 korpuses oli osalauseitega lausete osakaal tunduvalt suurem (substantiivide puhul 18%, adjektiivi puhul 58%, verbide ja adverbide puhul aga koguni 76%) (vt tabel 3).

Tabel 3. Kõrvallausetega lausete osakaal PSV-s, SS1-s ja korpuses etTenTen13

	Kõrvallausete osakaal (%)
Substantiivid	
PSV	0%
SS1	12%
etTenTen13	18%
Adjektiivid	
PSV	0%
SS1	14%
etTenTen13	58%
Verbid	
PSV	8%
SS1	10%
etTenTen13	76%
Adverbid	
PSV	20%
SS1	16%
etTenTen13	76%

Põhjuseks võib olla, et leksikograaf käsitab lauset definitsiooni täiendusena ega lisa infot, mis otseselt ei illustreeri vastava sõna kasutust. Kuid korpuse lausetes peegeldub soov anda rohkem konteksti (näide 5).

- (5) Teeme Pere ja Koduga lähemat koostööd ning nii mai, juuni kui juuli numbrist võib lugeda põhjalikke ja asjalikke artikleid lapse seksuaalse arengu teemal ning sellest, kuidas last sel arenguteel toetada. (etTenTen13)

Ilmnes ka, et kõik PSV ja SS1 laused olid süntaktilises mõttes predikatiiviga laused (kas liht- või liitõeldisega). Korpused lausete seas oli palju elliptilisi lauseid (näide 6).

- (6) Koht ise päris ilus. (etTenTen13)

Samuti iseloomustab korpuse lauseid suur hulk pärisnimesid ja numbreid (alla joonitud) (näide 7).

- (7) Kosmosesond Voyager 1 on praegu Maast 17,9 miljardi kilomeetri kaugusel – see tähendab, **väga, väga** kaugel, kolm kuni neli korda Päikesest kaugemal kui Pluuto. (etTenTen13)

Analüüsi andmetele toetudes töötasime välja GDEX-i eesti mooduli klassifikaatorid, mille põhjal eelistatakse lauseid, mis vastavad järgmistele parameetritele:

- lause algab suure tähega ja lõppeb kirjavahemärgiga;
- lause pikkus on 5–20 sõna;
- lemma ei kordu;
- ei esine sõnu, mis on pikemad kui 20 tähemärki või sisaldavad sümboleid;
- ei esine sõnu, mille sagedus on alla 5;
- lause ei alga sidesõnaga;
- lauses ei esine pärisnimesid, lühendeid, tagasiviiteid *mina, sina, tema, see, too* ning adverbe *siin, siia, siit, seal, sinna, sealt, siis, seejärel*.

Kuna KOLS on õppeotstarbeline sõnastik, rakendasime lausete genereerimisel musta nimekirja. Nimekirja aluseks oli OÜ Filosoofi¹⁵ koostatud loend, milles on sõnad, mida eesti keele speller ei tohi soovitada vigaste ja tundmatute sõnade asendajaks. Nimekirja täiendamiseks uurisime EKSS-i sõnu, mille stiil oli märgitud kui VULG (vulgaarne), HALV (halvustav), KÕNEK (kõnekeelne) või SLÄNG. Musta nimekirja lisasime nt *türa*, *narkots* jne. Peale selle täiendasime nimekirja interneti akronüümide (*omg*, *wtf*, *lol*, *irw*) ning inglisi- ja venekeelsete sõimusõnade (*fuck*, *pohui*) ja nende mugandatud variantidega (*fakk*, *pohh*). Lõplik nimekiri koosnes 446 sõnast.

Tulemuseks paranes GDEX-i väljund olulisel määral. Joonisel 7 on esitatud näitelauseid kollokatsioonile *korralik inimene*.

Tickbox Lexicography - Select Examples

Lemma: **inimene**
 Gramrel: **Adj_modifier**
 Template: **vanilla**

korralik

- Iga *korralik* inimene käib vähemalt jõulude ajal oma vanematel külas .
- Korralik* inimene on ka traktorile talveks soojapidava kihi peale pannud .
- Korralik* inimene vaatab enne pikki pühi oma ravimikapi ja rohuvaru üle .
- Tavakodanikule näivad lahkunud olevat pigemini olnud erakordselt *korralikud* inimesed .
- Isegi oma õnnestunud vargustest räägivad muidu justkui *korralikud* inimesed uhkusega .

Joonis 7. Kollokatsiooni *korralik inimene* automaatselt valitud näitelauseid

4.4. Andmebaasi genereerimine: parameetrid ja tulemused

Andmebaasi genereerimiseks vajalik rakendusprogramm on kirjutatud Pythonis. Andmed saadi korpuspäringusüsteemist XML-faili kujul. Genereerimisel oli kogu märksõnastik jagatud kahte sagedusklassi. Esimese klassi moodustasid märksõnastiku 5000 sagedasemat sõna (minimaalne esinemissagedus korpuses 5057), teise klassi jäid märksõnastiku ülejäänud sõnad. Mõlema sagedusklassi jaoks töötasime välja eraldi parameetrid. Esiteks määrasime kõikide sõnaliikide puhul, missugused kollokatsioonitüübid (grammatilised suhted) automaatselt ekstraheeritakse. Valisime välja 48 grammatilist suhet. Arvestasime vaid neid suhteid, kus kaasmoodustajateks olid substantiivid, adjektiivid, verbid, adverbid ja adpositsioonid. Valikust jäid välja suhted, kus kaasmoodustajaks olid näiteks arvsõnad, asesõnad või pärisnimed. Grammatilise suhte sagedus korpuses pidi olema minimaalselt 10 ja esilduvuse indeks pidi olema positiivne. Sõltuvalt grammatilise suhte tüübist ekstraheerisime iga suhte 5 kuni 20 sagedasemat kollokaati. Lisaks määrasime kollokaatide minimaalse sageduse määra (*minimal frequency*) ja esilduvuse indeksi (*score*). Kollokaadi sagedus korpuses pidi olema vähemalt 10 esimese sagedusklassi sõnade jaoks ja vähemalt 5 teise sagedusklassi jaoks, kollokatsiooni esilduvuse

¹⁵ Autorid tänavad Heiki-Jaan Kaalepit (OÜ Filosoofi) loendi eest.

indeks pidi olema positiivne. Kui kollokaat nendele kriteeriumidele ei vastanud, siis seda ei ekstraheeritud. Iga kollokaadi kohta valis programm viis näitelauseid. Lausete valikul toetuti GDEX-i väljundile. Joonis 8 illustreerib XML-andmete esitust kollokatsiooni *uus auto* näitel.

```

<?xml version="1.0"?>
<sr>
- <headword>
  <lemma>auto</lemma>
  <pos>s</pos>
  <freq>304721</freq>
- <gramrel>
  <grname>Adj_modifier</grname>
  <freq>30618</freq>
  <score>1.240256</score>
- <collocation>
  <collo>uus</collo>
  <freq>5498</freq>
  <score>6.830433</score>
- <example>
  Uus
  <b>auto</b>
  ja tundmatu võistlus, sunnivad mehi prognoosides ettevaatlikeks.
</example>
- <example>
  Kavatsen soetada uue
  <b>auto</b>
  ja mark oleks kindlalt Škoda Octavia.
</example>
- <example>
  Ford nõuab sõitjailt häid tulemusi ning panustab samal ajal uue
  <b>auto</b>
  ehitamisse.
</example>
- <example>
  Selle asemel hakatakse käibemaksuga maksustama otseselt uute
  <b>autode</b>
  isiklikku kasutust.
</example>
- <example>
  Eesti Raudtee on aga müügiturul siiski pigem vana kui uue
  <b>auto</b>
  seisuses.
</example>
</collocation>
- <collocation>

```

Joonis 8. XML-andmete esitus kollokatsiooni *uus auto* näitel

Jooniselt 8 on näha, et automaatselt genereeritud andmebaasi üksused on märksõna (*auto*), märksõna sõnaliik (*s*), märksõna esinemissageduses eesti keele ühendkorpuses (30 472), grammatilise suhte nimetus (*Adj_modifier*), grammatilise suhte sagedus (*freq* 30 618) ja esilduvuse indeks (*score* 1,240256), kollokaat (*uus*), kollokaadi esinemissagedus (*freq* 5498) ja esilduvuse indeks (*score* 6.830433), viis näitelauseid. Sel kujul olid andmed imporditud Eesti Keele Instituudi sõnastiku-süsteemi EELex. Tulemuseks on andmebaasis 10 939 märksõna, grammatiliste suhete üldarv on 82 678, kollokaate 493 971 ning näitelauseid 2 469 855. Andmebaas sisaldab rohkem üksusi, kui sõnastikus lõpuks esitatakse. See võimaldab toimetamise käigus täpsemat valikut teha.

5. Kokkuvõte

Artiklis analüüsisime tänapäeva korpusleksikograafia üldisi arengutendentse ja uusi suundi Eesti korpusleksikograafias. Automaattuvastamist ja -ekstraheerimist rakendatakse paljude leksikograafiliste infoüksuste puhul. Automaatselt on võimalik luua märksõnastikku, tuvastada kollokatsioone, valida häid näitelauseid, tõlkevasteid jm.

Tutvustasime hiljuti loodud veebikorpust etTenTen13 ning eesti keele ühendkorpuse ülesehitust ja kasutusvõimalusi. Eesti keele ühendkorpus (563 mln sõnet) on hetkel suurim ja žanriliselt mitmekesiseim eesti keele korpus. Ühendkorpuse sisu saab tarkvara Sketch Engine abil analüüsida kogu korpuse ulatuses, allkorpuste (nt ilukirjandustekstid, teadustekstid, ajakirjandustekstid, blogid, foorumid) või domeenide kaupa.

Eitasime valmiva eesti keele kollokatsioonisõnastiku (KOLS) koostamis- põhimõtteid ja näitasime, kuidas toimus sõnastiku andmebaasi automaatne genereerimine Sketch Engine'iga. Selleks kasutasime funktsioone: sõnaloend (Word List), sõnavisand (Word Sketch) ja hea näitelause (Good Dictionary Example). Sõnavisandite grammatika viimane versioon 1.6 sisaldab 116 reeglit ja võimaldab tuvastada kõiki KOLS-i jaoks olulisi kollokatsioonirühmi. Lisaks töötasime välja eestikeelsete heade näitelause automaatseks tuvastamiseks vajalikud parameetrid. Automaatselt genereeritud kollokatsioonisõnastiku andmebaasi maht on 10 939 märksõna, 493 971 kollokaati ja 2 469 855 näitelause. Järgmine etapp on sõnastiku andmebaasi toimetamine ja täiendamine.

KOLS-i projekti tulemusena valmib mahukas sõnaraamat, mis on suureks abiks eesti keele õppijale. Edaspidi saab sõnastiku andmebaasi kasutada uute leksikograafiliste ressursside loomisel ja eri tüüpi leksikograafiliste veebiportaalide täiendamisel.

Viidatud kirjandus

- EKSS = Eesti keele seletav sõnaraamat I–VI. [The Explanatory Dictionary of Estonian.] Margit Langemets, Mai Tiits, Tiia Valdre, Leidi Veski, Ülle Viks, Piret Voll (Toim.). Eesti Keele Instituut. Tallinn: Eesti Keele Sihtasutus, 2009.
- Jakubíček, Miloš; Kilgarriff, Adam; Vojtěch, Kovář; Rychlý, Pavel; Suchomel, Vit 2013. The TenTen corpus family. – 7th International Corpus Linguistics Conference CL 2013. Lancaster, 125–127.
- Kallas, Jelena 2013. Eesti keele sisusõnade süntagmaatilised suhted korpus- ja õppeleksikograafias. [Syntagmatic Relationships of Estonian Content Words in Corpus and Pedagogical Lexicography.] Tallinna Ülikooli humanitaarteaduste dissertatsioonid 32. Tallinn: Tallinna Ülikool. <http://e-ait.tlulib.ee/id/eprint/303>
- Kallas, Jelena; Tuulik, Maria 2011. Eesti keele põhisõnavara sõnastik: ajalooline kontekst ja koostamis põhimõtted. [The basic dictionary of Estonian: The historical context and the principles of compilation.] – Eesti Rakenduslingvistika Ühingu aastaraamat, 7, 59–75. <http://dx.doi.org/10.5128/ERYa7.04>
- Kilgarriff, Adam 2001. Web as corpus. – Proceedings of the Corpus Linguistics Conference (CL 2001), 13 (Special Issue), 342–344.
- Kilgarriff, Adam 2013. Using corpora as data source for dictionaries. – Howard Jackson (Ed.). The Bloomsbury Companion to Lexicography. London: Bloomsbury, 77–96.

- Kilgarriff, Adam; Husák, Milos; McAdam, Katy; Rundell, Michael; Rychlý, Pavel 2008. GDEX: Automatically finding good dictionary examples in a corpus. – E. Bernal, J. DeCesaris (Eds.). Proceedings of the 13th EURALEX International Congress. Barcelona: Institut Universitari de Linguística Aplicada, Universitat Pompeu Fabra, 425–432.
- Kilgarriff, Adam; Kovář, Vojtěch; Rychlý, Pavel 2010. Tickbox Lexicography. – S. Granger, M. Paquot (Eds.). eLexicography in the 21st Century: New Challenges, New Applications. Proceedings of eLex 2009, Louvain-la-Neuve, 22–24 October 2009. Louvain-la-Neuve: Presses universitaires de Louvain, 411–418.
- Kilgarriff, Adam; Kosem, Iztok 2012. Corpus tools for lexicographers. – S. Granger, M. Paquot (Eds.). Electronic Lexicography. Oxford: Oxford University Press, 31–55.
- Kilgarriff, Adam; Rychlý, Pavel; Jakubíček, Milos; Kovář, Vojtěch; Baisa, Vit; Kocincová, Lucia 2014. Extrinsic corpus evaluation with a collocation dictionary task. – Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC 2014).
- Kilgarriff, Adam; Rychlý, Pavel; Smrz, Pavel; Tugwell, David 2004. The Sketch Engine. – G. Williams, S. Vessier (Eds.). Proceedings of the 11th EURALEX International Congress. Lorient, France: Université de Bretagne Sud, 105–115.
- Kosem, Iztok; Gantar, Polona; Krek, Simon 2013. Automation of lexicographic work: An opportunity for both lexicographers and crowd-sourcing. – I. Kosem, J. Kallas, P. Gantar, S. Krek, M. Langemets, M. Tuulik (Eds.). Electronic Lexicography in the 21st Century: Thinking Outside the Paper. Proceedings of the eLex 2013, 17–19 October 2013, Tallinn, Estonia, 17–19.
- Kosem, Iztok; Husák, Milos; McCarthy, Diana 2011. GDEX for Slovene. – I. Kozem, K. Kosem (Eds.). Electronic Lexicography in the 21st Century: New Applications for New Users. Proceedings of the eLex 2011 conference, Bled, 10–12 November 2011, 151–159.
- Klosa, Annette 2013. The lexicographical process (with special focus on online dictionaries). – H. R. Gouws, U. Heid, W. Schweickard, H. E. Wiegand (Eds.). Dictionaries. An International Encyclopedia of Lexicography. Supplement Volume: Recent Developments with Focus on Electronic and Computational Lexicography. Berlin–Boston: de Gruyter, 517–524.
- Langemets, Margit; Mägedi, Marike; Viks, Ülle 2005. Süntaktiline info sõnastikus: probleeme ja väljavaateid. [Syntactic information in dictionaries: Problems and solutions.] – Eesti Rakenduslingvistika Ühingu aastaraamat, 1, 71–98. <http://dx.doi.org/10.5128/ERYa1.04>
- Langemets, Margit; Tiits, Mai; Valdre, Tiia; Voll, Piret 2010. *In spe*: ühekõiteline eesti keele sõnaraamat. [A prospective monolingual Estonian dictionary.] – Keel ja Kirjandus, 11, 793–810.
- Laufer, Batia 2011. The contribution of dictionary use to the production and retention of collocations in a second language. – International Journal of Lexicography, 24 (1), 29–49. <http://dx.doi.org/10.1093/ijl/ecq039>
- Leech, Geoffrey 2007. New resources, or just better old ones? The Holy Grail of representativeness. Corpus linguistics and the web. – M. Hundt, N. Nesselhauf, C. Biewer (Eds.). Language and Computers, Corpus Linguistics and the Web. Rodopi, 133–149.
- Lew, Robert 2004. Which Dictionary for Whom? Receptive Use of Bilingual, Monolingual and Semi-Bilingual Dictionaries by Polish Learners of English. Poznań: Motivex.
- MCD 2010 = Macmillan Collocations Dictionary for Learners of English. Australia: Macmillan Education, 2010.
- Metslang, Helena; Kibar, Triin 2012. Üldakadeemiline sõnavara. Abivahend eesti keele õppeks kõrgkoolis. [Estonian Academic Vocabulary.] Tallinn: Tallinna Ülikool.
- OCDSE 2002 = Oxford Collocations Dictionary for Students of English. Oxford: Oxford University Press, 2002.
- Pomikalek, Jan 2011. Removing Boilerplate and Duplicate Content from Web Corpora. PhD thesis. Brno: Masaryk University.

- Pomikalek, Jan; Suchomel, Vit 2012. Efficient web crawling for large text corpora. – Proceedings of the 7th Web-as-Corpus workshop, Lyon, France.
- PSV = Eesti keele põhisõnavara sõnastik. Jelena Kallas, Mai Tiits, Maria Tuulik (Toim.). Madis Jürviste, Kristina Koppel, Maria Tuulik (Koost.). Tallinn: Eesti Keele Sihtasutus, 2014.
- Rundell, Michael 2012. How the dictionary was created? <http://www.macmillandictionaries.com/features/how-dictionaries-are-written/macmillancollocations-dictionary/> (29.9.2014).
- Tiberius, Carole; Schoonheim, Tanneke (ilmumas). The Algemeen Nederlands Woordenboek (ANW) and its lexicographical process. – Vera Hildenbrandt (Ed.). Der lexikografische Prozess bei Internetwörterbüchern. 4. Arbeitsbericht des wissenschaftlichen Netzwerks "Internetlexikografie". Mannheim: Institut für Deutsche Sprache.

Jelena Kallas (Eesti Keele Instituut) on eesti keele kollokatsioonisõnastiku töörühma juht. Põhilised uurimisvaldkonnad: õppeleksikograafia, korpusleksikograafia, eesti keele kui teise keele õpetamise meetodika.

Roosikrantsi 6, 10119 Tallinn, Estonia
jelena.kallas@eki.ee

Kristina Koppel (Eesti Keele Instituut) on eesti keele kollokatsioonisõnastiku töörühma liige.

Põhilised uurimisvaldkonnad: leksikograafia, sõnamoodustus.

Roosikrantsi 6, 10119 Tallinn, Estonia
kristina.koppel@eki.ee

Maria Tuulik (Eesti Keele Instituut) on eesti keele kollokatsioonisõnastiku töörühma liige.

Põhilised uurimisvaldkonnad: leksikograafia, leksikaalne semantika, tekstilingvistika.

Roosikrantsi 6, 10119 Tallinn, Estonia
maria.tuulik@eki.ee

NEW POSSIBILITIES IN CORPUS LEXICOGRAPHY BASED ON THE EXAMPLE OF THE ESTONIAN COLLOCATIONS DICTIONARY

Jelena Kallas, Kristina Koppel,
Maria Tuulik

Institute of the Estonian Language

This article aims to introduce new resources and methods used in Estonian corpus lexicography to create monolingual Estonian dictionaries. Corpora can be used in many ways: headwords list development, grammatical and frequency labels, word sense division, identifying collocations, good dictionary examples, translation equivalents (Kilgarriff 2013). The paper focuses on features offered by Sketch Engine (Kilgarriff et al. 2004), a state-of-the-art lexicographic tool for corpus analysis. For Estonian, Sketch Engine contains different types of corpora, including the recently created 260 million-word web corpus etTenTen13 and the 463 million-word Estonian National Corpus.

Through the example of the Estonian Collocations Dictionary, we analyse how corpus data (headwords, collocations and example sentences) can be automatically extracted from the Estonian National Corpus.

The Estonian Collocations Dictionary contains approx. 10 000 headwords (nouns, adjectives, verbs and adverbs). The various collocates within each headword are grouped according to the lexico-grammatical structure formed by the collocational phrase, and for each collocation one or two example sentences are provided. The main elements needed to develop the algorithm for automatic data extraction are the Sketch Grammar and Good Dictionary Example (Kilgarriff et al. 2008) configurations. The new Sketch Grammar version 1.6 includes all of the lexico-grammatical structures that will be presented in the collocations dictionary. It contains 116 rules in total. For the extraction of dictionary examples, the first version of GDEX for Estonian was developed. Classifiers concerning optimum sentence length, optimum word length, number and type of punctuation marks, word frequency, tokens starting with capital letters, abbreviations etc. were proposed and implemented. The use of classifiers brought significant improvements to the output.

The data was extracted in XML format and imported into the EELEX dictionary-writing system, where it will be examined, edited and supplemented by lexicographers. The Estonian Collocations Dictionary will be published in 2018.

Keywords: corpus lexicography, collocations dictionary, corpus query system, dictionary writing system, Estonian

MORPHONOTACTICS IN L1 ACQUISITION OF LITHUANIAN: TD VS. SLI

Laura Kamandulytė-Merfeldienė

Abstract. The aim of the present study is to test the Strong Morphotactic Hypothesis (SMH), according to which speakers use morphotactic consonant clusters as morphological boundary signals (Korecky-Kröll et al. 2014). It is hypothesized that morphotactic clusters will be better retained during production than phonotactic clusters due to the function fulfilled by a morpheme. The study is based on experimental data collected from 60 Lithuanian TD children and 11 Lithuanian SLI children.

This study explores the impact of morphology on the acquisition of phonotactics. The findings suggest that TD children process morphotactic clusters more accurately than phonotactic clusters because morphotactic clusters have the function of co-signalling the existence of a morphological rule. In contrast to TD children, for SLI children prototypical morphotactic clusters are the most difficult as SLI children are not sensitive to morphological information which is carried by morphotactic clusters.*

Keywords: morphonology, consonant cluster, phonotactics, morpheme, morpheme boundary, Lithuanian

1. Introduction

This paper will discuss the acquisition of Lithuanian phonotactics and morphotactics, i.e. consonant clusters occurring within morphemes and across morpheme boundaries. Morphotactics is the area of interaction between morphotactics and phonotactics and represents a subfield of morphonology which in turn is the area of interaction between morphology and phonology (Dressler et al. 2010: 51). Morphotactics is a new research field which has been established in recent years (Dressler, Dziubalska-Kolaczyk 2006a, 2006b) and concerns the co-occurrence of sounds at morpheme boundaries. According to Wolfgang U. Dressler, Katarzyna

* The author is grateful to Prof. Ineta Dabašinskienė and all the colleagues who took part in the CLAD project. The author also thanks Prof. Wolfgang U. Dressler for his collaboration on the studies of Lithuanian morphotactics.

Dziubalska-Kołączyk and Lina Pestal (2010: 52), a morphonotactic consonant cluster is a cluster which occurs only through morphotactic operations. Phonotactic preferences hold for the basic forms of monomorphemic words; the less they are respected, the more marked phonotactic sequences arise. Morphonotactic sequences, on the other hand, which are due to morphological operations of inflection or word-formation, are much more likely to be marked (Dressler et al. 2010).¹ Based on these theoretical propositions, the paper focuses on the acquisition of phonotactic and morphonotactic consonant clusters in Lithuanian. The aim of the study is to test the Strong Morphonotactic Hypothesis (SMH), according to which speakers use morphonotactic consonant clusters as morphological boundary signals (Korecky-Kröll et al. 2014, Calderone et al. 2014: 59). It is hypothesized that morphonotactic clusters will be better retained in production than purely phonotactic clusters due to the function fulfilled by a morpheme.

Although studies exploring the acquisition of morphonotactics are rare (Kamandulytė 2006a, 2006b, Marshall, van der Lely 2006, Freiberger 2007, Zydorowicz 2007, 2010), they show that morphological information plays a significant role in the acquisition of morphonotactic clusters. Being a strongly inflected fusional language with a very rich inflectional and derivational morphology, Lithuanian represents an especially interesting case for the study of the acquisition of morphonotactics. Some preliminary research based on the longitudinal data of one Lithuanian child support the hypothesis that morphonotactic clusters are acquired earlier and processed faster and more accurately than purely phonotactic clusters (Kamandulytė 2006a, 2006b). This study presents experimental data collected from 60 typically developing (TD) children. In addition, experimental data of 11 children with specific language impairment² (SLI) were included in order to test the claim that SLI children perform better on monomorphemically legal clusters than on monomorphemically illegal clusters (Marshall, van der Lely 2006).

The paper is structured as follows. Section 2 focuses on the features of morphonotactics in Lithuanian and presents the types of Lithuanian (mor)phonotactic clusters. Section 3 describes the methodology of testing. Section 4 presents the results of the study. Section 5 discusses the role of morphology in the acquisition of morphonotactic clusters and provides conclusions.

2. Morphonotactics in Lithuanian

Based on the morphonotactic classification (Dressler, Dziubalska-Kołączyk 2006a: 83), Lithuanian consonant clusters can be graded according to the role of morphology and phonotactics:

- (a) **Prototypical morphonotactic clusters (monomorphemic illegal)**, i.e. clusters which are exclusively due to morphological operations and never occur in monomorphemic words. As we see in the Table 1³, many word final consonant clusters in Lithuanian are prototypical morphonotactic clusters.

¹ According to W. U. Dressler et al. (2010: 65) the prediction that morphonotactic consonantal clusters are more marked (either universally or language-specifically) than intramorphemic phonotactic ones appears to hold true for inflection and, largely, derivation, but not for compounding. The theoretical basis of universal markedness (or universal preferences) draws on models of Natural Phonology (see Hurch, Rhodes 1996) and Natural Morphology (see Dressler et al. 1987).

² Specific language impairment (SLI) is a developmental language disorder which can affect both comprehension and production of language and is not linked to any other developmental disorders or neurological conditions such as learning disabilities, autism, Williams Syndrome or brain injury (Leonard 1998).

³ The classification of Lithuanian clusters is based on the examples from the Corpus of Spoken Lithuanian <http://donelaitis.vdu.lt/sakytines-kalbos-tekstynas/> (24.2.2015) and the corpora of two Lithuanian children.

Usually they are formed by the second-person singular imperative in *-k*, e.g. *lip-k* (climb-IMP.2SG) and by the third-person future in *-s*, e.g. *rakin-s* (lock-FUT.3). Morphonotactic clusters that occur in medial position are formed by the plural imperative in *-k*, e.g. *lip-k-ite* (climb-IMP-2PL), the imperfect in *-d*, e.g. *šok-dav-o* (dance-IMP-3), and very rarely by the future in *-s*, e.g. *rakin-s-ite* (lock-FUT-2PL). Looking at the Table 1, we see that prefixing sometimes creates morphonotactic clusters, as in *at-neš-ti* (PFX-bring-INF).

Table 1. Morphonotactic clusters

Cluster	Position	Example
Imperative		
<i>pk</i>	medial	<i>lip-k-ite</i> (climb-IMP-2PL)
	final	<i>lip-k</i> (climb-IMP.2SG)
<i>bk</i>	medial	<i>dirb-k-ite</i> (work-IMP-2PL)
	final	<i>dirb-k</i> (work-IMP.2SG)
<i>mk</i>	medial	<i>im-k-ite</i> (take-IMP-2PL)
	final	<i>im-k</i> (take-IMP.2SG)
<i>sk</i>	final	<i>skris-k</i> (fly-IMP.2SG)
<i>lk</i>	final	<i>kel-k</i> (lift-IMP.2SG)
<i>rk</i>	final	<i>bar-k</i> (scold-IMP.2SG)
<i>žk</i>	final	<i>lauž-k</i> (break-IMP.2SG)
Future		
<i>ns</i>	medial	<i>rakin-s-ite</i> (lock-FUT-2PL)
	final	<i>rakin-s</i> (lock-FUT.3)
<i>ps</i>	final	<i>kep-s</i> (bake-FUT.3)
<i>bs</i>	final	<i>dirb-s</i> (work-FUT.3)
<i>ls</i>	final	<i>kel-s</i> (lift-FUT.3)
<i>ks</i>	final	<i>pyk-s</i> (be angry-FUT.3)
Imperfect		
<i>kd</i>	medial	<i>šok-dav-o</i> (dance-IMP-3)
<i>md</i>	medial	<i>stum-dav-o</i> (push-IMP-3)
Prefixes		
<i>tn</i>	medial	<i>at-neš-ti</i> (PFX-bring-INF)
<i>šp</i>	medial	<i>iš-pil-ti</i> (PFX-pour_out-INF)
<i>ts</i>	medial	<i>at-si-sės-ti</i> (PFX-REFL-sit_down-INF)

- (b) Clusters which are **mophonotactic** as a **strong default** or (c) as a **weak default**, i.e. morphologically motivated clusters with very few phonotactic exceptions in (b) and more phonotactic exceptions in (c). As it is difficult to establish a boundary between types (b) and (c), Lithuanian examples of both types are presented in one table. Lithuanian clusters that are morphonotactic as defaults occur in medial or final position (see Table 2). Medial clusters arise from the concatenation of a consonant-final root and the future suffix *-s*, e.g. *ger-s-i* (drink-FUT-2SG), the imperative suffix *-k*, e.g. *gul-k-ite* (lie_on-IMP-2PL), and the infinitive suffix *-ti*, e.g. *im-ti* (take-INF). Morphonotactic default clusters that occur in word-final position are formed by third person future forms in *-s*, e.g. *ger-s* (drink-FUT.3).

Table 2. Morphonotactic default clusters

Cluster	Position	Within morpheme (phonotactic)	Across morphemes (morphonotactic)
Future			
<i>rs</i>	medial	<i>gars-as</i> (sound-NOM.SG)	<i>ger-s-i</i> (drink-FUT-2SG)
	final	<i>nors</i> (though)	<i>ger-s</i> (drink-FUT.3)
<i>rš</i>	medial	<i>virš-us</i> (top-NOM.SG)	<i>pamir-š-i</i> (forget-FUT-2SG)
<i>ms</i>	medial	<i>tams-us</i> (dark-NOM.SG)	<i>im-s-i</i> (take-FUT-2SG)
	final	<i>draug-ams</i> (friend-DAT.PL)	<i>im-s</i> (take-FUT.3)
<i>ls</i>	medial	<i>bals-as</i> (voice-NOM.SG)	<i>gul-s-i</i> (lie-FUT-2SG)
<i>ps</i>	medial	<i>liepsn-a</i> (flame-NOM.SG)	<i>sup-s-i</i> (swing-FUT-2SG)
<i>ks</i>	medial	<i>keks-as</i> (cake-NOM.SG)	<i>susitik-s-i</i> (meet-FUT-2SG)
Infinitive			
<i>mt</i>	medial	<i>rimt-as</i> (serious-NOM.SG)	<i>gim-ti</i> (be born-INF)
<i>št</i>	medial	<i>aštuon-i</i> (eight-NOM.PL)	<i>pieš-ti</i> (draw-INF)
Imperative			
<i>lk</i>	medial	<i>vilk-as</i> (wolf-NOM.SG)	<i>gul-k-ite</i> (lie_on-IMP-2PL)
<i>žk</i>	medial	<i>kažk-as</i> (something-NOM.SG)	<i>lauž-k-ite</i> (break-IMP-2PL)

- (d) clusters which exist both due to morphology and without interaction with morphology (**monomorphemic legal**). Clusters that occur across morpheme boundaries usually are formed by plural imperatives in *-k*, e.g. *ger-k-ite* (drink-IMP-2PL) and infinitives in *-t*, e.g. *veik-ti* (do-INF) (see Table 3). These clusters occur in the medial position.

Table 3. Monomorphemic legal clusters

Cluster	Position	Within morpheme (phonotactic)	Across morphemes (morphonotactic)
Imperative			
<i>rk</i>	medial	<i>agurk-as</i> (cucumber-NOM.SG)	<i>ger-k-ite</i> (drink-IMP-2PL)
<i>nk</i>	medial	<i>rank-a</i> (hand-NOM.SG)	<i>skambin-k-ite</i> (call-IMP-2PL)
<i>sk</i>	medial	<i>visk-as</i> (all-NOM.SG)	<i>mes-k-ite</i> (throw-IMP-2PL)
Infinitive			
<i>kt</i>	medial	<i>nakt-is</i> (night-NOM.SG)	<i>veik-ti</i> (do-INF)
<i>st</i>	medial	<i>kopūst-as</i> (cabbage-NOM.SG)	<i>žais-ti</i> (play-INF)
<i>nt</i>	medial	<i>spint-a</i> (cupboard-NOM.SG)	<i>skin-ti</i> (pick-INF)
<i>lt</i>	medial	<i>gelton-as</i> (yellow-NOM.SG)	<i>gul-ti</i> (lie_on-INF)

- (e) Clusters which **never come into being due to morphology** (only phonotactic), e.g. initial clusters in a language which have neither monoconsonantal prefixes, nor morphological deletion of the first vowel of a word. A lot of Lithuanian onomatopoeic words belong to this class, e.g. **apči** 'atishoo'.

According to Wolfgang U. Dressler and Katarzyna Dziubalska-Kolaczyk (2006a: 83), prototypical morphonotactic clusters (a) have the function of co-signalling the existence of a morphological rule; morphonotactic default clusters (b) and (c) perform this function less adequately, while phonotactic clusters of the type (d)

and (e) cannot perform this function and therefore they may be called prototypical phonotactic clusters.

The aforementioned division of consonant clusters according to their position in words and morphemic compositions of words is essential when analysing the processes of language acquisition and studying whether the position of consonant clusters within words has any influence on their acquisition.

3. Testing methodology

To achieve the goals listed above, a test on phonotactic and morphonotactic clusters has been chosen. This test has been developed for the CLAD (Crosslinguistic Language Diagnosis) project⁴ initiative. The CLAD project focused on language acquisition in typically developing children and in children affected by language impairments. One of the goals of the project was to **develop** a set of clinical markers to identify children at risk of specific language impairments.

The methodology of the test on phonotactic and morphonotactic clusters was suggested by Wolfgang U. Dressler and has been adapted to Lithuanian by Laura Kamandulytė-Merfeldienė. The lexical content of the sentences in the test was carefully controlled. The Lithuanian words with consonant clusters, like the other words in the test, were selected from the corpora of Lithuanian children (age 1;8–4;3). The procedure of the test included 2 tasks, a production task and a repetition task. We conducted pre-experimental testing with both production and repetition tasks before conducting the final testing.

The participants in the final testing were 60 TD children (3–7 years old) and 11 SLI children (6–7 years old) from kindergarten divided into several age groups (summarized in Table 4). Unfortunately, the sample of SLI children was rather small. The criterion for selecting the children was lower than that used in most studies of SLI as SLI is not diagnosed yet in Lithuania; speech therapists are still not aware of it and consequently there are no standardized diagnostic tests. Besides, there is considerable heterogeneity among children that are diagnosed with language impairments, so to be able to select children with SLI and not some other disorder, we started from a large group of children in 3 different kindergartens, 2 of which specialize in helping children with special needs. The children that were expected to have SLI were directed to us by speech therapists as having some language impairment. Control subjects have been recruited from a group of TD children.

The grammar and phonology screening (GAPS) test (van der Lely et al. 2007)⁵ was used as a main tool for selecting SLI children, as in Lithuania there is no standardized test to diagnose SLI. Lithuanian speech therapists use a non-standardized test (Gaulienė et al. 2000), which was also used in our subject selection as the starting point since no better available testing materials were available. If the child's performance was poor on both GAPS-LT and the non-standardized Lithuanian test, such child was considered to have an SLI.

After selecting children according to the linguistic and general developmental criteria, each of the selected children was tested individually in a separate room. Each testing session lasted about ten minutes. The responses were audiotaped.

⁴ The European Project CLAD (2008–2011) was funded by the Lifelong Learning Programme and coordinated by Professor Maria Teresa Guasti of the Dept. of Psychology of the University of Milano-Bicocca. The project involved Austria, Belgium, Germany, Italy, Lithuania, and Romania. <http://www.cladproject.eu/> (24.2.2015)

⁵ The GAPS test has been adapted to Lithuanian by Ineta Dabašinskienė and Jūratė Ruzaitė (for information about the adaptation see Ruzaitė, Dabašinskienė 2010).

Table 4. Subjects for the morphonotactic test

Age	TD				SLI
	3–4 ⁶	4–5	5–6	6–7	6–7
Boys	7	8	5	5	11
Girls	8	7	10	10	0
Total	15	15	15	15	11
	60				11

3.1. Production task

The targeted categories of the production task included:

- morphonotactic (monomorphemic illegal) clusters *ns*, *pk*, *mk* (type (a) according to morphonotactic classification, see chapter 2),
- morphonotactic default clusters *št*, *mt*, *ps*, *ks*, *lk* (types (b), (c) according to morphonotactic classification),
- monomorphemic legal clusters *nt*, *lt*, *nk*, *sk* (type (d) according to morphonotactic classification).

Pictures and stimulus sentences were used in the production task. A picture on a computer screen was shown to the child for each experimental item. Furthermore, pre-recorded audio stimuli were played from the computer. The child's task was to use the verb that is produced in the pre-recorded audio in an appropriate form. The first part of the task required the participants to use infinitive forms containing morphologically motivated consonant clusters, e.g. while presenting the picture (see Figure 1) the following sentence (1a) was played:

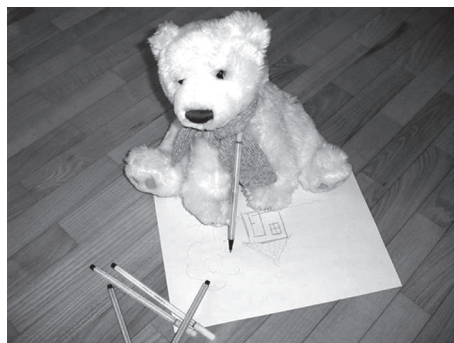


Figure 1. The picture used in the production task

- (1a) I: *Meškiukas* *piešia.*
 teddy-bear:NOM.SG draw:PRS.3
 Paklausk, ar jam patinka tai daryti?
 ‘The teddy-bear is drawing. Ask him if he likes to do that?’

The present form was used for audio stimuli, and participants were required to say the sentence with the infinitive form (1b).

- (1b) CH: Meškiuk, ar tau patinka *pieš-ti*?
 teddy-bear.VOC.SG Q YOU:DAT like:PRS.3 draw-INF
 ‘Teddy-bear, do you like drawing?’⁷

In the second part of the task the participants were required to use the future tense (2), and in the third part they were required to use the imperative (3):

- (2) I: Meškiukas *sutinka* draugus.
 teddy-bear:NOM.SG meet:PRS.3 friend:ACC.PL
 Paklausk, ar jis darys tai rytoj?
 ‘The teddy-bear is meeting friends. Ask him if he will do that tomorrow?’
- CH: Meškiuk, ar tu *sutik-s-i* draugus?
 teddy-bear:VOC.SG Q YOU.NOM meet-FUT-2SG friend:ACC.PL
 ‘Teddy-bear, will you meet friends?’
- (3) I: Meškiukai *nenori skambinti*.
 teddy-bear:NOM.PL want:PRS.3:NEG call:INF
 Paprašyk, kad jie *paskambintų*.
 ‘The teddy-bears don’t want to call. Ask them to call.’
- CH: Meškiukai, *pa-skambin-k-it(e)*.
 teddy-bear:VOC.PL PFX-call-IMP-2PL
 ‘Teddy-bears, please call!’

In total, 36 verbs with morphologically motivated clusters (across morpheme boundaries) were chosen for the production task. Transcription of the verbs used in the infinitive, future and imperative allowed us to compare the accuracy of production of morphonotactic, morphonotactic default and monomorphemic legal consonant clusters.

3.2. Repetition task

The following types of consonant clusters were included in the repetition task:

- morphonotactic default medial clusters *ks, ls, mt, št, lk* (types (b), (c) according to morphonotactic classification, see chapter 2),
- monomorphemic legal medial clusters *nk, sk, kt, st* (type (d) according to morphonotactic classification).

Clusters of both types were used within morphemes and across morpheme boundaries, as in (4)–(7).

- (4) Šitas *keks-as* skanus.
 this:NOM.SG cake-NOM.SG tasty-NOM.SG
 ‘This cake is tasty.’
- (5) Ką tu *veik-s-i* rytoj?
 what you do-FUT-2SG tomorrow
 ‘What are you doing tomorrow?’
- (6) *Nakt-į* visi miega.
 night-ACC.SG all sleep:PRS.3
 ‘Everyone sleeps at night’.

⁷ If the child used some kind of light verb construction or pronouns instead of the expected form (i.e. ‘Do you like to do it?’, ‘Do you like it?’) we asked him ‘To do what?’, requiring him to use the target verb.

- (7) Šok-ti man patinka.
 dance-INF I:DAT like:PRS.3
 ‘I like dancing’.

In total, 36 sentences with morphonotactic default and monomorphemic legal clusters were presented to the participants during the testing. The participants were asked to repeat every sentence aloud while their reproduction was recorded. The task included warm-up sentences and fillers without any clusters. By transcribing the answers and coding the errors, we were able to analyse phonological reproduction accuracy in children of different age groups and to compare the production of clusters within morphemes and across morpheme boundaries.

4. Results

4.1. Production task

As previously mentioned, the main aim of this paper is to test the statement that morphonotactic clusters are better retained during production than phonotactic clusters as they carry significant morphological information. Figure 2 indicates the percentage of incorrect responses for different cluster types according to the morphonotactic classification (see Chapter 2).

The results regarding the types of consonant clusters were as expected. It was observed that prototypical morphonotactic clusters (monomorphemic illegal) which are exclusively due to morphological operations were produced with high accuracy by TD children in various age groups. The percentage of errors made by 3–4 year-old children was 6%. More complicated for 3–4 year-old children were monomorphemic legal and morphonotactic default clusters (14% and 17% incorrect responses). 4–5 year-old TD children performed much better than the children of age group 3–4 in the production of all categories, especially prototypical morphonotactic clusters (3% incorrect responses) and monomorphemic legal clusters (3% incorrect responses). Nevertheless, they still made a lot of errors while producing morphonotactic default clusters (12%). Furthermore, when we compare the different TD groups, we can see that the number of errors decreases with age. At age 5–6 and 6–7, TD children are able to produce different types of consonant clusters rather well, that is, for morphonotactic clusters the percentage of incorrect responses is 3% in the 5–6 age group and 2% in the 6–7 age group. Incorrect production of monomorphemic legal clusters comprises 0% and 2% in the 5–6 and 6–7 age groups, respectively. 5–6 and 6–7 year-old participants produced morphonotactic default clusters with high accuracy; incorrect responses comprised only 1% and 0.5%, respectively.

In summary, these results suggest that the high ability to produce consonant clusters develops only at age 5–6. The prototypical morphonotactic clusters seem to be acquired earlier than the morphonotactic default and legal monomorphemic clusters. There are some errors observed in the youngest groups even in the production of morphonotactic clusters, however, 3–5 years old TD children produce them accurately in 95% of the trials.

Continuing the analysis of the results of the production task, we see that some more findings seem quite interesting. The results indicate that errors with different

types of clusters differ considerably between the SLI and TD children. It was shown that SLI children perform worse than all the other groups of children (even than children of the 3–4 age group) in the production of morphonotactic clusters (see Figure 2).

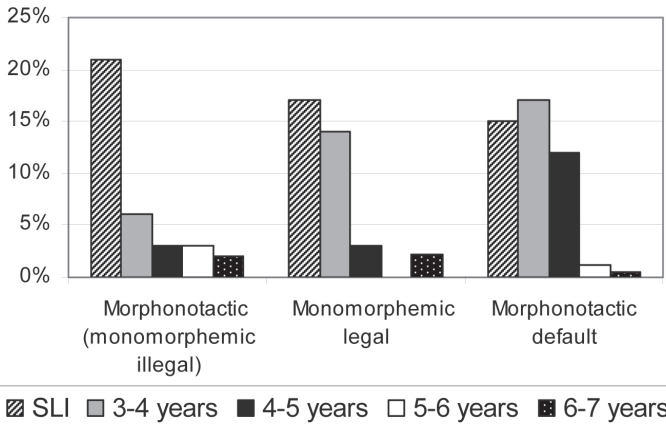


Figure 2. Distribution of errors in morphonotactic, morphonotactic default and monomorphemic legal clusters in the production task

The number of errors made by 6–7 year-old SLI children is 22%, far greater than the 6% of errors made by 3–4 year-old TD children. Monomorphemic legal clusters, which exist due to morphology as well as without interaction with morphology, also pose problems to SLI children. SLI children make more errors with monomorphemic legal clusters than do TD children of all age groups. For instance, the percentage of errors in the SLI group is 17%, whereas in the 3–4 year-old TD group it is 14%. The biggest difference is shown when we compare the results of 6–7 year-old SLI subjects (17% errors) with the results of 6–7 year-old TD subjects (2% errors). The simplest category for the SLI children is morphonotactic default clusters, i.e. the number of errors is 15%, similar to the results from the 3–4 year-old TD children (17%). Despite the fact that SLI children do make many errors with morphonotactic default clusters, this category of consonant clusters does not pose as many problems for them as prototypical morphonotactic clusters do. These results run counter to the results from the TD children, who find morphonotactic default clusters the most difficult to produce (4–5 year-olds still make errors producing them 12% of the time).

These findings justify the claim that SLI children perform better on monomorphemically legal clusters than on monomorphemically illegal clusters (Marshall, van der Lely 2006). Based on this statement, it is assumed that the SLI children, contrary to the TD children, are not that sensitive to the morphological information carried by morphonotactic clusters.

4.2. Repetition task

The results obtained during the repetition task reveal several main tendencies. Figure 3 indicates the percentage of incorrect responses for each of the target groups. The evidence suggests that SLI children performed worse than all the other groups of children in the repetition of both types of clusters. Reproduction errors of the 6–7 year-old SLI children are similar to those made by 3–4 year-old TDs. The most difficult category for SLI children was monomorphemic legal clusters existing both due to morphology and without interaction with morphology. As we see in Figure 3, monomorphemic legal clusters were incorrectly produced more often than morphonotactic default clusters (where there are only few cases of phonotactic usage) not only by SLI but by TD children as well. Although the difference between the production of the two cluster types is not significant (12% and 16% errors in the SLI group, 8% and 13% in the age 3–4 TD group, 2% and 4% in the age 4–5 TD group, 1% and 1% in the age 5–6 TD group, 1% and 0% in the age 6–7 TD group), it is hypothesized that a more detailed analysis would show a bigger difference in the production of morphonotactic and monomorphemic clusters.

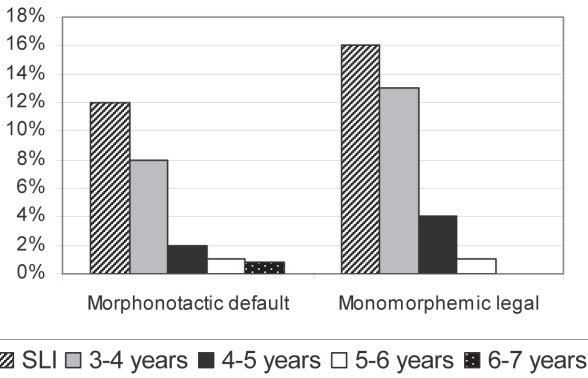


Figure 3. Distribution of errors in morphonotactic default and monomorphemic legal clusters in the repetition task

Tables 5 and 6 provide detailed comparisons of the cluster pairs that are of most interest to us, i.e. clusters within morphemes and across morpheme boundaries. Table 5 presents the morphonotactic default clusters within morphemes (phonotactic) versus across morpheme boundaries (morphonotactic), indicating the most complicated cases.

Table 5 shows the distribution of the participants' incorrect responses to the morphonotactic default clusters *ks*, *ls*, *mt*, *št*, and *lk*. It can be seen that the percentage of incorrectly produced phonotactic clusters (within morphemes) is much higher in comparison with morphonotactic clusters across morphemes. For phonotactic *ks* the percentage of incorrect responses reached about 30% and for morphonotactic *ks* 18%. Incorrect production of the consonant cluster *ls* within morphemes comprised about 15% whereas across morpheme boundaries the error rate was about 5%. The participants produced the morphonotactic cluster *mt* with maximum accuracy (0% errors), although the phonotactic cluster *mt* is more

difficult to master, as the incorrect responses comprise about 10%. The consonant cluster *št* was produced accurately (100% accuracy) when it occurred across morpheme boundaries. The phonotactic cluster *št* (within morphemes) was produced incorrectly in 5% of cases. For the cluster *lk*, however, the opposite tendency was observed. Incorrect responses with the phonotactic *lk* and morphonotactic *lk* comprised around 8% and 15% respectively. The latter item of evidence could be explained by the rare usage of the plural imperative second person in natural speech. In sum, these findings suggest that morphonotactic default clusters which occur within morphemes (phonotactic function) are more difficult to produce than morphonotactic default clusters across morphemes, except for *lk* being used for a grammatically complex and rare imperative form. Moreover, these sets of data confirm the Strong Morphonotactic Hypothesis stating that morphonotactic clusters stand a better chance of being articulated correctly than do phonotactic clusters.

Table 5. Morphonotactic default clusters

Cluster	Type	Error rate
Future		
<i>ks</i>	within morpheme	30%
	across morphemes	18%
<i>ls</i>	within morpheme	15%
	across morphemes	5%
Infinitive		
<i>mt</i>	within morpheme	10%
	across morphemes	0%
<i>št</i>	within morpheme	5%
	across morphemes	0%
Imperative		
<i>lk</i>	within morpheme	8%
	across morphemes	15%

Table 6 provides the percentages of incorrect usage of monomorphemic legal clusters which exist both due to morphology and without interaction with morphology.

Table 6. Monomorphemic legal clusters

Cluster	Type	Error rate
Imperative		
<i>nk</i>	within morpheme	15%
	across morphemes	20%
<i>sk</i>	within morpheme	25%
	across morphemes	25%
Infinitive		
<i>kt</i>	within morpheme	15%
	across morphemes	15%
<i>st</i>	within morpheme	5%
	across morphemes	5%

We can clearly see that the accuracy of production of clusters within morphemes and across morphemes is very similar (phonotactic *nk* 15% errors, morphonotactic *nk* 20% errors; phonotactic *sk* 25%, morphonotactic *sk* 25% errors; phonotactic *kt* 15%, morphonotactic *kt* 15% errors; phonotactic *st* 5%, morphonotactic *st* 5% errors). Such results are to be expected, because monomorphemic legal clusters exist both due to morphology and without interaction with morphology. This means that unlike morphonotactic default clusters, children using monomorphemic legal clusters across morphemes are not sensitive to the function fulfilled by a morpheme because the cluster occurs often in a monomorphemic context.

To summarize the results of the repetition task, children perform better at morphonotactic default clusters across morpheme boundaries than at morphonotactic default clusters within morphemes. Analysing monomorphemic legal clusters, no differences were found in children's performance at repeating clusters across morphemes and clusters within morphemes. Such results may imply that children are more sensitive to morphological information carried by clusters that appear very rarely in monomorphemic contexts.

5. Conclusion

The main goal of the present study was to demonstrate the interaction between morphology and phonology, as well as to explore the relationship between the acquisition of phonotactics and the acquisition of morphology. For this purpose, experimental data collected from 60 TD children of four age groups and 11 SLI children were analyzed.

Although this study was conducted on a relatively small number of subjects, it has explored the impact of morphology on the acquisition of phonotactics. A comparison of behaviour of morphonotactic (monomorphemic illegal), morphonotactic default and monomorphemic legal consonant clusters suggest that each of the categories poses different challenges to the child. It is demonstrated that TD children process morphonotactic clusters more accurately than morphonotactic default clusters and monomorphemic legal clusters because morphonotactic clusters have the function of co-signalling the existence of a morphological rule. Morphonotactic default clusters perform this function less adequately, as is shown by the results of the production task. As we have seen, morphonotactic default clusters pose more problems than morphonotactic clusters for TD children. Monomorphemic legal clusters do not co-signal the function of a morphological rule, thus the accuracy in producing these clusters within morphemes and across morphemes is very similar.

The findings that children with SLI perform worse than TD children on all categories of consonant clusters are most probably not surprising. As the results of repetition and production tasks indicate, 6–7 year-old SLI children match TDs who are 3–4 years old. However, children with SLI behave differently while producing consonant clusters. In contrast to TD children, for SLI children prototypical morphonotactic clusters are the most difficult, as SLI children are not sensitive to morphological information carried by morphonotactic clusters. These findings supplement the statement of Marshall and van der Lely (2006) that SLI children perform better on monomorphemically legal clusters than on monomorphemically illegal clusters.

In sum, this study strengthens and expands the Strong Morphonotactic Hypothesis, according to which speakers use morphonotactic consonant clusters as morphological boundary signals. The study supports the assumption that morphonotactic clusters are better retained in production than phonotactic clusters, as they carry significant morphological information which facilitates the processing of consonantal structures. These findings are related to the theory of coherent feature selection which claims that children, when listening to the speech of adults, first notice natural elements and those elements of the language that denote a particular function or have an exceptional feature (Dressler et al. 1995: 19, Wurzel 1984: 21).

The results also suggest directions for future research, e.g. checking whether the accuracy of production of morphologically motivated clusters holds steady when measured using bigger and more varied samples. Furthermore, some additional factors in the acquisition of consonant clusters were not explored here at all, namely the position of the cluster in the word, frequency or phonotactic markedness. Further work is planned to test these factors and to develop the test on (mor)phonotactics so that it can become a useful tool for both clinical and research purposes.

Abbreviations

2, 3	2nd, 3rd person
ACC	accusative
CH	child
CLAD	Crosslinguistic Language Diagnosis
DAT	dative
DIM	diminutive
FUT	future
GAPS	grammar and phonology screening
GAPS-LT	grammar and phonology screening test for Lithuanian
I	interviewer
IMP	imperative
INF	infinitive
IMPF	imperfect
NEG	negative
PFX	prefix
PL	plural
PRES	present tense
Q	question particle
REFL	reflexive particle
SG	singular
SLI	specific language impairment
SMH	Strong Morphonotactic Hypothesis
TD	typically developing
VOC	vocative

References

- Calderone, Basilio; Celata, Chiara; Korecky-Kröll, Katharina; Dressler, Wolfgang U. 2014. A computational approach to morphonotactics: Evidence from German. – *Language Sciences*, 46 (Part A), 59–70. <http://dx.doi.org/10.1016/j.langsci.2014.06.007>
- Dressler, Wolfgang U.; Mayerthaler, Willi; Wurzel, Wolfgang U.; Panagl, Oswald 1987. *Leitmotifs in Natural Morphology*. Studies in Language Companion Series 10. Amsterdam: Benjamins. <http://dx.doi.org/10.1075/slcs.10>

- Dressler, Wolfgang U.; Drażyk, Roberta and Danuta; Dziubalska-Kołaczyk, Katarzyna; Jagła, Ewelina 1995. On the earliest stages of acquisition of Polish declension. – *Wiener Linguistische Gazette*, 53–54, 1–21.
- Dressler, Wolfgang U.; Dziubalska-Kołaczyk, Katarzyna 2006a. Proposing morphonotactics. – *Rivista di Linguistica*, 18 (2), 249–266.
- Dressler, Wolfgang U.; Dziubalska-Kołaczyk, Katarzyna 2006b. Proposing morphonotactics. – *Wiener Linguistische Gazette*, 73, 69–87.
- Dressler, Wolfgang U.; Dziubalska-Kołaczyk, Katarzyna; Pestal, Lina 2010. Change and variation in morphonotactics. – *Folia Linguistica*, 31, 51–67.
- Freiberger, Eva Maria 2007. Morphonotaktik im Erstspracherwerb des Deutschen. – *Wiener Linguistische Gazette*, 74, 1–23.
- Gaulienė, Veronika; Giedrienė, Regina; Grikainienė, Loreta 2000. *Kalbos tyrimas*. [Language screening.] Vilnius.
- Hurch, Bernhard; Rhodes, Richard (Eds.) 1996. *Natural Phonology: The State of the Art*. Berlin: Mouton de Gruyter. <http://dx.doi.org/10.1515/9783110908992>
- Kamandulytė, Laura 2006a. Morfonotaktikos įsisavinimas. [The acquisition of Morphonotactics.] – *Žmogus ir žodis*, 8 (1), 84–88.
- Kamandulytė, Laura 2006b. The acquisition of morphonotactics in Lithuanian. – *Wiener Linguistische Gazette*, 73, 88–96.
- Korecky-Kröll, Katharina; Dressler, Wolfgang U.; Freiberger, Eva Maria; Reinisch, Eva; Mörth, Karlheinz; Libben, Gary 2014. Morphonotactic and phonotactic processing in German-speaking adults. – *Language Sciences*, 46 (Part A), 48–58. <http://dx.doi.org/10.1016/j.langsci.2014.06.006>
- Lely, Heather K. J. van der; Gardner, Hilary; Froud, Karen; McClelland, Alastair G. R. 2007. *The grammar and phonology screening (GAPS) test*. London: www.DLDCN.com.
- Leonard, Laurence B. 1998. *Children with Specific Language Impairment*. Cambridge, Mass: The MIT Press.
- Marshall, Chloe R.; Lely, Heather K. J. van der 2006. A challenge to current models of past tense inflection: The impact of phonotactics. – *Cognition*, 100 (2), 302–320. <http://dx.doi.org/10.1016/j.cognition.2005.06.001>
- Ruzaitė, Jūratė; Dabašinskienė, Ineta 2010. Specific language impairment: Adaptation of a screening test for Lithuanian. – *Darbai ir dienos*, 54, 277–300.
- Zydorowicz, Paulina 2007. Polish morphonotactics in first language acquisition. – *Wiener Linguistische Gazette*, 74, 24–44.
- Zydorowicz, Paulina 2010. Consonant clusters across morpheme boundaries: Polish morphonotactic inventory and its acquisition. – *Poznań Studies in Contemporary Linguistics*, 46 (4), 565–588. <http://dx.doi.org/10.2478/v10010-010-0028-0>
- Wurzel, Wolfgang U. 1984. *Flexionsmorphologie und Natürlichkeit*. Berlin: Akademie-Verlag.

Web references

- CLAD = Crosslinguistic Language Diagnosis. European project on the study of Specific Language Impairment and Dyslexia. <http://www.cladproject.eu/> (24.2.2015).
- Corpus of Spoken Lithuanian. <http://donelaitis.vdu.lt/sakytines-kalbos-tekstynas/> (24.2.2015).

Laura Kamandulytė-Merfeldienė (Vytautas Magnus University). Her research interests include corpus linguistics, psycholinguistics, first and second language acquisition, and specific language impairments.

K. Donelaičio 58, 44248 Kaunas, Lithuania
l.kamandulyte-merfeldiene@hmf.vdu.lt

MORFONOTAKTIKA LEEDU KEELE OMANDAMISEL ESIMISE KEELENA: TÜÜPILISE ARENGUGA VS. SPETSIIFILISE KÕNEARENGUPUUDEGA LAPSED

Laura Kamandulytė-Merfeldienė

Vytautas Magnuse Ülikool

Artikkel käsitleb fonotaktiliste ja morfonotaktiliste konsonantühendite omandamist leedu keeles. Eesmärgiks on kontrollida tugevat morfonotaktilist hüpoteesi, mille järgi keelekõnelejad kasutavad morfonotaktilisi konsonantühendeid morfoloogiliste piiride signaalidena (Korecky-Kröll jt 2014, Calderone jt 2014: 59). On oletatud, et morfonotaktilised konsonantühendid on morfeemide funktsionaalsuse tõttu keeikasutusel püsivamad kui puhtfonotaktilised ühendid.

Rikka tuletus- ja muutemorfoloogiaga fusiiivse keelena pakub leedu keel erilist huvi morfonotaktika omandamise uurimisele. Varasem, ühe leedu lapse keeleomandamise pikiuuring kinnitas hüpoteesi, et morfonotaktilised konsonantühendid omandatakse varem ja kergemini kui puhtfonotaktilised. Siinne uuring põhineb katseandmetel 60 eakohase arenguga ja 11 spetsiifilise kõnearengupuudega (SKAP) lapselt. Ülesandeks on uurida morfoloogia mõju fonotaktika omandamisele ja kontrollida seisukohta, et SKAP-lapsed tulevad monomorfeemsete legaalsete konsonantühenditega paremini toime kui monomorfeemsete illegaalsete ühenditega (Marshall, van der Lely 2006).

Uurimistulemused näitavad, et eakohase arenguga lapsed töötlevad morfonotaktilisi konsonantühendeid täpsemalt kui fonotaktilisi, sest morfonotaktilised ühendid annavad märku morfoloogiareeglist. Seevastu SKAP-lastele on morfonotaktilised ühendid kõige raskemad; põhjuseks võib pidada seda, et SKAP-lapsed ei tunne ära nende ühenditega kaasnevat morfoloogilist informatsiooni.

Võtmesõnad: morfoloogia, konsonantühend, fonotaktika, morfeem, morfeemi-piir, leedu keel

MILLEKS MULLE EESTI KEEL? RIIGIKEELE OSKUSE VAJALIKKUSEST VENE KOOLINOORTE PILGU LÄBI

Birute Klaas-Lang, Kristiina Praakli

Ülevaade. Artikli eesmärk on analüüsida vene õppekeele üldharriduskoolide 9. klasside õpilaste hoiakuid riigikeele oskuse vajalikkuse suhtes. Õpilastega läbi viidud poolstruktureeritud rühmaintervjuud toovad riigikeele õppe senistesse diskussioonidesse keeleõppija, st 15–16-aastase vene koolinoore hääle. Ainestiku diskursiivne analüüs näitab, et koolinoorte hoiakutes riigikeele õppimise ja oskuse suhtes tõusevad esile arusaamad eesti keele väiksusest, piiratud kasutusvõimalused lähikeskkonnas, eri keelte roll tööturul, aga ka õpilaste rahvuslik ja keeleline enesemääratlemine. Ühe põhitulemusena võib välja tuua, et 9. klassi koolinoore silmis on eesti keelel peamiselt vaid instrumentaalne väärtus, st keelt nähakse vahendina eneseteostuse saavutamisel. Vene koolinoor ei näe eesti keelt oma identiteedi ühe komponendina ning seab võrdluses maailma suurkeeltega eesti keele jätkusuutlikkuse (ning seeläbi eesti keele õppimise otstarbekuse) kahtluse alla.*

Võtmesõnad: motivatsioon, hoiakud, vene kool, eesti keel

1. Sissejuhatus

1.1. Rahvusvaheline taust ja teoreetiline raamistik

Käesolevas artiklis on vaatluse all põhikooli kolmanda kooliastme 9. klasside õpilaste argumentatsioon riigikeele õppe teemadel (st kuidas kõneldakse riigikeelest, selle õppimisest, kasutamisest ning tulevikuväljavaadetest) poolstruktureeritud rühmaintervjuudes. Kuigi eesti keele õpe on esmajoones Eesti kui rahvusriigi küsimus, ei saa eesti keele õpet ning sellega seotud hoiakuid vaadata lahus globaalsest keelekeskkonnast, tõstatades ka mikrotasandil küsimuse eri keelte vajalikkusest,

* Artikkel põhineb uuringul "Arvamusi ja hinnanguid riigikeele õppe korraldamise kohta vene õppekeele koolides" (Klaas-Lang jt 2014), mille eesmärk oli avada riigikeele õppimise tulemuslikkuse lahknevuse tagamaid. Uuringuga kontrollis töörühm muuhulgas ka hüpoteesi, kas riigikeele eksamitulemuste ligi kahekordne erinevus n-õ edukate ning vähem edukate koolide lõikes põhikooli lõpus võib kooli asukoha (eesti- vs. venekeelne keskkond) ning sotsiaal-majanduslike tegurite kõrval johtuda ka riigikeele õppe suhtes valitsevate hoiakute erinevusest. Artikli autorite kõrval osalesid uuringu läbiviimisel veel Alla Lašmanova ning Anni Peedisson.

nende rollidest ning tulevikuväljavaadetest. Keeleline mitmekesisustumine, keelte võimuhete ja prestiiži muutumine ning ümberhindamine, migratsioon ja turism, infotehnoloogia ning sotsiaalmeedia esiletõus on oluliselt mõjutanud ka Eesti keelekeskkondi ning nende kõnelejate hoiakuid eri keelte suhtes. Tulenevalt ühe või teise keele geopoliitilisest positsioonist ja arengust, kujunevad ühed keeled olulisemaks kui teised. Tingimustes, kus keelte varasemad rollid ja paiknemine rahvusvahelistel suhtlusväljadel on ümber mängitud (vrd inglise, hiina, hindi, araabia vs. saksa ja prantsuse keel), jäävad paljud väikekeeled senisest haavatavamasse positsiooni. Eeskätt satub ohtu nende atraktiivsus konkurentsisuutlike sihtkeeltena võõrkeeleõppes, isegi kui keel on riigikeel. Uue võõrkeele õppimine ei tähenda ainult keele struktuuri ja sõnavara omandamist, vaid tagab ka juurdepääsu ülemaailmsetesse selle keele võrgustikesse. Võrdluses maailma suurkeeltega on eestikeelsed võrgustikud oluliselt piiratumad nii globaalses, lokaalses kui ka individuaalses võtmes. Niisugustes muutunud keeleoludes on võõrkeeleõpe, õppimise motivatsioon ja strateegiad ning õppijate-kõnelejate hoiakud ühe või teise keele suhtes üha enam uurijate huviorbiidis. Eelpool toodu valguses tõusebki senisest rohkem esile küsimus keeleõppe motivatsioonist ning hoiakutest, mida käesolevas artiklis püütakse avada 15–16-aastase vene koolinoore silme läbi.

Keeleõppeprotsessi mõjutavate tegurite hulk on tohutu: need hõlmavad individuaalseid, sotsiaal-majanduslikke, kultuurilisi, kognitiivseid, psühholoogilisi ning mitte-keelelisi faktoreid (detailne ülevaade Jalkanen, Ruuska 2007: 27–44). Nimetatuist olulisimad on tundetegurid ehk motivatsioon, hoiakud, õppija keeleline minapilt ning keelehirm, mis mõjutavad kõige otsesemalt seda, milliseks kujuneb õpilase suhe õpitavasse keelde ja õppeprotsessi ning valmisolekusse keelt suhtluses kasutada (Jalkanen, Ruuska 2007: 27, vt ka Pihko 2007). Viimast rõhutab ka “Euroopa keeleõppe raamdokument” (EKR 2007: 110): “Keelekasutaja või -õppija suhtlustoimingut ei mõjuta mitte ainult tema teadmised, arusaamad ja oskused, vaid ka tema eneseteadvusega seotud tegurid – hoiakud, motiivid, väärtushinnangud, veendumused, kognitiivsed stiilid ja isiksustüüp...”.

Motivatsioon ning hoiakud ongi võõrkeele omandamist mõjutavad põhitegurid. Mõistetena on mõlemad üksjagu problemaatilised ning nende defineerimine pole alati selge (vt Klaas 1992, Dörnyei 1998: 117–122, Gardner 2010: 1–29). Võõrkeele õppimise, õpetamise ning keeleõppe motivatsiooni süstemaatiline uurimine algas 1970. aastatel ning sellealase kirjanduse hulk on väga suur (vt nt Dörnyei 1998, 2001, Dörnyei, Schmidt 2001, Dörnyei 1998, 2001a, 2001b, Mitchell, Myles 2004, Ellis, Barkhuizen 2005, Ellis 2008). Valdkonnale aluse pannud Kanada teadlase Robert Gardneri (1985) keeleomandamise sotsiaalhariduslik mudel (ingl *socio-educational model*) on endiselt enim rakendatav teoreetiline raamistik, kuigi seda on ka rohkesti kritiseeritud. Gardner (1985, 2010) eristab integratiivset ja instrumentaalset motivatsiooni. Integratiivne motivatsioon tähendab õppija positiivset huvi sihtkeele kõnelejate ja kultuuri vastu, keelt õpitakse keelelise ja kultuurilise lõimumise soovist. Keeleõppe eelduseks on õppija positiivne hoiak ning soov olla interaktsioonis selle keele kõnelejaskonnaga. Seevastu instrumentaalne motivatsioon on seotud keeleõppe praktiliste eesmärkidega, mille taustal on uue keele oskusest saadav praktiline kasu ja soodustused. Keele oskamine on seega vahend, mille kaudu võib keeleõppija püüda saavutada majanduslikku kasu või paremat sotsiaal-majanduslikku positsiooni.

Gardneri oponendid näevad mudelit piiratud raamistikuna, mis jätab kõrvale muud keeleõpet mõjutavad tegurid (nt Dörnyei 2003: 4–23), samuti on kriitikat pälvinud instrumentaalse-integratiivse motivatsiooni suhe ning on esitatud kaalukaid vastuargumente integratiivse motivatsiooni kui keeleõppeprotsessi ainuliselt tulemuslikult suunava motivatsiooni suhtes. Mitmed uurimustulemused on näidanud (vt nt Knight 2013: 18–23), et võõrkeele õpimotivatsiooni taustal on oluliselt rohkem tegureid, muuhulgas füüsiline ja sotsiaalne keskkond, kus õpe aset leiab, õppematerjalid, õpetaja isik jne. Keeleõppekeskkonnas, kus õppijate kontaktid seda keelt emakeelena kõnelejatega on piiratud või puuduvad sootuks – Eesti näitel mitmed keelekeskkonnad Ida-Virumaal – võibki integratiivne motivatsioon jääda tagaplaanile, kuna keeleõppijal puudub tema keelelist arengut toetav “keelesild” emakeelekõnelejatega, kuid keeleõppeprotsess võib sellele vaatamata edukas olla. Artikli tulemustest ette rutates võib väite kinnituseks tuua näiteks üldhariduskooli Ida-Virumaal, mille õpilastel puuduvad kooliväliselt igasugused kontaktid eesti keele kõnelejatega, ometi on see kool riigieksami tulemuste poolest esirinnas. Seega ei pruugi integratiivse motivatsiooni puudumine olla keeleõpet pärssiv tegur.

Niisama oluline mõiste on ka demotivatsioon (lähemalt Dörnyei 2001a: 50–51, 2001b: 141–155) ehk õppimist takistavad tegurid, nagu näiteks oskamatu õpetaja, viletsad kooliruumid, igav õpik ja halvasti koostatud õppematerjalid, aga ka negatiivne hoiak õpitava keele ja selle kõnelejate suhtes, õppeaine kohustuslikkus jne. Keeleõppe kohustuslikkuse avalikud diskussioonid on ilmselt tuntuimad Soome riigi keelevaidluste näitel, kus teise riigikeele ehk rootsi keele, Soomes tuntud kui *pakkoruotsi*, kohustuslik õppimine on Soome avalikkuses olnud intensiivseks aruteluküsimuseks viimastel aastakümnetel.¹ Käesoleva artikli autoritele teadaolevalt pole demotivatsioonitegureid vene koolide kontekstis Eestis uuritud, küll aga võib intervjuuainestikule tuginevalt välja tuua õpilaste kriitika kasutatavate õppematerjalide, grammatikakeskse õpetamise ning ootuspäraselt ka aine kohustuslikkuse ning riigikeele pealesunnitud õppe kohta, kui nimetada mõningaid keeleõppimist pärssivaid tegureid (vt lähemalt Klaas-Lang jt 2014).

Seega pole keeleõppes vähem olulised ühiskonnas või vahetus õpikeskkonnas valitsevad hoiakud ning arusaamad keeleõppe tähtsuse ja vajalikkuse kohta ning meedia roll arvamuse kujundajana, kinnistajana ning taastootjana, mis ulatuvad järellainetena ka kooliruumi ning keeletundi. Teine oluline mõiste, mida artiklis käsitletakse, ongi hoiak (ingl *attitude*). Ingliskeelses kirjanduses on hoiakut defineeritud näiteks kui tunnet mingi psühholoogilise objekti poolt või vastu, rõhutades sellega positiivseid ja negatiivseid emotsionaalseid vastuseid, mida hoiak väljendab (vt Garrett 2010: 19 ja sealsed viited). “Euroopa keeleõppe raamdokument” mõistab hoiakut keelekasutaja või -õppija avatusena uutele kogemustele, teistele inimestele, ideedele, rahvastele, ühiskondadele ja kultuuridele ning huvile nende vastu, valmisolekut suhestada oma kultuurilised seisukohad ja väärtushinnangute süsteem ning olla valmis ja suuteline kaugenema juurdunud hoiakutest ja saada aru kultuurierinevustest (EKR 2007: 123–124). Kui motivatsioonist rääkides rõhutatakse selle dünaamilisust ja situatsioonikesksust, siis hoiakute puhul on küsimus püsivamast ilmingust, mis võib küll aja jooksul muutuda.

¹ Teemast lähemalt: Kaksi kantaa pakkoruotsista. <http://www.hs.fi/sunnuntai/a1305708043006> (10.10.2014).

1.2. Eesti vene õppekeelega koole käsitlev diskursus

Viimased viisteist aastat on Eesti vene õppekeelega koolide olukord leidnud mitmekülgset kajastamist nii laiale avalikkusele suunatud meediaartiklites kui ka rohketes uurimustes, mis analüüsivad riigikeele õpet, selle teoreetilist ja keelepoliitilist alust, realiseerumist praktikas, samuti suhtumisi ja hoiakuid ning muid keeleõpet mõjutavaid tegureid. Tähelepanu keskmesse on tõusnud vene kool pärast üleminekut osalisele eestikeelsele aineõppele gümnaasiumiastmes, mille tulemuslikkus on viimasel paaril aastal leidnud käsitlemist mitmetes raportites (nt Klaas-Lang jt 2014, Metslang jt 2013).

Vene õppekeelega koolide olukorda ja riigikeele õpet on käsitletud mitmetes uurimustes (vt ülevaated Vaiss 2009, Klaas-Lang jt 2014). Eraldi tasub nimetada eesti keele seisundit ja õpetamise olukorda analüüsivat ülevaadet (Vare 2004), TNS Emori uuringuid eestikeelse aineõppe hetkeolukorrast ja vajadustest ning mitte-eestlaste teadlikkusest ja suhtumisest vene õppekeelega koolide üleminekusse eestikeelsele aineõppele (TNS Emor 2006, TNS Emor 2008), Tallinna Ülikooli uurijate koostatud venekeelse kooli kakskeelse õppe uuringut (Metslang jt 2013), eestivenelaste keeleoskuse ja gümnaasiumireformi suhtumiste uuringut (Proos 2005), eestivene noorte positsiooni ja ootuste kajastust (Proos, Pettai 2008).

Keelehoiakute uurimisest eesti keeleteaduses väärib esiletõstmist Kadri Koreiniku doktoritöö (2011) ning artiklid, milles autor analüüsib keeleideoloogiaid ning keelelisi representatsioone lõunaeesti keele avalikus diskursuses. Koreiniku ja Praakli ühisartiklis (2013) oli vaatluse all keele kestlikkuse ja kao diskursiivne konstrueerimine Saksamaa ja Soome eestlaste ning setu ja võru keele kasutajate poolstruktureeritud intervjuude näitel. Martin Ehala ning Anni Tammemägi (2012) on uurinud eesti koolinoorte hoiakuid inglise keele kasutamise suhtes, valmisolekut muulaste keeleliseks integratsiooniks ning eesti keele kasutamist segaperekonnas. Natalia Vaiss (2009) on vaadelnud vene õppekeelega põhikooli õpilaste ja õpetajate hoiakuid eestikeelse aineõppe suhtes enne gümnaasiumireformi.

2. Uuringu meetod, informandid ja aineestik

Motivatsiooni ning hoiakute uuringutes rakendatakse aineestiku kogumisel nii kvantitatiivseid kui ka kvalitatiivseid meetodeid, põhiliselt kas küsimustikke või intervjuusid (vt nt Vaiss 2009, Ehala, Tammemägi 2012, Koreinik, Praakli 2013). Käesoleva artikli aineestik on kogutud otsese meetodiga (ingl *direct approach*, vt lähemalt Garrett 2010: 37–38), milleks on poolstruktureeritud rühmaintervjuud. Üks tavalisemaid intervjuude analüüsimise meetodeid on intervjuudes osalenute öeldu taandamine sisukategooriateks ning informatsiooni koondamine erinevate teemade alla (lähemalt Koreinik, Praakli 2013: 135). Artikli analüüsisosas lähtumegi sellest, mida ühe või teise teemaga seoses intervjuudes räägiti, eesmärgiga näidata, millisenäeb vene koolinoor eesti keele oskust ja vajalikkust.

Uuringu informandid on 15–16-aastased koolinoored, kes õpivad vene õppekeelega põhikoolis, st koolis, kus suurem osa õppetööst toimub vene keeles. Õpilased olid valitud õppeasutuse 2013. a riigieksami tulemuste järgi kaheksast põhikoolist (neli edukamat ja neli vähem edukamat kooli), mis paiknesid Tallinnas, Tartus

ning Ida-Virumaal. Uuringus osalenud õppeasutused tegutsevad geograafilise asukoha ning keskkonna rahvusliku struktuuri võrdluses üksteisest vägagi erinevas keelekeskkonnas. Uuringu tulemustest rääkides toome eraldi välja edukate ning vähem edukate koolide õpilaste seisukohad. Kokkuleppel õppeasutustega säilitame artiklis koolide ning õpilaste anonüümsuse ning õppeasutuse nimesid ei maini. Eesti keele teise keelena lõpueksami tulemuste detailse statistika koolide lõikes leiab sihtasutuse Innove veebilehelt.² Artikli autorid rõhutavad, et artikli eesmärk on tuua kuuldavale vene noore häälel, mitte koole omavahel võrrelda.

Uuringuks vajalik aineistik koguti poolstruktureeritud rühmaintervjuude vormis, järgides kvalitatiivse uuringu tavapäraseid protseduure. Intervjuude pikkus oli 90 minutit ning nendes keskenduti neljale põhiteemale:

- 1) riigikeele kasutus õpilaste argielus;
- 2) õpilaste hinnang oma riigikeele oskuse kohta;
- 3) riigikeele õppimise olulisus;
- 4) õpetaja, kooli ja riigi roll riigikeele õppes.

Intervjuud viidi läbi vene keeles ning need toimusid klassiruumis. Kuigi kõik intervjuud lähtusid samadest põhimõtetest, ei saa välistada mõningast vestlusteemade varieeruvust. Nii nähtub intervjuuainestikku koolide kaupa kõrvutades, et eeskätt vähem edukate koolide õpilased toovad omaalgatuslikult sisse eesti ja vene noortevahelised suhted, keelekeskkonna olulisuse keeleõppes ning eesti keele tulevikuväljavaated. Samuti peatusid nad teiste koolide eakaaslastest rohkem perede sotsiaal-majanduslikul olukorral ning erinevatel võimalustel koolivälise keeleõppe toetamiseks. Kuigi õpilaste sotsiaal-majanduslik taust ei kuulunud otseselt intervjuude teemade hulka, näitab küsimuse esiletõus probleemi seost keeleõppe edukusega. Kõneka vestlusteemana tõuseb esile eesti keele tuleviku ning Eestist emigreerumise argumentatsioon. Kuna empiiriline aineistik on kogutud rühma-intervjuude vormis, tuleb arvesse võtta võimalikku rühmadünaamikat küsimustele vastamisel (klassikaaslastevahelised suhted, mõne õpilase domineerimine või tagaplaanile jäämine jne).

3. Tulemused

3.1. Riigikeele oskus ja kasutus õpilaste igapäevaelus

Informantidele esitati kolmeosaline põhiküsimus “Kellega, kui tihti ja kus räägite eesti keelt?” eesmärgiga hinnata eesti keele kasutamise sagedust õpilaste argisuhtluses ning selgitada välja eesti keele kasutusväljad. Kuna küsimuse eesmärk oli saada üldpilt koolinoorte keelekeskkonna olemusest, piirduti selle määratlemiseks rühma-intervjuuga. Detailsem analüüs, mis võtaks arvesse keelekasutusväljade võimalikult laia spektrit (nt keelevalikud kodus ja suhtlusvõrgustikes ning avalikus ruumis), vajab ankeetküsitluse vormi ning kvantitatiivseid meetodeid. Põhikoolis õppiva vene koolinoore eesti keele kasutusvõimalustest võib üldistada tüüpilised keelekasutusväljad, kus riigikeele kasutamine piirdub üldjuhul õpilase eesti keele tundidega, vahel ka ainetundidega (v.a Tartu ja Tallinna koolinoorte puhul).

² <http://www.innove.ee> (10.10.2014).

- (1) Eesti keeles räägime eesti keele tunnis.
- (2) Mõnikord saab poes kasutada, aga mitte Narvas.

Eesti keele kasutamise võimalused väljaspool koolikeskkonda johtuvad suuresti haridusasutuse geograafilisest asukohast. Sestap on ootuspärane, et kõige rohkem võimalusi eesti keele kasutamiseks on Tartu õpilastel, kelle tüüpvastuseks eesti keele kasutuse kohta väljaspool klassi- ja kooliruumi oli *kasutame igal pool*. Ka Tallinna kooliõpilastel on oluliselt rohkem eestikeelseid kontakte kui näiteks Narva õpilastel. Eesti keelt kasutatakse Tallinnas ja Tartus ka huviringide keelena (*Käin võrkpallitrennis, seal saan keelega hästi hakkama; Käin tantsuringis ja seal suhtlen. Trennis suhtleme palju*). Samas on ka Tallinnas oluliseks mõjuriks kooli geograafiline asukoht ning õpilase elukoht, mis tingib ja kujundab õpilase suhtlusvõrgustike struktuuri (*Meie linnaosas on ikka eestlasi rohkem, eesti keelt on igal pool kuulda*). Võiks eeldada, et eestikeelne meedia, mis on ju kättesaadav elukohast sõltumata, pakub samuti keelekeskkonda. Paraku õpilaste vastustest eestikeelse meedia tarbimine ei nähtu.

Samas ei pruugi keelekeskkonna puudumine väljaspool kooli tingimata välistada riigikeele head oskust. Edukamate koolide hulka kuulub näiteks Ida-Virumaa kool, kus 9. klasside õpilastel puuduvad kooliväliselt peaaegu igasugused eestikeelsed kontaktid. Selle kooli õpilaste vastustest nähtub, et eesti keele peaaegu ainus kasutusruum on küll kool, kuid peale eesti keele tundide ka ainetunnid (nt inimese-õpetus), kus aine mõistestik esitatakse paralleelselt eesti ja vene keeles. Samuti pakub mitmekesine projektitegevus õpilastele lisavõimalusi riigikeele suhtlusruumi laiendamiseks ja õpimotivatsiooni tekkeks: *Mul oli selline juhtum, et meil koolis olid külalised Tartust ja üks poiss elas minu juures. Kui ma sain aru, et ma ei räägi hästi eesti keelt, siis kohe tekkis motivatsioon keelt õppida*.

Riigieksami tulemuste põhjal vähem edukate koolide õpilastel puuduvad intervjuudes öeldule tuginedes võimalused eesti keele kasutamiseks väljaspool koolitundi.

- (3) Räägime ainult eesti keele tunnis, saame rääkida erinevatel teemadel, mitte ainult sellest, mida oli vaja õppida.
- (4) Eesti keeles räägime ainult õpetajaga koolis. Kellega meil siis veel rääkida? Rohkem polegi kellegagi rääkida. Võib-olla ainult poes mõnikord läheb vaja.
- (5) Eesti keeles räägime ainult koolis. Mul on kaks eestlasest sõpra, nendega üritan eesti keeles rääkida.
- (6) Eestlastest sõpradega räägime ka inglise keeles, sest seda keelt oskame nii mina kui tema paremini kui mina eesti keelt või tema vene keelt.

3.2. Õpilaste enesehinnang eesti keele oskuse kohta

Teine põhiküsimus hõlmas 9. klasside õpilaste hinnangut enda eesti keele oskuse kohta. Küsimuse eesmärk oli uurida, millisel tasemel olevana näevad õpilased enda eesti keele oskust ning mil määral nad on suutelised eesti keeles suhtlema. Küsimus esitati vormis “Kuidas hindad enda eesti keele oskust?”. Rõhutame, et õpilastega vesteldi vene keeles ning nende eesti keele oskust eraldi ei testitud. Seega põhinevad vastused õpilaste subjektiivsetel hinnangutel. Peamiseks arutelüküsimuseks

kujunes eesti keele valdamine osaoskuste kaupa (kõnelemine ja kirjutamine), millest omakorda kasvasid välja intensiivsed arutelud suhtlusoskuse arendamise tähtsusest ning kriitilised seisukohad eesti keele õppe liigse grammatikakesksuse kohta. On märkimisväärne, et õpilased hindavad eesti keele oskust üsnagi heaks, kõrge enesehinnang aga ei pruugi olla alati adekvaatne.

Edukamate koolide 9. klasside õpilased hindavad oma eesti keele oskuse väga heaks (*Oleme oma eesti keele oskusega väga rahul*). Rahulolu keeleoskusega selgitatakse tugevas koolis õppimisega, eesti keele õppe alustamisega 1. klassis, eestikeelse aineõppe ning nõudlike õpetajatega, aga ka õpilaste enda valmisolekuga eesti keele intensiivseks õppimiseks.

- (7) Meie eksamitulemused on head, sest meil on väga nõudlikud õpetajad, nad nõuavad palju. Meie ise tahame kooli hästi lõpetada, sellepärast õpime palju.
- (8) Oleme väga rahul, õpime eesti keelt esimesest klassist saadik, meil on ka teised ained eesti keeles.
- (9) Kui me eestlastega suhtleme, siis nad imestavad, et me ei ole eestlased ja et meie eesti keel on nii hea. Oleme selle üle uhked.

Eesti keele oskuse head taset seostavad õpilased piisava praktikaga. Eeskätt Tartu õpilased näevad eestikeelses keelekeskkonnas teiste piirkondade noorte ees selget eelist riigikeele paremaks omandamiseks (*Meil on kindlasti lihtsam ka tulevikus tööle saada kui teistel*). Riigikeele eksami häid tulemusi seostatakse eesti keele "kohaloluga" kooliruumis ning eestikeelse aineõppega.

- (10) Meil siin on head eksamitulemused, sest meil on koolis palju eesti keelt kuulda ja algkoolis ka õpitakse eesti keelt palju ja teised ained on ka eesti keeles.

Riigieksami tulemuste põhjal vähem edukate koolide õpilased hindavad samuti oma eesti keele oskuse heaks, kuid arutledes kriitiliselt keele osaoskuste teemadel, nimetavad ka nemad eesti keele õppes liigset grammatikakesksust, mis jätab tagaplaanile suulise eneseväljenduse arendamise. Õpilaste keeleoskust varjutab suhtlusvõimaluste ning toetava eestikeelse keskkonna puudumine. Kõik õpilased tõdesid riigikeele kasutusvõimaluste piiratust (*Oskame küll, aga kus me seda ikka räägime*) ning olid sellepärast ka oma suulise eneseväljendusoskuse suhtes pigem kriitilised.

- (11) Keeleoskusega ei saa rahul olla ainuüksi selle pärast, et me küll õpime eesti keelt, aga me ei räägi seda kusagil ega saa ka harjutada.
- (12) Ei ole rahul [eesti keele oskusega], sest kui trammis sõidan ja kuulen, kuidas eesti memmekesed räägivad, siis aru saan, aga vastu rääkida ei oska, ainult naeratan. Kui midagi küsitakse, siis ainult naeratan.

Intervjuudest võib järeldada, et keelekeskkonna puudumise kõrval ei toeta vähem edukate koolide õpilaste eesti keele õpet ka nende kodune keeleolukord, kus vanematel puudub laste kooliõpingutes aitamiseks piisav eesti keele oskus (*Kodus keegi ei aita; Mul isa aitab mõnikord; Mõnikord aidatakse, aga vähe*). Intervjuudes osalenud nelja kooli õpilastest vaid üksikute õpilaste kodus ja sugulasvõrgustikus suheldakse mingilgi määral eesti keeles (nt ühega vanematest või vanavanematega).

3.3. Riigikeele õppimise olulisus õppija vaatenurgast

Õpilaste hoiakute väljaselgitamiseks esitati põhiküsimus sõnastuses “Miks on vaja eesti keelt õppida?”, “Milleks läheb Sinu arvates vene noorel eesti keele oskust vaja?”. Arutelude käigus kasvas esitatud põhiküsimusest välja mitu alateemat, nt eesti keele “väiksus” ja keele tulevik, väljaränne, vene identiteet Eestis ja selle säilitamine, Ida-Virumaa koolides ka eestlaste ja venelaste suhted.

Riigikeele eksami tulemuste põhjal edukamate ja vähem edukate koolide 9. klasside õpilaste argumentid eesti keele õppimise vajalikkuse ning kasulikkuse teemadel ei erine üksteisest. On märkimisväärne, kuivõrd selgelt pragmaatiline on õpilaste suhtumine riigikeele oskuse vajalikkusesse ja kuidas õppimise kannustajaks on peaaegu eranditult kõigis vastustes instrumentaalne motivatsioon. Siinjuures tuleks juhtida tähelepanu, et ka artikli sissejuhatuses viidatud uuringus osalenud õpetajad ja koolijuhid ei pööranud tähelepanu integratiivsele motivatsioonile riigikeele õppes (vt Klaas-Lang jt 2014).

Edukamate koolide õpilaste vastused demonstreerivad riigikeele õppimise kohustust (*See on ju riigikeel, me peame seda oskama*), eesti keele õppimise vajalikkust seostatakse ka lähituleviku plaanide ning gümnaasiumisse astumisega (kuigi mitte kõik potentsiaalsed gümnaasistid ei plaani haridusteed jätkata Eestis). Eesti keele õppimises nähakse kõige selgemat “kasu” selles, et edukas koolilõpetamine ning võimalikult heal tasemel eesti keele oskamine avab ukSED õpingute jätkamiseks gümnaasiumiastmes või ülikoolis.

- (13) Et eksam ära teha ja edasi õppida.
- (14) Et gümnaasiumis või kutsekoolis õppida.
- (15) Et saada gümnaasiumis hakkama.
- (16) Et pärast ülikoolis saaks õppida.
- (17) Et suhelda selles keeles ja/või õppida edaspidi samas keeles.

Ka vähem edukate koolide õpilaste argumentatsioonist nähtub selge pragmaatilisus, kus riigikeele õpe on üks kohustuslikest ainetest teiste seas, milles eksami sooritamine on vajalik põhikooli lõpetamiseks ning õpingute jätkamiseks gümnaasiumis. Eesti keele õppimisel on esiplaanil eeskätt instrumentaalne motivatsioon, kus riigikeele õppimist nähakse kohustusliku ning põhikooli lõpetamiseks vajaliku õppeaine või keelena.

- (18) Teha eesti keele eksam ära, et saaks kooli lõpetada, ja kuna meil valikut pole, me peame selle eksami tegema.
- (19) Et eksam ära teha ja edasi õppida.
- (20) Et gümnaasiumis hakkama saada.
- (21) Et paremasse gümnaasiumisse sisse saada.

Vaid vähesed õpilased seostavad eesti keele vajalikkust elamisega Eesti riigis. Eesti keel on Eesti territooriumil kõneldav põhikeel, mistõttu on vaja laiema üldsusega või avalikus ruumis suhelda eesti keeles: *Kui selles riigis elada, siis peaks keelt oskama; Eesti keelt on vaja selleks, et suhelda müüjatega*. Oluliseks motivaatoriks riigikeele õppimisel peetakse ka Eesti kodakondsuse saamist: *Tahan saada Eesti kodakondsust ja sellepärast õpin eesti keelt*. Ka eelnenud vastused demonstreerivad siiski pigem vajadust ja kohustust, mitte aga tahtmist ja võimalust enesearenguks.

Siiski töid üksikud õpilased välja eri kultuuridega suhtlemise aspekti, mida võiks mõõndustega seostada integratiivse motivaatoriga: *Erinevates keeltes rääkides saab huvitavate inimestega suhelda.*

Eesti keele valdamist peetakse oluliseks ka tööturul parema positsiooni saavutamiseks (*Keel on tähtis, et tööd saada; Et saada paremat töökohta. Keelest sõltub palju, tutvused on ka tähtsad*), kuid seda varjutab 9. klasside õpilaste selge kahtlemine vene emakeelega isikute konkurentsivõimes tööle kandideerimisel. Selgelt on kinnistunud tõekspidamine, et vene keele kõnelejatele seatakse Eesti tööturule pääsemiseks piiranguid. Tüüpkommentaarina kõigest intervjuudest kõlama jäänud arvamus *Kui on valida eestlasest või venelasest töötaja vahel, siis valitakse eestlane, hoolimata keeleoskusest* markeerib selgelt 9. klasside õpilaste ettekujutust Eesti tööturu toimimisest. Eesti keele vajalikkust seostatakse ka isikliku majandusliku kasuga ning selle saavutamiseks peetakse oluliseks alustada õpinguid kohe eestikeelses koolis.

- (22) Et pärast saaks parema töökoha, aga seal on ka tutvused tähtsamad. Meid sinna tööle nagunii ei võeta, venelasi ei võeta heameelega tööle. Isegi kui sa eesti keelt hästi oskad ja sul on kõrge kateegooria, sind ei võeta tööle, kui on valida sinu ja eestlase vahel.
- (23) Olen kindel, et mind ei valita, sest mul on vene perekonnanimi. See on kindlasti takistuseks. Mul tahtsid vanemad perekonnanime vahetada, aga ma ei taha. Kindlasti võetakse panka tööle see inimene, kelle perekonnanimi on Sibul, mitte Petrova. Minu peres on selline kogemus.

Samas nähtub argumentatsioonist ka inglise keele olulisus, mida peetakse (tasuva) töökoha saamisel eesti keele valdamise kõrval paralleelselt oluliseks.

- (24) Hea töökoha saab, kui oskad eesti ja inglise keelt, muidu seda keelt pole vaja, kui ära lähed. Praegu kõik lähuvad ära – eestlased ja venelased, sest tööd ei ole. Need, kes tahavad siin elada ja keelt hästi osata, peaksid kohe eesti kooli minema, siis on palju lihtsam.

Õpilaste argumentatsioonis tõuseb esile ka eesti keele väiksuse diskursus ning sellega seoses eesti keele väljasuremise või kadumise oht. Hoolimata eesti keele väga heast valdamisest, seadsid ka edukamate koolide õpilased eesti keele õppimise olulisuse keele väiksuse tõttu kahtluse alla. Samas kasutatakse intervjuudes adjektiivi *väike* üldsõnaliselt, toomata välja konkreetseid mõjureid, mis eesti keele kestlikkuse ohtu seaksid. Eesti keele kadumist nähakse selle “väiksuse” tõttu kui paratamatut protsessi, mis *nagunii juhtub* ning millega seatakse kahtluse alla keele õppimise otstarbekus (*Milleks õppida midagi, mis nagunii ära kaob?*). Niisugused seisukohad olid pigem üldised, sõltumata üldhariduskooli asukohast ja riigikeele eksamil saavutatud tasemest. Õpilaste argumentatsioonis peegeldub keeleõppe instrumentaalne funktsioon: kui keeles ei nähta otsest kasu, ei peeta selle õppimist vajalikuks.

- (25) Eesti keelt pole vaja, see on nii väike keel, et see keel kaob varsti ära. Pole erilist mõtet seda keelt õppida.

Eesti keeles nähakse ka vene identiteedi ohustajat. Palju arutelu tekitab põhiküsimusest välja kasvanud küsimus, kas vene noor peaks õppima eesti- või venekeelses

üldhariduskoolis. Kooli valikus nähakse seost rahvusliku identiteedi ning sellest võimaliku loobumise vahel. Eesti koolis õppimises nähakse ohtu enda vene identiteedi säilimisele ning eestikeelse kooli valikut peetakse võimalikuks vaid siis, kui *tahad eestlaseks hakata*. Õpilased on ka arvamusel, et

- (26) Selline vene kool, kus eesti keelt on palju, on parem kui eesti kool. Laps peaks jääma ikka vene kultuuriruumi, sest mentaliteet on teine.

Sellelaadsed õpilasi vägagi erutavad arutelud toovad esile veel ühe dimensiooni, nimelt küsimuse vene noorte keelelisest ja rahvuslikust kuuluvusest ning mitmekeelsusest laiemas mõttes. Tõstatub ka küsimus, kas vene noore keelelise mitmekeelsuse üheks komponendiks võiks olla eesti keel. Intervjuud osutavad, et eesti keele õppimist või kõnelemist ei osata näha õpilase mitmekeelset ning -kultuurilist identiteeti kujundava komponendina. Enamiku intervjuudes osalenute jaoks pole eesti keelel vene ja inglise keele kõrval muud rolli kui üks kohustuslik õppeaine teiste seas. Õpilaste argumentatsioonist võib aimata, et võimalikuks peetakse vaid ükskeelsust ning ühekultuurilisust, mis välistab kuulumise kahte kultuuri. *Eestlaseks hakkamine* seostub vene emakeelega noore jaoks identiteedi ja rahvuse vahetamise, mitte aga kodanikuks olemisega. Võimalus, et *eestlane* võib tähistada ka vene emakeelega isikut, kes samastab ennast nii vene kui ka eesti kultuuriga, on mudel, mille peale vene koolinõor pole ilmselt mõelnud. Vene emakeelega noore kakskeelsus kui identiteedi komponent ja mitmekultuurilisus kui väärtus on leidnud küll tee koolide visioonidokumentidesse ja arengukavadesse, kuid tegelikus elus seda ei tunnetata (vt Klaas-Lang jt 2014). Vähemalt intervjuudes ei pööranud ükski vastanute grupp (õpilased, õpetajad, koolijuhid) taoliste eesmärkidele peaaegu mingit tähelepanu.

Olulise erinevusena eespool analüüsitud edukate koolide rühmast tõusevad vähem edukate koolide õpilaste vastustes kohati esile arvamused, kus eesti keele oskamist seostatakse õigusega elada Eestis. Eesti keele õppimises nähakse kohustusi Eesti riigi ees ning riigikeele oskuses Eestisse jäämise alust: *See keel on siin ja kui sa selles keeles aru ei saa, siis kuidas sa siin elad?* Riigikeele mitteoskamine võib õpilaste arvates kaasa tuua lausa drastilisi tagajärgi.

- (27) [Keelt on vaja õppida ja osata selleks,] et meid ei pekstaks Eestist välja.

Eesti keele oskust peetakse aga tööturukeelena igati oluliseks. Ühtlasi ei pea vene emakeelega teismelised riigisektoris töötamist venekeelse inimese jaoks üldse võimalikuks.

- (28) Eestis ilma eesti keeleta eriti hakkama ei saa, kui tahad enamat kui Sillamäel kükitada.
(29) Selleks et pärast hästi elada – hästi teenida, saada head tööd, kui sa siin Eestis elada tahad.
(30) Kui keelt ei oska, siis saad miinimumpalka.
(31) Eesti keele oskusest sõltub ka tulevane töökoht, aga see on vähem tähtis kui head tutvused. Siis kui sul on häid tuttavaid, siis saab ka hea töökoha äris. Riigiasutuses küll keegi töötada ei taha, ja meid sinna ei võeta ka.

Samas on ka Ida-Virumaa 9. klasside õpilased veendunud selles, et vene emakeelega inimestel on Eesti tööturule ilma tutvusteta võimatu siseneda, eriti riigisektoris.

Õpilaste argumentatsioonist nähtuvad ka selged ettekujutused era- ja riigisektori toimimise erinevustest, mida seostatakse samuti riigikeele oskusega.

- (32) Kui on tutvused, siis saab igale poole tööle minna, siis pole ka keeleoskus tähtis. Erasektorisse saab igale poole tööle minna, kui tutvused on. Riigi-asutustes kontrollitakse kõike.

Eesti keele õppimise vajalikkust või mittevajalikkust seostatakse ka võimaliku väljäreändega. Eesti keele õppimine pole oluline, kuna *välismaal ei lähe seda nagunii vaja*.

- (33) Paljud tahavad Eestist ära minna, Euroopasse või Venemaale, siis ikka inglise keelt läheb rohkem vaja. Seal eesti keelt ei lähe nagunii vaja.

3.4. Eesti versus inglise keel

Empiiriline aineestik kõneleb selgelt inglise keele kõrgemast prestiižist ning olulisemast rollist koolinoore argielus. Vene koolinoorte keelevelikud markeerivad laiemalt kahe rühma – eesti- ja venekeelse – suhtlusmallide muutumist, kus kahe rühma suhtlust võib vahendada hoopiski inglise keel. Pole ilmselt vale väita, et vene keele kõrval on noortel inglise keelega tihedamad kokkupuuted; võimalik, et seda osatakse suhtluskeele tasandil ka paremini kui eesti keelt ning inglise keelega ollakse ainuüksi popkultuuri ning sotsiaalmeedia mõju tõttu vahetumas kontaktis kui eesti keelega. Intervjuudes antud vastustest aimdub vene noorte valmisolek suhelda eesti eakaaslastega inglise keeles ning oletatavasti on see valmisolek vastastikune. Seega ei pruugi inglise keele eelistamine johtuda vaid vene noorte soovist, vaid ka eesti koolinoorte silmis võib inglise keelel olla suurem väärtus (noorte keelehoiakutest lähemalt Ehala, Tammemägi 2012), mistõttu ühiseks suhtluskeeleks valitakse “suurem” keel.

- (34) Alati on nii, et kui kohtuvad eesti ja vene rahvusest noored, siis nad räägivad omavahel inglise keeles. Üritan öelda midagi eesti keeles, aga eestlane näeb, et ma pole eestlane ja kohe läheb inglise keelele üle.
- (35) Inglise keel on lihtsam, seda läheb ka rohkem vaja, oleme ju Euroopas. Kui oleks selline olukord, et ümberringi ongi ainult eesti või mingi muu keel, siis ikka õpid ära, käidki sõnaraamatuga ringi ja õpid.

Paljud õpilased vahendavad intervjuudes vanemate arvamusi eesti keele kui mitte-olulise keele kohta ning inglise keele kõrgemale positsioonile asetamise kohta. Intervjuudes öeldut tõlgendades nähtub, et õpilaste vanemad näevad sageli eesti keelt eeskätt vahendina millegi saavutamiseks (nt põhikooli lõpetamine, hea riigikeele oskuse abil paremasse gümnaasiumisse pääsemine jne).

- (36) Vanemad ütlevad, et inglise keel on tähtsam.
- (37) Ema ütleb, et pean kooli lõpetama ja selleks on eksam vaja teha.
- (38) Minu peres ei räägi keegi eesti keelt, ema ütleb, et õpi eesti keelt, siis saad hea eksamihinde ja saad kooli lõpetada ja siit ära sõita.

Eesti keele olulisuse ning vajalikkuse temaatikast kasvas õpilastega tehtud intervjuudest kohati välja õpilaste hoiakuid detailsemalt avav alateema, nimelt küsimus,

kas vene keelt emakeelena kõnelev õpilane ise paneks oma lapse eestikeelsesse kooli. Õpilaste antud vastused on eitavad ning esitatud väidete ning põhjendustega välistatakse järeltuleva põlvkonna õppimine eestikeelses koolis. Eesti keeles nähakse konkurenti vene keelele, mis seab ohtu nii vene keele kui ka vene identiteedi säilitamise laiemas mõttes. Eri keelte rollidest oma argielus kõneldes nähtub, et vene kooliõpilasele on ilmselt võõras mõelda endast kui mitme keele kõnelejast ning eesti keele kasutamises nähakse pigem liikumist identiteedi kaotamise, mitte aga enda (mitmekeelse ja -kultuurilise) identiteedi tugevdamise suunas. Artikli autorid peavad siinkohal vajalikuks rõhutada, et mitmekeelsuse tähenduse, selle defineerimise ning mõistmise teemadel intervjuudes pikemalt ei peatunud ning seega jääb praegu vastuseta küsimus, miks vene koolinoored just niimoodi mõtlevad. Võib arvata, et see johtub paljudest eri teguritest, mida on põhjust eraldi uurida.

- (39) Õppisin neli aastat eesti koolis, aga tulin vene kooli üle, sest olen vene inimene ja tahan vene keeles õppida, suhelda. See oli minu, mitte minu vanemate valik.
- (40) Ei paneks enda last eesti kooli, sest siis ta hakkab ainult eestlastega, mitte venelastega suhtlema.
- (41) Eesti kooli keegi lapsi ei paneks, sest eesti keelt saab koolis õppida, aga emakeelt peab ikka oskama. Kui laps panna juba eesti lasteaeda, siis laps õpib eesti keele paremini selgeks kui oma emakeele [vene keele].

Õpilastega nende tulevikuplaanidest kõneldes nähtub, et enamik intervjuudes osalenuist ei seosta enda tulevikku Eestiga ning plaanib riigist lahkuda. Kuigi enamik kavatses õpinguid siiski gümnaasiumis jätkata, soovitakse pikemas perspektiivis Eestist lahkuda. Vene keeles hariduse omandanute väga suur soov jätkuõpinguteks Eestist lahkuda on selgunud ka näiteks Praxise 2012. a uuringust “Koolilõpetajad ja nende karjäärivalikud” (Mägi, Nestor 2012), mille kohaselt vähemalt 30% vene emakeelega gümnaasiumilõpetajatest soovib edasi õppida välismaal.

4. Kokkuvõte

Vene õppekeelega koolide 9. klasside õpilaste eesti keele kasutusväljad on suhteliselt ahtad ja seda just Ida-Virumaal, piirdudes valdava osa intervjuudes osalenutel koolikeskkonna ja/või eesti keele teise keelena tundidega. Kui väljaspool Ida-Virumaad (st Tallinnas ja Tartus) elavad õpilased kasutavad eesti keelt iga päev ka väljaspool kooli (nt huviringides, trennides, kaubanduses) ning isiklikes suhtlusvõrgustikes, siis Ida-Virumaa koolide õpilastel on piirkonna keelelisest-rahvuslikust eripärast tingituna eestikeelseks suhtluseks ootuspäraselt vähem võimalusi. Kui üldiselt hindavad noored oma riigikeele oskust heaks ja väga heaks (iseküsimus on sellise enesehinnangu adekvaatsus), siis üsnagi kriitilised ollakse oma eestikeelse suhtluspädevuse osas. Põhjusena tuuakse enamasti välja just väga vähene võimalus kooliväliselt eesti keelt kasutada.

Vene noore jaoks on eesti keel eeskätt kohustuslik õppeaine, keel, mille õppimise ja oskamise kohustus tuleneb eesti keele riigikeele positsioonist. Eesti keelt nähakse kui vahendit, et gümnaasiumis või kutsekoolis edasi õppida. Väga vähe (kui üldse) ilmneb eesti keele õppimises selle integratiivne motivatsioon ehk keeleõppija

isiklikum suhe õpitavasse keelde ja kultuuri, eesti keele tunnetamine identiteedi osana. Kuigi koolide visioonidokumendid väärtustavad õpilaste kasvatamist mitmekeelseks ja -kultuuriliseks, ei pea õpilased selle väljatoomist oluliseks.

Intervjuudes esitatud argumentidest nähtub ilmne pragmaatilisus keeleõppes ja -valikutes ning oluliseks keeleks peetakse keelt, mis on “suurem” ning “tuntum” ning mis võimaldab nende omaduste kaudu keelekasutajale laiemaid isiklikke hüvesid, nii praegu kui ka tulevikku vaatavalt. Eesti keel oma väiksuse ja “kadumisohuga” võib samuti pärssida õpimotivatsiooni. Õpilastega nende tulevikuplaanidest kõneldes nähtub, et enamik intervjuudes osalenuist ei seosta enda tulevikku Eestiga ning plaanib riigist lahkuda. See on ilmselt ka üks põhjustest, miks rõhutatakse inglise keele õppimise vajalikkust.

Ka eesti koolinoorte suhtumine keelde on kohati pragmaatiline, mis ilmnes Ehala ja Tammemäe (2012) uuringus, kus instrumentaalsest aspektist lähtuvalt väärtustasid õpilased enim inglise keele oskust. Küll aga väärtustasid eesti koolinoored eesti keelt just tunnetuslikust ja kultuurilisest ehk siis integratiivsest aspektist. On mõistetav, et sellist tunnetuslikku väärtust kannab venekeelse noore jaoks tema emakeel ehk siis vene keel. Intervjuude põhjal võib ka öelda, et vene-eesti kakskeelsus ja mitmekultuurilisus ei kuulu, vähemalt praegusel momendil, venekeelse noore tunnetuslike väärtuste paradigmasse. Uuringu tulemuste kontekstis on uudne just mitmekeelsuse- ja kultuurilisuse probleemistik, mis avardab oluliselt vene koolinoore riigikeele hoiakute tagamaid. Kas ja kuivõrd mitmekultuuriline identiteet võiks olla õpimotivatsiooni soodustavaks teguriks, väärrib edasistes uuringutes detailsemat vaatluse alla võtmist.

Viidatud kirjandus

- Dörnyei, Zoltán 1998. Motivation in second and foreign language teaching. – *Language Teaching*, 31 (3), 117–135. <http://dx.doi.org/10.1017/S026144480001315X>
- Dörnyei, Zoltán 2001a. New themes and approaches in second language motivation research. – *Annual Review of Applied Linguistics*, 21 (1), 43–59. <http://dx.doi.org/10.1017/S0267190501000034>
- Dörnyei, Zoltán 2001b. Teaching and researching motivation. *Applied Linguistics in Action Series*. Essex: Pearson Education Limited.
- Dörnyei, Zoltán 2003. Attitudes, orientations, and motivations in language learning: Advances in theory, research, and applications. – *Language Learning*, 53 (S1), 3–32. <http://dx.doi.org/10.1111/1467-9922.53222>
- Dörnyei, Zoltán; Schmidt, Richard (Eds.) 2001. *Motivation and Second Language Acquisition*. University of Hawai'i at Mānoa: Second Language Teaching & Curriculum Center.
- Ehala, Martin; Tammemägi, Anni 2012. Eesti õpilaste keelehoiakud 2011. aastal. [Language attitudes of secondary school students 2011.] – *Keel ja Kirjandus*, 4, 241–260.
- EKR 2007 = Euroopa keeleõppe raamdokument: õppimine, õpetamine, hindamine. [Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR).] Tartu: Haridus- ja Teadusministeerium.
- Ellis, Rod 2008. *The Study of Second Language Acquisition*. Oxford: Oxford University Press.
- Ellis, Rod; Barkhuizen, Rod 2005. *Analysing Learner Language*. Oxford: Oxford University Press.
- Gardner, Robert C. 1985. *Social Psychology and Second Language Learning: The Role of Attitudes and Motivation*. London: Arnold.
- Gardner, Robert C. 2010. *Motivation and Second Language Acquisition: The Socio-Educational Model*. New York: Peter Lang.

- Garrett, Peter 2010. Attitudes to Languages. Key Topics in Sociolinguistics. Cambridge: Cambridge University Press.
- Jalkanen, Liisa; Ruuska, Johanna 2007. Affektiiviset tekijät vieraan kielen opiskelussa. Tapaustutkimus alakoulun englannin tunnista oppilaiden kokemana. [The Affective Factors in Studying a Foreign Language: A Case Study of the English Lessons in the Primary School as Experienced by the Pupils.] Pro gradu -tutkielma. Jyväskylä: Jyväskylän yliopisto.
- Kaksi kantaa pakoruotsista. [Two opinions on mandatory Swedish in Finland.] <http://www.hs.fi/sunnuntai/a1305708043006> (12.10.2014).
- Klaas, Birute 1992. Mõtteid võõrkeeleõpetuse psühholoogiast. [Some thoughts about the psychology in foreign language teaching.] – Lähivertailuja 5: Suomalais-virolainen kontrastiivinen seminaari Viroilla 27. ja 28. syyskuuta. Maia Väkräm (Toim.). Tampereen yliopiston suomen ja yleisen kielitieteen laitoksen julkaisuja 15. Tampere: Tampere University Press, 29–33.
- Klaas-Lang, Birute; Praakli, Kristiina; Peedisson, Anni; Lašmanova, Alla 2014. Arvamusi ja hinnanguid riigikeele õppe korraldamise kohta vene õppekeele koolides. Uuringu lõpparuanne. [Opinions and Assessments on Teaching of Estonian in Russian-Medium School. Final Report.] Tartu: Tartu ülikooli eesti ja üldkeeleteaduse instituut.
- Knight, Tiina 2013. Pelkkää pakoruotsia? Tamperelaisten yliopisto-opiskelijoiden ruotsin kielen opiskelimoivaatio. [Mandatory Swedish Only? Students Motivation to Study Swedish at the University of Tampere.] Pro gradu -tutkielma. Tampere: Tampereen yliopisto.
- Koreinik, Kadri 2011. Language Ideologies in the Contemporary Estonian Public Discourse: With a Focus on South Estonian. Dissertationes sociologicae universitatis Tartuensis 5. Tartu: Tartu University Press.
- Koreinik, Kadri; Praakli, Kristiina 2013. Keele kestlikkuse ja kao konstrueerimine poolavalikes diskursustes. [The construction of language maintenance and loss in semi-public discourses.] – Eesti Rakenduslingvistika Ühingu aastaraamat, 9, 131–148. <http://dx.doi.org/10.5128/ERYa9.09>
- Metslang, Helena; Kibar, Triini; Kitsnik, Mare; Koržel, Jevgenia; Krall, Ingrid; Zabrodskaia, Anastassia 2013. Kakskeelne õpe vene õppekeele koolis. [Bilingual Learning in Russian-Medium Schools. Final Report.] Tallinn: Tallinna Ülikooli eesti keele ja kultuuri instituut.
- Mitchell, Rosamond; Myles, Florence 2004. Second Language Learning Theories. Second edition. London–New York: Routledge.
- Mägi, Eve; Nestor, Mihkel 2012. Koolilõpetajad ja nende karjäärivalikud. Keskkharidusastme lõpetajate valikute uuringu lõpparuanne. [Survey of Upper-Secondary Graduates' Educational Choices.] SA Poliitikauuringute Keskus Praxis. http://praxis.ee/fileadmin/tarmo/Publikatsioonid/Koolilõpetajad_uuring.pdf (12.10.2014).
- Pihko, Marja-Kaisa 2007. Minä, koulu ja englanti. Vertaileva tutkimus englanninkielisen sisällönopetuksen ja perinteisen englannin opetuksen affektiivista tuloksista. [Me, School and English: A Comparison of Affective L2 Learning Outcomes in CLIL Classrooms and EFL Classrooms.] Opettajankoulutus-laitoksen tutkimuksia 85. Jyväskylä: Jyväskylän yliopisto.
- Proos, Ivi 2005. Eestivenelaste keeleoskus ja suhtumine 2007. aasta gümnaasiumireformi. Uuringu “Integratsiooni monitooring 2005” aruanne. [Language Skills of Estonian Russians and the Attitude Toward the Educational Reform in 2007.] Tallinn: Tallinna Ülikooli rahvusvaheliste ja sotsiaaluuringute instituut.
- Proos, Ivi; Pettai, Iris 2008. Eestivene noored: uue põlvkonna positsioon ja ootused. [Estonian Russian Youths: The Position and Expectations of the New Generation.] Tallinn: Eesti Avatud Ühiskonna Instituut. <http://uuringud.tallinnlv.ee/document.aspx?id=11178> (17.2.2015).

- TNS Emor 2006 = Eestikeelne aineõpe vene õppekeele koolides: hetkeolukord ja vajadused. [Content and Language Integrated Learning (CLIL) in Russian-Medium Schools: Current Situation and Needs.] <https://dspace.utlib.ee/dspace/handle/10062/40647> (2.1.2015).
- TNS Emor 2008 = Mitte-eestlaste teadlikkus ja suhtumine vene õppekeele koolide gümnaasiumiastmes eestikeelsele aineõppele üleminekust. [Non-Estonians' Awareness of and Attitude towards the Transition of Russian-Medium Schools to Teaching Subjects in Estonian at the Upper Secondary School Level.] <https://dspace.utlib.ee/dspace/handle/10062/40654> (2.1.2015).
- Vaiss, Natalia 2009. Vene õppekeele põhikooli õpilaste ja õpetajate hoiakud eestikeelse aineõppe suhtes vahetult enne 2007. aasta gümnaasiumireformi algust. [Attitudes of Russian-medium school learners and their teachers towards partial Estonian language instruction before implementation of the education reform in 2007.] – Eesti Rakenduslingvistika Ühingu aastaraamat, 5, 287–301. <http://dx.doi.org/10.5128/ERYa5.19>
- Vare, Silvi 2004. Eesti keel vene koolis. [Estonian Language in Russian-Medium School.] Tallinn: Eesti Keele Sihtasutus.

Birute Klaas-Lang (Tartu Ülikool) on erialalt eesti filoloog, kelle peamised uurimisteemad on keeletüpoloogia, keelekontaktid, keelepoliitika, väliseesti keel.
Tartu Ülikool, eesti ja üldkeeleteaduse instituut, Jakobi 2, 51014 Tartu, Estonia
birute.klaas-lang@ut.ee

Kristiina Praakli (Tartu Ülikool) on erialalt eesti filoloog, kelle peamised uurimisteemad on väliseesti keel, keele kestlikkus ja keelekontaktid.
Tartu Ülikool, eesti ja üldkeeleteaduse instituut, Jakobi 2, 51014 Tartu, Estonia
kristiina.praakli@ut.ee

WHY DO I NEED ESTONIAN? THE IMPORTANCE OF OFFICIAL LANGUAGE COMPETENCE IN THE EYES OF RUSSIAN TEENAGERS

Birute Klaas-Lang, Kristiina Praakli

University of Tartu

The aim of the article is to analyze the attitudes of 9th grade students at Russian-medium general education schools in Estonia toward the importance of competence in the official language. Semi-structured group interviews allow the voice of the language learner – in this case 15–16-year-old Russian school students – to be heard in the discussion of official language instruction. The analysis of these interviews reveals that Russian students' attitudes toward the official language are shaped by their perception of Estonian as a small language, limited opportunities to use the language in their immediate environment, and the role of different languages on the job market, but also by the students' national and linguistic identity. One primary conclusion is that in these students' eyes, the Estonian language has only instrumental value, that is, it is seen merely as a tool needed for self-actualization. Russian youths do not see the Estonian language as part of their identity, and question its sustainability (and therefore the benefit of studying it) in comparison to larger world languages.

Keywords: motivation, attitude, Russian-medium school, Estonian

LÄBIRÄÄKIMISTE MODELLEERIMINE ARVUTIL

Mare Koit

Ülevaade. Artiklis käsitletakse dialooge, kus osalejad esitavad läbirääkimiste käigus argumente selle poolt ja vastu, et üks neist teeks või jätaks tegemata teatava tegevuse. Kui osalejatel on üks ja sama eesmärk, siis otsivad nad ühiselt argumente, mis näitaksid, mida positiivset toob kaasa tegevuse tegemine või vastavalt mittetegemine ja kuidas kõrvaldada võimalikud takistused eesmärgi saavutamisel. Kui osalejate eesmärgid on teineteisele vastupidised, siis peab vähemalt üks neist argumentide ja vastuargumentide esiletoomise järel oma algsest eesmärgist loobuma. Artiklis arendatakse dialoogimudelit, mille koosseisu kuulub argumentimudel. Tuuakse näiteid dialoogikorpuse analüüsist, et anda esialgne hinnang mudeli sobivusele inimestevaheliste argumenteerimisdialoogide analüüsimiseks. Mudeli piiratud versioon on rakendatud arvutil, täielik realisatsioon on kavandatud edaspidiseks.*

Võtmesõnad: läbirääkimised, argument, dialoogimudel, teadmuse esitus, dialoogikorpus, eesti keel

1. Sissejuhatus

Läbirääkimised on selline suhtlemise vorm, kus osalejate grupp, keda ühendab soov koostööd teha, püüab jõuda ühiselt aktsepteeritava kokkuleppeni (vrd Rahwan jt 2004). Suhtlejad esitavad argumente oma seisukohtade toetuseks ja vajadusel partnerite omade ümberlukkamiseks. Suhtlemisel võivad, aga ei tarvitse eristuda selgelt fikseeritud pooled erinevate eesmärkidega. On võimalik, et ühel poolel õnnestub oma eesmärk vastaspoolele peale suruda. Samas võib kumbki pool jääda oma esialgse eesmärgi juurde ja läbirääkimised lõpevad kokkulepet saavutamata. Samuti võivad osalejad suhtluse käigus oma eesmarke modifitseerida ja neid teistele lähendada.

Paljud uurijad on modelleerinud läbirääkimisi arvutil.

Dung (1995) vaatleb argumenteerimissüsteemi kui paari $\langle A, R \rangle$, kus A on (lõplik) argumentide hulk ja R on nn rünnaku relatsioon argumentide vahel ($R \subseteq A \times A$).

* Artikli valmimist on toetanud Euroopa Regionaalarengufond Eesti Arvutiteaduse Tippkeskuse kaudu ning Eesti Teadusagentuur (projektid IUT20-56 "Eesti keele arvutimudelid", ETF9124 "Suhtlusagendi modelleerimine ja Eesti dialoogikorpus" ja ETF8558 "Eestikeelse spontaanse dialoogi struktuuri loomise keelelised vahendid").

Wagner (1998) käsitleb argumenteerimisel kasutatavaid mõisteid ja defineerib argumenti väite G kasuks kujul

$$\langle \{f_1, \dots, f_m\}, \langle r_1, \dots, r_n \rangle \rangle,$$

kus $\{f_1, \dots, f_m\}$ on n baasväidete ehk faktide hulk ja r_1, \dots, r_n reeglite järjend. Sealjuures on r_n nn tippreegel, millest järeljub väide G . Wagner vaatleb tehisagente kui üleminekute süsteeme, mis osalevad läbirääkimistes ja debattides.

Rahwan jt (2004) käsitlevad kolme viisi läbirääkimiste formaliseerimiseks: mänguteoreetilist, heuristilist ja argumentipõhist. Neist esimene, nn dialoogimäng, modelleerib reeglite abil vestlust, kus kaks osalejat võtavad ühise eesmärgi nimel vaheldumisi voore ja esitavad argumente (Yuan jt 2008, Caelen, Xuereb 2011). Klassikalisel mänguteoreetilisel lähenemisel on aga tõsiseid arvutuslikke piiranguid, mis tegelikes rakendustes ei tarvitse kehtida, nt eeldus, et agentide ressursid on piiramatud ja et olekute ruum, kus sooritatakse otsing, on täielikult teada. Heuristilise lähenemisviisi puhul otsitakse lähendusi otsustustele, mida vestluses osalevad agendid vastu võtavad. Heuristilised mudelid kasutavad aga üksnes ligikaudset ratsionaalsuse mõistet, ei uuri läbi kogu võimalike olekute ruumi ja annavad seetõttu sageli lahendeid, mis pole optimaalsed (Rahwan jt 2004). Mõlemal juhul – nii mänguteoreetiliselt kui ka heuristilise lähenemise korral – eeldatakse, et agentide eelistused on fikseeritud. Kumbki agent ei saa otseselt mõjutada teise agendi eelistusi ega sisemisi mentaalseid hoiakuid (arvamusi, soove, eesmärgi jms), mis moodustavad tema eelistuste mudeli. Ratsionaalne agent üksnes modifitseerib oma eelistuste mudelit saadud uue info põhjal.¹

Argumentipõhise lähenemise korral võivad agendid läbirääkimiste protsessis põhjendada oma arvamusi ja muid mentaalseid hoiakuid. Argumenti käsitletakse läbirääkimistes kui infohulka, mis võimaldab agendil: a) tunnistada oma seisundit või b) mõjutada teise agendi seisundit (Jennings jt 1998). Seega võib agent lisaks ettepaneku aktsepteerimisele või tagasilükkamisele seda kritiseerida.

Besnard ja Hunter (2008) formaliseerivad argumenteerimist klassikalise loogika abil ja defineerivad argumenti kui paari $\langle \Phi, \alpha \rangle$, kus Φ on selline valemite hulk (teadmusbaasi alamhulk) ja α on selline valem, et

1. Φ ei ole vasturääkiv,
2. Φ implitseerib α ,
3. Φ on teadmusbaasi minimaalne alamhulk, mis rahuldab 2.

Öeldakse, et α on väide (ingl *claim*), $\langle \Phi, \alpha \rangle$ on argument väite α toetuseks ja Φ on väite α toetus (ingl *support*).

Rahwan ja Larson (2011) uurivad loogilise tuletamise protseduuride kavandamise ja formaalse loogika seost juhul, kui teadmus on jagatud mitme osaleja vahel.

Hadjinikolis jt (2012) esitavad veenmisdialoogide argumenteerimispõhise raamistiku, kus agendid osalevad erilises dialoogimängus, esitades argumente ja kasutades ka partneri mudelit.

Kogu valdkonnast annavad ülevaate nt Chesñevar jt (2000), Prakken ja Vreeswijk (2002), Besnard ja Hunter (2008).

¹ Uue lähenemise dialoogi modelleerimisele võib pakkuda uut tüüpi mänguteooria – lingvistiline geomeetria (LG). LG kasutab formaalsete keelte hierarhiat mängustrateegiate ja geomeetriat olekute ruumi esitamiseks (Stilman 2011, Stilman jt 2010). Mängija järgmise käigu otsing olekute ruumis asendatakse strateegia konstrueerimisega. Olekute ruum jaotatakse alamruumideks, konstrueeritakse mängijate võimalike strateegiate hulga ja jäetakse kõrvale need strateegiad, mida ei saa rakendada. Alles jäänud strateegiatest konstrueeritakse probleemi optimaalne lahend. LG abil on modelleeritud mitmeid keerulisi probleeme, mida tavaliste arvutusmeetoditega ei õnnestu inimeksperdi tasemel lahendada, sh malemäng, kübersõda jms. (Autor tänab anonüümset retsensenti viite eest LG-le.)

Siinses artiklis uuritakse dialooge, kus osalejad *A* ja *B* peavad läbirääkimisi selle üle, kas üks neist (*B*) peaks vastu võtma otsuse teha või tegemata jätta teatud tegevus *D*. Otsuse vastuvõtmisele eelneb arutus, kus subjekt kaalub *D* tegemise või tegematajätmise positiivseid ja negatiivseid aspekte. Mõlemal osalejal võib olla üks ja sama eesmärk, näiteks nii *A* kui ka *B* taotleavad, et *B* teeks *D*. Sellisel juhul otsivad nad dialoogi käigus koos argumente selle otsuse toetuseks, sh viise võimalike *D* tegemist takistavate asjaolude kõrvaldamiseks. Osalejatel võivad aga olla hoopis teineteist välistavad eesmärgid, nt *A* taotleb, et *B* teeks *D*, aga *B* tahaks *D* tegemata jätta. Sellisel juhul püüab *A* suhtluse käigus mõjutada *B* arutlusprotsessi nii, et *B* otsustaks siiski teha *D*. Ainus viis partnerit suhtluses mõjutada on esitada argumente, mis näitavad, millised meeldivad, kasulikud jms positiivsed aspektid on *D* tegemisel *B* jaoks ning et need kaaluvad üles *D* tegemise võimalikud negatiivsed aspektid.

Varem on välja töötatud dialoogimudel, mille koosseisu kuulub arutlusmudel (Koit, Õim 2014a, Koit 2010). Käesolevas artiklis arendatakse seda mudelit edasi ja esitatakse läbirääkimiste formaalne mudel, kus osalejad toovad vaheldumisi esile argumente *D* tegemise poolt ja vastu. Samuti võivad nad esitada teineteisele küsimusi, selleks et teha valikuid võimalike argumentide vahel ja tõrjuda partneri vastuargumente.

Artikli ülesehitus on järgmine: 2. osa on pühendatud läbirääkimiste mudelile; 3. osas tutvustatakse inimestevaheliste läbirääkimisdialoogide analüüsi tulemusi; 4. osas arutletakse mudeli sobivuse üle läbirääkimiste või vaidluse modelleerimiseks ja argumendi kirjeldamiseks; 5. osas tehakse kokkuvõtteid.

2. Läbirääkimiste koostisosad

Siin tutvustatakse dialoogimudelit ja rakendatakse seda läbirääkimistele, kus leiab muuhulgas aset argumentide edastamine ühelt osalejalt teisele.

2.1. Dialoogimudel

Vaatleme dialoogi loomulikus keeles kahe osaleja *A* ja *B* (inimese või tehisagendi) vahel (vt Koit, Õim 2014a). Olgu dialoogi algataja *A* ja tema suhtluseesmärk mõjutada partnerit *B* nii, et emba-kumba: kas *B* otsustab teha tegevuse *D* või *B* otsustab mitte teha tegevust *D*.

B suhtluseesmärk võib olla seesama mis osalejal *A* või hoopis sellele vastupidine. Suhtluses püüab *A* mõjutada partnerit *B* nii, et too võtaks tegevuse *D* suhtes vastu just sellise otsuse, mis vastab *A* suhtluseesmärgile. Võimalikud on järgmised juhtumid:

- (1) *A* suhtluseesmärk on “*B* teeb *D*” ja *B* eesmärk on “*B* teeb *D*”;
- (2) *A* eesmärk on “*B* teeb *D*” ja *B* eesmärk on “*B* ei tee *D*”;
- (3) *A* eesmärk on “*B* ei tee *D*” ja *B* eesmärk on “*B* teeb *D*”;
- (4) *A* eesmärk on “*B* ei tee *D*” ja *B* eesmärk on “*B* ei tee *D*”.

Juhtumitel (1) ja (4) on *A* ja *B* suhtluseesmärk üks ja seesama ning nad otsivad vestluse käigus üheskoos põhjendusi (argumente), miks teha (vastavalt mitte teha) *D* ja kuidas ületada võimalikke takistusi *D* tegemise ees või, vastupidi, vältida võimalikke ebasoovitavaid tagajärgi, mida *D* tegematajätmine võib kaasa tuua.

Juhtumitel (2) ja (3) on *A* ja *B* eesmärgid teineteist välistavad, initsiaator *A* esitab argumente, mis peaksid mõjutama partnerit *B* aktsepteerima *A* eesmärki ja loobuma omaenese algsest eesmärgist. Samas võib *B* esitada vastuargumente, mis peaksid sundima hoopis partnerit *A* aktsepteerima *B* eesmärki ja loobuma oma algsest eesmärgist.

A käsutuses on partneri mudel – tema ettekujutus *B* kohta, mis annab talle alust arvata, et *B* soostub võtma omaks *A* suhtluseesmärgi. Konstrueerides oma esimest vooru, peab *A* planeerima dialoogiaktid (nt ettepanek, soov, küsimus, ettepanek koos argumentidega jne, sõltuvalt oma ettekujutusest *B* kohta) ja määrama nende verbaalse vormi (st lausungid). Partner *B* interpreteerib *A* vooru ja enne kui anda vastus, käivitab oma peas arutlusprotseduuri, et võtta vastu otsus – teha või mitte teha *D*. Siis planeerib ta dialoogiaktid (nt nõustumine, keeldumine, keeldumine koos argumentidega jne) ja nende verbaalse vormi, selleks et teatada partnerile oma otsusest. Kui *B* vastus on keeldumine, siis peab *A* muutma olemasolevat partneri mudelit (see ei vastanud tegelikkusele, sest *A* eeldas, et *B* nõustub omaks võtma *A* eesmärki) ja leidma uusi argumente, et panna *B* siiski vastu võtma taotletavat otsust.

B võib oma keeldumisele lisada argumente. Need (vastu-)argumentid annavad *A*-le infot *B* peas toimunud arutlusprotsessi kohta, mis lõppesid negatiivse otsusega.

2.1.1. Arutlusmudel

Siin kasutatavat arutlusmudelit on detailselt kirjeldatud artiklis (Koit, Õim 2014a). Üldjoontes järgib see ideid, mis on realiseeritud BDI (ingl *belief-desire-intention*) mudelis, kuid tal on spetsiifiline iseärasus – modelleeritakse nn “naivset” arutlusteooriat, mida inimesed kasutavad igapäevaelus, kui nad suhtlevad teiste inimestega ning püüavad prognoosida ja mõjutada nende otsuseid.

Arutlusmudel koosneb kahest osast: 1) inimese motivatsioonisfääri mudelist ja 2) arutlusprotseduuridest. Motivatsioonisfääris saab esile tuua põhilised tegurid, mis reguleerivad subjekti arutlust teatava tegevuse *D* kohta. Juhul kui subjekt arutleb selle üle, kas **teha** tegevus *D*, siis, esiteks, võib subjekt olla *soov* teha *D*, kui selle tegevuse meeldivad aspektid ületavad ebameeldivaid; teiseks, tal võib olla *vaja* teha *D*, kui see on kasulik mingi kõrgema eesmärgi saavutamiseks ja *D* tegemise kasulikud aspektid ületavad kahjulikke; kolmandaks, arutlev subjekt võib olla olukorras, kus ta *peab* (on kohustatud) tegema *D* – kui tegematajätmine toob kaasa karistuse. Neid tegureid nimetatakse mudelis vastavalt SOOV-, VAJA- ja PEAB-determinantideks.

Kui subjekt arutleb selle üle, miks **mitte teha** *D*, siis on arutlust käivitavad põhifaktorid analoogilised: esiteks, subjekt *ei soovi* teha *D*, kui *D* ebameeldivad aspektid kaaluvad üles meeldivad; teiseks, tal *pole vaja* teha *D*, kui *D* kahjulikud aspektid ületavad kasulikke; ning kolmandaks, *D* tegemine on talle *keelatud* (ei ole lubatud), tuues kaasa teatava karistuse. Neid faktoreid nimetatakse vastavalt POLE-SOOVI-, POLE-VAJA- ja POLE-LUBATUD-determinantideks.

Siinses arutlusmudelil esitatakse subjekti motivatsioonisfääri mudel vektorina, mille komponentidel on arvulised väärtused, nn kaalud: $w = (w(\text{ressursid}), w(\text{meeldiv}), w(\text{ebameeldiv}), w(\text{kasulik}), w(\text{kahjulik}), w(\text{kohustuslik}), w(\text{keelatud}), w(\text{karistus-teha}), w(\text{karistus-mitte-teha}))$. Seejuures $w(\text{ressursid}) = 1$, kui subjektil on olemas kõik D tegemiseks vajalikud ressursid (vastasel korral 0); $w(\text{kohustuslik}) = 1$, kui D on kohustuslik (muidu 0); $w(\text{keelatud}) = 1$, kui D on keelatud (muidu 0). Edasi, $w(\text{meeldiv})$ jne tähistavad D meeldivate jne aspektide kaalu; $w(\text{karistus-teha}) - \text{keelatud}$ tegevuse D tegemisele järgnevat karistuse suurust ja $w(\text{karistus-mitte-teha}) - \text{kohustusliku}$ D tegematajätmisele järgnevat karistuse suurust. Nende aspektide kaalud on mittenegatiivsed täisarvud.

Arutlusmudeli teine osa koosneb arutlusprotseduuridest, mis eeldatavasti reguleerivad inimese arutelusid tegevuste kohta. Iga arutlusprotseduur koosneb sammudest, mida subjekt läbib oma arutlusprotsessis; igal sammul arvutab ja võrdleb ta D erinevate aspektide summaarseid kaale ning tulemuseks on otsus: teha või mitte teha D .

Iga arutlusprotseduur sõltub determinandist, mis selle käivitab (siinses mudelis on nendeks SOOV, VAJA ja PEAB või vastavalt POLE-SOOVI, POLE-VAJA, POLE-LUBATUD). Esitatagu siin näitena arutlusprotseduur, mille käivitab determinant POLE-VAJA, st D tegemine ei ole arutlevale subjektile kasulik.

Eeltingimus: $w(\text{kasulik}) < w(\text{kahjulik})$.

- 1) Kas D tegemiseks on piisavalt ressursse? Kui ei, siis 9.
- 2) Kas $w(\text{kasulik}) < w(\text{ebameeldiv}) + w(\text{kahjulik})$? Kui jah, siis 9.
- 3) Kas D on kohustuslik? Kui ei, siis 5.
- 4) Kas $w(\text{kasulik}) + w(\text{karistus-mitte-teha}) < w(\text{ebameeldiv}) + w(\text{kahjulik})$? Kui jah, siis 9. Kui ei, siis 8.
- 5) Kas D on keelatud? Kui jah, siis 9.
- 6) Kas $w(\text{kasulik}) < w(\text{ebameeldiv}) + w(\text{kahjulik}) + w(\text{karistus-teha})$? Kui jah, siis 9.
- 7) Kas $w(\text{meeldiv}) + w(\text{kasulik}) < w(\text{ebameeldiv}) + w(\text{kahjulik}) + w(\text{karistus-teha})$? Kui jah, siis 9.
- 8) Otsus: teha D . Lõpp.
- 9) Otsus: mitte teha D .

Mudelil kasutatakse kahte kaalude vektorit. Vektor w^{AB} (A arvamused B hinnangute kohta tegevusele D , kus B tähistab suhtluspartnerit) kujutab endast partneri mudelit ning vektor w^B , kus komponentide väärtused on B tegelikud hinnangud D aspektidele (mille täpseid väärtusi A ei tea), on B iseenda mudel.

2.1.2. Suhtlusstrateegiad ja -taktikad

Suhtlusstrateegia all mõistetakse siin algoritmi, mida suhtluses osaleja kasutab oma suhtluseesmärgi saavutamiseks. Suhtluse algataja (A) võib realiseerida oma suhtlusstrateegiat erinevatel viisidel: rõhutada D meeldivust või vastavalt ebameeldivust (st *ahvatleda* partnerit B tegevust tegema või vastavalt tegemata jätma), rõhutada D kasulikkust või vastavalt kahjulikkust B jaoks (st *veenda* partnerit), rõhutada karistust, mis järgneb kohustusliku D tegematajätmisele või vastavalt keelatud tegevuse tegemisele (*ähvardada*) jne. Neid suhtlusstrateegia realiseerimise konkreetseid

viise nimetatakse suhtlustaktikateks. Osaleja *A*, püüdes suunata partneri *B* arutlust soovitava otsuse poole (tegema või vastavalt mitte tegema *D*), esitab argumente *D* tegemise (mittetegemise) poolt, aga *B*, kui ta vaidleb vastu, esitab vastuargumente.

Taotledes *B* otsust teha *D*, võib *A* rõhutada

- *D* meeldivust (st püüda käivitada partneris arutlusprotseduuri determinandiga SOOV),
- *D* kasulikkust (käivitada arutlusprotseduuri determinandiga VAJA),
- (kohustusliku) *D* tegematajätmisele järgnevat karistust (käivitada arutlusprotseduuri determinandiga PEAB).

Saavutamaks *B* otsust mitte teha *D*, võib *A* rõhutada

- *D* ebameeldivust (käivitamiseks arutlusprotseduuri determinandiga POLE-SOOVI),
- *D* kahjulikkust (käivitamiseks arutlusprotseduuri determinandiga POLE-VAJA),
- (keelatud) *D* tegemisele järgnevat karistust (käivitamiseks arutlusprotseduuri determinandiga POLE-LUBATUD).

Kui *A* ja *B* suhtluseesmärgid on vastandlikud, siis lihtsaim taktika, mida *A* saab kasutada, on nn kaitse: suhtluse algataja *A* on loovutanud initsiatiivi ega rõhuta ühtegi *D* positiivset (vastavalt negatiivset) aspekti, vaid ainult tõrjub *B* võimalikke vastuargumente. *B* jaoks on lihtsaim taktika tõrkumine: keeldumine omaks võtmast *A* eesmärki, ühtegi argumenti lisamata.

Igal suhtlustaktikal, mida *A* saab rakendada, on oma “nimiaspekt”, milleks on vastavalt meeldivus tegevuse tegemisele või ebameeldivus tegevuse tegematajätmisele *ahvatlemise* korral, kasulikkus tegevuse tegemisele või kahjulikkus tegevuse tegematajätmisele *veenmise* korral ja keelatud *D* tegemisele või kohustusliku *D* tegematajätmisele järgnev karistus *ähvardamise* korral. *A* püüab esitada argumente, rõhutades oma valitud taktika nimiaspekti (see on kooskõlas arutlusprotseduuriga, mida *A* püüab *B* peas käivitada).

2.1.3. Teadmusbaas

Osaleja (agendi) *A* teadmusbaas sisaldab 1) arutlusalgoritmid, 2) suhtlusstrateegiad ja -taktikad, 3) partneri mudeli w^{AB} , 4) dialoogiaktide nimestiku, mida *A* saab kasutada (ettepanek, küsimus, väide jne), 5) lausungite nimestiku, mida *A* saab kasutada dialoogiaktide verbaliseerimiseks. Siin eeldatakse lihtsuse mõttes, et kumbki suhtluses osaleja kasutab lõplikku hulka etteantud lauseid. Lauseid ei analüüsita morfoloogiliselt ega süntaktiliselt; semantiline analüüs ja genereerimine on lihtsustatud sel viisil, et need laused on eelnevalt klassifitseeritud. Näiteks on olemas laused partnerile suhtluseesmärgi teatamiseks (*Lähme kinno*), laused tegevuse meeldivate, ebameeldivate jne aspektide rõhutamiseks (*Film on põnev; Kinopilet on odav* jne), nõustuvad laused (*Olen nõus*) jne. Igal sellisel lausel on (kasutatavas mudelis numbriline) kaal – mõned laused “kaaluvad” rohkem kui teised. Kaalud sõltuvad nii tegevusest kui ka suhtluses osalejatest. Näiteks lausel *Peole tuleb kümme külalist*, mida *A* kasutab tegevuse *D* meeldivuse rõhutamiseks, võib ühe suhtluspartneri jaoks olla kaal 10 ja teise jaoks 1.

B teadmusbaasis sisalduvad analoogilised teadmised, välja arvatud see, et partneri mudeli w^{AB} asemel on iseenda mudel w^B .

Siin eeldatakse, et iga lausungit saab kasutada ainult ühe korra. Seega, kui A rakendab näiteks tegevuse tegemisele ahvatlemise taktikat ja tema käsutuses ei ole enam lauseid D meeldivuse rõhutamiseks (mis on ahvatlemise nimiaspekt), siis peab A valima ahvatlemise asemel uue taktika või loobuma oma suhtlusesmärgist.

2.1.4. Argumendi struktuur

Pidades läbirääkimisi tegevuse tegemise või mittetegemise üle, edastavad A ja B teineteisele oma argumente. Osaleja A argumendi üldine struktuur on järgmine (vrd Besnard, Hunter 2008, Koit, Öim 2014b):

$$\langle \{R, T, w^{AB}_i, \text{fakt}^A\}, \text{eesmärk}^A \rangle,$$

kus

- R on arutlusprotseduur, mida A üritab käivitada partneris B ,
- T on kasutatav suhtlustaktika,
- $w^{AB}_i = (w^{AB}_i(\text{ressursid}), w^{AB}_i(\text{meeldiv}), w^{AB}_i(\text{ebameeldiv}), w^{AB}_i(\text{kasulik}), w^{AB}_i(\text{kahjulik}), w^{AB}_i(\text{kohustuslik}), w^{AB}_i(\text{keelatud}), w^{AB}_i(\text{karistus-teha}), w^{AB}_i(\text{karistus-mitte-teha}))$ on jooksev partneri mudel (ajahetkel i),
- fakt^A tähistab A valitud lauset partneri mudelis esinevate kaalude mõjutamiseks, pärast mida annab arutlusprotseduur R muudetud mudelil eeldatavasti B positiivse otsuse (kooskõlas A eesmärgiga); tema kaal on $w(\text{fakt}^A)$,
- $\text{eesmärk}^A = "B$ teeb $D"$ (või vastavalt " B ei tee $D"$).

Argumendi koosseisu võib kuuluda ka mitu erinevat fakti/lauset (mitte ainult üksainus).

Siin eeldatakse, et võimalikud laused (lausungid) on jagatud klassidesse, nt $P_{\text{on_ressursid}}$ tähistab lausete hulka, mida A saab kasutada, näitamaks partnerile B , et D tegemiseks leiduvad ressursid; $P_{\text{suurenda_meeldivust}}$ tähistab lausete hulka D meeldivuse suurendamiseks jne.

Suhtluses A esiletoodud fakt^A "toodab" uue partneri mudeli w^{AB}_{i+1} (järgmisel ajahetkel $i+1$):

kui $\text{fakt}^A \in P_{\text{on_ressursid}}$, siis $w^{AB}_{i+1}(\text{ressursid}) = 1$;

kui $\text{fakt}^A \in P_{\text{puuduvad_ressursid}}$, siis $w^{AB}_{i+1}(\text{ressursid}) = 0$;

kui $\text{fakt}^A \in P_{\text{suurenda_meeldivust}}$, siis $w^{AB}_{i+1}(\text{meeldivus}) = w^{AB}_i(\text{meeldivus}) + w(\text{fakt}^A)$ jne.

Partneri B (vastu-)argument on analoogilisel kujul:

$$\langle \{R^B, T^B, w^B, \text{fakt}^B\}, \text{eesmärk}^B \rangle,$$

kus arutlusalgoritm R^B toodab mudelil w^B otsuse " B teeb D " või vastavalt " B ei tee D " (s.o eesmärk^B), fakt^B näitab D aspekti, mille (liiga väike või liiga suur) väärtus põhjustab selle otsuse, ja T^B on B jooksev suhtlustaktika. Siin eeldatakse, et B iseenda mudel w^B ei muutu, st ainult A saab suhtluse käigus järk-järgult lähendada oma ettekujutust partnerist tegelikkusele.

Partneri B fakt^B annab osalejale A infot tema järgmise lause/fakti valimiseks suhtluses. Näiteks kui A suhtluseesmärk on “ B teeb D ” ja fakt^B $\in P_{\text{puuduvad_ressursid}}$, siis on w^{AB}_i (ressursid) tegelik väärtus 0 ning A peab oma järgmise lause valima hulgast $P_{\text{on_ressursid}}$ (pärast seda lauset kehtib eeldatavasti w^{AB}_{i+1} (ressursid) = 1) ja järgmine lause tuleb valida arutlusprotseduuri R nimiaspekti väärtust suurendavate lausete hulgast, mida A üritab käivitada partneris B , kasutades suhtlustaktikat T . Teisiti öeldes, A reageerib küll B toodud vastuargumentidele (tõrjudes seda), aga samas jätkab ka oma valitud taktikat T , esitades järgmise fakti, mis rõhutab T nimiaspekti.

Kuidas valib partner B oma järgmise lause/fakti? Ta käivitab parajasti kehtiva arutlusprotseduuri R^B oma mudelil w^B . (Nii see arutlusprotseduur kui ka B iseenda mudel võivad muidugi erineda A kasutatavast arutlusprotseduurist R ja partneri mudelist w^{AB} .) B rakendab oma arutlusprotseduuri ja selle tulemusel saab ta määrata D aspekti, mille väärtus põhjustas tema (negatiivse) otsuse. Näiteks võib ta valida lause, mis väljendab, et D on ebameeldiv: *[Ma ei tule kinno, sest] mulle ei meeldi põnevusfilmid*, aga ta võib ka lihtsalt teatada: *Ma ei tule kinno*. Viimasel juhul ei saa A rünnata ühtki vastuargumenti, tal ei jää üle muud, kui valida järjekordne lause jooksva suhtlustaktika T nimiaspekti väärtust suurendavate lausete hulgast, lootuses, et see muudab partneri mudelit positiivse otsuse saavutamiseks vajalikul määral.

Kui A käsutuses ei ole rohkem lauseid T nimiaspekti väärtuse suurendamiseks, siis ta kas 1) valib B mõjutamiseks uue suhtlustaktika ja sellele vastava arutlusalgoritmi või 2) loobub oma suhtluseesmärgist.

2.2. Läbirääkimiste struktuur

Kui A ja B suhtluseesmärgid on vastandlikud, siis alustavad nad debatti ehk vaidlust (nt A suhtluseesmärk on “ B teeb D ”, aga B eesmärk on “ B ei tee D ”). Üks osaleja saavutab vaidluse tulemusel oma suhtluseesmärgi (“võidab”), teine aga peab oma eesmärgist loobuma (“kaotab”).

Kui A ja B suhtluseesmärgid ühtivad, siis otsivad nad kooperatiivses suhtluses üheskoos argumente, mis toetaksid selle ühise eesmärgi saavutamist. Suhtluses võib emb-kumb neist, näiteks B , viidata takistustele eesmärgi saavutamisel. Siis peab A leidma argumente, näitamaks, kuidas neid takistusi saab kõrvaldada. Läbirääkimiste lõpptulemuseks on kas eesmärgi saavutamine (mõlemad osalejad “võidavad”) või sellest loobumine, kui teatavaid takistusi ei õnnestu kõrvaldada (mõlemad osalejad “kaotavad”).

Praeguses mudelis ei ole ette nähtud suhtluseesmärgi modifitseerimist, st otsus on kas B teeb D või B ei tee D (välistatud on näiteks D tegemine või tegematajätmine teatud tingimustel). Siiski võivad osalejad lõpetada läbirääkimised ka kokkuleppega, et otsustamine lükatakse edasi kuni järgmiste läbirääkimisteni.

Siin eeldatakse, et A ja B arutlusalgoritmide hulk on ühine. Samuti eeldatakse, et A ja B saavad mõlemad kasutada fikseeritud hulka dialoogiakte ja vastavaid lauseid, mis on liigitatud semantiliselt, nagu eespool juba mainitud.

Alustades läbirääkimisi, fikseerib algataja A (või moodustab oma eelteadmiste alusel) partneri mudeli w^{AB} ja määrab suhtlustaktika T , mida kasutada, st koos sellega määrab ta ka arutlusalgoritmi R , mida püüab käivitada B peas. Partneril B

on olemas iseenda mudel w^B . Ta määrab arutlusprotseduuri R^B , mida tema kasutab, otsustamaks, kas teha D või mitte, ja suhtlustaktika T^B (näiteks püüda haarata initsiatiivi ja rünnata partneri argumente).

Läbirääkimiste üldine struktuur on järgmine (sulgudes olevad dialoogiaktid võivad puududa):

A: ettepanek (+ argument)

KORRATA

(

B: küsimus

A: vastus/info andmine

)

B: nõustumine VÕI keeldumine (+ argument)

(

A: küsimus

B: vastus/info andmine

)

A: argument

KUNI lõpetamistingimused on täidetud

Mõlemad osalejad, nii A kui ka B , võivad osutada, et lõpetamistingimused on täidetud: 1) suhtluseesmärk on saavutatud, 2) loobuda oma suhtluseesmärgist (anda alla), hoolimata sellest, et leidub veel lauseid/fakte uute argumentide väljendamiseks, 3) ei leidu enam lauseid fikseeritud suhtlustaktika jätkamiseks, aga ei valita uut taktikat, kuigi see oleks veel võimalik, 4) kõiki suhtlustaktikaid on juba proovitud ja kõik laused/faktid edastatud, kuid ometi ei ole õnnestunud saavutada suhtluseesmärki.

Osalejad võivad esitada suhtluse käigus teineteisele küsimusi, selleks et teha valikuid erinevate lausete vahel, mida nad saavad kasutada argumenteerimisel.

Arvutil on realiseeritud kirjeldatud mudeli erijuht – vaidlus, kus A suhtluseesmärk on “ B teeb D ” ja B eesmärk “ B ei tee D ” (Koit, Õim 2014a, 2014b, Koit 2012).

3. Inimestevaheliste dialoogide analüüs

Selleks et anda esialgne hinnang mudelile, võeti ette inimestevaheliste argumenteerimisdialoogide analüüs. Tartu Ülikooli eesti dialoogikorpuses leidub kolme liiki suulisi dialooge, kus esineb läbirääkimisi: müügivestlused ning mõned reisibüroodialoogid ja argivestlused (Hennoste jt 2009). Kõik dialoogid on salvestatud autentsetes olukordades ja litereeritud, kasutades vestlusanalüüsi transkriptsiooni. Käesoleva uurimuse jaoks analüüsiti 52 telefoni teel toimunud müügivestlust (neist 32 sisaldasid läbirääkimisi; ülejäänutes oli tegu kas soovitud isiku puudumisega, kõne katkemisega enne läbirääkimisteni jõudmist vms tõrkega), 44 argivestlust (pooled telefoni- ja pooled silmast-silma vestlused), kus leidis lõike, milles peeti läbirääkimisi mingi tegevuse üle, ning 28 reisibüroodialoogi (millest vaid viies vestluses argumenteeris ametnik teatud reisi kasuks; ülejäänud olid infodialoogid).

Järgnevas tuuakse kaks näidet eesti dialoogikorpusest. Esimene on katkend silmast-silma vestlusest reisibüroos ametniku (A) ja kliendi (B) vahel, kus kliendi eesmärk on reisida Hispaaniasse ja ametnik teeb temaga koostööd selle eesmärgi

saavutamise nimel. Teine näide on ühe koolitusfirma (mille varjunimi on Tiritamm) esindaja (A) telefonikõne, kes pakub koolituskursusi teise firma juhile ehk kliendile (B).

- (1) Ametnik põhjendab, miks pakutav reis on huvitav ja kasulik. Klient esitab lisaküsimuse.

/---/

A: siin on: ' paar sellist=õ (0.6) ' päris huvitavat prog' rammi.

fakt^A₁

0.5) näiteks=õ (2.2) näiteks väga ' iseloomulik on Hispaaniale se ' Kordoba: (0.9) ' Kordoba ((Cordoba)) piirkond, (0.4) siin=on se=on ' keskajal keskaegne ' linn?

fakt^A₂

ja=sis=õõ (0.5) seal on ' võimalik: ka seda (0.8) õ= ' härjavõitlust vaadata.

fakt^A₃

(.) et see ' etendus on=ka=nagu=juba seal hinna ' sees, (0.5) ja see maksab mingi ' kaheksasada ' kuus krooni.

fakt^A₄

(1.5)

B: need võib kõik [' eraldi ' juurde] tellida.

küsimus

A: ' jaa (.) jaa

(.) ja neid saab juba ' siit koha- ' siit broneerida.

vastus

/---/

- (2) Siin toob klient B mitmeid vastuargumente koolitusfirma esindaja A pakutud kursustele. A esitab küsimuse ja B vastus võimaldab tal leida uue argumendi: ta osutab, et firma suudab siiski õpetada midagi kliendile vajalikku, vastupidiselt kliendi oletusele.

/---/

B: ' teate mis ' te: te olete valesse f(h)irmasse \$ sattund. \$

keeldumine

.hhh põhjus on lihtsalt selles et minu mõlemad müüja=konsultandid nad ' väga hästi haritud.

fakt^B₁

(0.2) nad ' on (0.4) & eluaeg inimestega suheld ' kõrgetel kohtadel.

fakt^B₂

.hhh mis ' minusse puutub (0.3) olen ' mina õppinud ära ' firma juhtimise \$ mul on ' topelt ' kõrgharidus. \$

fakt^B₃

/---/

A: .hhh et ' tehakse sisuliselt presentatsi ' oo[ne kus siis=ee]

küsimus

B: [' just ' nimelt.]

vastus

/---/

A: .hhhh ?jaa?jaa.

.h SEDA PAKUB KA 'TIRITAMM muidu 'välja: ee kuidas siis=õõ 'läbi
[viia 'presentatsi'oonne.]

fakt^A

/---/

Korpuse analüüs toetab arvamust, et toodud mudel on üldjoontes sobiv läbirääkimisdialoogi kirjeldamiseks. Seni on eesti dialoogikorpuse dialoogides märgendatud dialoogiaktid, kasutades Tartu Ülikoolis väljatöötatud dialoogiaktide tüpoloogiat (Hennoste jt 2009), ning spetsiaalse tarkvara abil määratud alamdialoogid ja suhtlusstrateegiad (Aller jt 2014), kuid ei ole lähemalt uuritud läbirääkimisdialoogide argumentidstruktuuri. Reaalsed dialoogid ei ole muidugi “puhtad” läbirääkimised vaadeldava mudeli mõttes, vaid sisaldavad inimestevahelistele vestlustele omaseid kõrvalekaldumisi pealiinilt, nt ühe või teise osaleja algatatud parandussekventse.

4. Arutelu

Artiklis on vaatluse all dialoogid, kus osalejad *A* ja *B* toovad kordamööda esile argumente teatava tegevuse *D* tegemise poolt ja vastu. Dialoogi algataja *A* suhtluseesmärgiks on, et partner *B* võtaks vastu teatava otsuse tegevuse *D* tegemise kohta. Osalejate eesmärgid võivad olla omavahel vastuolus (nt *A* eesmärk on, et *B* teeks *D*, aga *B* eesmärk on mitte teha *D*) või kooskõlas (nt nad taotleavad ühiselt *B* otsust teha *D*, olgugi et *D* tegemisel võib olla takistusi, mille kõrvaldamisvõimalusi nad otsivad läbirääkimiste käigus, esitades mitmesuguseid argumente, edastades ja kaitses sellega oma seisukohti).

Artikli 2. osas esitati läbirääkimisdialoogi formaalne mudel. Peab rõhutama, et see mudel hõlmab üksnes teatud kitsalt piiritletud läbirääkimisi: kõne all on mingi tegevuse tegemine või tegematajätmine. Läbirääkimiste tulemuseks on *B* otsus (*jah* või *ei*) tegevuse tegemise kohta. Kui *A* ja *B* on üksmeelsed, st nende algsed eesmärgid on samad, siis jõuavad nad läbirääkimiste lõpuks ühise kokkuleppeni. See ei tähenda tingimata, et nad saavutavad oma (ühise) suhtluseesmärgi. Näiteks võib juhtuda, et kuigi nii *A* kui ka *B* on püstitanud eesmärgi, et *B* teeks *D*, selgub suhtluse käigus, et *B* ressursid *D* tegemiseks on puudulikud ja kumbki, ei *A* ega *B*, ei suuda leida argumente, mis avaksid võimalusi ressursse juurde hankida. Sellisel juhul jõutakse küll ühise otsuseni (“*B* ei tee *D*”), kuid suhtluseesmärk jääb saavutamata. Mudel ei käsitle olukordi, kus algset suhtluseesmärki modifitseeritakse, näiteks seatakse *D* tegemisele teatavad eeltingimused.

Kui *A* ja *B* alustavad vastandlike suhtluseesmärkidega, siis on tegu vaidlusega. Vaidlus lõpeb tavaliselt ühe osaleja “võidu” ja teise “kaotusega”, st üks osaleja saavutab oma eesmärgi ja teine peab sellest loobuma. On ka võimalik, et n-ö “otsad jäävad lahtiseks”, st *B* lükkab otsustamise edasi. Sellisel juhul ei ole vaidluses võitjat ega kaotajat. Mudel ei käsitle olukordi, kus läbirääkimiste käigus jõutakse kompromissini algsete eesmärkide vahel, st kus mõlemad pooled teevad järeleandmisi ja muudavad oma eesmärgi.

Argumendi struktuur, mida mudelis kasutatakse, on samuti kohandatud kitsalt siin vaadeldavat liiki läbirääkimistele. Võttes eeskujuks Besnardi ja Hunteri (2008) pakutud esituse $\langle \Phi, \alpha \rangle$, kus Φ on valemite hulk (teadmusbaasi alamhulk

ehk eeldused) ja α on valem (väide, järeldus), on siin Φ rollis järgmine teadmiste hulk: suhtlustaktika, arutlusprotseduur ja tegevuse D aspektide kaalud (A puhul oletatavad, B puhul tegelikud), mis kehtivad argumenti esitamise hetkel, ning fakt või faktide hulk tegevuse D teatud aspekti(de) kaalu(de) muutmiseks. Väite α rollis on argumenti autori suhtluseesmärk (mille toetuseks ta argumenti esitab). Suhtluses esitatakse argumentid ilmutatult üksnes fakt (või faktid), ülejäänud osad on varjatud (vt näited osas 3).

5. Kokkuvõte

Siin uuriti suhtlust, kus algataja A suhtluseesmärk on, et partner B võtaks vastu otsuse teatava tegevuse D suhtes (“teha D ” või “mitte teha D ”). B eesmärk võib olla sama mis A oma või vastupidine. Viimasel juhul alustavad A ja B vaidlust. Vaidluses püüab A mõjutada B arutlusprotsesse sellisel viisil, et B loobuks oma algsest eesmärgist ja võtaks vastu otsuse, mis vastab A suhtluseesmärgile (sellisel juhul A võidab ja B kaotab vaidluse). Kui A -l ei õnnestu suhtluses jõuda B soovitava otsuseni, siis ta ei saavuta oma suhtluseesmärki (A kaotab ja B võidab vaidluse).

Kui A ja B taotleavad ühte ja sedasama eesmärki, siis alustavad nad diskussiooni, et selgitada, kas D tegemisel on takistusi või vastavalt D mittetegemisel on ebasoovitavaid tagajärgi.

Artiklis esitati dialoogimudel, mis sisaldab ühe osana argumentide edastamise ühelt osalejalt teisele. Esitati argumentimudel, mis suhtluse algataja A puhul sisaldab partneri B mudeli (või B puhul iseenda mudeli), arutlusprotseduuri, mida A üritab käivitada B peas (või vastavalt, mida B ise rakendab), suhtlustaktika ja lausete/faktide hulga, mis kogumina peaksid A ja/või B tooma soovitava järelduseni. Järeldust (otsust D tegemise kohta) tõlgendatakse argumenti struktuuris kui väidet (mis samas on argumenti autori suhtluseesmärk).

Mudelit hinnati dialoogikorpusest võetud tegelikel inimestevahelistel dialoogidel. Korpuse analüüs annab alust oletada, et esitatud mudelit saab kasutada inimestevaheliste läbirääkimisdialoogide analüüsimiseks ja modelleerimiseks.

Arvutil on realiseeritud mudeli erijuht – vaidlus, kus A suhtluseesmärk on “ B tee D ” ja B eesmärk “ B ei tee D ”. Kogu mudeli rakendamine jääb aga edaspidiseks.

Viidatud kirjandus

- Aller, Sven; Gerassimenko, Olga; Hennoste, Tiit; Kasterpalu, Riina; Koit, Mare; Mihkels, Krista; Laanesoo, Kirsi; Rääbis, Andriela 2014. Dialoogide pragmaatilise analüüsi tarkvara. [Software for pragmatic analysis of dialogues.] – Eesti Rakenduslingvistika Ühingu aastaraamat, 10, 23–36. <http://dx.doi.org/10.5128/ERYa.1736-2563>
- Caelen, Jean; Xuereb, Anne 2011. Dialogue and game theory. – Corneliu Burileanu, Horia-Nicolai Teodorescu (Eds.). Proc. of the 6th Conference on Speech Technology and Human-Computer Dialogue (SpeD), Braşov, Romania, 1–10. <http://dx.doi.org/10.1109/SPED.2011.5940726>
- Besnard, P.; Hunter, A. 2008. Elements of Argumentation. Cambridge, MA: MIT Press. <http://dx.doi.org/10.7551/mitpress/9780262026437.001.0001>
- Chesñevar, C.; Maguitman, A.; Loui, R. 2000. Logical models of argument. – ACM Computing Surveys, 32 (4), 337–383. <http://dx.doi.org/10.1145/371578.371581>

- Dung, Phan Minh 1995. On the acceptability of arguments and its fundamental role in non-monotonic reasoning, logic programming and n-person games. – *Artificial Intelligence*, 77 (2), 321–358. [http://dx.doi.org/10.1016/0004-3702\(94\)00041-X](http://dx.doi.org/10.1016/0004-3702(94)00041-X)
- Hadjinikolis, C.; Modgil, S.; Black, E.; McBurney, P.; Luck, M. 2012. Investigating strategic considerations in persuasion dialogue games. – *Frontiers in Artificial Intelligence and Applications* 241: STAIRS 2012. IOS Press, 137–148. <http://dx.doi.org/10.3233/978-1-61499-096-3-137>
- Hennoste, Tiit; Gerassimenko, Olga; Kasterpalu, Riina; Koit, Mare; Rääbis, Andriela; Strandson, Krista 2009. Suulise eesti keele korpus ja inimese suhtlus arvutiga. [Corpus of spoken Estonian and human-computer interaction.] – *Eesti Rakenduslingvistika Ühingu aastaraamat*, 5, 111–130. <http://dx.doi.org/10.5128/ERYa5.07>
- Jennings, N. R.; Parsons, S.; Noriega, P.; Sierra, C. 1998. On argumentation-based negotiation. – *Proc. of the International Workshop on Multi-Agent Systems*, Boston, 1–7.
- Koit, Mare 2012. Konversatsiooniagendi modelleerimine argumenteerimisdialogis: suhtlus kui infoseisundite värskendamine. [Modelling of conversational agents in argumentation: Conversation as updating of information states.] – *Eesti Rakenduslingvistika Ühingu aastaraamat*, 8, 109–122. <http://dx.doi.org/10.5128/ERYa8.07>
- Koit, Mare 2010. Eesti dialoogikorpus ja argumenteerimisdialogi arvutil modelleerimine. [Estonian dialogue corpus and computer modelling of argumentation dialogue.] – *Keel ja Kirjandus*, 4, 241–262.
- Koit, Mare; Õim, Haldur 2014a. A computational model of argumentation in agreement negotiation processes. – *Argument & Computation*, 5 (2–3), 209–236. <http://dx.doi.org/10.1080/19462166.2014.915233>
- Koit, Mare; Õim, Haldur 2014b. Modelling debates on the computer. – Joaquim Filipe, Jan Dietz, David Aveiro (Eds.). *Proc. of the 6th International Conference on Knowledge Engineering and Ontology Development (KEOD)*. Rome: SciTePress, 361–368. <http://dx.doi.org/10.5220/0005129303610368>
- Prakken, H.; Vreeswijk, G. A. W. 2002. Logics for defeasible argumentation. – D. M. Gabbay, F. Guenther (Eds.). *Handbook of Philosophical Logic*, vol. 2. 2nd edition. Dordrecht–Boston–London: Kluwer Academic Publishers, 219–318.
- Rahwan, I.; Larson, K. 2011. Logical mechanism design. – *The Knowledge Engineering Review*, 26 (1), 61–69. <http://dx.doi.org/10.1017/S0269888910000421>
- Rahwan, I.; Ramchurn, S. D.; Jennings, N. R.; Mcburney, P.; Parsons, S.; Sonenberg, L. 2004. Argumentation-based negotiation. – *The Knowledge Engineering Review*, 18 (4), 343–375. <http://dx.doi.org/10.1017/S0269888904000098>
- Stilman, B. 2011. Thought experiments in linguistic geometry. – *COGNITIVE 2011: Proc. of the Third International Conference on Advanced Cognitive Technologies and Applications*, 76–83.
- Stilman, B.; Yakhnis, V.; Umanskiy, O. 2010. Discovering role of linguistic geometry. – *Proc. of the 9th Mexican International Conference on Artificial Intelligence Conference on Advances in Soft Computing MICAI 2010*, Pachuca, Mexico, Part II, 1–21. http://dx.doi.org/10.1007/978-3-642-16773-7_1
- Yuan, T.; Moore, D.; Grierson, A. 2008. A human-computer dialogue system for educational debate, a computational dialectics approach. – *International Journal of Artificial Intelligence in Education*, 18 (1), 3–26.
- Wagner, G. 1998. *Foundations of Knowledge Systems with Applications to Databases and Agents*. Kluwer Academic Publishers. <http://dx.doi.org/10.1007/978-1-4615-5723-4>

Mare Koit (Tartu Ülikool) on uurinud dialoogi modelleerimist arvutil.
Liivi 2, 50409 Tartu, Estonia
mare.koit@ut.ee

MODELLING NEGOTIATIONS ON THE COMPUTER

Mare Koit

University of Tartu

In the paper, a model of negotiation is developed which includes a model of argument. The topic of negotiation between two participants is an action to be performed by one of them. The initial communicative goals of the participants can match or not. In the first case, they are cooperatively looking for reasons why performing (or not performing) the action is useful, pleasant, etc. In the second case, they are involved in debate. They present arguments and counter-arguments and in the end, one of them will abandon his or her initial communicative goal. An analysis of human-human negotiations is carried out in order to give a preliminary evaluation of the introduced model. A limited version of the model is implemented on the computer.

Keywords: negotiation, argument, dialogue model, knowledge representation, dialogue corpus, Estonian

NARRATIVE PRODUCTION WEAKNESS IN RUSSIAN DYSLEXICS: LINGUISTIC OR PROCEDURAL LIMITATIONS?

Aleksandr N. Kornev, Ingrida Balčiūnienė

Abstract. The study deals with the impact of non-linguistic factors on narrative production in Russian-speaking dyslexic children. The experimental group consisted of 12 children (age 9–10) with dyslexia and the control group comprised 12 peers without any developmental disorders. The sample was counterbalanced from the perspective of narrative mode, story complexity, and task order. One of the classic methodologies for narrative analysis, i.e. story grammar, was extended in our study by a novel dynamic approach, enabling us to evaluate procedural features of narrative production.

The results of our study highlight limitations in dyslexic narrative language underlined by two different causes. The first one can be defined as inefficiency in developing logical (temporal/causal) relationships between events; the other is difficulties in structuring an episode description. The high flexibility and dynamic changes in the episode structure in dyslexics anticipated the evidence that limitations in dyslexic narrative language are related to the deficit in procedural functions rather than to the primary language limitations; however, linguistic shortcomings in dyslexic narrative production still remain.

Our experience with the dynamic approach to narrative assessment lends support to its value as a research tool. The novel dynamic approach to episode completeness analysis proved to be an effective and informative method that might highlight new mechanisms of narration and thus extend the classic narrative analysis by the addition of qualitative information.*

Keywords: first language acquisition, language pathology, Russian

* The research was carried out with the financial support of the Russian Foundation for Humanities, grant No. 14-04-00509 *Formirovanie jazykovyh podsistem u detej s normoj i otstavaniem v razvitii reči: korpusnoe i eksperimental'noe issledovanie tekstov* [Development of linguistic sub-systems in typically-developing and language-impaired children: Corpus-based and experimental study].

1. Introduction

It is well known that for some children learning to read proceeds very slowly and inefficiently (Pressley 2006). The struggling readers are usually classified into those with specific and nonspecific reading disabilities (Snowling 1996, Voeller 2004). The prevalence of specific reading disabilities or developmental dyslexia among a population reaches 11% in languages with shallow literacy and as much as 3–5% in languages with transparent literacy (Lindgren et al. 1985, Shaywitz, Shaywitz 2003). In Russia, about 5% of school-age children are dyslexics (Kornev 1995, Kornev, Išimova 2010). Generally, dyslexia can be defined as a selective disability to learn reading in spite of regular education and normal intelligence (Lyon et al. 2003). From the perspective of the Russian-speaking population, the following difficulties are usually recognized: a) immature recoding and decoding skills, indicating that the child has difficulty in transitioning from letter- to syllable- to word reading; b) low reading fluency; c) lack of accuracy while reading aloud; e) sometimes a lack of comprehension, both at the word and sentence levels (Kornev et al. 2010).

Skills in reading and text comprehension are highly correlated with those in narrative production and comprehension. However, the direction of the correlation is still not clear. Does experience in reading affect the quality of narrative production? Or – vice versa – are skills in discourse necessary prerequisites for reading development? Following modern theoretical models, reading is considered an active and creative process of so called ‘text image generation’ (van Dijk, Kintsch 1983, Kintsch 1988, Leontyev 1997, 2004) that includes a) (re-)encoding (also called ‘decoding’), and b) elaboration of connection (coherence) between words and relating of word units in a logically coherent structure, i.e. text base (Gorin, Embretson 2006). According to Walter Kintsch and Teun A. van Dijk (1978), text processing can be interpreted as an iteration of the construction and integration processes. A text, as a set of propositions, is continuously integrated through the reader’s experience. According to the later theoretical model (Kintsch 1988, 1998), text comprehension can be interpreted as a cyclical process. During one cycle, a reader extracts propositions from the text and binds them into a logically coherent network. In the integration phase, the network is activated with the maximum activity level at the points of the highest correlations. After each of the cycles, the most activated propositions are transferred from the working memory to the next cycle for the further process of analysis and synthesis.

General reading ability is related to tacit knowledge of the structural importance of story units (Smiley et al. 1977), and knowledge about stories may also influence reading performance. For example, Charles A. Perfetti (1994) has proposed that a possible source of comprehension failure is inadequate reading experience; Carole Peterson and Pamela Dodsworth (1991) have noted that narrative production is used in school to develop children’s reading and writing skills. In a number of studies in narrative production, components similar to those involved in the text comprehension have been mentioned (e.g. Cain 2003, Mar 2004).

In this respect it should be reasonable to use narrative analysis as a tool to identify the mechanisms underlying text processes in reading in both the normal and language-impaired population.

During the last few decades, narrative analysis has been applied as a language assessment and/or intervention tool in many different fields, including but not limited to developmental psycholinguistics, psychology, linguistics, speech therapy, and education. The methodology of narrative analysis is considered ecologically valid and flexible, and thus applicable for different linguistic, cultural, and social populations. The great variety of methodological approaches to narrative analysis (*The Cat Story*, *The Horse Story* (Hickmann 2003), different versions of *The Frog Story* (e.g., Mayer 1967), *The Bus Story* (Renfrew 1969), *TNL* (Gillam, Pearson 2004), *The Stone Story* (Veneziano, Hudelot 2009), *ENNI* (Schneider et al. 2005), *MAIN* (Gagarina et al. 2012)) naturally leads to debates on the pros and cons of different methods and tools. After many years of discussions, personal narratives are considered to be informative when comparing different cultures and linguistic communities (McCabe, Bliss 2003) and also to serve as a clinical tool (Bliss et al. 1998, Bliss, McCabe 2008), while fiction stories (both telling and retelling) might be easily applicable for child language studies. However, a variety of different approaches to fiction narratives (e.g., task mode such as telling, retelling, or model story; visual stimuli such as picture, picture sequence, or video; shared vs. non-shared attention) not only enables but also encourages to search for the most relevant assessment tool for a particular linguistic, cultural, and social population. A number of studies have been based on comparative analysis between different presentations of stimuli in both typically-developing (TD) and language-impaired (LI) populations (Schneider 1996, Botting 2002, Schneider, Dubé 2005, Soodla et al. 2010, Soodla 2011); however, it must be noted that the majority of the studies focused on only linguistic measures such as macro- and microstructural characteristics, while the impact of cognitive factors (e.g. the subject's working memory, serial ordering, attention) still seems to be ignored (but see Mar 2004). The traditional static approach focuses on the result of narration, but the process-based approach (also called 'dynamic assessment') might significantly extend narrative analysis (Gillam et al. 1999).

There is an urgent need for complex methodology that combines linguistic and cognitive approaches and thus enables an analysis of both the **process** and the **result** of narration in the context of studies in language impairments. Language-impaired children, despite pure linguistic limitations, demonstrate emotional disorders (Kornev 2006), impaired executive function (Im-Bolter et al. 2006), limitations in visuo-spatial processing (Marton 2008), ADHD (Bellani et al. 2011), and low self-esteem (McAndrew 1999), thus the impact of non-linguistic factors on narrative comprehension/production should be particularly taken into account when assessing language-impaired populations.

There are relatively few studies devoted to cognitive experimental evaluation of the discourse (including narrative) process in both production and comprehension (Mar 2004). A story is, in part, a coherent causal-temporal ordering of selected information. It can be argued that a person must possess an ability to distinguish between story-significant and story-insignificant elements in order to construct a story. The capacity to properly sequence these events with regard to time and influence appears to be of fundamental importance for this process. Story grammar knowledge has been associated with the super-structural level of discourse organization, involving the integration of conceptual semantic and pragmatic information. This level of discourse processing is thought to be reliant on executive

functions (e.g., working memory, sequencing, and planning); see Coelho (2002), Tucker and Hanlon (1998). Experiments have provided evidence suggesting “that when evaluating narrative discourse ability, the integration of organizational and content measures yields an index of story goodness that was found to be sensitive and reliable in discriminating individuals with no brain injury and individuals with brain injury” (Lê et al. 2011: 124). The same seems reasonable for distinction between developmentally disordered individuals.

Results of our previous studies (Kornev, Balčiūnienė 2014) have confirmed that narrative macro- and microstructure are influenced by non-linguistic **procedural** variables such as cognitive resources that include working memory, serial ordering, etc. According to the competing model of speech processing in utterance programming, different components compete with each other for the cognitive resources necessary for parallel execution (MacWhinney, Bates 1989).

In our study, we attempted to evaluate the impact of non-linguistic factors such as **task mode** (telling vs. retelling), **story complexity** (relatively easy vs. complex story), and **session** (task order) on the comprehension and production of narrative macrostructure in dyslexic children. While comparing experimental and control groups, we expected to find the most sensitive linguistic measures that are directly dependent on non-linguistic factors and the conditions of narrative assessment in dyslexic children.

2. Methodology

2.1. Subjects

The subjects of the experiment were 12 Russian-speaking monolingual dyslexic children (mean age 9 years 9 months) living in Saint Petersburg and attending state schools. The inclusion criteria were extremely low indices of accuracy and fluency of reading, namely, ≥ 1.5 SD (standard deviation) below the average for the target age group, according to the standardized reading test for Russian-speaking children (Kornev 2003, Kornev, Išimova 2010). Children with mental retardation and vision or hearing impairments were excluded from the experiment. The control group consisted of 12 Russian-speaking monolingual TD peers attending the same schools in Saint Petersburg.

2.2. Visual stimuli

Two picture sequences, the *Baby Birds* and the *Baby Goats* (Gagarina et al. 2012), were selected for eliciting the narratives. Each sequence consisted of six coloured pictures (10 x 10 cm), without a text (see Appendix). In order to achieve comparability across narratives, the authors of the visual stimuli have aimed for congruence between the scripts and pictorial content by creating parallel storylines for the picture sequences. However, after a pilot study (Kornev, Balčiūnienė 2014), the *Baby-Goats* sequence is considered more complex to perceive because of the parallel development of two episodes and more numerous protagonists.

2.3. Procedure

Subjects from both the experimental and control groups performed two tasks, i.e. story generation (so called story-telling) and story retelling; both of the tasks were followed by ten comprehension questions. Each child was tested individually; the 1st and the 2nd session were separated by a few minutes of free talk between the interviewer and the child. The order of tasks was counterbalanced with regard to story complexity and task mode. Namely, half of the sample started with story-telling, while the others started with retelling. Half of the sample performed story-telling according to the *Baby Birds* sequence and retold a story according to the *Baby Goats* sequence, while the others did the opposite, i.e. they told a story according to the *Baby Birds* sequence and performed retelling according to the *Baby Goats* sequence.

2.4. Measures

In our study, both the micro- and macrostructure of narratives were analysed, but in the current paper only the **macrostructure**, namely, story structure and episode completeness, will be discussed.

Story structure (also called ‘story grammar’ (Stein, Glenn 1979), ‘narrative quality’ (Fey et al. 2004), or ‘plotline’ (Duinmeijer et al. 2012), etc.) has been analysed in a great number of studies. It can be generally interpreted as a set of structural elements that are logically connected to each other by temporal/causal means and thus create a coherent story. A concept of story structure was initially suggested by Vladimir Propp (1968) based on Russian folktales, and later on it was modified by a number of researchers and applied as a valuable tool for analysing fictional stories. According to the methodology of story grammar, a good narrative begins with a setting, continues with an initiating event or some explicit problem, the protagonist’s attempt to solve a problem, and the result of that attempt (McCabe, Bliss 2003). Table 1 presents the set of structural elements of both the *Baby Birds* and the *Baby Goats* sequences.

Table 1. Macrostructural framework (a set of structural elements) of the sequences

Episode	Element	<i>Baby Birds</i>	<i>Baby Goats</i>
	Setting	<i>One day...</i>	<i>Once upon a time...</i>
1	Goal	<i>The mother bird wants to feed her chicks.</i>	<i>The mother goat wants to help her baby goat.</i>
	Attempt	<i>The mother bird flies away.</i>	<i>The goat runs into the water.</i>
	Outcome	<i>The mother bird brings a worm.</i>	<i>The goat saves the baby goat.</i>
2	Goal	<i>The cat wants to catch the chicks.</i>	<i>The fox wants to catch the other baby goat.</i>
	Attempt	<i>The cat starts climbing the tree.</i>	<i>The fox grabs the baby goat.</i>
	Outcome	<i>The dog stops the cat.</i>	<i>The bird stops the fox.</i>
3	Goal	<i>The dog wants to help the chicks.</i>	<i>The bird wants to help the baby goat.</i>
	Attempt	<i>The dog grabs the cat’s tail.</i>	<i>The bird grabs the fox’s tail.</i>
	Outcome	<i>The cat runs away.</i>	<i>The fox runs away.</i>

During the experiment, each of the structural elements (i.e. setting, goal, attempt, and outcome of each episode) was scored 1 point, thus the story structure was scored 0–10 points in total. The original scoring version was suggested by Gagarina et al. (2012), but later on it was modified by the authors of this paper (Kornev, Balčiūnienė 2014), in order to develop a more flexible and sensitive tool for scoring narrative structure.

Episode completeness is a relatively novel measure in narrative studies. In addition to the story structure (i.e. a purely quantitative characteristic), episode complexity/completeness (i.e. a qualitative characteristic) has been measured in a few studies (Coggins et al. 1998). Despite treating Goal, Attempt and Outcome as merely classical components of event structure, these elements have been evaluated from the perspective of the narrator’s programming ability. Following the latter methodology, all episodes are classified into complete and incomplete ones. A complete episode consists of 1) an initiating event that prompts a character to act (Goal), 2) an Attempt related to the goal, and 3) a direct consequence (Outcome) of the attempt. A partial (incomplete) episode consists of one or two components (e.g., goal and attempt, or bare attempt). Unfortunately, just a few studies (Reilly et al. 1998, Grice, Wright-Harp 2004, Lê et al. 2011) have attempted to analyse the incomplete episodes, thus almost nothing is known about their structure and functions. It should be considered that separate episode completeness is based on some processes distinct from the whole story structure organisation. Episode completeness thus appears to be a useful macrostructural measure for narrative evaluation.

In our study, the episodes were classified into complete and incomplete ones and all of them were scored accordingly (see Table 2).

Table 2. Scoring episode completeness

Episode	Structure	Example of an episode	Points
Complete	Goal-Attempt-Outcome	<i>The mother bird wanted to feed her chicks. She flew away and brought them a worm.</i>	4
Incomplete	Goal-Attempt	<i>The mother bird wanted to feed her chicks, thus she flew away.</i>	3.5
	Goal-Outcome	<i>The mother bird wanted to feed her chicks and bring them a worm.</i>	3
	Attempt-Outcome	<i>The mother bird flew away and brought a worm to her chicks.</i>	2
	Bare Goal	<i>The mother bird wanted to feed her chicks.</i>	1.5
	Bare Attempt	<i>The mother bird flew away.</i>	1
	Bare Outcome	<i>The mother bird brought a worm to her chicks.</i>	1.5

As can be seen, the most valuable (4 points) episodes were the complete ones, while bare Attempts were scored only 1 point. Since for both the *Baby Birds* and the *Baby Goats* sequences three episodes (the *mother bird*, the *cat*, and the *dog*; the *mother goat*, the *fox*, and the *bird*) were designed, the episode completeness was scored 0–12 points in total.

3. Results

At the first step of analysis, the 1st session results were recognised as the basic level of achievements in narrative production. A one-way ANOVA¹ (Mertler, Vannata 2002, Larsen, Marx 2011) analysis revealed the only distinction between the groups, in episode completeness ($F = 4.31$; $P < 0.05$). A factorial ANOVA² evidenced that the indices of story structure complexity and episode completeness were influenced in participants by different variables: structure complexity was influenced by the task mode ($F = 6.05$; $P < 0.026$; $\eta^2 = 0.274$) while episode completeness was influenced by the group the participant belonged to ($F = 5.45$; $P < 0.033$; $\eta^2 = 0.25$). Namely, story-telling episode completeness was significantly worse in the dyslexic group than in the TD children ($F = 10.15$; $P < 0.01$). Story-telling structure complexity was also less advanced in the dyslexics than in the TD group but the difference was not statistically significant ($F = 4.39$; $P < 0.06$).

Turning to the particular non-linguistic factors, it should be noted that the task mode significantly influenced the story structure complexity as was mentioned above; the story structure complexity index was significantly higher in retelling than in telling ($F = 5.57$; $P < 0.028$). Episode completeness was significantly influenced by the task mode only in the *Baby Goats* story ($F = 8.02$; $P < 0.022$; $\eta^2 = 0.50$). Both dyslexics and TD children had the highest score in this index. In story retelling, dyslexics did not differ from the TD children in any indices.

At the second step of analysis, the influence of the task order on narrative macrostructure was estimated. In this study, the main point of our interests was **dynamic changes** of narrative production quality influenced by recent experience. Hence, we attempted to evaluate the quality of a) story telling in the 2nd session preceded by a story retelling in the 1st session, and b) story retelling in the 2nd session preceded by a story telling in the 1st session. Our results confirmed the preliminary expectation that the execution of story retelling in the 1st session should influence the macrostructural features of story-telling in the 2nd session. Namely, story-telling structure and episode completeness were more advanced in the 2nd session than in the 1st one; however, this result was obtained within only the dyslexic group (see Figures 1a, 1b). Because of the small size of the groups, a significant difference was found only in the episode completeness indices ($F = 4.08$; $P < 0.05$). This distinction was valid for only the *Baby Birds* story (see Figure 2a, 2b).

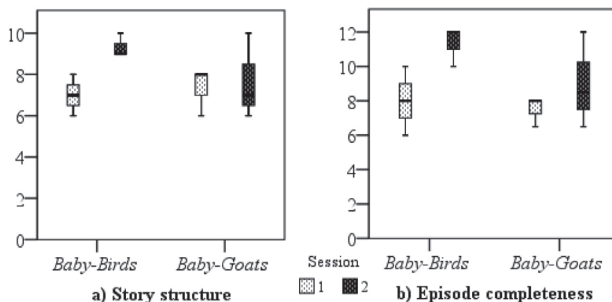


Figure 1. Story structure and episode completeness in storytelling in dyslexics

¹ The one-way analysis of variance (ANOVA) is used to determine whether there are any significant differences between the means of independent (unrelated) groups.

² Factorial ANOVA or a general linear model is a general procedure for analysis of variance and covariance as well as regression.

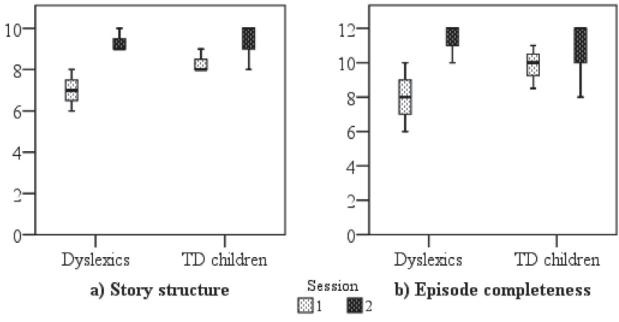


Figure 2. Story structure and completeness in telling the *Baby-Birds*

As a novel approach in narrative assessment, a qualitative analysis of episode completeness was carried out, i.e. the distribution of structural types of episodes was estimated. The most prevalent episode structural types were the following: ‘Goal-Attempt-Outcome’, ‘Goal-Outcome’, ‘Goal-Attempt’, ‘Attempt-Outcome’, ‘Bare goal’, ‘Bare Attempt’ and ‘Bare Outcome’. From the perspective of a dynamic approach, we attempted to evaluate the impact of the non-linguistic factors, i.e., ‘session’ (task order), ‘story’ (story complexity), and ‘mode’, on episode completeness. In order to achieve the most reliable results, story-telling and story retelling were analysed separately.

A factorial ANOVA analysis revealed that structure complexity was influenced by the variable ‘story’ and the interacting variables ‘story’ x ‘session’ (respectively, $F = 7.0$; $P < 0.018$; $\eta^2 = 0.30$; and $F = 5.14$; $P < 0.038$; $\eta^2 = 0.24$); episode completeness was influenced by the interacting variables ‘group’ x ‘session’ ($F = 5.36$; $P < 0.034$; $\eta = 0.251$). In the *Baby Goats* story presented in the 1st session, the groups performed differently: the dominant episode structure pattern in TD children was ‘GAO+G+A’ (the difference between the groups was significant; $F = 12.31$; $P < 0.006$), while the dyslexic children preferred the ‘GO+AO+O’ structure (respectively, $F = 3.91$; $P < 0.076$) (see Figure 3). Moreover, in dyslexics, structurally incomplete episodes were prevalent both in telling and retelling.

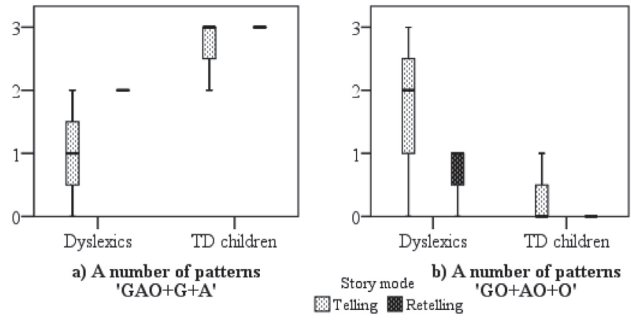


Figure 3. Episode structure patterns in telling the *Baby Goats*

One more significant factor for story telling was session (the *task order*). In the TD children, complete episodes were more numerous in the 1st session (see Table

4 and Figure 3), while the dyslexics demonstrated the reverse tendency (although the difference was not significant).

As for story retelling, the task order did not have any influence on episode completeness in either dyslexic or TD children (see Table 3).

Table 3. Macrostructural measures

Measure	DY children N= 12		TD children N= 12		df	F	Sig.
	Mean	Std Dev	Mean	Std Dev			
Telling							
Story structure	7.83	1.47	8.0	1.21	1	0.092	0.76
Episode completeness	8.96	2.21	9.67	1.76	1	0.753	0.39
Internal state terms	4.00	1.91	4.92	3.09	1	0.77	0.39
Comprehension questions	8.42	1.08	9.33	1.15	1	4.02	0.06
Retelling							
Story structure	8.0	1.59	8.58	1.0	1	1.01	0.33
Episode completeness	9.62	1.94	10.46	1.50	1	1.02	0.32
Internal state terms	4.17	3.01	4.83	2.44	1	0.36	0.56
Comprehension questions	9.00	0.85	8.75	1.36	1	0.29	0.59

Table 4. Impact of non-linguistic factors on macrostructural measures in dyslexic and TD children's storytelling

Independent variable	Dyslexic children N = 12			TD children N = 12		
	F	Sig.	η^2	F	Sig.	η^2
Story Structure						
Story	0.76	0.41	0.087	14.29	0.005	0.641
Story x Session	1.71	0.227	0.176	5.14	0.053	0.391
Episode completeness						
Session	5.07	0.054	0.388	0.85	0.384	0.096

4. Discussion and conclusions

A multitude of studies in dyslexia have focused on different aspects of linguistic deficits (Nation, Snowling 2004, Casalis, Cole 2004, Hogan et al. 2005, Robertson, Joanisse 2010). Similarly, narrative limitations found in dyslexics have usually been recognized in the scope of narrative language weakness (Davenport et al. 1986, Westerveld et al. 2008). However, according to the resource deficit hypothesis, the core limitations in dyslexics are caused by non-linguistic factors, namely by a cognitive resource deficit (van der Schoot et al. 2000, Kibby et al. 2004). In some studies, the applicability of the cognitive approach to narrative analysis has been confirmed (Mar 2004, Wright et al. 2011). The present study aimed at testing the dynamic model in application to dyslexia by the means of narrative assessment.

In the current paper, we focused mainly on macrostructural measures and their dependency on the experimental setting, task presentation mode, and personal new

experience gained during the experiment. Although the results of our experiment generally coincide with the findings of previous studies in dyslexic narrative language (Plaza 2000, Westerveld et al. 2008), the linguistic deficit in our dyslexic group was not as dramatic as initially expected. In the first session when performing the relatively easy task, i.e. the *Baby Birds* story, the dyslexic children did not differ from the TD children in either telling or retelling. Significant differences between the groups were revealed only in the more complex task, i.e. the *Baby Goats* story; the episode completeness was much less advanced in the dyslexic children than in the TD peers. The dyslexics produced the same amount of structural elements as TD children but in each event description they tended to form structurally simpler and/or incomplete schemes without Goal or Attempt elements. Again, in the 2nd session in the less complex story they did not differ from the TD peers.

The discussion of the results should be started by stating that limitations in narrative text production are presumed to be caused by several different underlying mechanisms. One of them is inefficiency in developing logical (temporal and causal) relationships between events. Another one is difficulties in structuring an episode description. The episode is the central unit in the majority of story grammar models (e.g., Stein, Glenn 1979, Thorndyke, Yakovitch 1980). Episode components are defined as statements bearing information about some characters' goal, their attempts to solve the problem, and the consequences of their attempts (Liles et al. 1989). The production and structuring of episode descriptions is thought to involve processes that are not exclusively linguistic (Coelho et al. 1994).

Theoretically, narrative text limitations in dyslexics might be recognized as a consequence of their low reasoning capacity. This may be one of the expected factors that influenced the observed poverty in episode structure completeness. However, the dynamic approach indicated that a short experience of story retelling in the 1st session provided a sufficient base for the dyslexic children to improve their story structure and episode completeness in the 2nd session. Consequently, the limitations in narrative structure should be treated rather as a procedural inefficiency related to a cognitive resource deficit (Coelho et al. 1994, Ferstl, Cramon 2001, Mar 2004). The influence of the 1st session task might be recognised as a priming effect (Snedeker, Thothathiri 2008, Eisenbeiss 2010). Execution of the retelling task in the 1st session probably activates a cerebral network that underlies the story production process and enables structural composition. This effect was extremely evident when the more complex picture sequence (i.e. the *Baby Goats*) was presented for retelling in the 1st session and followed by a telling of the less complex the *Baby Birds* story.

The other problem the dyslexics faced was the qualitative organization of episode structure. As was mentioned above, episodes are the basic elements of story grammar. Following van Dijk and Kintsch (1983), we recognize episodes as the macro-propositions that compose the plot. In our study, each of the episodes included 3 propositions, i.e. a goal, attempt, and outcome. Hence, when telling a story, children had to recognise each of the propositions, to relate them into a logical sequence, and to verbalise the sequence. The results of our study highlighted that children with dyslexia generally tended to produce more structurally incomplete episodes in comparison to their TD peers. Most often in dyslexics the Goal and the Attempt were omitted. The omission of the Goal should be related to the absence of

direct pictorial signs for this structural component. The child has to derive the Goal from the whole sense of the event. The omission of the Attempt could be explained by the unstructured global event recognition of dyslexics. They covertly imply the Attempt in the description of the Outcome (e.g. *The baby birds were saved and happy. The fox was scared by the bird.*). This tendency was extremely evident in the 1st session. In the inverse task order, i.e. story telling in the 2nd session preceded by story retelling in the 1st session and answering comprehension questions, episode completeness significantly increased and became equal to that of TD children (see Figure 2b). Taking into account this flexibility and the dynamic changes in episode structure, it should be reasonable to argue that this limitation was caused by procedural factors, not by a deficit in reasoning or language. Some previous studies have shown that the larger utterance units are explored in discourse planning, the more cognitive resources are required (Swets et al. 2014). Consequently, the cognitive resource deficit should result in simple and short phrases as well as structurally less complex narrative text.

These explanations do not exclude the fact of linguistic drawbacks in the dyslexic narrative production. In our previous paper (Kornev, Balčiūnienė 2014), we have argued for evidence of a semantic deficit in the denomination of the protagonists. Hence, our future studies in dyslexic children narratives will be focused not only on the macro- but also on the microstructural analysis.

In conclusion, it should be noted that our experience with the dynamic approach to narrative assessment lends support to its value as a research tool. The novel episode completeness analysis proved to be an effective and informative method that can highlight new mechanisms of narration and thus extend the classic narrative analysis by the addition of qualitative information.

Abbreviations

A	attempt
ADHD	attention deficit hyperactivity disorder
DY	dyslexic
G	goal
LI	language-impaired
O	outcome
TD	typically developing

References

- Bellani, Marcella; Moretti, Anna; Perlini, Cinzia; Brambilla, Paolo 2011. Language disturbances in ADHD. – *Epidemiology and Psychiatric Sciences*, 20 (4), 311–315. <http://dx.doi.org/10.1017/S2045796011000527>
- Bliss, Lynn S.; McCabe, Allyssa 2008. Personal narratives. Cultural differences and clinical implications. – *Topics in Language Disorders*, 28 (2), 162–177. <http://dx.doi.org/10.1097/01.TLD.0000318936.31677.2d>
- Bliss, Lynn S.; McCabe, 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)
- Botting, Nicola 2002. Narrative as a tool for the assessment of linguistic and pragmatic. – *Child Language Teaching and Therapy*, 18 (1), 1–21. <http://dx.doi.org/10.1191/0265659002ct2240a>

- Cain, Kate 2003. Text comprehension and its relation to coherence and cohesion in children's fictional narratives. – *British Journal of Developmental Psychology*, 21 (3), 335–351. <http://dx.doi.org/10.1348/026151003322277739>
- Casalis, Severine; Pascale, Cole 2004. Morphological awareness in developmental dyslexia. – *Annals of Dyslexia*, 54 (1), 114–138. <http://dx.doi.org/10.1007/s11881-004-0006-z>
- Coelho, Carl A. 2002. Story narratives of adults with closed head injury and non-brain-injured adults: Influence of socioeconomic status, elicitation task, and executive functioning. – *Journal of Speech, Language, and Hearing Research*, 45 (December), 1232–1248. [http://dx.doi.org/10.1044/1092-4388\(2002/099\)](http://dx.doi.org/10.1044/1092-4388(2002/099))
- Coelho, Carl A.; Liles, Betty Z.; Duffy, Robert 1994. Cognitive framework: A description of discourse abilities in traumatically brain-injured adults. – Ronald L. Bloom, Loraine K. Obler, Susan De Santi, Jonathan S. Ehrlich (Eds.). *Discourse Analysis and Applications: Studies in Adult Clinical Populations*. Hillsdale, NJ: Erlbaum, 96–110.
- Coggins, Truman E.; Friet, T.; Morgan, T. 1998. Analysing narrative productions in older school-age children and adolescents with fetal alcohol syndrome: An experimental tool for clinical applications. – *Clinical Linguistics & Phonetics*, 12 (3), 221–236. <http://dx.doi.org/10.3109/0269920980808985223>
- Davenport, Linda; Yingling, Charles D.; Fein, George; Galin, David; Johnstone, Jack; Porter, Langley 1986. Narrative speech deficits in dyslexics. – *Journal of Clinical and Experimental Neuropsychology*, 8 (4), 347–361. <http://dx.doi.org/10.1080/01688638608401326>
- Dijk, Teun A. van; Kintsch, Walter 1983. *Strategies of Discourse Comprehension*. NY: New York Academic Press.
- Duinmeijer, Iris; de Jong, Jan; Scheper, Annette 2012. Narrative abilities, memory and attention in children with a specific language impairment. – *International Journal of Language & Communication Disorders*, 47 (5), 542–555. <http://dx.doi.org/10.1111/j.1460-6984.2012.00164.x>
- Eisenbeiss, Sonja 2010. Production methods in language acquisition research. – Elma Blom, Sharon Unsworth (Eds.). *Experimental Methods in Language Acquisition Research*. Amsterdam: John Benjamins Publishing, 11–34.
- Ferstl, Evelyn C.; von Cramon, D. Yves 2001. The role of coherence and cohesion in text comprehension: An event-related fMRI study. – *Cognitive Brain Research*, 11 (3), 325–340. [http://dx.doi.org/10.1016/S0926-6410\(01\)00007-6](http://dx.doi.org/10.1016/S0926-6410(01)00007-6)
- Fey, Marc E.; Catts, Hugh W.; Proctor-Williams, Kerry; Tomblin, J. Bruce; Zhang, Xuyang 2004. Oral and written story composition skills of children with language impairment. – *Journal of Speech, Language, and Hearing Research*, 47 (December), 1301–1318. [http://dx.doi.org/10.1044/1092-4388\(2004/098\)](http://dx.doi.org/10.1044/1092-4388(2004/098))
- Gagarina, Natalia; Klop, Daleen; Kunnari, Sari; Tantele, Koula; Välimaa, Taina; Balčiūnienė, Ingrida; Bohnacker, Ute; Walters, Joel 2012. MAIN: Multilingual Assessment Instrument for Narratives. *ZAS Papers in Linguistics* 56. Berlin: ZAS.
- Gillam, Ronald B.; Pearson, Nils A. 2004. *Test of Narrative Language*. Austin, TX: Pro-Ed.
- Gillam, Ronald B.; Peña, Elisabeth D.; Miller, Lynda 1999. Dynamic assessment of narrative and expository discourse. – *Topics in Language Disorders*, 20 (1), 33–47. <http://dx.doi.org/10.1097/00011363-199911000-00005>
- Gorin, Joanna; Embretson, Susan E. 2006. Item difficulty modeling of paragraph comprehension items. – *Applied Psychological Measurement*, 30 (5), 394–411. <http://dx.doi.org/10.1177/0146621606288554>
- Grice, Angela M.; Wright-Harp, Wilhelmina 2004. The narrative performance of adolescents and emerging adults. – Paper presented at the NIDCD Research Symposium in Aphasiology. Park City, UT.
- Hickmann, Maya 2003. *Children's Discourse: Person, Space and Time Across Languages*. Cambridge: Cambridge University Press.

- Hogan, Tiffany P.; Catts, Hugh W.; Little, Todd D. 2005. The relationship between phonological awareness and reading: Implications for the assessment of phonological awareness. – *Language, Speech, and Hearing Services in Schools*, 36 (4), 285–293. [http://dx.doi.org/10.1044/0161-1461\(2005/029\)](http://dx.doi.org/10.1044/0161-1461(2005/029))
- Im-Bolter, Nancie; Johnson, Janice; Pascual-Leone, Juan 2006. Processing limitations in children with specific language impairment: The role of executive function. – *Child Development*, 77 (6), 1822–1841. <http://dx.doi.org/10.1111/j.1467-8624.2006.00976.x>
- Kibby, Michelle Y.; Kroese, Judith M.; Morgan, Allison E.; Hiemenz, Jennifer R.; Cohen, Moris J.; Hynd, George W. 2004. The relationship between perisylvian morphology and verbal short-term memory functioning in children with neurodevelopmental disorders. – *Brain and Language*, 89 (1), 122–135. [http://dx.doi.org/10.1016/S0093-934X\(03\)00310-9](http://dx.doi.org/10.1016/S0093-934X(03)00310-9)
- Kintsch, Walter 1988. The role of knowledge in discourse comprehension construction-integration model. – *Psychological Review*, 95 (2), 163–182. <http://dx.doi.org/10.1037/0033-295X.95.2.163>
- Kintsch, Walter 1998. *Comprehension: A Paradigm for Cognition*. NY: Cambridge University Press.
- Kintsch, Walter; Dijk, Teun A. van 1978. Toward a model of text comprehension and production. – *Psychological Review*, 85 (5), 363–394. <http://dx.doi.org/10.1037/0033-295X.85.5.363>
- Kornev, Aleksandr N. 1995. *Disleksija i disgrafija u detej*. Sankt-Peterburg: Gippokrat.
- Kornev, Aleksandr N. 2003. *Narušenija čtenija i pis'ma*. Sankt-Peterburg: Reč'.
- Kornev, Aleksandr N. 2006. *Sistemnyj analiz psihologičeskogo razvitija detej s nedorazvitiem reči. Avtoreferat dissertacii na soiskanie učennoj stepeni doktora psihologičeskijh nauk*. Sankt-Peterburg: Psihonevrologičeskij institut im. B. M. Behtereva.
- Kornev, Aleksandr N.; Balčiūnienė, Ingrida 2014. Story (re-)telling and reading in children with dyslexia: Language or cognitive resource deficit? – *Book of Abstracts: LSCD-2014*. London: UCL, 23–26.
- Kornev, Aleksandr N.; Išimova, Olga A. 2010. *Metodika diagnostiki disleksii u detej. Metodičeskoe posobie*. Sankt-Peterburg: Izdatel'stvo Politehničeskogo universiteta.
- Kornev, Aleksandr N.; Rakhlin, Natalia; Grigorenko, Elena L. 2010. Dyslexia from a cross-linguistic and cross-cultural perspective: The case of Russian and Russia. – *Learning Disabilities: A Contemporary Journal*, 8 (1), 51–78.
- Larsen, Richard J.; Marx, Morris L. 2011. *Introduction to Mathematical Statistics and Its Applications*. 5th edition. Prentice Hall.
- Lê, Karen; Coelho, Carl A.; Mozeiko, Jennifer; Grafman, Jordan 2011. Measuring goodness of story narratives. – *Journal of Speech, Language, and Hearing Research*, 54 (February), 118–126. [http://dx.doi.org/10.1044/1092-4388\(2010/09-0022\)](http://dx.doi.org/10.1044/1092-4388(2010/09-0022))
- Leontyev, Aleksej A. 1997. *Osnovy psiholingvistiki*. Moskva: Smysl.
- Leontyev, Aleksej A. 2004. *Jazyk i rečevaja dejatel'nost' v obščej i pedagogičeskijh psihologii*. Moskva: Izdatel'stvo Moskovskogo psihologo-social'nogo instituta.
- Liles, Betty Z.; Coelho, Carl A.; Duffy, Robert; Zalagens, Mary Rigdon 1989. Effects of elicitation procedures on the narratives of normal and closed head-injured adults. – *Journal of Speech and Hearing Disorders*, 54 (August), 356–365. <http://dx.doi.org/10.1044/jshd.5403.356>
- Lindgren, Scott D.; de Renzi, Ennio; Richman, Lynn C. 1985. Cross-national comparisons of developmental dyslexia in Italy and the United States. – *Child Development*, 56 (6), 1404–1417. <http://dx.doi.org/10.2307/1130460>
- Lyon, G. Reid; Shaywitz, Sally E.; Shaywitz, Bennet A. 2003. Defining dyslexia, comorbidity, teachers' knowledge of language and reading. – *Annals of Dyslexia*, 53 (1), 1–14. <http://dx.doi.org/10.1007/s11881-003-0001-9>

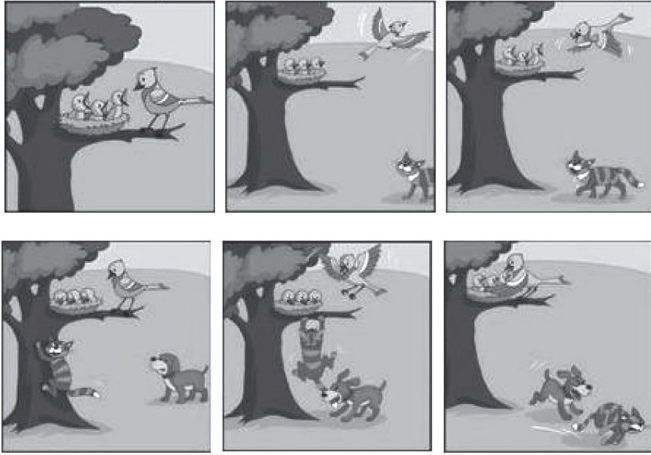
- MacWhinney, Brian; Bates, Elisabeth 1989. Functionalism and the competition model. – B. MacWhinney, E. Bates (Eds.). *The Crosslinguistic Study of Sentence Processing*. New York: Cambridge University Press, 3–73.
- Mar, Raymond A. 2004. The neuropsychology of narrative: Story comprehension, story production and their interrelation. – *Neuropsychologia*, 42 (10), 1414–1434. <http://dx.doi.org/10.1016/j.neuropsychologia.2003.12.016>
- Marton, Klara 2008. Visuo-spatial processing and executive functions in children with specific language impairment. – *International Journal of Language & Communication Disorders*, 43 (2), 181–200. <http://dx.doi.org/10.1080/16066350701340719>
- Mayer, Mercer 1967. *A Boy, a Dog and a Frog*. New York: Penguin.
- McAndrew, Edell 1999. The relationship between self-esteem and language disordered children. – *Child Language Teaching and Therapy*, 15 (3), 219–232. <http://dx.doi.org/10.1191/026565999666834880>
- McCabe, Allyssa; Bliss, Lynn S. 2003. *Patterns of Narrative Discourse: A Multicultural, Life Span Approach*. Boston, MA: Allyn and Bacon.
- Mertler, Craig A.; Vannata, Rachel A. 2002. *Advanced and Multivariate Statistical Methods. Practical Application and Interpretation*. 2nd edition. McNaughton and Gunn, Inc.
- Nation, Katie; Snowling, Margaret J. 2004. Beyond phonological skills: Broader language skills contribute to the development of reading. – *Journal of Research in Reading*, 27 (4), 342–356. <http://dx.doi.org/10.1111/j.1467-9817.2004.00238.x>
- Perfetti, Charles A. 1994. Psycholinguistics and reading ability. – Morton A. Gernsbacher (Ed.). *Handbook of Psycholinguistics*. San Diego, CA: Academic Press, 849–894.
- Peterson, Carole; Dodsworth, Pamela 1991. A longitudinal analysis of young children's cohesion and noun specification in narratives. – *Journal of Child Language*, 18 (2), 397–415. <http://dx.doi.org/10.1017/S0305000900011120>
- Plaza, Monique 2000. Reference and evaluation in narrative speech of a group of French-speaking children. – *Psycholinguistics on the Threshold of the Year 2000: Proceedings of the 5th International Congress of the International Society of Applied Psycholinguistics*, 665–668.
- Pressley, Michael 2006. Children who experience problems in learning to read. – M. Pressley (Ed.). *Reading Instruction That Works. Third Edition: The Case for Balanced Teaching*. The Guilford Press, 66–95.
- Propp, Vladimir 1968. *Morphology of the Folktale*. Austin, TX: University of Texas Press.
- 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>
- Renfrew, Catherine E. 1969. *The Bus Story: A Test of Continuous Speech*. North Place, Old Headington: Oxford.
- Robertson, Erin K.; Joanisse, Marc F. 2010. Spoken sentence comprehension in children with dyslexia and language impairment: The roles of syntax and working memory. – *Applied Psycholinguistics*, 31 (1), 141–165. <http://dx.doi.org/10.1017/S0142716409990208>
- Schneider, Phyllis 1996. Effects of pictures versus orally presented stories on story retellings by children with language impairment. – *American Journal of Speech-Language Pathology*, 5 (February), 86–96. <http://dx.doi.org/10.1044/1058-0360.0501.86>
- Schneider, Phyllis; Dubé, Rita V. 2005. – *American Journal of Speech-Language Pathology*, 14 (February), 52–60. [http://dx.doi.org/10.1044/1058-0360\(2005/007\)](http://dx.doi.org/10.1044/1058-0360(2005/007))
- Schneider, Phyllis; Dubé, Rita V.; Hayward, Denise V. 2005. *The Edmonton Narrative Norms Instrument*. Retrieved from University of Alberta Faculty of Rehabilitation Medicine website: <http://www.rehabresearch.ualberta.ca/enni> (27.2.2015).
- Shaywitz, Sally E.; Shaywitz, Bennett A. 2003. Dyslexia (specific reading disability). – *Pediatrics in Review*, 24 (5), 147–153.

- Smiley, Sandra S.; Oakley, Drew D.; Worthen, David; Campione, Joseph C.; Brown Ann L. 1977. Recall of thematically relevant material by adolescent good and poor readers as a function of written versus oral presentation. – *Journal of Education Psychology*, 69 (4), 381–387. <http://dx.doi.org/10.1037/0022-0663.69.4.381>
- Snedeker, Jesse; Thothathiri, Malathi 2008. What lurks beneath: Syntactic priming during language comprehension in preschoolers (and adults). – Irina A. Sekerina, Eva M. Fernández, Harald Clahsen (Eds.). *Developmental Psycholinguistics: On-line methods in Children's Language Processing. Language Acquisition and Language Disorders 44*. Amsterdam: John Benjamins Publishing Company, 137–168. <http://dx.doi.org/10.1075/lald.44>
- Snowling, Margaret J. 1996. Dyslexia: A hundred years on. – *BMJ*, 313 (2), 1096–1097. <http://dx.doi.org/10.1136/bmj.313.7065.1096>
- Soodla, Piret 2011. *Picture-Elicited Narratives of Estonian Children at the Kindergarten-School Transition as a Measure of Language Competence*. Tartu: Tartu Ülikooli Kirjastus. <http://hdl.handle.net/10062/16715>
- 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.] – *Eesti Rakenduslingvistika Ühingu aastaraamat*, 6, 277–296. <http://dx.doi.org/10.5128/ERYa6.17>
- Stein, Nancy L.; Glenn, Christine G. 1979. An analysis of story comprehension in elementary school children. – R. Freedle (Ed.). *Discourse Processing: Multidisciplinary Perspectives*. Norwood, NJ: Ablex.
- Swets, Benjamin; Jacovina, Matthew E.; Gerrig, Richard J. 2014. Individual differences in the scope of speech planning: Evidence from eye-movements. – *Language and Cognition*, 6 (1), 12–44. <http://dx.doi.org/10.1017/langcog.2013.5>
- Thorndyke, Perry W.; Yakovitch, Frank R. 1980. A critique of schema-based theories of human story memory. – *Poetics*, 9 (1–3), 23–49. [http://dx.doi.org/10.1016/0304-422X\(80\)90011-X](http://dx.doi.org/10.1016/0304-422X(80)90011-X)
- Tucker, Frances M.; Hanlon, Robert E. 1998. Effects of mild traumatic brain injury on narrative discourse production. – *Brain Injury*, 12 (9), 783–792. <http://dx.doi.org/10.1080/026990598122179>
- Schoot, Menno van der; Licht, Robert; Horsley, Tako M.; Sergeant, Joseph A. 2000. Inhibitory deficits in reading disability depend on subtype: Guessers but not spellers. – *Child Neuropsychology*, 6 (4), 297–312. <http://dx.doi.org/10.1076/chin.6.4.297.3139>
- 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>
- Voeller, Kytja K. S. 2004. Dyslexia. – *Journal of Child Neurology*, 19 (10), 740–744.
- Westerveld, Marleen F.; Gillon, Gail T.; Moran, Catherine 2008. A longitudinal investigation of oral narrative skills in children with mixed reading disability. – *International Journal of Speech-Language Pathology*, 10 (3), 132–145. <http://dx.doi.org/10.1080/14417040701422390>
- Wright, Heather H.; Capilouto, Gilson J.; Srinivasan, Cidambi; Fergadiotis, Gerasimos 2011. Story processing ability in cognitively healthy younger and older adults. – *Journal of Speech, Language, and Hearing Research*, 54 (3), 900–917. [http://dx.doi.org/10.1044/1092-4388\(2010/09-0253](http://dx.doi.org/10.1044/1092-4388(2010/09-0253)

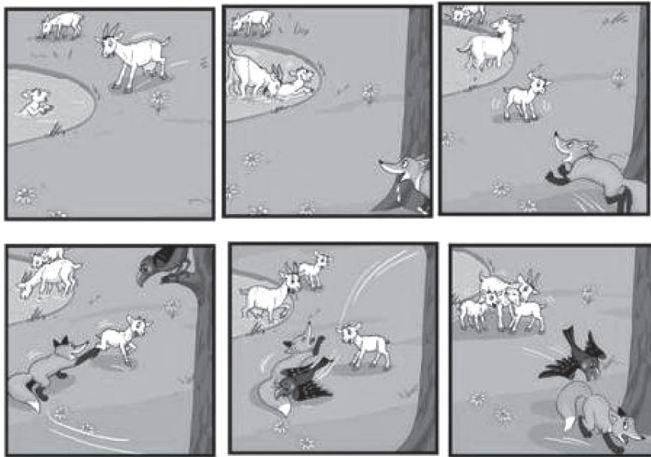
Appendix.

Stimulus material for the study (based on Gagarina et al. 2012)

The *Baby Birds* sequence



The *Baby Goats* sequence



Ingrida Balčiūnienė (Vytautas Magnus University). Her research interests include first language acquisition, narrative analysis, and discourse analysis.

K. Donelaičio 58, 44248 Kaunas, Lithuania

i.balciuniene@hmf.vdu.lt

Aleksandr N. Kornev (Saint-Petersburg State Pediatric Medical University). His research interests include child clinical psychology, speech/language pathology, and dyslexia.

Litovskaya 2, 194100, Saint Petersburg, Russia

k1949@yandex.ru

NARRATIIVILOOME NÕRKUS VENE DÜSLEKTIKUTEL: KEELELINE VÕI PROTSEDUURILINE PIIRATUS?

Aleksandr N. Kornev¹, Ingrida Balčiūnienė²

Peterburi Riiklik Pediaatriameditsiini Ülikool¹, Vytautas Magnuse Ülikool²

Uuring käsitleb keeleväliste tegurite mõju venekeelsete düslektiliste laste narratiiviloomele. Katserühm koosnes 12 düsleksiaga lapsest vanuses 9–10 aastat, kontrollrühmas oli 12 samaealist arenguhäireta last. Valim oli tasakaalustatud narratiivi laadi, jutustuse kompleksuse ja ülesannete järjekorra aspektist. Laiendamine klassikalist narratiivianalüüsi uudse dünaamilise lähenemisega, mis võimaldas hinnata narratiiviloome protseduurilisi karakteristikuid.

Uurimistulemused tõid välja kaks düslektikute narratiivikeele piiratuse põhjust: raskused inimeste ja sündmuste vaheliste loogiliste (ajaliste, põhjuslike) seoste kujundamises ja jutustuse episoodstruktuuri organiseerimises. Episoodstruktuuri suur paindlikkus ja dünaamilised muutused selles annavad põhjust järeldada, et düslektikute narratiivide keele piiratust tingivad pigem protseduurilised kui keelised tegurid. Siiski leidub ka puhtkeelalisi vajakajäämisi.

Dünaamiline lähenemine narratiivi hindamisele on meie kogemuse põhjal vajalik uurimisviis. Episoodide täielikkuse dünaamiline käsitus näitas end meie uurimuses tõhusa ja informatiivse meetodina, mis võib esile tuua uusi jutustamise mehhanisme ja seega avardada klassikalist narratiivianalüüsi kvalitatiivse informatsiooni abil.

Võtmesõnad: esimese keele omandamine, narratiivianalüüs, keele patoloogia, düsleksia, vene keel

DIREKTIIVSED KÕNEAKTID NING NENDE KASUTAMISE DÜNAAMIKA ISA JA EMA KÕNES: JUHTUMIUURING

Helen Kõrgesaar

Ülevaade. Artikli eesmärk on anda ülevaade direktiividest eesti lapsele suunatud kõnes. Vaatluse all on kaks sama pere kooliealist last ja nende dialoogid mõlema vanemaga eraldi. Meetodina on kasutatud pikiuuringut (üht last on lindistatud regulaarselt vanuses 7–10 eluaastat, teist vanuses 10–14 eluaastat), mis hõlmab 16 tunni jagu audiomaterjali. Vanemate kõne direktiivsed lausungid on liigitatud otsesteks ja kaudseteks käskudeks, keeldudeks, kõneleja soovitud tegevuse väljendusteks, palveteks, soovideks, soovitusteks, ettepanekuteks, hoiatusteks ja kultuurinormi edastusteks. On võrreldud eri tüüpi direktiivide kasutamist ema ja isa kõnes ning vaadeldud direktiivide kasutamise dünaamikat laste kasvades ning suhtlusolukorriti. Selgub, et ehkki direktiivide hulk hoidjakeeles aja jooksul väheneb, kasutavad nii isa kui ka ema kõige sagedamini siiski otseseid käske.

Võtmesõnad: hoidjakeel, lapsele suunatud kõne, eesti keel

1. Sissejuhatus

Eesti lapsele suunatud kõne¹ uurimine on pikka aega keskendunud emade keelekasutusele, alles viimasel ajal on mõnevõrra uuritud ka isade kõnet (Rääbis 2012) ja võrreldud seda emade omaga (Kõrgesaar 2014b). Enamik seniseid hoidjakeele uurimusi keskendub vanemate kõnele väikelastega, kes on alles keeleomandamise algstaadiumis (Argus 2010, Argus, Kõrgesaar 2012, Argus, Kõrgesaar 2014, Kõrgesaar 2014a). Arengupsühholoogia valdkonnas on mõnevõrra uuritud ka lasteaia- või kooliealistele lastele suunatud kõnet (Tulviste 2001, Tulviste 2003, Tulviste jt 2003, Tulviste jt 2004), ent keeleteaduse vaatenurgast on niisuguseid uurimusi vaid üksikuid (Tamm jt 2014, Kõrgesaar 2014b).

Kui varem väideti hoidjakeeleuurimustes üsna kindlalt, et lastega räägivad kõige rohkem just emad ja nimelt emade hoidjakeel on lastele keele omandamise

¹ Laste- ja hoidjakeele terminite kohta vt Kõrgesaar, Kapanen 2015 siinses kogumikus.

juures kõige suuremaks toeks, siis tänapäevasemad uurimused kinnitavad, et ka isade hoidjakeelel on laste kõne arengus oma suur osa. Isade ja emade hoidjakeele keelekoguseid² uurides on leitud, et isad kõnelevad lapsega samasuguses mahus nagu emad, samuti on väidetud, et isad ärgitavad lapsi rohkem vestlema ning arendavad ja laiendavad nende sõnavara. (Snow 1977, Olson 1986, Matychuk 2005, Leech jt 2013, Pratt jt 1992)

Elizabeth Hill väidab, et üldiselt on varasemad uurimused kinnitanud, et emad ilmutavad lastega kõneldes suuremat tundlikkust ja vastavad lastele rohkem. Ta viitab ka Campbell Leaperi, Kristin J. Andersoni ja Paul Sandersi uurimusele (1998: 3), milles on leitud, et emad on lastega suheldes jutukamad, samal ajal kui isad kasutavad pigem direktiivsemaid vestlusstrateegiaid ja vastavalt rohkeid imperatiivivorme. Oma töös on Hill muuhulgas võrrelnud isa ja ema kasutatud keelekoguseid ning kasutatud käske ja keelde. Ühe kuueaastase lapse ja tema mõlema vanemaga peetud kahe dialoogi kõneandmete analüüsi põhjal leidis kinnitust oletus, et emalt lapsele suunatud keele kogus on märksa suurem kui isa oma, hõlmates ligi kaks korda rohkem lausungeid, mis on ka isa lausungitest pikemad. Küll aga ei leidnud kinnitust seisukoht, nagu oleks isa lapsega direktiivsem kui ema. (Hill 2009: 163–165, 168)

Kuivõrd siinkirjutaja varasemast uurimusest (Kõrgesaar 2014b) ilmnes, et isad ja emad kasutavad lastega kõneldes ühesuuruseid keelekoguseid ning ka kõne funktsionaalne jaotus on üsnagi sarnane (nii isade kui ka emade kõnes esineb laias laastus ühepalju küsimusi, väiteid ja käske), keskendub seekordne käsitlus kitsamalt direktiividele ja püüab välja selgitada, kuidas need isade ja emade kõnes jagunevad. Artikli teises osas antakse ülevaade direktiivi mõistest, kolmas osa kirjeldab uurimismeetodit ja andmestikku ning seejärel vaadeldakse direktiivide kasutussagedust suhtlusolukorriti, võrreldakse eri tüüpi direktiivide osakaalu isade ja emade kõnes, samuti direktiivide kasutamise dünaamikat eri vanuses lapsega kõneldes.

Vanema kõne direktiividel on kindel eesmärk eelkõige suunata ja juhtida lapse käitumist, kuid mitte ainult – direktiivide abil saab kaudselt kaasa aidata ka lapse sõnavara arengule (nt *Too emmele see suur punane pall sealt issi töölaua alt!*) ja neil on oluline roll lapse ühiskonda sotsialiseerimise protsessis. Eri tüüpi direktiivide kasutamine viitab suhtluse komplekssemaks muutumisele. Võib oletada, et väikese lapsega kasutatakse pigem otseseid käske ja lapse vanuse kasvades võib suureneda kaudsete direktiivide hulk, nt edastatakse käsk hoopis üldise kultuurinormina (*enne sööki pestakse käsi*) või hoiatusena (*kui sa seal kaua turnid, kukud alla*).

2. Direktiivi mõistest ja varasematest käsitlustest

Direktiivi mõiste üle on palju vaieldud. Alexandra Craven ja Jonathan Potter (2010: 420) defineerivad direktiivi kui käsku, öeldes, et direktiiv on “kõneakt, mille abil üks vestluse osapool käsib [*tells*] teisel midagi teha”. Samamoodi sõnastab direktiivi Hill (2009: 165), lisades sõnale ‘käsib’ ka sõna ‘nõuab’ (“the speaker is telling or ordering someone to do something”). Neile kahele definitsioonile vastanduvad Traci Curl ja Paul Drew, kelle järgi palve (ingl *request*) on “kõneakt, mille abil üks vestluse osapool palub [*asks*] teisel midagi teha” (Curl, Drew 2008: 12). Kõneaktiteooria

² Keelekoguse (ingl *amount of language*) all on siinses artiklis silmas peetud vestluse ühe osalise lausungite hulka ühe tunni pikkuses spontaanse kõne lindistuses.

looja John Searle käsitleb direktiivi ülemmõistena, mille alla koondab soovi, käsu, käskluse, küsimuse, anumise, palve jne. Direktiivi eesmärk on Searle'i (1979: 5) järgi panna kuulaja midagi tegema. Susan Ervin-Tripp (1976: 29) on pakkunud välja oma lahenduse, milles toob ära kuut tüüpi direktiivid. Esitan need järgnevalt koos eesti näidetega.

- Vajadust väljendavad deklaratiivid (*need statements*), nt *Mul on tikke vaja*.
- Otsesed käsud e imperatiivid (*imperatives*), nt *Anna mulle tikud!* või vaeglause *Tikud!*
- Varjatud imperatiivid (*embedded imperatives*), nt *Kas sa saaksid mulle tikke anda?* – juhud, kus tegevuse subjekt, objekt, tegevus ja sageli ka kasusaaja on väljendatud niisama selgesõnaliselt kui otseses käsus, kuid need on varjatud muude semantiliste ja süntaktiliste tunnuste taha.
- Lubavad/võimaldavad direktiivid (*permission directives*), nt *Kas ma tohiks sinu tikke paluda?* – suhtlusolukord, mille keeleline vormistus ei jäta vestluspartnerile muud võimalust kui nõutud tegevus teha.
- Küsimusekujulised direktiivid (*question directives*), nt *Tikke on?*, mis ei täpsusta soovitud tegevust.
- Vihjed (*hints*), nt *Kõik tikud on otsas*.

Nagu eeltoodust nähtub, iseloomustab kõiki nelja direktiivi definitsiooni kõneleja soov vestluspartnerit midagi tegema panna, erinevused ilmnevad pigem selles, mida eri autorid on tegemapanemise all silmas pidanud. Kui Craven ja Potter (2010) ning Curl ja Drew (2008) on välja toonud vaid ühe viisi või võimaluse (käsu ja palve) oma eesmärgi saavutamiseks, siis oluliselt varasemast ajast pärinevad Searle'i (1979) ja Ervin-Trippi (1976) käsitlused võimaldavad mitmekülgsemat lähenemist, kuivõrd hõlmavad kõiki kõneakte, millel on üks eesmärk (panna vestluse teine osapool midagi tegema), ehkki keelelised vormistusvahendid on erinevad (küsimuse vormis direktiivid, vihjed jne). Mainitud põhjusel näivad viimased lähenemised hoidjakeele direktiivide uurimiseks rohkem sobivat.

Direktiivide liigitamist märksa laiemalt kui vaid käskudeks-palveteks toetab ka värske lähenemine, mille on välja pakkunud Ursula Stephany, Ayhan Aksu-Koç ja Maria D. Voeikova (2013: 11). Direktiivid liigitatakse selles järgmiselt (lisatud on taas eesti näited):

- käsud (*commands*), nt *Tule siia!*
- keelud (*prohibitions*), nt *Ära puutu!*
- soovid (*speaker's wish*), nt *Ma tahan seda!*
- soovituselised (*suggestions*), nt *Sa pead ema sõna kuulama. Sa võiks nüüd koju minema hakata.*
- ettepanekud (*hortatives*), nt *Hakkame nüüd sööma!*
- hoiatused (*warnings*), nt *Kui sa auto maha pillad, siis see kaob ära.*
- kultuurinormi edastamine (*statements of social rules*), nt *Pärast tualetis käimist pestakse käsi! Väikeseid lapsi ei tohi kiusata!*
- kõneleja soovitud tegevuse väljendus (*statements of an addressee's desired action*), nt *Seda ei puutu!*³
- kõneleja soovitud tegevus või seisund, mis on seotud kolmanda osalisega (*statement of state of the third part*), nt *Las see olla seal!*

³ Erinevus keelust (*Ära puutu!*) seisneb vormistuses. Kõneleja soovitud tegevust on väljendatud deklaratiiviga, keeldu aga eitava imperatiivvormiga.

“Eesti keele grammatika” (EKG II: 174–176) järgi loetakse käsklauseks (silmas on peetud pigem kirjalikku teksti ja lauset, mitte niivõrd suulise kõne lausungit) lausetüüpi, “mille grammatiline ehitus on kohandatud ülesandele väljendada käsku (palvet, üleskutset, nõuet, kategoorilist käsku) lausega märgitud sündmuse teostamiseks”. Eristatakse ka negatiivset käsku ehk keeldu (*Ära muutu jämedaks!*) ja tuuakse välja viisid, kuidas käsku vormistada: imperatiivne verbivorm otsese käsu puhul (*Osta poest leiba!*) ja jussiiivne vahendatud käsu puhul (*Rohelised kogunegu plaanikomitee ette*). Samuti on mainitud võimalust esitada käsku küsimuse (*Kas sa tõused juba ükskord?*) või hüüdfraaslausena (*Stopp!*) ja keeldu üttelausena (*Ma lähen nüüd lõplikult minema. – Armas laps!*).

Helle Metslang (2004) on uurinud suulise kõne (argikõne) direktiivide väljendamise võimalusi ja esitanud direktiivide väljendamise keelelised vahendid. Metslang selgitab, et otseseid käsked moodustatakse enamasti imperatiivse (*Anna see nuga siia*), vahendatud käsked aga jussiiivse verbiga (*Pangu obune ette*). *me*-lõpulise verbivormi kasutamise kohta (*Tule Tallinna ja lähme Märdi juurde sauna!*) ütleb Metslang, et selle semantiline väli hõlmab käsku, soovitust ja kavandamist. Eelmiste kõrval on tähelepanu pööratud veel direktiivi moodustamise modaalsetele konstruktsioonidele (moodustatud modaalverbi abil, nt *Sa pead tõesti kodus kõvasti harjutama*) ja verbita käsklustele (*Mulle ka võid!*), samuti mittemodaalsetele küsimusekujulistele direktiividele (*Ise sa pärast kirjutad need sõnad ülesse, eks?*) ja konditsionaali abil moodustatud soovidele, palvetele ja kaudsetele käskudele.

Karl ja Renate Pajusalu töö on keskendunud tingiva kõneviisi uurimisele, mille käigus selgus, et tingiva kõneviisi abil väljendatakse sageli soove ja palveid (nt *Kas te võiksite mulle ulatada...? Nii väga tahaks...*), ühtlasi on sellele, kellele palve esitatakse, tingiva kõneviisi puhul justkui kergem palve täitmisest keelduda. Tuumverbide *peaks* ja *tahaks* jmt kasutamine aitab autorite sõnul aga oma soovi pehmendada (K. Pajusalu, R. Pajusalu 2004: 265, 268). Renate Pajusalu eestisoome-vene palvete võrdlusest selgub aga, et täiskasvanute omavahelises kõnes leidub otsese käsuna vormistatud palveid vaid väga vähesel hulgal (2 palvet/küsimust 200-st) (R. Pajusalu 2014: 251).

Tiia Tulviste on uurinud direktiive emade hoidjakeeles erinevates kodustes suhtlussituatsioonides (mh söögilauavestlused, puslede kokkupanek) ning võrrelnud tulemusi Soome, Rootsi ja Ameerika emade keelekasutusega (Tulviste 2001, Tulviste jt 2003, Tulviste jt 2004). Isa direktiive on uurinud Andriela Rääbis (2012). Rääbise analüüs põhineb ühel 42-minutilisel lindistusel ja vaatluse all on isa vestlus oma 5-aastase tütrega. Keskendutud on sellele, kuidas kõnes direktiivset vooru moodustatakse ja kuidas vestluspartner sellele reageerib; materjali ei ole võrreldud ema hoidjakeele andmetega.

3. Uurimismeetod ja andmestik

Uuring põhineb spontaanse kõne lindistustel, vaatluse alla on võetud ühe pere kaks last ja mõlemad vanemad. Valim ei ole küll suur, kuid juhtumipõhine pikiuuring võimaldab hästi jälgida täiskasvanu keelekasutust ja selle dünaamikat, teha tähelepanekuid, kui palju kasutab vanem direktiive noorema, kui palju vanema lapsega kõneldes ja mis tüüpi direktiividega on tegemist. Noorem last (laps 1)

on lindistatud vanuses 7, 8, 9 ja 10 aastat, vanemat last (laps 2) vanuses 10, 11, 13 ja 14 aastat, lindistatud on igal aastal üks kord ühe tunni jooksul. Iga-aastased lindistused on toimunud eraldi nii isa kui ka emaga, puudu on lindistus vanema lapse 12. eluaastast. Uurimismaterjal hõlmab 16 tunni jagu lindistusi, milles on kokku 10 950 lausungit. Lindistuste näol on tegemist argidialoogidega igapäevastes kodustes situatsioonides nagu koduste ülesannete tegemine, koos meisterdamine, koristamine, üheskoos toidu valmistamine, lauamängu mängimine jne.

Lindistused on transkribeeritud programmiga CLAN (MacWhinney 1991) ja on kättesaadavad rahvusvahelise lastekeele andmebaasi CHILDES eesti keele korpuse Kõrgesaare alamkorpuses.⁴ Materjali kirjapanekul on kasutatud kuuldeortograafiat (vt Hennoste 2000), kõneühikuna on aluseks võetud lausung ehk kõnevooru niisugune osa, mida piiritlevad pausid või vooruvahetus (vt täpsemalt Hennoste, Rääbis 2004: 28) ja mida suulise transkriptsiooni tavas defineeritakse selle järgi, et kahe lausungi vahele jääb vähemalt kolmesekundiline paus (vt nt Goldin-Meadow jt 2010).

Eesti viimase aja käsitlustes (R. Pajusalu 2014, Argus 2014) on enamasti lähtutud Searle'i kõneaktiteooriast (Searle 1979: 11); ka siinse artikli kontekstis on direktiiviks loetud igasugune kõneakt, mille eesmärk on panna kaasvestleja (laps) midagi tegema või mitte tegema, esindagu siis kõneakti vorm mis tahes intensiivsuseastet – käsku, keeldu, palvet, soovi, ettepanekut või muud. Transkribeeritud failidest on kogutud kõik direktiivsed lausungid ja liigitatud need tüübi järgi. Uurimismaterjali analüüsimise asudes selgus, et täiskasvanute kõnes esineb väga suurel hulgal eri tüüpi direktiive ja nende liigitamisel ainult ühe või teise jaotuse alusel jääks hulk direktiive kõrvale, mistõttu otsustas siinse artikli autor Searle'i, Ervin-Trippi ning Stephany jt jaotusi kombineerida. Stephany käsitlusest on pärit keelud, kõneleja soovitud tegevuse väljendused, kultuurinormi edastused, hoiatused ja ettepanekud, Stephany käsud on jagatud Ervin-Trippi eeskujul otsesteks ja varjatud käsklusteks, kusjuures viimaste hulka on arvatud ka vihjed. Ülejäänud liigid (palve ja soov) esinevad ühe või teise nimetuse all kõigi kolme autori käsitlustes. Niisugune jaotus sobis olemasoleva keelematerjali liigitamiseks kõige paremini, kuna võimaldas määrata kõik materjalis esinenud direktiivsed kõneaktid. Siinse materjali liigitusalus on järgmine (kirjeldamiseks on kasutatud konstrueeritud näiteid):

- otsene käsk (*Võtmed siia! Võtmed! Pane see ära! Anna siia! Palun anna mulle..., Ulata mulle...*);
- keeld (*Ära puutu! Ei tohi võtta! Ei karju!*);
- kõneleja soovitud tegevuse väljendus (*See jääb nüüd sinna! Las see olla seal!*);
- palve (*Kas ma saaksin...? Kas ma tohiksin...?*);
- soov (*Ma tahan/tahaksin seda.*);
- soovitus (*Sa peaksid nüüd selle sinna panema. Sa ei pea karjuma. Sa võiksid nüüd õppima hakata.*);
- ettepanek (*Aga paneme selle siia, eks? Läheme nüüd hoopis õue.*);
- hoiatus (*Kui sa selle maha pillad, siis see läheb katki. Tuline! Tass on serva peal!*);

⁴ <http://childes.psy.cmu.edu/browser/index.php?url=Other/Estonian/Korgesaar/> (23.2.2015).

- kaudne käsk / vihje (*Kui sa selle ära paneksid, siis... Kust ma saaksin...? See kukub sealt maha.*);
- kultuurinormi edastamine (*Sa pead ema sõna kuulama! See ei ole mänguasi. Nii ei tehta! Jalgu lauale ei panda!*).

Järgnevalt on võrreldud eri tüüpi direktiivide kasutussagedust isa ja ema kõnes (kogu andmestiku sagedused on toodud lisades 1 ja 2) ning vaadeldud, kuidas see kasutussagedus vanemate kõnes aja jooksul ja lapse vanemaks saades muutub.

4. Ülevaade vaatlusperioodi lindistuste suhtlusolukordadest ja neis enim kasutatud direktiividest

Vaatlustulemused näitasid, et peale ühe lindistuse (laps 1 vanuses 8 a, isaga) domineerisid kõigis ülejäänutes otsesed käsud (vt direktiivide jagunemist lähemalt osas 5).

- (1) nii paberit paberit paberit.
- (2) käi vetsus leia teksad ja mine ütle vennale ka et minek.
- (3) näpud eemale.

Eri tüüpi direktiivide kasutamine lapsega kõneldes sõltub suurel määral suhtlusolukorrast. Et anda lugejale ülevaade uurimuses kasutatud eri suhtlussituatsioonidest, on järgnevalt esitatud nii isa kui ka ema kõigi lindistuste vestlusteemad koos mõne ilmekama näitega enim kasutatud või huvitavamatest direktiividest. Suhtlusolukorra järel on sulgudes esitatud lindistuses kasutatud direktiivitüübid ja nende hulk.

7 a, isaga – mängivad lauamängu (otseseid käske 57, keelde 24, kõneleja soovitud tegevuse väljendusi 9, soovitusi 5, ettepanekuid 4, hoiatusi 1, kaudseid käske / vihjeid 13); nt kaudne käsk (4), (5); keeld (6)–(8).

- (4) kui me niimoodi venitama hakkame siis tal on palju raskem ma arvan.
- (5) aa mäletad see tuli panna nii et otsad kokku sa saad panna siia sa saad panna siia sa võid ükskõik kuhu panna siis kui sa keerad teda.
- (6) ära lollita.
- (7) ära pane neid järjest kõiki palun.
- (8) sina ära minu eest käi ma käin ise enda eest.

7 a, emaga – tehakse koduseid ülesandeid (otseseid käske 99, keelde 14, kõneleja soovitud tegevuse väljendusi 13, soove 2, soovitusi 4, ettepanekuid 14, hoiatusi 2, kaudseid käske / vihjeid 8, kultuurinormi edastusi 3); nt ettepanek (9), kaudne käsk / vihje (10), (11), soov (12).

- (9) ei sobi jah ja ma arvan et tead me kustutame terve selle hõhhide rea ära siit.
- (10) mid ma ei saa aru kui sa nii kiiresti räägid ma ei saa mitte midagi aru.
- (11) kui me sama tempoga jätkame et iga asja viis korda kustutame siis sa võibolla saad ommikuks valmis selle.
- (12) ma tahan et sa teeks õigesti mitte valesti.

8 a, isaga – täidavad lugemispäevikut (otseseid käske 5, keelde 1, kõneleja soovitud tegevuse väljendusi 1, ettepanekuid 1, kaudseid käske / vihjeid 6); nt kaudne käsk / vihje (13).

(13) see võib ära määrada.

8 a, emaga – ema teeb süüa, laps abistab teda (otseseid käske 60, keelde 13, kõneleja soovitud tegevuse väljendusi 16, soove 4, soovitusi 8, ettepanekuid 2, hoiatusi 2, kaudseid käske / vihjeid 10, kultuurinormi edastusi 1); nt kõneleja soovitud tegevuse väljendus (14).

(14) no las ta istub seal pausil mis taga juhtub.

9 a, isaga – valmistavad sünnipäevakaarti (otseseid käske 38, keelde 2, kõneleja soovitud tegevuse väljendusi 26, soovitusi 1, ettepanekuid 9, hoiatusi 1, kaudseid käske / vihjeid 3); nt kõneleja soovitud tegevuse väljendus (15), soovitus (16).

(15) no las ta olla siis mõttme midagi muud välja.

(16) noh sa võiksid akata tegema õisi.

9 a, emaga – ema toimetab kodus ja aitab lapsel paberist langevarjurit meisterdada (otseseid käske 35, kõneleja soovitud tegevuse väljendusi 6, soovitusi 7, ettepanekuid 2, kaudseid käske / vihjeid 3); nt soovitus (17).

(17) no aga sinna võiks ju vähemalt kui sa üleni teda ära värvida ei viitsi sa võiks vähemalt mingid mm mustrid peale joonistada.

10 a, isaga, laps 1 – sorteerivad enne kooli algust koolitarbeid (otseseid käske 37, keelde 3, kõneleja soovitud tegevuse väljendusi 9, palveid 1, soovitusi 1, ettepanekuid 13, hoiatusi 1, kaudseid käske / vihjeid 1); nt kõneleja soovitud tegevuse väljendus (18), ettepanek (19).

(18) need ei pea ka siin olema lähevad teise sahtlisse.

(19) teeme nimodi teeme sinna unniku asjadest mis ei ole kooliasjad.

10 a, emaga, laps 1 – hakivad talveks tilli sügavkülma ja meisterdavad käepaelu (otseseid käske 28, keelde 8, kõneleja soovitud tegevuse väljendusi 12, soove 2, soovitusi 3, ettepanekuid 4, hoiatusi 1, kaudseid käske / vihjeid 2); nt kaudne käsk / vihje (20), (21).

(20) sinna mahub ikka veel.

(21) näe nüüd saad käärid jälle ära pesta.

10 a, isaga, laps 2 – joonistavad lapse palvel suurele paberile rallirada (otseseid käske 8, keelde 1, kõneleja soovitud tegevuse väljendusi 2, palveid 1, soovitusi 1, ettepanekuid 7); nt ettepanek (22), (23).

(22) nii keerame teistpidi vao vot nii nüüd meil on palju suurem.

(23) kuule räägime mulle nüüd kuskohas mis meil on.

10 a, emaga, laps 2 – laps esmalt värvib, seejärel tehakse koos eesti keele koduseid ülesandeid (otseseid käske 75, keelde 5, kõneleja soovitud tegevuse väljendusi 4, soovitusi 10, ettepanekuid 6, hoiatusi 1, kultuurinormi edastusi 1); nt soovitus (24), hoiatus (25).

(24) sa pead poolitama vist.

(25) kui sa nii laialt kirjutad siis sul ei mahu peale err tähe midagi ära enam.

11 a, isaga – tehakse koos matemaatika koduseid ülesandeid (otseseid käske 44, keelde 5, kõneleja soovitud tegevuse väljendusi 8, soovitusi 4, ettepanekuid 9); nt ettepanek (26), kõneleja soovitud tegevuse väljendus (27).

(26) jätame selle kevadilma praält lähme edasi mis järgmine on?

(27) tähed lähevad järgmisse ruutu juba jah nii miinus kuuskend kaheksa kaheksakend kaks.

11 a, emaga – ema toimetab kodus, vestlevad argiteemadel (otseseid käske 16, keelde 6, kõneleja soovitud tegevuse väljendusi 2, soovitusi 3, ettepanekuid 2, kaudseid käske / vihjeid 2); nt kaudne käsk / vihje (28).

(28) ma ei kuule kui sa seal räägid.

13 a, isaga – valmistavad lapse palvel koos internetiajakirja (otseseid käske 11, kõneleja soovitud tegevuse väljendusi 2, ettepanekuid 9); nt ettepanek (29), kaudne käsk / vihje (30).

(29) vaatame kas siin on veel midagi.

(30) google peaks igal juhul ütlema sulle ju millal ta välja tuleb.

13 a, emaga – lobisevad ja joovad teed (otseseid käske 13, keelde 5, kõneleja soovitud tegevuse väljendusi 2, palveid 2, soovitusi 4, ettepanekuid 2, kaudseid käske / vihjeid 1, kultuurinormi edastusi 1); nt kõneleja soovitud tegevuse väljendus (31), palve (32).

(31) aa siis natuke las tõmbab veel.

(32) nii kuule kas ma saaksin kuskilt paberit ja pliiatsit?

14 a, isaga – varahommikune vestlus enne tööle- ja kooliminekut (otseseid käske 8, keelde 1, kõneleja soovitud tegevuse väljendusi 1, soovitusi 2, kaudseid käske / vihjeid 1); nt keeld (33), soovitus (34).

(33) tormad ära torma.

(34) sa pead kääd ära pesema pärast su käsi on must aga ennem ei ole mõtet kui sa viid selle rula tagasi sinna kuhu ta meil jääb praägult ootama.

14 a, emaga – ema koristab laste riidekappi, laps oma tuba (otseseid käske 40, keelde 5, kõneleja soovitud tegevuse väljendusi 1, soovitusi 10, ettepanekuid 1, kaudseid käske / vihjeid 4); nt kaudne käsk / vihje (35), otsene käsk (36).

(35) jah mis iganes nad on sellepärast et tegelikult tuleks nüüd võtta niiske lapp ja näed teha siit ümbert kõik puhtaks telekas ka ja see on nii tolmune.

(36) meie ei vaidle üldse sel teemal.

Ülaltoodu illustreerib ilmekalt tõsiasja, et otseste käskude hulk sõltub suurel määral suhtlussituatsioonist. Otseseid käske on märgatavalt vähem niisugustes situatsioonides, kus lapsel ei ole õppija või abistaja rolli (siinse materjali hulgas näiteks lindistused, kus lapse soovil täidetakse lugemispäevikut, joonistatakse rallirada, toimetatakse kodus, koostatakse internetiajakirja, juuakse teed või vesteldakse niisama päeva hakul), kõige direktiivirohkemad on aga olukorrad, kus

lapsega lahendatakse koduseid ülesandeid. Ootamatult direktiivirohke on ka isa ja lapse lauumängu-lindistus, milles isal esines rohkem direktiive kui üheski teises lindistuses.

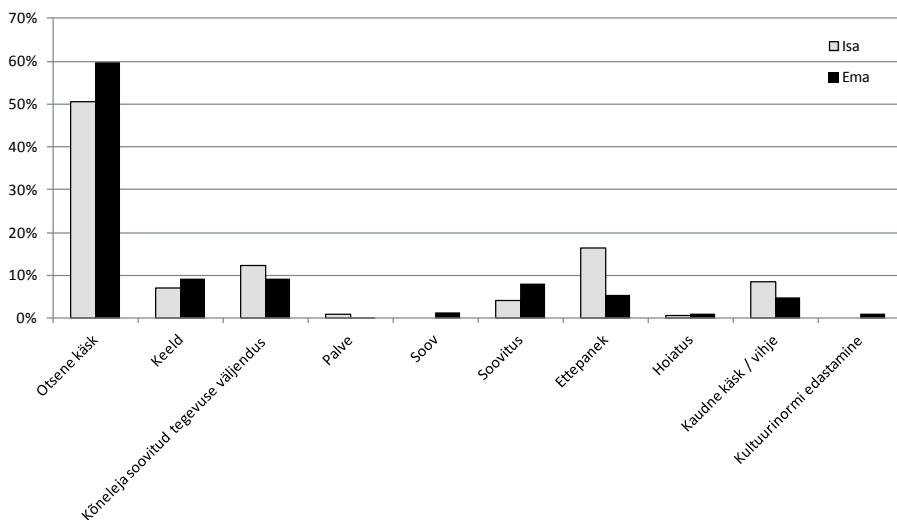
Tasub tähele panna, et nii mõnigi näitelause ei jäta ehk muljet direktiivsest kõneaktist põhjusel, et kirjalik tekst ei suuda edastada kõneleja hääletooni, nt (10), (21), (28), (30), (32), (35). Näite (35) puhul andis ema hääletoon selgelt mõista, et ta ootab just lapselt, et too võtaks nüüd kohe lapi ja puhastaks teleri ümbruse ära, mitte ei öelnud lihtsalt suusoojaks, et keegi peaks millalgi selle ära tegema.

Selleks, et täpselt aru saada, mida mingi lausungiga on silmas peetud, on oluline teada konteksti ja mõnel puhul ka kõnelejate tausta. Näidetes (10) ja (28) väljendab ema hääletoon samuti, et ta ootab, et laps kordaks öeldut selgemini või valjemini, ent taust lisab näites (10) teadmise, et laps on võrdlemisi kehva diktsiooniga ja tal on korduvalt palutud selgemini rääkida. Teisel puhul (28) saame varasema dialoogi põhjal aru, et laps läks ruumist ära ja räägib nüüd kusagil eemal. Ema vihje, et ta ei kuule, kui laps seal räägib, on selge märguanne, et laps tuleks ruumi tagasi.

5. Eri tüüpi direktiivid isa ja ema lapsele suunatud kõnes

5.1. Direktiivide jagunemine ning nende dünaamika isa ja ema kõnes

Kirjeldan esmalt eri tüüpi direktiivide jagunemist keelematerjalis kogu vaatlusperioodi jooksul üldiselt. Isa kõne 2300 lausungis esines kokku 398 ja ema kõne 3228 lausungis 612 direktiivset kõneakti, eri tüüpi direktiivide jaotusest annab ülevaate joonis 1. (Arvandmed on toodud lisades 1 ja 2 ning tabelis 1.)



Joonis 1. Eri tüüpi direktiivide jagunemine andmestikus

Nagu jooniselt näha, on mõlema vanema kõnes kõige sagedasemad otsesed käsud. Need domineerivad enamikus lindistustest, otseste käskude osakaal on isa kõnes keskmiselt 50 ja ema kõnes 60 protsenti ning jääb lindistuste lõikes isa kõnes 36 ja 63, ema kõnes aga 43 ja 74 protsendi vahele. Veel hakkab silma, et ema kõnes jagunevad kaudsed direktiivid märksa ühtlasemalt, kaetud on ka need direktiivitüübid, mida isa kõnes ei esinenud: oma soovi väljendab ema kokku kaheksal korral (1,3% juhtudest) ja kultuurinormi edastab ta direktiivsel moel kuuel korral (1% juhtudest).

Tulemused võimaldavad järeldada, et mõlemad vanemad on suhtluses lastega võrdlemisi direktiivsed, kasutades selleks, et last midagi tegema või mitte tegema panna, kõige enam (üle 50%) otsesteid käskude. Muude direktiivide jagunemise põhjal võib öelda, et ema kasutab samal määral keeldusid ja kõneleja soovitud tegevuse väljendamist (mõlemat 9%) ning annab lapsele soovitusi (8%). Isa väljendab seevastu võrdlemisi palju (12%) soovi lapse tegevuse kohta (nt *Need lähevad ruuduliste vihikute hunnikusse*), kuid peaaesjalikult (16%) esitab ta pigem ettepanekuid edasise tegevuse kohta (*Aga teeme nii, et...*), samas kui kogu vaatlusperioodi jooksul ei esine isa kõnes ainsatki soovi väljendust. Samuti on huvitav fakt, et isa ei edasta lapsele direktiive kultuurinormi kaudu. Palveid ja hoiatusi esineb mõlema vanema kõnes võrdselt vähe, alla 1%.

Kui vaadelda muutusi, mis ema ja isa direktiivikasutuses aja jooksul aset leiavad (tabel 1; absoluutarvud on toodud lisades 1 ja 2), tuleb märkida, et otseste käskude hulk mõlema vanema kõnes läbib küll mõningaid tõuse ja langusi, ent siiski ei saa väita, et nende osakaal kõigi direktiivitüüpide hulgas aja jooksul selgelt kasvaks või kahaneks. Otseseid keeldusid kasutab isa kõige rohkem (21%) just väiksema lapsega vesteldes, ema kõnes nende hulk varieerub. Kõneleja soovitud tegevust väljendavate direktiivide hulk isa kõnes on vaatlusperioodi algul sama suur kui ema kõnes – mõlemal 8% – ja see kasvab mõlema vanema lindistustes, saavutades isa kõnes maksimumi (33%) 9-aastase ja ema kõnes (20%) 10-aastase lapsega kõneldes. Pärast seda hakkab näitaja mõlema vanema kõnes langema ning jõuab isa kõnes 14-aastase lapsega vestlemisel tagasi alguspunkti (8%), ema kõnes aga koguni 1,6%-ni. Soovitude hulk ema kõnes läbib vaatlusperioodi kestel üsnagi kindla tõusu (vaatlusperioodi algul 2,5%, lõpus 16%), kuid isa kõnes leiab viimase lindistusega aset soovitude hulga järsk kasv (15%), pärast seda kui see on varem kõikunud 1% ja 6% vahel. Ettepanekud on isa kõnes otseste käskude järel levinuimad direktiivid, mille hulk ajaga hüppeliselt kasvab, jõudes 13-aastase lapsega suhtlemisel koguni 41%-ni, mistõttu on ootamatu, et viimases lindistuses ei esine ilmselt vestlussituatsioonist tingituna ühtki ettepaneku vormis direktiivi. Ema kõne ettepanekute hulk püsib aga vaatlusperioodi vältel võrdlemisi stabiilne, tehes läbi vaid paar langust. Kõige rohkem ettepanekuid (9%) kasutas ema 7-aastase lapsega vesteldes.

Veel hakkab silma, et hoiatusi andsid mõlemad vanemad direktiivsel moel edasi pigem nooremale lapsele, pärast 10. eluaastat neid kummagi vanema kõnes enam ei esinenud. Ka kaudsete käskude ja vihjete hulk oli ema kõnes võrdlemisi stabiilne (3–8%), kuid isa kõnes oli see 7-aastase lapsega kõneldes 11% ja 8-aastasega kõneldes koguni 42%. Pärast niisugust järsku tõusu aga nende hulk vähenes niisama järsult ja püsis kuni vaatlusperioodi lõpuni samuti 4% ja 8% vahel, puududes mõnes lindistuses sootuks.

Tabel 1. Eri tüüpi direktiivide osakaal protsentides isa/ema kõnes

Laps ja vanus	Otsene käsk	Keeld	Kõneleja soovitud tegevuse väljendus	Palve	Soov	Soovitus	Ettepanek	Hoia-tus	Kaudne käsk / vihje	Kultuuri-normi edastamine
Laps 1, 7-aastane	50,4 / 62,3	21,2 / 8,8	8,0 / 8,2	-	- / 1,3	4,4 / 2,5	3,5 / 8,8	0,9 / 1,3	11,5 / 5,0	- / 1,9
Laps 1, 8-aastane	35,7 / 51,7	7,1 / 11,2	7,1 / 13,8	-	- / 3,4	- / 6,9	7,1 / 1,7	- / 1,7	42,9 / 8,6	- / 0,9
Laps 1, 9-aastane	47,5 / 66,0	2,5 / -	32,5 / 11,3	-	-	1,3 / 13,2	11,3 / 3,8	1,3 / -	3,8 / 5,7	-
Laps 1, 10-aastane	56,1 / 46,7	4,5 / 13,3	13,6 / 20,0	1,5 / -	- / 3,3	1,5 / 5,0	19,7 / 6,7	1,5 / 1,7	1,5 / 3,3	-
Laps 2, 10-aastane	40,0 / 73,5	5,0 / 4,9	10,0 / 3,9	5,0 / -	-	5,0 / 9,8	35,0 / 5,9	- / 1,0	-	- / 1,0
Laps 2, 11-aastane	62,9 / 51,6	7,1 / 19,4	11,4 / 6,5	-	-	5,7 / 9,7	12,9 / 6,5	-	- / 6,5	-
Laps 2, 13-aastane	50,0 / 43,3	- / 16,7	9,1 / 6,7	- / 6,7	-	- / 13,3	40,9 / 6,7	-	- / 3,3	- / 3,3
Laps 2, 14-aastane	61,5 / 65,6	7,7 / 8,2	7,7 / 1,6	-	-	15,4 / 16,4	- / 1,6	-	7,7 / 6,6	-
Keskmine	50,5 / 59,8	6,9 / 9,2	12,4 / 9,2	0,8 / 0,3	- / 1,3	4,2 / 8,0	16,3 / 5,4	0,5 / 1	8,4 / 4,9	- / 1

5.2. Muutused direktiivide kasutussageduses vaatlusperioodi jooksul

Tabelis 2 on esitatud nii isa kui ka ema lausungite koguhulk igas linnistuses, direktiivide arv ja nende osakaal protsentides.

Tabel 2. Direktiivide arv ja osakaal lapsele suunatud kõne lausungites

Laps ja vanus	Isa lausungeid kokku	Isa direktiivide koguhulk	Isa direktiivide osakaal	Emal lausungeid kokku	Emal direktiivide koguhulk	Emal direktiivide osakaal
Laps 1, 7-aastane	574	113	20%	387	159	41%
Laps 1, 8-aastane	217	14	7%	398	116	29%
Laps 1, 9-aastane	238	80	34%	411	53	13%
Laps 1, 10-aastane	346	66	19%	382	60	16%
Laps 2, 10-aastane	156	20	13%	487	102	21%
Laps 2, 11-aastane	356	70	20%	401	31	8%
Laps 2, 13-aastane	163	22	13%	391	30	8%
Laps 2, 14-aastane	250	13	6%	371	61	16%
Kokku	2300	398	17%	3228	612	19%

Materjalis ei ilmne direktiivide osakaalu järjekindlat vähenemist lapse vanemaks saades. Isa kõnes näitab direktiivide kasutussagedus märgatavat langust kahel viimasel aastal, ema kõnes toimub samasugune langus lapse vanuses 11 ja 13 eluaastat, kuid ema viimase lindistuse direktiivide hulga kasv kahekordseks annab põhjust oletada, et muutused on tingitud pigem vestlussituatsioonist. Ehkki kõik lindistused on tehtud tavapäraistes kodustes situatsioonides, tuleb siiski eristada suhtlusolukordi, kus täiskasvanud on direktiivsemad. Nimetatud olukorrad ei hõlma ainult ühist koduste ülesannete tegemist (mis ilmselt kaldub küll olema kõige direktiivirohkem), vaid ka muid koostööolukordi nagu toa koristamine, koolitarvete sorteerimine, sünnipäevakaardi meisterdamine ja elektroonilise ajakirja koostamine, tilli lõikamine, vormeliraja joonistamine jne, kus täiskasvanul on täita õpetaja ja juhendaja roll. Üsnagi direktiivirohke oli aga ka lindistus, kus isa ja laps mängisid koos lauamängu (*Sinu kord! Tõmba kaarti ka! Kää nüüd! Veereta!* jne). Vähem direktiive esines seevastu lindistustes, kus sihipärane tegevus puudus. Näiteks isa viimases, 14-aastase lapsega salvestatud lindistuses puudusid üllatuslikult isale varem nii omased olnud ettepanekud, seevastu kasutas ta seni harvemini esinenud soovitusi. Täiskasvanu ja laps viibisid selle lindistuse ajal lihtsalt hommikul enne tööle- ja kooliminekut samas toas, mõlemad olid parasjagu unised ja kordamööda püüti vestlust ülal hoida. Igapäevaelus on niisugused vestlused pigem haruldased.

Kui siinkirjutaja viimast uurimust isade ja emade lastele suunatud kõne koguhulga ja eri tüüpi lausungite kasutamise kohta kinnitas, et isad ja emad kõnelevad lastega üldjoontes ühepalju ja ka eri lausungitüüpe (mh direktiive) kasutatakse peaaegu võrdselt (Kõrgesaar 2014b), siis siinsed tulemused on mõnevõrra vastuolulised, näidates, et kõigi lindistuste peale kokku esines isa kõnes 2300 ja ema kõnes 3228 lausungit. Sama üllatav oli tulemus direktiivide kasutussageduse kohta, millest selgus, et isa lausungite hulgas oli direktiive 398, ema kõnes aga esines enam kui 200 direktiivi rohkem – 612. Samas kinnitab direktiivide osakaal kõigist lausungitest siiski varasemaid tulemusi, jäädes isa kõnes 17% ja ema kõnes 19% juurde (vt Kõrgesaar 2014, kus kahe eri isa ja ema lapsele suunatud kõne direktiivide osakaal kõigist lausungitest jäi 10% ja 15% piiresse).

6. Kokkuvõtteks

Uurimuse ülesandeks oli selgitada, kuidas jagunevad eri tüüpi direktiivid isa ja ema kõnes ning kuidas muutub direktiivikasutus üldiselt sedamööda, mida vanema lapsega täiskasvanu kõneleb. Ehkki analüüsitulemused kinnitavad, et direktiivide hulk täiskasvanu kõnes aja jooksul üldiselt väheneb, tuleb siiski märkida, et otsesed käsud on 9–14-aastase lapsega suheldes nii isa kui ka ema kõnes levinuimad direktiivid, mis moodustavad enam kui poole vestluses kasutatavatest direktiividest.

Direktiivikasutuse suhtlusolukorrapõhine vaatlus näitab, et kõige enam otseseid käskide esines neis suhtlussituatsioonides, kus lapsel oli õppija või abistaja roll, täiskasvanul aga õpetaja-instrueerija roll. Vähem otseseid käskide leidis olukordades, kus tehti küll midagi koos (meisterdamine, joonistamine), kuid lapse palvel. Isa kõnes esines arvuliselt kõige rohkem otseseid käskide vaatlusperioodi esimeses lindistuses, kus mängiti 7-aastase lapsega lauamängu, ja osakaaluna kõigist direktiividest lindistuses, kus valmistati 9-aastase lapsega sünnipäevakaarti. Ema kõne puhul

langesid nii arvuline ülekaal kui ka osakaal kokku samuti esimeses lindistuses, kus lahendati 7-aastase lapsega koduseid ülesandeid.

Tulemused kinnitavad autori varasemat väidet, et emad ja isad kasutavad direktiive üsnagi samal määral, mistõttu ei saa väita, et ema kõne oleks selgelt direktiivsem kui isa oma või vastupidi. Kindlasti võib aga väita, et ema kõnes jagunevad eri direktiivituübid ühtlasemalt ja erinevalt isast kasutab ema direktiive vähesel määral ka selleks, et anda lapsele edasi üldisi kultuuri- ja käitumisnorme. Kuivõrd siinse uurimuse näol oli tegemist juhtumianalüüsiga, mis põhines vaid ühe pere vanemate keelekasutusel, oleks üldistuste tegemiseks otstarbekas viia edaspidi läbi järgmine, laia valimi põhjal tehtud uurimus, milleks praegu CHILDES-i andmebaasi eesti lastekeele korpusel piisav võrdlusmaterjal puudub. Samuti tasuks järgmistes hoidjakeeleuurimustes pöörata rohkem tähelepanu isade hoidjakeelele, et selgitada välja, kas üldse ja kui, siis mille poolest erineb isade hoidjakeel emade omast.

Viidatud kirjandus

- Argus, Reili 2010. Mida teeb *tegema*-verb hoidjakeeles. [Constructions with the verb *tegema* 'do, make' in child directed speech.] – Eesti ja soome-ugri keeleteaduse ajakiri ESUKA / Journal of Estonian and Finno-Ugric Linguistics JEFUL, 1 (2), 17–34.
- Argus, Reili 2014. Deontilise modaalsuse omandamine. Palved. Esimene sissevaade. – Ettekanne TLÜ EKKI teadusseminaris 6. mail 2014.
- Argus, Reili; Kõrgesaar, Helen 2012. *Hakkama*-verb lapsekeeles: kas aspektilisuse või tuleviku tähistaja? [“Hakkama”-verb in child language: A marker of aspectuality or future?] – Emakeele Seltsi aastaraamat, 57 (2011), 9–24. <http://dx.doi.org/10.3176/esa57.01>
- Argus, Reili; Kõrgesaar, Helen 2014. Sõnaliigid eesti lapse kõnes ja lapsele suunatud kõnes. [Word classes in the child's speech and in the child-directed speech.] – Eesti Rakenduslingvistika Ühingu aastaraamat, 10, 39–53. <http://dx.doi.org/10.5128/ERYa10.03>
- 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), 129–53. <http://dx.doi.org/10.1080/08351810802028613>
- EKG II = Ereht, Mati; Kasik, Reet; Metslang, Helle; Rajandi, Henno; Ross, Kristiina; Saari, Henn; Tael, Kaja; Vare Silvi 1993. Eesti keele grammatika II. Süntaks. [Estonian Grammar II. Syntax.] Tallinn: Eesti TA Keele ja Kirjanduse Instituut.
- Ervin-Tripp, Susan 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>
- Goldin-Meadow, Susan; Hedges, Larry; Huttenlocher, Janellen; Levine, Susan; Raudenbush, Stephen; Small, Steven 2010. Language Development Project 2002–2012. University Chicago. <http://joyrex.spc.uchicago.edu/ldp/docs/index.html> (24.9.2014).
- Hennoste, Tiit 2000. Suulise eesti keele uurimine: transkriptsioon, taust ja korpus. [Research into spoken Estonian: Transcription, background, corpus.] – Keel ja Kirjandus, 2, 91–106.
- Hennoste, Tiit; Rääbis, Andriela 2004. Dialoogiaktid eesti infodialoogides: tüpologia ja analüüs. [Dialogue Acts in Estonian Information Dialogues: Typology and Analysis.] Tartu: Tartu Ülikooli kirjastus. <http://hdl.handle.net/10062/18995>
- Hill, Elizabeth 2009. Do mothers and fathers differ in their speech styles when speaking to their children? Are there speaker gender effects on parents' language behaviour? – Innervate: Leading Undergraduate Work in English Studies, 2, 163–182.

- Kõrgesaar, Helen 2014a. Küsimused eesti lapsele suunatud kõnes leedu ja vene keele taustal. [Interrogatives in Estonian child-directed speech.] – Eesti Rakenduslingvistika Ühingu aastaraamat, 10, 193–207. <http://dx.doi.org/10.5128/ERYa10.12>
- Kõrgesaar, Helen 2014b. Eesti isade-emade hoidjakeel: kes küsib, kes käsutab, kes räägib rohkem? [Estonian fathers' and mothers' child-directed speech: Who asks more questions, who gives orders, and who speaks more?] – Krista Kerge (Toim.). Uurimusi keele omandamisest, õppimisest ja korpustest. Tallinna Ülikooli eesti keele ja kultuuri instituudi toimetised 16. Tallinn, 59–82.
- Kõrgesaar, Helen; Kapanen, Airi 2015. Kui lapsega ei räägi üksnes ema: valik termineid eesti laste- ja hoidjakeele kohta. – Eesti Rakenduslingvistika Ühingu aastaraamat, 11, 177–188. <http://dx.doi.org/10.5128/ERYa11.11>
- Leaper, Campbell; Anderson, Kristin J.; Sanders, Paul 1998. Moderators of gender effects on parents' talk to their children: A meta-analysis. – *Developmental Psychology*, 34 (1), 3–27. <http://dx.doi.org/10.1037/0012-1649.34.1.3>
- Leech, Kathryn A.; Salo, Virginia C.; Rowe, Meredith L.; Cabrera, Natasha J. 2013. Father input and child vocabulary development: The importance of wh-questions and clarification requests. – *Seminars in Speech and Language*, 34 (4), 249–259. <http://dx.doi.org/10.1055/s-0033-1353445>
- MacWhinney, Brian 1991. The CHILDES project: Tools for Analyzing Talk. Hillsdale, NJ: Lawrence Erlbaum Associates Inc.
- Matychuk, Paul 2005. The role of child-directed speech in language acquisition: A case study. – *Language Sciences*, 27 (3), 301–397. <http://dx.doi.org/10.1016/j.langsci.2004.04.004>
- Metslang, Helle 2004. Imperative and related matters in everyday Estonian. – *Linguistica Uralica*, 40 (4), 243–256.
- Olson, Sheryl L. 1986. Mother-child interaction and children's speech progress: A longitudinal study of the first two years. – *Merrill-Palmer Quarterly*, 32 (1), 1–20.
- Pajusalu, Renate 2014. Palved eesti, soome ja vene keeles: Grammatika pragmaatika teenistuses. [Requests in Estonian, Finnish and Russian: Grammar and pragmatics.] – Eesti Rakenduslingvistika Ühingu aastaraamat, 10, 241–257. <http://dx.doi.org/10.5128/ERYa10.15>
- Pajusalu, Karl; Pajusalu, Renate 2004. The conditional in everyday Estonian: It's form and functions. – *Linguistica Uralica*, 40 (4), 257–269.
- Pratt, Michael W.; Kerig, Patricia K.; Cowan, Philip A.; Pape Cowan, Carolyn 1992. Family worlds: Couple satisfaction, parenting style, and mothers' and fathers' speech to young children. – *Merrill-Palmer Quarterly*, 38 (2), 245–262.
- Rääbis, Andriela 2012. Direktiivisekvensid isa ja tütre suhtluses: juhtumianalüüs. [Directive sequences in a conversation of father and daughter: A case study.] – Eesti Rakenduslingvistika Ühingu aastaraamat, 8, 213–230. <http://dx.doi.org/10.5128/ERYa8.14>
- Searle, John 1976. A classification of illocutionary acts. – *Language Society*, 5 (1), 1–23. <http://dx.doi.org/10.1017/S0047404500006837>
- Searle, John 1979. *Expression and Meaning: Studies in the Theory of Speech Acts*. Cambridge: Cambridge University Press. <http://dx.doi.org/10.1017/CBO9780511609213>
- Snow, Catherine E. 1977. The development of conversation between mothers and babies. – *Journal of Child Language*, 4 (1), 1–22. <http://dx.doi.org/10.1017/S0305000900000453>
- Stephany, Ursula; Aksu-Koç, Ayhan; Voeikova, Maria D. 2013. The category of modality: A crosslinguistic study of its development. – *Ettekanne rahvusvahelisel pre- ja protomorfoloogia projekti aastakonverentsil Viinis*, 21.–23. veebruaril 2013.
- Tamm, Anne; Argus, Reili; Suurmäe, Kadri 2014. Evidentsiaalid: kas uskuda või mitte? Teabe vormi mõju eesti koolieelikute käitumisele. [Evidential sentences guiding Estonian preschoolers' exploration of novel objects and their properties.] – *Emakeele Seltsi aastaraamat*, 59 (2013), 244–262. <http://dx.doi.org/10.3176/esa59.13>

- 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>
- Tulviste, Tiia 2003. Mothers' conversational styles across cultures. The case of Estonia, Finland, Sweden and the U.S. – *Proceedings of the IACCP XVIth Congress*, Yogyakarta, Indonesia, 15.–19. juuli 2002.
- Tulviste Tiia; Mizera, Luule; De Geer, Boel; Tryggvason, Marja-Terttu 2003. A comparison of Estonian, Swedish, and Finnish mothers' controlling attitudes and behaviour. – *International Journal of Psychology*, 38 (1), 46–53. <http://dx.doi.org/10.1080/00207590244000278>
- 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>

Võrgumaterjalid

CHILDES, Kõrgesaare alamkorpus. <http://childes.psy.cmu.edu/browser/index.php?url=Other/Estonian/Korgesaar/> (5.2.2015).

Helen Kõrgesaar (Tallinna Ülikool) uurib eesti hoidjakeelt, selle dünaamikat ja mõju lapse kõne arengule.

Narva mnt 25, 10120 Tallinn, Estonia

helen.korgesaar@gmail.com

Lisa 1. Isa kõne eri tüüpi direktiivid

Laps ja vanus	Otsene käsk	Keeld	Kõneleja soovitud tegevuse väljendus	Palve	Soov	Soovitus	Ettepanek	Hoiatus	Kaudne käsk / vihje	Kultuurinormi edastamine	Kokku
Laps 1, 7-aastane	57	24	9	0	0	5	4	1	13	0	113
Laps 1, 8-aastane	5	1	1	0	0	0	1	0	6	0	14
Laps 1, 9-aastane	38	2	26	0	0	1	9	1	3	0	80
Laps 1, 10-aastane	37	3	9	1	0	1	13	1	1	0	66
Laps 2, 10-aastane	8	1	2	1	0	1	7	0	0	0	20
Laps 2, 11-aastane	44	5	8	0	0	4	9	0	0	0	70
Laps 2, 13-aastane	11	0	2	0	0	0	9	0	0	0	22
Laps 2, 14-aastane	8	1	1	0	0	2	0	0	1	0	13
Kokku	208	37	58	2	0	14	52	3	24	0	398

Lisa 2. Ema kõne eri tüüpi direktiivid

Laps ja vanus	Otsene käsk	Keeld	Kõneleja soovitud tegevuse väljendus	Palve	Soov	Soovitus	Ettepanek	Hoiatus	Kaudne käsk / vihje	Kultuurinormi edastamine	Kokku
Laps 1, 7-aastane	99	14	13	0	2	4	14	2	8	3	159
Laps 1, 8-aastane	60	13	16	0	4	8	2	2	10	1	116
Laps 1, 9-aastane	35	0	6	0	0	7	2	0	3	0	53
Laps 1, 10-aastane	28	8	12	0	2	3	4	1	2	0	60
Laps 2, 10-aastane	75	5	4	0	0	10	6	1	0	1	102
Laps 2, 11-aastane	16	6	2	0	0	3	2	0	2	0	31
Laps 2, 13-aastane	13	5	2	2	0	4	2	0	1	1	30
Laps 2, 14-aastane	40	5	1	0	0	10	1	0	4	0	61
Kokku	366	56	56	2	8	49	33	6	30	6	612

DIRECTIVE SPEECH ACTS AND THE DYNAMICS OF THEIR USAGE IN FATHER'S AND MOTHER'S SPEECH: A CASE STUDY

Helen Kõrgesaar

Tallinn University

The aim of the article is to give an overview of directive speech acts in Estonian child-directed speech. Two school-age children from the same family and their dialogues with both parents individually are under observation. A longitudinal study has been used as the method (one child has been regularly recorded from ages 7 to 10, the other from ages 10 to 14), including 16 hours of audio material. All directive speech acts have been singled out from the parents' speech and classified into imperatives, hints, prohibitions, statements of an addressee's desired actions, permissions, speaker's wishes, suggestions, hortatives, warnings and statements of social rules. Based on the analysis, comparisons are made between the mother's and father's speech with regard to the usage of different types of directives, and the dynamics of using directives when the child is growing as well as in situations of communication have been examined.

Although the results of the analysis confirm that the amount of directives in adults' speech decreases over time, it should be pointed out that imperatives are the most common directives in both the father's and mother's speech when communicating with a child aged between 7 and 14, comprising more than half of the directives used in conversation. The results confirm the author's previous assertion that mothers and fathers use comparable amounts of directives, ruling out the claim that the mother's speech has clearly more directives than that of the father or vice versa. However, it can surely be said that different types of directives are divided more equally in the mother's speech and unlike the father, the mother uses directives (although in a very small amount) in order to teach the child cultural norms and social rules.

Keywords: caregivers' speech, child-directed speech, Estonian

KUI LAPSEGA EI RÄÄGI ÜKSNES EMA: VALIK TERMINEID EESTI LASTE- JA HOIDJAKEELE KOHTA

Helen Kõrgesaar, Airi Kapanen

Ülevaade. Paralleelselt lastekeele ja selle omandamise uurimisega on Eestis viimastel kümnenditel uuritud lastekeelega tihedalt seotud hoidjakeelt. Erinevalt eesti keelest on neid kaht keelevarianti võõrkeelses teaduskirjanduses nimetatud mitme eri terminiga, eesti keeles on piiratud üksnes *lastekeele*, *hoidjakeele* ja *lapsele suunatud kõnega*, mis aga ei anna laste- ja hoidjakeeleuurijatele piisavat hulka vajalikke ja täpseid töövahendeid. Terminite ja nende mõistesisu paljude tõlgenduste tõttu on praeguseks kujunenud vajadus määrata kindlaks lapsele suunatud kõne eri kasutusjuhud, kuna eri vanuses ja eri lähedusastmes olevate lastega räägitakse isemoodi.

Et vältida segadust, kus autori eelistustest lähtuvalt on ühe terminiga tähistatud mitut mõistesisu, on siinse artikli eesmärk ühelt poolt avada võõrkeelses teaduskirjanduses käibel olevate terminite mõistesisu ning teiselt poolt neid täpsustada ja anda soovitusi, kuidas võiks neid termineid edaspidi eesti laste- ja hoidjakeele uurimisel kasutada.*

Võtmesõnad: esimese keele omandamine, lapsele suunatud kõne, isade kõne, emade kõne, terminivara, eesti keel

1. Sissejuhatus

Kas lastega kõneldes kasutatakse üksnes mugandatud keelevarianti? Kas igas vanuses lastega kõneldakse alati ühtmoodi? Kas kõik täiskasvanud räägivad lastega ühtmoodi? Kas kaheaastasele lapsele ja täiskasvanud lapsele suunatud kõnet võiks tähistada ühe ja sama terminiga? Kes on täiskasvanud laps? Kuhu tõmmata piir, mis vanusest alates laps ei ole enam laps? Nii laste- kui ka hoidjakeele uurimisse süvenedes tekib uurijal rohkelt küsimusi, mis annab tunnistust erialaselt tunnetatud vajadusest erinevate, sisult täpsete terminite järele. Siit ka küsimus – kas selleks, et kirjeldada, kuidas laps keelt omandab ja täiskasvanu seda toetab, piisab ainuüksi

* Autorid tänavad retsensente hoolika süvenemise ja asjakohaste märkuste eest.

termineist *lastekeel* ja *hoidjakeel*, kui vaadeldavad lapsed võivad olla mis tahes vanuses (eeldusel, et ka täiskasvanud laps on kellegi laps), suhtluskontekst võib olla isesugune, lapsega vestlejaks võib peale ema olla ka isa, tädi-onu, õde-vend, vanavanem, lapsehoidja, peretuttav?

Laste endi kõne kohta kasutatav termin *lastekeel* keeleuurijatele probleeme ei valmista – lastekeel on laste endi produtseeritav keel, mis muutub vastavalt lapse vanusele.¹ Inglise keeles on lastekeele kohta paralleelselt kasutusel terminid *child language* ja *first language* (õdede-vendade omavahelise vestluse puhul ka *peer talk*), mille mõistesisu on üsna üksühene. Lastekeele uurimise juures pole seni olnud märgata vajadust lisada lastekeele mõistele kõrval- või alammõisteid, kuivõrd (erinevalt hoidjakeelest) seda keelevarianti kõneleb ainult laps ise, olenemata sellest, et see muutub ja täiustub aja jooksul, läbides eri etappe ja sarnanedes lõpuks üha enam täiskasvanute kõnega. Lastekeele uurimisel on lapse bioloogiline vanus kõige täpsem eristaja.

2. Lastekeele kohta varem kasutatud termineid

Terminid *lastekeel* kasutatakse enamasti keelevariandi kohta, mis pole veel jõudnud sihtkeele ehk täiskasvanupärase keele kõiki tunnuseid omandada. Lastekeele mõistesisu ei hõlma üksnes lihtsustatud leksikat, vaid kindlasti kuuluvad lastekeele alla ka žestid, vokaliseerimised, samuti eriomane loov ja väljendusrikas keelekasutus, mille abil laps täiskasvanu tähelepanu püüab ja kontakti hoiab, ning laste kuuldavaks tehtud sisekõne (ingl *private speech*).

Üks esimesi, kes Eestis pani tähele keelekasutuse erinevust sõltuvalt kõneleja vanusest, oli Julius Mägiste² (1924). Kuigi ta on märganud, et on olemas lastekeel ja ammede keel, teeb ta seda muu jutu seas ega erista kaht keelevarianti termini tasandil, liiatigi kasutab ta saksa keele termineid *Kindersprache* ja *Ammensprache* üksnes lastekeele tähistamiseks, kuigi tänapäeva keeleteaduses tähistab esimene laste- ja teine hoidjakeelt. Mägiste eristab eaklasside alusel *lastekeele*, *noorema-põlve keele* ja *vanema-põlve keele* (viimase all on silmas peetud täiskasvanukeelt) ning jätab targu tõmbamata täpse vanusepiiri, millal üleminek ühelt keelevariandilt teisele aset leiab. Pöörates tähelepanu lapse keele üleminekule noorema-põlve keeleks ütleb Mägiste, et “kõne õppimisel esimesi samme tegeva lapse ja puberteedile ligineva lapse vahel on määratu vahe, mis peegeldub kõige teravamalt nende keeles kui väljendusvahendis” (Mägiste 1924: 3). Keelte eespool mainitud jagunemist eaklasside alusel silmas pidades leiab Mägiste, et kõige iseseisvam neist on lastekeel (sks *Kindersprache*, *Ammensprache*). Siit nähtub, et Mägiste paigutab lastekeele ja hoidjakeele ühe nimetuse alla (*Kindersprache* ‘lastekeel’, *Ammensprache* ‘hoidjakeel, ammekeel’) (Mägiste 1924: 2).

Mägiste kasutab terminiga *lastekeel* paralleelselt ka terminid *lastemurre*, viidates teiste hulgas ka Saksamaa kasvatusteadlase Berthold Otto artiklile ajakirjas *Kindesmundart*, kus viimane on püüdnud lastekeelt pedagoogikas tähtsale kohale seada ja on öelnud: “Ligemas tulevikus ei tohi pedagoogi aunime peale

¹ Artikli autorid on kursis nii terminite kui ka definitsioonidega, mida on kasutatud õigekeelsussõnaraamatus, eesti keele seletavas sõnaraamatus ning muudes üld- ja oskuskeele sõnastikes. Kuigi osas neis allikais on öeldud, et *lastekeel* ja *hoidjakeel* on sünonüümid, on see siiski eksitav, kuivõrd tegu on kahe eri mõistesisuga, mida vastavalt tähistavaidki kaks eri terminid.

² Julius Mägiste kutsus Emakeele Seltsi ajakirjas Eesti Keel avaldatud artiklis “Paar sõna lastekeelest” avalikkust üles tegema jõudetundidel ülestähendusi oma lapsepõlvemälestuste ja hilisemate kokkupuudete põhjal *laste- ja ammede keele*ga ning saatma neid Tartu Emakeele Seltsile (Mägiste 1924: 9).

pretendeerida keegi, kes ei valda *lastemurret*". Termini *lastekeel* kohta leiab Mägiste, et see "võistluseta läbilõõnud nimetus" pole küll päris täpne, kui pidada silmas autoriteetide (ta nimetab Hermann Pauli, Wilhelm Wundti, Hugo Schuchardi ja Adolf Gotthard Noreeni) seisukohta, kelle sõnul "pole lastekeel – nimelt just kui kõrvale jätta kõige varasema lapsepõlve instinktiivsete refleksiide keel, nn *Lallwörter* ('lalin') jm – nii palju laste eneste, kui nende hoidjate, ammede ja muu ümbruse loodud. Need kohandavad oma kõne lastega viimaste kõnekatsete puhul tehtud tähelepanekuile vastavalt lastepärasemaks, ja lastel pole muud kui oma matkimisandi kasutada" (Mägiste 1924: 3). Nii selgubki, et juba ligi sada aastat tagasi on märgatud, et keelevariandid, mida kasutavad lapsed ja mida räägivad nendega teised kõnekogukonna liikmed, on erinevad, ja neid on püütud tähistada eri terminitega.

Väärrib ka märkimist, et alates Mägistest ei ole senimaani määratud mingit kindlat vanusepiiri, millal lastekeel lõpeb, kuna tegemist on individuaalse, areneva keelekasutusega ega ole võimalik kindlaks teha mingit ajapiiri, millal see enam ei arene. Eesti keeleomandamisuuringutes on terminiga *lastekeel* tähistatud enamasti väikelaste ja koolieelikute kõnet, aga mõeldav on kasutada sama terminit ka vanemate laste keele kohta, seda muidugi juhul, kui räägitakse veel arenevast keelest.

3. Hoidjakeele kohta seni kasutusel olnud termineid

Tiiu Erelt on "Eesti keele käsiraamatus" maininud lastekeelset ehk *ninnutussõnavara* (lapsega rääkimine – *ninnutamine*, kõne – *ninnutis*), mille alla ta paigutab nii laste endi kui ka lapsehoidjate keeles kasutatavad sõnad (lähemalt vt Erelt jt 2007: 620–621). Siinse artikli autorid leiavad, et mugandatud keelevariante kasutatakse ka vanemate kui kolmeaastaste lastega kõneldes, ehkki täiskasvanu konkreetset mugandused sõltuvad lapse vanusest. Erelti kirjeldatud ninnutist võiks ehk pidada hoopis kuni aastasele lapsele suunatud keelekasutuseks, milles tõepoolest on rohkem hüüdsõnu ja -lausungeid ning mille nn pudisõnad kõlavad samuti pigem lapse fonoloogiliselt lihtsate sõnade imitatsioonidena. Patricia Kuhl (eri valdkondade ühisuurimus, vt Gopnik jt 1999), kes on lastekeele uurimisel küll pigem intonatsioonile keskendunud, kirjeldab ema häält kui akustilist konksu, mille eesmärk on püüda beebi tähelepanu ja fokuseerida see kõnelejale. Kuhl arutleb, kas emad kasutavad seda spetsiifilist kõnevarianti ainult beebide tähelepanu võitmiseks või on sellel veel mõni oluline eesmärk, ja kinnitab samas, et ema kõneldud keel aitab esmajärgus lahendada beebide n-ö keeleprobleemi, st võimaldab beebil hakata kõnevoolus sõnu eristama. Kui väikelapsega kõneldakse, on laused lühemad ja lihtsamad, samu asju korratakse (sageli ka eri variatsioonidega) üle, alateadlikult on häälikud moodustatud selgemalt ja neid hääldatakse korrektsemalt kui täiskasvanuga suheldes, konsonandid ja vokaalid on pikendatud ja argikeelega võrreldes hääldatud liialdatud selgusega. Kuhl on seisukohal, et sellise keelevariandi omadused, mida emad kasutavad beebiga suheldes, võivad aidata neil aru saada emakeele sõnadest ja grammatikast (Gopnik jt 1999).

Kui vaadelda laste- ja hoidjakeele uurimist sugulaskeelte hulgas, võib märgata, et näiteks Soome lastekeeleuurijad on möödunud sajandi 1990. aastatel näinud samuti vajadust täiskasvanult lapsele suunatud keelevariante märkivat terminit

täpsustada. 1993. aastal arutleb Kirsti Toivainen soome lastekeele uurimise juures kasutatavate terminite *hoivakieli* ('hoidjakeel') ja *hoivasana* ('hoidjasõna') üle. Tema sõnul võivad need terminid hoolimata oma semantilisest selgusest harilikule keekekasutajale harjumatud tunduda, kuid märgivad lingvistikatermineina täiskasvanult lapsele suunatud n-ö pudikeelt ja pudisõnu (sm *lepertelykieli* ja *lepertelysana*). Termin *hoivakieli* loodi tähistama vastet ingliskeelsele terminile *baby talk*. Toivainen leiab, et terminiga *hoivakieli* ei tähistata laste eneste arenevaid keelt, vaid tegemist on keelevariandiga, mida kasutavad täiskasvanud, kui nad räägivad lapsega. Seetõttu tuleb eristada *lastekeelt*, *hoidjakeelt* ja *täiskasvanukeelt*. (Toivainen 1993)

Sari Kunnari ja Tuula Savinainen-Makkonen (2004) sõnul tähistatakse terminiga *hoivakieli* ('hoidjakeel') keelevarianti, mida täiskasvanu räägib lapsega ja mis on n-ö hoidjavärvinguga. Hoidjakeele ülesehitus on lihtsustatud ja olulisi asju korratakse aina üle (nt "Tule siia, tule!"), mis peaks kindlustama, et teade jõuab lapsele kindlasti päralt. Hoidjakeelele on iseloomulikud kõrgem hääletoon ja nähtus, kus üht ja sama mõtet antakse lapsele edasi mitmel eri moel. Hoidjakeele kasutamist põhjustavad ühelt poolt lapse välised omadused (lühike nägu, esiletõusev laup, ümarad silmad, paksud põsed, tema üleüldine väikus), teiselt poolt tema puudulik suhtlemisoskus, samuti võib hoidjakeele kasutamise põhjuseks olla alateadlik soov õpetada lapsele keelt. Täiskasvanult lapsele suunatud kõne kaudu õpib laps tundma kultuuriomast vastastikust mõjutamist ja ümbritsevat keskkonda. (Kunnari, Savinainen-Makkonen 2004: 88–89)

Tänapäeval on soome lastekeeleuurimustes kasutusel ka termin *vauvakieli* (ingl *baby talk*), mida erinevad soome-inglise oskussõnastikud³ käsitlevad kui emade kõne (ingl *motherese*) vastet, lisades täpsustuseks, et see on "täiskasvanult lapsele suunatud nn pudikõne".

Ka ingliskeelses teadussõnavaras tähistab lapsega suhtlemist hulk eri termineid ning nende kasutamises valitseb teatav ebaühtlus, et mitte öelda segadus. Reili Argus on lisaks terminile *baby talk* samuti nimetanud termineid *parentese* ja *motherese* ning leiab kõige sagedamini kasutatavama termini olevat sisendkõnet või -keelt tähistava *input*'i, lisades, et sageli räägitakse ka *lapsele suunatud kõnest* (ingl *child-directed speech*, CDS). (Argus 2008: 9) Termin *motherese* on ingliskeelsetes lastekeeleuurimustes mõningast segadust külvanud, kuna sellega on sageli üldistatult tähistatud igasugust lapsele suunatud keelevarianti ja kasutatud seda hoidjakeele sünonüümina. Catherine E. Snow loobub sel põhjusel 1986. aastal "neologismide" *baby talk* ja *motherese* kasutamisest ja kasutab nende asemel terminit *child directed speech*, kuna termin *motherese* jätab tema sõnul väärast muljet, otsekui räägiks lastega sel erilisel moel ainult emad (Snow 1986, 1997: 69). Kahtlemata tuleb siinkohal Snow'ga nõustuda. Samal seisukohal on Jay Belsky ja Brenda Volling, kes nimetavad terminit *motherese* eksitavaks ja teevad ettepaneku lisada keeleuurijate professionaalsesse sõnavarasse ka termini *fatherese* või siis sooga vähem seotud termini *parentese* (Belsky, Volling 1984), mida mõnel määral on hiljem ka kasutatud, ehkki mitte järjekindlalt.

Põhjendusel, et laps ei suhtle ainult emaga, on lapsele suunatud ja mugandatud kõnet nimetatud ka eesti keeles alates 1990. aastate keskpaigast *hoidjakeeleks*, on kasutatud ka terminit *hoidekeel* (nt Pajusalu 1996).

Niisiis tundub Ereleti pakutud termin *ninnutamine* käivat vaid keelevariandi ühe tasandi, sõnavara kohta, ent samas ei muganda vanem mitte üksnes leksikat, vaid mugandusi leidub keele muudelgi tasanditel, fonoloogiast süntaksini (Lieven 2008: 187–200). Emade kõne ninnutamiseks nimetamine on aga pigem hinnang. See keelend sobib leksikoni hinnanguliselt laetud sõnade hulka ega kvalifitseeru neutraalseks teadusterminiks, seetõttu leiavad autorid, et edaspidi on kohane kasutada pigem terminit *beebikõne*⁴, mis tähistab kuni aastasele lapsele suunatud kõnet, kusjuures laps ise ei pruugi osata veel vastu rääkida. Termin kasutuselevõtu poolt kõneleb ka selle laiem tuntuus rahvusvahelises lastekeeleuurimises (ingl *baby talk*, nt Brown 2013).

Nagu lastekeele puhul, nii tuleb ka hoidjakeele kohta märkida, et ei ole võimalik tõmmata täpset vanuselist piiri, kus hoidjakeel lõpeb. Ka täiskasvanud laps on oma vanemate laps ja temaga räägitakse teistmoodi kui teiste sama vanade inimestega. Märkimisväärne on aga see, et mingis (samuti täpsustamata) vanuses leiab vanema ja lapse vahelises suhtluses aset teatav rollide vahetumine ning hoopis laps hakkab kasutama hoidjakeele registrit, kui kõneleb oma vanade vanematega. Niisugune keelevariant on nii gerontoloogia- kui ka neuropsühholoogiauurimustes tuntud kui *elderspeak* (ka *caregivers' speech*) ja seda iseloomustavad ühelt poolt valjem hääl, kõrgendatud hääletoon ja vanemlik suhtlusstiil (Balsis, Carpenter 2006) ning teiselt poolt lühemad laused, lihtsamad sõnad, rohked kordused, aeglasem kõne, pikemad pausid jm (Kemper 1994).

4. Lastekeel ja hoidjakeel kui sotsiolektid

Siinse artikli autorid, nagu paljud eelviidatud, on seisukohal, et lastekeelt ja hoidjakeelt tuleb kindlasti eristada, ehkki mõlemat keelevarianti iseloomustavad ühtmoodi mitteformaalne kõne ja intiimne, emotsionaalselt lähedane kõnesituatsioon. Hoidjakeelt on vast sobilik nimetada kõneregistriks, millel on oma kindlad tunnused (eripärane intonatsioon, täiskasvanule suunatud kõne lausungitega võrreldes lühemad lausungid, aeglasem kõne, rohked reduplikatsioonid, kordused ja ümberütlemised (Ferguson 1964, Snow, Ferguson 1977, 2013), täiskasvanukõnest erinev lausungite sõnaliigiline koostis, samuti teistsugune pragmaatiline funktsioon (Clark 2011, Argus, Kõrgesaar 2014, Kõrgesaar 2014a jne)) ja mida võib lapsega rääkida ükskõik kes, alates emast ja lõpetades hambaarstiga. Lastekeelt seevastu räägib vaid laps ise – tegemist on lapse enda üha areneva ja muutuva keelekasutusega, mistõttu on lastekeelt sobivam käsitada kui vanuselist dialekti (Mägistel *lastemurre*). Selle termini ingliskeelset vastet *age dialect* kasutab inglise laste- ja hoidjakeelt vaadelnud heebrea keeleuurija Anat Ninio (2011), kes nimetab lapse varajast kõnet kui *early child speech*, mis tema sõnutsi esindab nimelt vanuselist dialekti. Samuti kasutab ta termineid *parental register* ja *child dialect*. Ninio ütleb, et on selge, et termin *register* on varajase lapsekeele tähistamiseks vähem sobilik ja pigem sobib see vanemalt lapsele suunatud kõne (*parental child-directed speech*) jaoks, nimetades terminiga *child directed speech* tähistatud inglise keele “eriliseks registriks” (Ninio 2011: 5, 48, 51, 62, 105, 167, Ninio 2014: 1256).

⁴ Termin *beebikõne* kasutuselevõtu poolt räägib eesti keele kokku- ja lahkukirjutuse üks põhimõtteid – tähendusühendite (Erelt jt 2007: 104) –, kuid ka sõnamoodustusprintsipi, mille kohaselt kirjutatakse sihtslikku täiendosaga omaette tähenduse moodustavad ühendid kokku. Viimatinimetatud tüüpi kokkukirjutistes märgib täiendosa objekti, millele põhiosaga väljendatud tegevus on suunatud (Erelt jt 2007: 106). Seega kui kasutada terminit *beebikõne*, siis väljendab liitsõna täiendosa *beebi* objekti, kellele põhiosaga *kõne* väljendatud tegevus on suunatud.

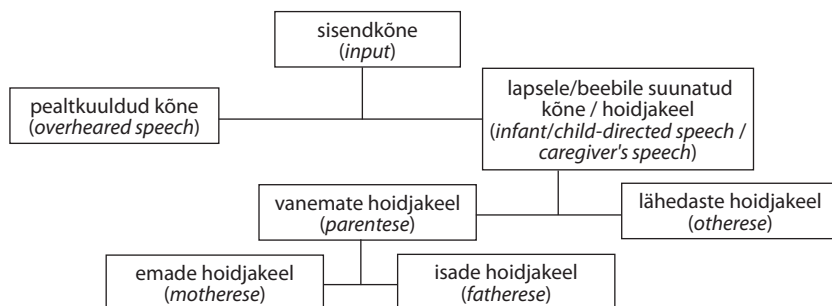
Terminit *vanusedialekt* kasutab ka Tiit Hennoste (2006), kui selgitab dialekti mõiste eri tähendusi: kohamurret ja sotsiolekti. Sotsiolekti moodustab Hennoste sõnul eri sotsiaalsete kihtide (vanus, sugu, sotsiaalne staatus vms) keelekasutus, mis väljendub peamiselt sõnavaras ja (kõige rohkem) häälduses. Peamiseks erinevuseks laste ja täiskasvanute keelekasutuse vahel ongi, et täiskasvanu mugandab kõnet sõltuvalt lapse vanusest jne, teine aga ei tee seda mitte. Väikelaps ei ole kognitiivselt, sh keeleliselt sama võimekas kui täiskasvanu ja tema suutlikkust keelt kasutada kammitsevad esialgu fonoloogilised, kognitiivsed jm arengupiirangud.

5. Terminite võimalik jaotus mõistehierarhias

1994. aastal ilmunud kogumiku “Input and Interaction in Language Acquisition” eessõnas on Catherine Snow loobunud oma varasemast seisukohast (piisab terminist *child directed speech*, kuna muud terminid külvavad liigset segadust), kirjeldades, kuidas paljud keeleuurijad, teiste hulgas ka kognitivistid, tunnistavad nüüd vajadust info järele sisendkõne kõigi andmete ja lapse keelelise keskkonna kohta, et paremini aru saada, kuidas lapsed keeleüksusi ja -nähtusi õpivad, mõistavad ja kasutavad, ning toob sealsamas näiteid uurimustest *beebikõne* (*baby talk*), *emade kõne* (*maternal input*) ja *õpetajate kõne* (*teachers’ input*) kohta (Snow 1994: 3–7).

David Crystal ingliskeelsest lingvistika- ja foneetikasõnaraamatust (2008) leiab aga märksõnade *caregiver’s speech* ja *fatherese* juurest viite, mis suunab lugeja otsima märksõna *motherese*. *Motherese* märgib Crystal järgi terminit, mida kasutati varajastes keeleomandamisuurimustes. Ta ütleb, et pakutud on ka terminit *fatherese*, kuid seda kasutatakse harva ja mõlemad terminid on koondunud ülemmõiste *lapsele suunatud kõne* (*child-directed speech*) alla, mis hõlmab lisaks vanematele ka vanavanemaid, lapsehoidjaid ja teisi, kes lapse eest hoolt kannavad (*other carers*). (Crystal 2008: 66, 74–75, 186, 316) Samas kõnelevad Heidelbergi Pedagoogikakõrgkooli teadustöõaruanded⁵ vajadusest isade ja emade kõnet siiski eristada ning nende keelevariantide tähistamiseks on kasutusel nimelt ingliskeelsed terminid *die Motherese* (aruandes alates 2003. aastast) ja *die Fatherese* (aruandes alates 2005. aastast).

Ülaltoodud terminirohkust ja mõistesegadust silmas pidades on siinse artikli autorid koostanud laste- ja hoidjakeele valdkonna edasise uurimise hõlbustamiseks ja terminivaras orienteerumiseks järgmise hierarhilise süsteemi (vt ka soovituslikku terminiloendit artikli lõpus).



Joonis 1. Soovituslik terminikasutus laste- ja hoidjakeele uurimiseks

⁵ Vt <https://www.ph-heidelberg.de/presse-und-kommunikation/publikationen/forschungsbericht.html> (2.1.2015).

Terminiga *sisendkõne* tähistatakse ülemmõistet, mis märgib igasugust kõnet, mida laps kuuleb, olgu see siis otse temale suunatud või mitte. Termin *pealtkuuldud kõne* all on silmas peetud kõike, mis ei ole otse lapsele suunatud, kuid mida laps siiski kuuleb: tele- ja raadiosaated, vestlused bussis, tänaval, kaupluses või arsti ooteruumi ukse taga, ema vestlus külalistega jne. Esineb ka kultuure (nt kaluli rahvas Paapua Uus-Guineas ja Lääne-Samoa), milles ei eksisteeri lapsele suunatud kõnet, vaid lapsed osalevadki suhtluses kui teiste kogukonna liikmete mugandamata keelevariandi pealtkuul(a)jad (ingl *overhearers*⁶) (Ochs, Schieffelin 1995: 78). Pealtkuuldud kõne osatähtsust sisendkõnes ei tohiks alahinnata, kuna see moodustab suurema osa kõnest, mida laps päeva jooksul kuuleb. Joost van de Weijeri (2002) uurimusest selgub, et vaid 14% kõnest, mida beebi päeva jooksul kuuleb, on otse temale suunatud, 30% moodustab tema vanematele õdedele-venedele suunatud kõne ning ülejäänud (enam kui poole) täiskasvanute omavaheline kõne (van de Weijer 2002). Yuriko Oshima-Takane uurimisrühm on uurinud sünnijärjekorra rolli keeleomandamises ja leidnud, et peresse hiljem sündinud lapsed arenevad palju keerukamas keelekeskkonnas kui esmasündinud, kuna kuulevad pealt ka vanematele õdedele-venedele suunatud kõnet, mis sisaldab nt rohkem isikulisi asesõnu, ja kõne iseenesest, mida nad kuulevad, on samuti märksa täiskasvanu-pärasem ja komplekssem kui see, mida kuulevad pere esmasündinud (Oshima-Takane jt 1996: 631).

Teise alajaotusena (vt joonis 1) on esitatud paralleeljaotised, kus

- a) *beebile/lapsele suunatud kõne* tähistab lapsele suunatud, kuid spetsiaalsete mugandusteta keelt, ning
- b) *hoidjakeel* (ka eesti *beebikõne/beebikeel*) tähistab lapsele suunatud ja mugandatud keelekasutust. Niisugune eristamine võiks aset leida põhjendusel, et vanem alati ei muganda oma keelekasutust lapsega rääkides ja argiolukorras võib ta suhelda lapse kui täiskasvanud vestluspartneriga, seda eriti suurema lapse puhul.

Mõlema, nii mugandatud kui ka mugandamata keelevariandi saab omakorda jagada kaheks: vanemate ja muude lähedaste (hoidja)keeleks. Vanemate (hoidja) keel omakorda jaguneb emade (hoidja)keeleks (vt Snow, Ferguson 1977, 2013, Newport 1975, Gleitman jt 1984) ja isade (hoidja)keeleks. Autorid on teadlikud, et ingliskeelses teaduskirjanduses ei tehta neil kolmel terminil – vanemate, emade või isade hoidjakeelel (ingl *motherese, fatherese* ja *parentese*) – alati kindlat vahet, sageli sel põhjusel, et neid polegi tarvis eristada, kui eesmärk on uurida lapsele suunatud kõnet üldisemalt. Kuna eesti lastekeeleuurimises on teemade valik mitmekesistunud⁷, võiks, olenevalt uurija eesmärgist, vajalikud tööriistad terminite näol eesti keeles siiski olemas olla.

Muude lähedaste all (vt joonis 1) mõeldakse kõiki neid, kes lapsega sagedamini kokku puutuvad ja temaga rohkem suhtlevad, avaldades sel moel lapse kõne arengule ka suuremat mõju. Selle jaotuse alla võivad kuuluda õed-vennad⁸ ja lähedased omavanused sõbrad (ingl *peer talk*, vt Blum-Kulka, Snow 2004) ning onud-tädid, vanavanemad, peretuttavad, kellega tihemini kohtutakse jpt. Laias laastus võib siia paigutada ka lapsehoidjad, lasteaia- ja hiljem kooliõpetajad, treenerid jts, ehkki nii pedagoogika- kui ka keeleuurimustes vaadeldakse eraldi ka õpetaja-õpilase vahelist suhtlust (ingl *teachers' speech*, vt Gleason 2013, Owen 1996, Snow 1994).

⁶ Erinevate kuulajaskonna tüüpide kohta, nagu *addressee, auditor, overheard* ja *eavesdropper*, vt Meyerhoff 2006: 43.

⁷ Eestis on isa ja tütre vestlust uurinud Andriela Rääbis (2012) ning erinevusi isade-emade keelekasutuses Helen Kõrgesaar (2014b, 2014c).

⁸ Õe ja venna omavahelist suhtlust on uurinud Tuuli Paalandi (2013).

Kuigi ingliskeelses lingvistikakirjanduses käsitatakse enamikku mainitud termineid sünonüümidenä, ei tarvitse see peegelduda eesti teaduskeeles. Selleks et eesti oskuskeel ja eriti selle tuumosana teaduskeel oleks konkreetne ja täpne, tuleb termineid kasutada võimalikult järjekindlalt (Nemvalts 2011: 165). Ingliskeelse teaduse terminisegaduse on tinginud tohutu suur uurijate-publitseerijate hulk ja nii on mõistagi keeruline ühtsust saavutada. Eestis on laste- ja hoidjakeele uurijate ring üsna kitsas, mistõttu selgete, täpsustatud terminite kasutuselevõtt oleks ühtaegu hõlpus ja valdkonnale igati kasulik. Otstarbekas oleks täpsustada ka muid valdkonnas rohkem või vähem kasutatavaid termineid (nt *keelekogus*, ingl *amount of language*) ning selgitada põhjalikumalt erinevusi kirjaliku teksti ja suulise kõne uurimisel kasutatavate terminite vahel.

6. Kokkuvõtte asemel

Artiklis on lähtunud Uno Mereste ületuspõhimõttest, mille kohaselt teistest hiljem mingil erialal termineid korrastav keel on soodsamas olukorras, sest saab teiste keelte piisamatuse avastada ja vältida, lähtuda erialade mõistesüsteemidest ja kasutada nii oma keele loomevõimsust kui ka laenuvõimalusi (Uno Mereste 2000: 77). Nagu siinse artikli ülevaates viidatud, on inglise keeles kasutatavaid termineid nii laste- kui ka hoidjakeele eri variantide kohta palju, kuid terminikasutuses pole järjepidevust, osa terminite mõistesisu kattub, kuid osa ei peegelda täpselt sedasama eesti keeles. Lisaks on kasutusel palju paralleelvorme, mida valdkonda süvenematu inimene tõlgendab tihtipeale valesti, kuna pole toetavat konteksti (nt *hoidjakeele* vaste *motherese* märgib sageli ka isade hoidjakeelt või kelle tahes kõnet, mis on suunatud lapsele).

Kokkuvõtte asemel esitavad autorid soovitusliku laste- ja hoidjakeele terminite loendi emma-kumma valdkonna uurijate töö edaspidiseks hõlbustamiseks.

7. Soovituslik terminiloend

beebikõne (ingl *baby talk*) – täiskasvanult beebile suunatud mugandatud keelevariant

beebile suunatud kõne (ingl *infant directed speech*) – täiskasvanult beebile suunatud mugandamata keelevariant

emade hoidjakeel (ingl *motherese, maternal speech, mother-child interaction/communication*) – emalt lapsele suunatud mugandatud keelevariant

hoidjakeel (ingl *caregivers' speech, motherese, maternal speech, mother-child interaction/communication*) – täiskasvanult lapsele suunatud mugandatud keelevariant

isade hoidjakeel (ingl *fatherese, paternal speech, father-child interaction/communication*) – isalt lapsele suunatud mugandatud keelevariant

lähedaste hoidjakeel (ingl *otherese*) – onude-tädide, vanavanemate, peretutavate, aga ka lapsehoidjate, lasteaia- ja kooliõpetajate, treenerite jt kõne

lapsele suunatud kõne (ingl *child-directed speech*) – täiskasvanult lapsele suunatud mugandamata keelevariant

lastekeel (ingl *child language*) – lapse enda arenev keelevariant

pealtkuuldud kõne (ingl *overheard speech*) – kõne, mida laps kuuleb, kuid mis ei ole mugandatud ega otseselt temale suunatud (ühissõidukites, tänaval, telerist jm)

sisendkõne (ingl *input*) – igasugune kõne, mida laps kuuleb, olgu see siis otse temale suunatud ehk mugandatud või mitte

õdede-vendade ja lähedaste omavanuste sõprade kõne (ingl *peer talk*) – keelevariant, mida laps kasutab suhtluses õe-venna või sõpradega

õpetaja-õpilase vahelise suhtluse keel (ingl *teachers' speech*)

Viidatud kirjandus

- Argus, Reili 2008. Eesti keele muutemorfoloogia omandamine. [Acquisition of Morphology in Estonian.] Tallinna Ülikooli humanitaarteaduste dissertatsioonid 19. Tallinn: Tallinna Ülikooli Kirjastus.
- Argus, Reili; Kõrgesaar, Helen 2014. Sõnaliigid eesti lapse kõnes ja lapsele suunatud kõnes. [Word classes in the child's speech and in the child-directed speech.] – Eesti Rakenduslingvistika Ühingu aastaraamat, 10, 37–53. <http://dx.doi.org/10.5128/ERYa10.03>
- Balsis, Steve; Carpenter, Brian B. 2006. Evaluations of elderspeak in a caregiving context. – *Clinical Gerontologist*, 29 (1), 79–96. http://dx.doi.org/10.1300/JO18v29n01_07
- Belsky, Jay; Volling, Brenda L. 1987. Mothering, fathering, and marital interaction in the family triad during infancy. – Phyllis W. Berman, Frank A. Pedersen (Eds.). *Men's Transitions to Parenthood: Longitudinal Studies of Early Family Experience*. Psychology Press, 37–63.
- Blum-Kulka, Shoshana; Snow, Catherine E. 2004. Introduction: The potential of peer talk. – *Discourse Studies*, 6 (3), 291–306. <http://dx.doi.org/10.1177/1461445604044290>
- Brown, Roger 2013 [1977]. Introduction. – Catherine E. Snow, Charles A. Ferguson (Eds.). *Talking to Children: Language Input & Acquisition*. New York: Cambridge University Press, 1–27.
- Clark, Eve V. 2011 [2003]. *First Language Acquisition*. Cambridge: Cambridge University Press.
- Crystal, David 2008. *A Dictionary of Linguistics and Phonetics*. 6th ed. Oxford: Blackwell. <http://dx.doi.org/10.1002/9781444302776>
- Erelt, Tiit; Erelt, Mati; Ross, Kristiina 2007. *Eesti keele käsiraamat*. 3. täiendatud trükk. Tallinn: Eesti Keele Sihtasutus.
- Ferguson, Charles 1964. Baby talk in six languages. – *American Anthropologist*, 66 (6), 103–114. http://dx.doi.org/10.1525/aa.1964.66.suppl_3.02a00060
- Gleason, Jean Berko 2013 [1977]. Talking to children: Some notes on feedback. – Catherine E. Snow, Charles A. Ferguson (Eds.). *Talking to Children: Language Input and Acquisition*. New York: Cambridge University Press, 199–205.
- Gleitman, Lila; Newport, Elissa L.; Gleitman, Henry 1984. The current status of the motherese hypothesis. – *Journal of Child Language*, 11 (1), 43–79. <http://dx.doi.org/10.1017/S0305000900005584>
- Gopnik, Alison; Meltzoff, Andrew; Kuhl, Patricia 1999. *The Scientist in the Crib: Minds, Brains, and How Children Learn*. New York: William Morrow & Co.
- Hennoste, Tiit 2006. Suuline keel 1. Keelekõrv. Eetris 4.3.2006. ERR-i arhiivis <http://arhiiv.err.ee/vaata/keelekorv-keelekorv-suuline-keel-1> (16.10.2014).
- Kemper, Susan 1994. Speech accommodations to older adults. – *Aging, Neuropsychology, and Cognition*, 1 (1), 17–28. <http://dx.doi.org/10.1080%2f09289919408251447>
- Kunnari, Sari; Savinainen-Makkonen, Tuula 2004. Mistä on pienten sanat tehty. Lasten äänteellinen kehitys. Helsinki: WSOY.

- Kõrgesaar, Helen 2014a. Küsimused eesti lapsele suunatud kõnes leedu ja vene keele taustal. [Interrogatives in Estonian child-directed speech.] – Eesti Rakenduslingvistika Ühingu aastaraamat, 10, 193–207. <http://dx.doi.org/10.5128/ERYa10.12>
- Kõrgesaar, Helen 2014b. Eesti isade-emade hoidjakeel: kes küsib, kes käsutab, kes räägib rohkem? [Estonian fathers' and mothers' child-directed speech: Who asks more questions, who gives orders, and who speaks more?] – Krista Kerge (Toim.). Uurimusi keele omandamisest, õppimisest ja korpustest. Tallinna Ülikooli eesti keele ja kultuuri instituudi toimetised 16. Tallinn, 59–82.
- Kõrgesaar, Helen 2014c. Direktiivsed kõneaktid ja nende kasutamise dünaamika isa-ema kõnes: juhtumiuuring. [Directive speech acts and the dynamics of their usage in father's and mother's speech: A case study.] – Rakenduslingvistika Ühingu aastaraamat, 11, 159–175. <http://dx.doi.org/10.5128/ERYa11>
- Lieven, Elena 2008. Language development: Overview. – Marshall M. Haith, Janette B. Benson (Eds.). Encyclopedia of Infant and Early Childhood Development. Oxford: Elsevier, 187–200.
- Mereste, Uno 2000. Oskuskeele üldprobleeme kaug- ja lähivaates. – U. Mereste. Oskuskeel ja seaduste keeleline rüü. Tallinn: Eesti Keele Sihtasutus, 57–83.
- Metsmägi, Iris; Sedrik, Meeli; Soosaar, Sven-Erik 2012. Eesti etimoloogiasõnaraamat. [Etymological Dictionary of Estonian.] Tallinn: Eesti Keele Sihtasutus.
- Meyerhoff, Miriam 2006. Introducing Sociolinguistics. London–New York: Routledge.
- Mund, Argo 2004. Mitmuse omastav luubi all. [Focus on plural genitive.] – Maire Raadik (Toim.). Keelenõuanne soovitab 4. Tallinn: 138–141.
- Mägiste, Julius 1924. Paar sõna lastekeelest. [A few words on child-language.] – Eesti Keel, 1, 1–9.
- Nemvalts, Peep 2011. Terminiteadusest eesti teaduskeeleni. [From terminology science to academic Estonian.] – P. Nemvalts (Toim.). Eesti teaduskeel ja terminikorrastus. Tallinna Ülikooli eesti keele ja kultuuri instituudi toimetised 13. Tallinn: 136–171.
- Newport, Elissa L. 1975. Motherese: The Speech of Mothers to Young Children. La Jolla: Center for Human Information Processing, University of California, San Diego.
- Ninio, Anat 2011. Syntactic Development, Its Input and Output. Great Britain: Oxford University Press. <http://dx.doi.org/10.1093/acprof:oso/9780199565962.001.0001>
- Ninio, Anat 2014. Learning a generative syntax from transparent syntactic atoms in the linguistic input. – Journal of Child Language, 41 (6), 1249–1275. <http://dx.doi.org/10.1017/S0305000913000470>
- Oshima-Takane, Yuriko; Goodz, Elizabeth; Derevensky, Jeffrey L. 1996. Birth order effects on early language development: Do secondborn children learn from overheard speech? – Child Development, 67 (2), 621–634.
- Ochs, Elinor; Schieffelin, Bambi 1995. The impact of language socialization on grammatical development. – Paul Fletcher, Brian MacWhinney (Eds.). The Handbook of Child Language. Oxford: Basil Blackwell, 73–94.
- Owen, David 1996. Do teachers modify their speech according to the proficiency of their students? – English Language Teacher Education and Development, 2 (1), 31–51.
- Paalandi, Tuuli 2013. Õe ja venna omavahelise spontaanse suhtlustegevuse keelelised erijooned mängusituatsioonis: küsimused ja direktiivid. Bakalaureusetöö. Tallinna Ülikool.
- Pajusalu, Karl 1996. Eesti hoidekeele kujunemine. – Suomalaiskielten omaksumista tutkimassa. Turun yliopiston suomalaisen ja yleisen kielitieteen laitoksen julkaisuja 53. Turku: Turun yliopisto, 117–124.
- Rääbis, Andriela 2012. Direktiivskventsid isa ja tütre suhtluses: juhtumianalüüs. [Directive sequences in a conversation of father and daughter: A case study.] – Rakenduslingvistika Ühingu aastaraamat, 8, 213–230. <http://dx.doi.org/10.5128/ERYa8.14>

- Snow, Catherine E. 1994. Introduction. – Clare Gallaway, Brian J. Richards (Eds.). *Input and Interaction in Language Acquisition*. Cambridge University Press, 3–12.
- Snow, Catherine E. 1997 [1986]. *Conversations with children*. – Paul Fletcher, Michael Garman (Eds.). *Language Acquisition*. Cambridge University Press, 69–89.
- Snow, Catherine E.; Ferguson, Charles A. 2013 [1977]. *Talking to Children: Language Input and Acquisition*. New York: Cambridge University Press.
- Toivainen, Kirsti 1993. Hoivakielen tutkimuksesta ja suomen murteiden hoivasanastosta. – Ettekande käsikiri soome-ugri keelte omandamise uurimise II seminarilt Nauvo Seilis 25. sept 1993.
- Weijer, Joost van de 2002. How much does an infant hear in a day? – Ettekanne. GALA'2001 Conference on Language Acquisition. Lisboa.

Helen Kõrgesaar (Tallinna Ülikool) uurib eesti hoidjakeelt, selle dünaamikat ja mõju lapse kõne arengule.

Narva mnt 25, 10120 Tallinn, Estonia

helen.korgesaar@gmail.com

Airi Kapanen (Tallinna Ülikool) uurib eesti laste keeleomandamist, täpsemini süntaksi arengut ja kompleksust.

Narva mnt 25, 10120 Tallinn, Estonia

airikapanen@gmail.com

WHEN THE MOTHER IS NOT THE ONLY ONE TALKING TO A CHILD: SELECTION OF TERMS IN ESTONIAN CHILD LANGUAGE AND CAREGIVERS' SPEECH

Helen Kõrgesaar, Airi Kapanen

Tallinn University

The article focuses on specifying and describing the terms used in Estonian child language and child-directed speech. The terms used now do not offer enough necessary and precise tools for researchers of child language. Since children of different ages and proximities are talked to differently, a need has arisen to define the different types of child-directed speech, due to the multiple interpretations of terms and their concepts. Although most of the terms describing speech directed from an adult to a child have been considered synonyms in the English linguistics literature, this might not be the case in Estonian scientific language. The confusion of terms in English science is due to the massive amount of researchers and publishers which obviously makes it difficult to achieve uniformity. Taking into consideration this multitude of terms and confusion of concepts, the authors of this article offer recommended terms to be used in the future research of child and child-directed speech, in order to simplify the research and facilitate understanding of the terminology. The circle of researchers of child and child-directed speech in Estonia is quite narrow, which not only makes the implementation of clear, specified terms easy to agree on but would also greatly benefit the field of research itself. The authors are aware that in English scientific literature the terms used (e.g. *motherese*, *fatherese*, *parentese*) are not strictly distinctive, often due to the fact that it is not necessary, if the aim is to research child-directed speech in general. Since Estonian child language research is diversified, the necessary tools in the form of Estonian-language terms should be provided to be used according to the aim of the researcher.

Keywords: first language acquisition, child-directed speech, fatherese, motherese, lexicon of terms, Estonian

EVENT COREFERENCE DETECTION IN ESTONIAN NEWS ARTICLES: PRELIMINARY EXPERIMENTS

Siim Orasmaa

Abstract. Event coreference detection is a task of automatically determining which fine-grained textual descriptions of events (e.g. sentences) corefer. The task is important in organizing information in large collections of news articles, as extracted coreferring event mentions can provide the user with an overview of the events discussed in articles, and can also provide a glimpse into the factual data related to the events. In this article, we survey previous approaches to automatic event analysis, discuss theoretical considerations related to event coreference detection and also outline a motivation for experimenting with event coreference detection in the context of limited linguistic resources. In the experimental part of our work, we consider the task of event coreference detection in the subset of articles mentioning a specific person in a fixed publishing period, and we use the experiments to outline possible general factors that are influencing the results.*

Keywords: natural language processing, text analysis, Estonian

1. Introduction

Event coreference detection is a task of automatically determining which textual event mentions corefer. In particular, the task is concerned with fine-grained text analysis: how to determine that two sentences or two words/phrases (two event mentions) are referring to the same event. For example, consider the following two sentences extracted from Estonian news:

- (1) Soome sideminister Matti Aura andis eile hommikul lahkumispalve, sest tunnistas tehtud viga, kui ta hiljuti õigustas Vennamo käitumist.
‘The Finnish Minister of Communications, Matti Aura, gave his resignation yesterday morning, as he admitted that his recent justification of Vennamo’s behaviour was a mistake.’ (ERC, Postimees 5.1.1999)

* This work was supported by Estonian Ministry of Education and Research (grant IUT 20-56 “Computational models for Estonian”), and by the European Regional Development Fund through the Estonian Centre of Excellence in Computer Science (EXCS). The author also wishes to thank anonymous reviewers for helpful comments.

- (2) Soome sideminister Matti Aura teatas eile oma tagasiastumisest seoses Sonera aktsiate müügi ümber puhkenud skandaaliga.
‘The Finnish Minister of Communications, Matti Aura, announced his stepping down yesterday, related to the scandal of selling Sonera’s shares.’
(ERC, SL Õhtuleht 5.1.1999)

Both sentences (1) and (2) mention the same event – an announcement of resignation, which is described by two event mentions: ‘gave his resignation’ and ‘announced his stepping down’.

From the perspective of a human reader, say, a historian trying to collect factual information regarding some events from contemporary newspapers, it is useful to compare different sources describing the same event in order to obtain a better picture of that event. This is because an event description from a single source can be incomplete (e.g. missing some factual information), or inaccurate (e.g. presenting wrong factual information). Factual information that is compatible across multiple/different sources can also be considered accurate with somewhat higher confidence (although the reliability of the source also plays an important role in determining the confidence, see Fokkens et al. 2014).

The task of automatically detecting coreferring event descriptions from news represents an interesting challenge in Natural Language Processing. As one can often assume that during the same publishing period (e.g. a day or a week), news articles from different sources tend to describe the same events, the search space of sentences potentially containing coreferring events can be effectively narrowed down by time constraints. Within such constrained search space, even a simple method matching lemmas of event mentions could be sufficient for solving the task (Cybulska, Vossen 2014). Note that while many past attempts at sentence-level event analysis have relied on conceptual modelling of events (e.g. relying on a set of predefined semantic frames for event detection), event coreference detection allows one to explore more simple and general event models, such as models where an event is mainly identified by arguments describing its participants, time and location (Glavaš, Šnajder 2013, Cybulska, Vossen 2013). Still, the task is very difficult, and difficulties seem to stem from the abstractness of the notion of event. It is not clear, how the notion of event should be established linguistically: which size text units serve best as concise descriptions of events (words/phrases/clauses/sentences/...?). Furthermore, even if the notion of event is established at some level, the coreference between events can still be hard to decide, as one needs to consider other relations between events, such as subevent and membership relations (Hovy et al. 2013).

The present work explores event coreference detection in Estonian news articles. We consider a realistic information retrieval setting, where the user wants to find information about events related to a specific person from a corpus of daily newspapers. In the experiments, we test out simple heuristics (lemma matching, as well as matching of event arguments of location and time) for the task of finding pairs of sentences mentioning the same event. Our work is preliminary as we do not aim to provide a thoroughly evaluated method for solving the problem. Instead, the aim is to explore how general factors – lexical homogeneity within the set of articles under analysis, and the media coverage of the events related to the person – influence the results.

2. A few notes on the notion of event

It is difficult to give a good definition for the notion of **event**, as many definitions seem to involve synonymous abstract words, such as *occurrence* or *happening*, and using these words can ultimately lead to a circular definition. The question about the essence of all events has been studied in philosophy, and there it eventually led to the question of event identity: when can two events be considered identical (Schneider 2005)? Donald Davidson (1969) first proposed that two events having the same causes and effects should be considered as identical. Later he rejected the account and agreed with Willard Van Orman Quine (1985) that two events occurring at the same time and in the same space should be considered identical (Davidson 1985).

In this work, we try to follow the Quinean-Davidsonian notion of event (an event being identified by its spatiotemporal location) in making judgments concerning coreference relations between events mentioned in natural language texts. As this is a preliminary work, we make no further attempt to divide events into classes, and we also regard states (such as *Jaan is hungry*, *the oven is hot*) as events.

3. Different levels of event analysis in information retrieval and extraction

As the concept of event is an abstract one, it allows a number of different views on how events are expressed in natural language. In this work, we focus mainly on information retrieval and extraction context, where one has a practical goal of finding documents (news articles) talking about the same event(s), and extracting fine-grained details about the events of interest. Previous work in this field can be roughly divided into three event analysis levels, each considering a different size text unit as a description of an event: **document-level**, **sentence-level**, and **mention-level** analysis.¹

Document-level event analysis mostly relies on an assumption that an article/news story has been built around one main event (the topic), although it also describes relevant events related to the main event. In Topic Detection and Tracking research (Allen et al. 1998) the goal is to analyze a stream of news stories and to detect stories that are discussing the same or directly related events. Because an event has a rather large textual description in this view (the whole article), lexical similarity between articles can be relatively successfully utilized for solving the task. A limitation of this processing level is that articles are often discussing many events, and knowing that two articles are discussing the same event does not immediately tell us which event exactly is the common one, nor does it reveal any interesting details of the event (e.g. who did what to whom, when and where?).

Sentence-level event analysis focuses on detecting an event mention along with a detailed description of its structure: event participants and circumstances. It is, perhaps, inspired by a prototypical simple sentence, in which the main verb refers to the event and other complements refer to participants and circumstances of the event (event arguments). On this level of event analysis, it is often assumed that

¹ This division draws inspiration from a presentation by TimeML Working Group (2007).

the conceptual structures of events are predefined, e.g. in the form of FrameNet semantic frames (Baker et al. 1998) or ACE² event types, and the goal is to detect presence of the event in the sentence, and to fill the argument slots in the corresponding conceptual structure. However, matching conceptual semantic arguments with their actual realizations in text has proven to be a very difficult task, and so far, successful automatic event analysis on this level has required setting the focus on a narrow set of event types (e.g. only events of *attack*, *death* and *injury*) (Naughton 2009). The obvious limitation is that many other events that are mentioned in the text will be left unanalysed (including relevant events not well-described in the given set of event types).

Mention-level event analysis focuses on the smallest possible linguistic units describing an event – event mentions such as verbs (e.g. *married*), nouns (e.g. *wedding*) and adjectives (e.g. *(was) pregnant*). Because the focus is only on detection of event mentions, without accompanying event argument structures, this makes it possible to define a rather general event model, which has the advantage of not being restricted to a list of predefined event types (like FrameNet frames or ACE event types). The disadvantage is, however, that it is difficult to give a precise definition of which linguistic units should be considered events. Complex problems include analysing multiword event units (e.g. *does held a conference* indicate one or two events, as the act of holding/organizing a conference can be distinguished from the actual event – the conference – as the former usually takes a longer time span), and distinguishing between event and non-event readings of nominals (e.g. *dinner* as an event and as a food) (Sprugnoli, Lenci 2014). TimeML (Pustejovsky et al. 2003) is an example of a framework that focuses on event annotation at the mention level.

4. Event coreference detection

4.1. Theoretical and empirical concerns regarding the event coreference detection

The task of event coreference detection can be divided into **within-document event coreference detection** (i.e. finding coreferring event mentions within the same document) and **cross-document event coreference detection** (i.e. finding coreferring event mentions from different documents) (Naughton 2009).

Eduard Hovy, Teruko Mitamura, Felisa Verdejo, Jun Araki and Andrew Philpot (2013) note that the difference between coreference and non-coreference is not clear-cut, and partial coreference of events can be distinguished. They study partial coreference in detail, arguing that partial coreference could indicate either a membership or a subevent relation between two events. A membership relation holds when one event mention describes a set of events of the same type, and the other event mention refers to a single member of that set. For example, in the sentence “I had *three meetings* last week, but only *the last one* was constructive” the event referred to as *the last one* can be considered a member of the set of events *three meetings*. In the case of a subevent relation, one event is considered to be a stereotypical sequence of events having a common goal (a script, such as *eating at the restaurant*, *visiting a doctor*), and the other event (the subevent) is one of

² ACE (Automatic Content Extraction) programme – a programme that was established in United States for developing automatic Information Extraction technologies, see also <https://www ldc.upenn.edu/collaborations/past-projects/ace> (4.9.2014)

the actions/events executed as part of that script (e.g. *ordering the meal* is part of *eating at the restaurant*, and *making an appointment with the doctor* is part of *visiting a doctor*). Authors also argue that the time, location, and participants of events provide key information for distinguishing between different types of coreference (and non-coreference). However, their inter-annotator agreement results on distinguishing partial coreference relations indicate that these concepts are not yet well understood.

The idea that time, location, and participants could play an essential role in event coreference detection has also been explored by Goren Glavaš and Jan Šnajder (2013). The authors use a generic event model where an event is defined by an event anchor (an event mention, usually a single word), and one or more arguments of four broad semantic types (agent, target, time, and location). They note and also confirm by inter-annotator agreement experiments that when humans are asked to decide event coreference considering only mentions and four arguments, their agreement and certainty on the task is rather low, as opposed to when they can examine the full context (both documents where event mentions occurred) before making the decision. This suggests that event coreference detection is inherently a composite task, where all levels of event analysis – the document, sentence, and mention levels – should be considered together.

4.2. Event coreference detection in the context of limited linguistic resources

Most recent works on automatic event coreference detection have considered English or other languages well-equipped with linguistic resources, such as Spanish, Italian and Dutch (Glavaš, Šnajder 2013, Cybulska, Vossen 2013, Vossen et al. 2014). The task in its full complexity seems to require that a number of advanced language analysis steps are implemented: 1) detection of event mentions, 2) named entity recognition and semantic role labelling for detection of event arguments, 3) normalization of temporal and locational expressions³ and named entity coreference resolution for aligning event argument structures, 4) aligning event mentions along with their argument structures. However, it is not clear to what extent these fine-grained language analysis steps must be (and can be) solved, and some of the recent research suggests that light-weight approaches to event coreference detection are also worth trying out.

Firstly, it is likely that tasks of within-document event coreference and cross-document event coreference have different levels of difficulty. Martina Naughton (2009) notes that two sentences mentioning the same event within the same document are likely to have heterogenous vocabulary, as the factual information (e.g. location, participants) is rarely repeated on the second mention of the event. In contrast, two cross-document sentences mentioning the same event have likely more homogeneous vocabulary, because the important factual information (location, time, participants) is repeated in both documents. Thus, the task of cross-document event coreference can potentially be approached even with simple methods relying on lexical similarity.

³ E.g. finding calendar dates corresponding to temporal expressions, and finding coordinates of geographical regions corresponding to locational expressions.

Secondly, if the set of documents under analysis can be narrowed down to the subset of documents discussing the same event(s) (such as in the Topic Detection and Tracking task), even a simple method – matching event mentions by their lemmas – can yield relatively good results, which are roughly comparable to the results obtained with complex methods matching event argument structures. This suggests that document clustering plays an important role in event coreference detection, and if one can obtain clusters of documents discussing the same events (e.g. by using lexical similarity and time constraints), one can resolve the amount of mention-level event coreference within these clusters by using simple lexical similarity methods. (Cybulska, Vossen 2014)

In this work, we explore the problem of event coreference detection in Estonian. As this is a preliminary work on the task in Estonian, we also start out experimenting with simple methods.

5. A case study: finding person-related events from news articles

As a case study, we consider an exploratory search setting, where the user wants to find events related to a specific person from a corpus of daily newspapers. We make a crude assumption that such events can be found from articles where the person is mentioned by full name, and the goal is to find pairs of sentences mentioning the same event within this set of articles.⁴ We add two additional constraints. Firstly, we only consider articles from a fixed time period: one week. And secondly, we focus on finding pairs of sentences from different articles, i.e. on cross-document event coreference.

5.1. Document-level considerations regarding event coreference

Before we can begin exploring sentence-level event coreference, we likely want to have some rough characterization of the given set of articles regarding the possible difficulty of sentence-level event coreference. Lexical similarity within the set of articles (more specifically, the average lexical similarity between all pairs of articles within the set of articles) can serve as such a characteristic:

- Lexically homogeneous articles (with high lexical similarity) are likely describing the same events.
- Lexically heterogeneous articles (with low lexical similarity) can describe different events, or can describe the same events using rather different vocabularies.

Naturally, the lexical heterogeneity is affected by the timespan chosen for analysis: the longer timespan typically means that there are more events reported, thus a more lexically heterogeneous set of articles is obtained. However, as we consider only articles mentioning a specific person, the increase of reports also depends on how much the media covers the given person's activities, and how this coverage changes over time.

⁴ An alternative would be to attempt to detect clusters of sentences, so that each cluster contains sentences referring to same event. However, this approach would be more difficult to evaluate (considering partial coreference relations and the fact that one sentence can refer to multiple events), so we chose to explore the pair-wise detection approach instead.

In the present work, we use the vector space model and calculate cosine similarity over the word lemmas to measure lexical similarity between articles. We also apply tf-idf weighting for lemmas (Manning et al. 2008), which means that lemmas occurring many times in few articles have the highest discriminative power (and sharing such lemmas contributes the most to the similarity between articles), while lemmas occurring almost in all articles of the corpus have the lowest discriminative power (thus sharing such lemmas contributes only little to the similarity between documents). Lemma counts and occurrences in articles are calculated over the whole corpus (one week of news articles).

5.2. Sentence-level event coreference

In our experimental setup (details discussed in the next section), we consider 3 different sets of articles with varying degrees of lexical similarity, and we test two different methods for sentence-level event coreference detection: 1) a simple lexical similarity measure (measuring the amount of lemmas overlapping between two sentences), and 2) an overlap of event argument structure components (location and time). It is expected that if the set of articles is lexically homogeneous, a simple lexical similarity measure (method 1) yields high-precision results, which potentially outperforms (in terms of precision) a more sophisticated event model involving argument structure similarity (method 2). On the other hand, in a lexically heterogeneous corpus, the simple lexical similarity measure is likely ineffective, and the model matching event location and time will likely provide better results.

As a similarity measure, we use the Jaccard Similarity Coefficient (Jaccard 1901), which is defined as a similarity between two sets: the size of the intersection of two sets divided by the size of the union of two sets.

In order to find the simple lexical similarity between two sentences (method 1), we extract word lemmas from both sentences, convert to sets (i.e. remove duplicate lemmas) and calculate the Jaccard Similarity Coefficient between these sets of lemmas. All pairs of sentences that had a similarity coefficient value greater than or equal to 0.5 were selected as pairs potentially containing coreferring event mentions.

Note that one of the possible limitations of using the Jaccard Similarity Coefficient over whole sentences is that the measure is sensitive to the contrast between sentence lengths: if a short sentence entirely overlaps with a long one, and the short one is less than half the length of the longer one, a coefficient value below 0.5 is obtained. This problem can be alleviated by using a method less sensitive to the difference between sentence lengths: the Second Kulczynski Coefficient (Pecina 2010). This method finds an average of two ratios: the size of the intersection divided by the cardinality of the first set, and the size of the intersection divided by the cardinality of the second set. However, as our experiments showed, this measure introduces another problem: a short sentence containing mainly non-content words (such as “Aga kes siis veel” ‘But who else then’) can match with many long sentences containing these non-content words, thus introducing some degree of noise into the results. Although filtering of non-content words could potentially alleviate this problem, we chose to stay with the Jaccard Similarity Coefficient in order to keep our models simple.

In the second similarity method, we consider overlap of argument structure components of the event mentions: temporal expressions and location names mentioned in sentences. Because we don't have syntactic structure or semantic role annotations available, we use a very crude approximation and consider all temporal expressions and locations appearing in one sentence to belong to one event argument structure. For temporal expressions, we consider normalized calendrical values of the expressions instead of lemmas, which allows us to match (lexically) different date expressions, such as *täna* 'today' and *eelmisel neljapäeval* 'last Thursday' if they refer to the same normalized date (e.g. 11.9.2014; for more on annotation of temporal expressions in Estonian, see Orasmaa 2012). We focus on date expressions containing day granularity (e.g. *two days ago*), month granularity (e.g. *in April*), or year granularity (e.g. *last year*) temporal information. For location names, we only consider lemma matches as we currently do not have means for normalizing location expressions to coordinates of geographical regions.

Similarly to the first method, we apply the Jaccard Similarity Coefficient and calculate two scores: a coefficient for matching calendrical values of temporal expressions, and a coefficient for matching lemmas of location expressions. If both coefficient values were greater than or equal to 0.5, then the pair of sentences was considered as potentially containing coreferring event mentions.

6. Experiments

6.1. The corpus

In our experiment, we selected all news articles from one week (from 4.1.1999 to 10.01.1999) of three Estonian daily newspapers (Postimees, Eesti Päevaleht, and SL Õhtuleht), as they can be found in the Estonian Reference Corpus⁵ (Kaalep et al. 2010). The corpus has been automatically annotated for sentence boundaries and morphological information (word lemmas, part of speech tags, morphological case and conjugation information). In addition, there are two layers of automatically added factual/semantic annotations: named entities (persons, organizations, locations, addresses, quantities), and temporal expressions, and the latter are normalized according to the TimeML annotation format (Pustejovsky et al. 2003).

For our experiment, we have chosen three persons – Jüri Mosin, Pekka Venamo and Mart Siimann, as articles mentioning these persons are characterized by different lexical similarity levels (measured as an average of cosine similarities between all pairs of articles), and all of these article sets contain articles from three different newspapers. Table 1 reports the statistics related to the three article sets.

Table 1. Statistics of article sets mentioning the three given persons

Person	Articles	Sentences	Words	Avg cosinus similarity (tf-idf weighting)
Jüri Mosin	4	64	1080	0.56
Pekka Vennamo	9	142	2610	0.45
Mart Siimann	27	624	10507	0.13

We will briefly describe the events mentioned in given article sets. Articles mentioning Jüri Mosin were mostly focusing on a single event – a trial over criminal Jüri Mosin and his accomplices – although they also discussed background events, such as the criminal history of the persons under trial, and possible future appeals. In terms of events, the articles were rather focused, which is also reflected in high lexical homogeneity. Articles mentioning Pekka Vennamo were discussing a scandal related to the person: a possible abuse of his official position for purchasing shares. The scandal had a longer development history, and on the given week, it culminated in the firing of Pekka Vennamo from his position, and the resignation of a Finnish minister related to the scandal. In terms of events, the articles were still quite focused on the two events (the firing and resignation of two high officials), although the history of the scandal, possible future developments and effects were more widely discussed than in the first case. Articles mentioning Mart Siimann, who was the prime minister of Estonia at that time, were least focused on certain events. There were few stories where his actions were at the focus, such as meetings with prime ministers of neighbouring countries; however, many articles just mentioned him once or twice, in a discussion of events not directly related to him (e.g. an article describing a sports event mentioned once that the prime minister was supporting the event). This lack of focus on a narrow set of events is also reflected in low lexical similarity between articles.

6.2. Evaluation of sentence-level event coreference

Because we do not have a corpus annotated for event coreference available, we can only measure and compare methods in terms of precision (that is, the number of correct sentence pairs divided by the number of all pairs of sentences returned by the method).⁶

Sentence pairs found by both methods were evaluated manually. If two sentences under comparison contained multiple event mentions, it was required that only one pair of event mentions should corefer for the sentence pair to be correct. For example, consider the following pair of sentences:

- (3) Läti peaminister Vilis Krishtopans lubas eile kohtumisel Eesti valitsusjuhi Mart Siimanniga ühtlustada kahe riigi vahelist viisarezhimi. (ERC, Eesti Päevaleht 8.1.1999)
 ‘Latvian prime minister Vilis Krishtopans promised to homogenise the visa regime between the two countries at yesterday’s meeting with Estonian prime minister Mart Siimann.’

⁶ A limitation of measuring only precision is that we do not have an overview of the completeness of the results (which can be measured by recall). However, measuring recall requires creating a corpus annotated for event coreference, which is not a trivial task (due to the complexity of the phenomenon). We believe that it requires a separate study, which is outside the scope of this preliminary work.

- (4) Eesti-Läti sealihatüli oli peaminister Mart Siimanni ja tema Läti kolleegi, peaminister Vilis Krishtopansi eilse kohtumise üks põhiteemasid. (ERC, Postimees, 8.1.1999)

‘The Estonian-Latvian dispute over pork trade was one of the main subjects of yesterday’s meeting between prime minister Mart Siimann and his Latvian colleague Vilis Krishtopans.’

Despite the fact that event mentions such as *promised*, *to homogenise*, and *dispute* are not common between sentences (3) and (4), we considered this pair to be correct because the *meeting* (of two prime ministers) is mentioned in both sentences.

The cases of uncertainty on deciding event coreference, which in our case also included the partial coreference discussed by Hovy et al. (2013), were counted separately from correct/incorrect cases.

Results of manual evaluation of the simple lexical similarity method (method 1) are listed in Table 2. As expected, the method obtains high precision on lexically homogeneous sets of articles (Jüri Mosin, Pekka Vennamo), and obtains more problematic results on the lexically most heterogeneous set of articles (Mart Siimann). However, the total number of found pairs is too small to draw any strong conclusions.

Table 2. Results of similarity method 1 (Jaccard Similarity Coefficient for measuring lemma overlap between two sentences) on finding sentence pairs referring to the same event

Person	Correct pairs	Incorrect pairs	Pairs with uncertain coreference	Precision
Jüri Mosin	4	0	0	100%
Pekka Vennamo	12	0	0	100%
Mart Siimann	4	0	2	66.6%
Total	20	0	2	90.9%

The results of manual evaluation of method 2 (overlap of calendric values of temporal expressions and lemmas of location expression) are listed in Table 3. Similarly to method 1, this method returns high-precision results on the lexically most homogeneous set of articles (Jüri Mosin), but it does not outperform method 1 on the lexically more heterogeneous sets of articles (Pekka Vennamo, Mart Siimann) as was expected. However, it is notable that method 2 makes rather different kinds of mistakes on these datasets – on the Pekka Vennamo dataset it returns more pairs with uncertain coreference, while in the Mart Siimann dataset it returns more pairs with incorrect coreference.

Table 3. Results of similarity method 2 (Jaccard Similarity Coefficient for measuring overlap of temporal and locational expressions of two sentences) on finding sentence pairs referring to the same event

Person	Correct pairs	Incorrect pairs	Pairs with uncertain coreference	Precision
Jüri Mosin	3	0	0	100%
Pekka Vennamo	6	2	5	46.1%
Mart Siimann	11	10	2	47.8%
Total	20	12	7	51.2%

We also examined the degree of overlap between the results returned by the two methods. It turned out that only 2 sentence pairs (out of 61 pairs returned by the two methods together) were overlapping, which shows the potential for these two methods to complement each other.

7. Discussion and conclusion

The dataset used in our evaluation is too small to draw any strong conclusions, but the general impression is that the results of fine-grained event coreference detection are dependent on how focused the articles are on a narrow set of events. This narrow set of events is difficult to describe formally, but it seems to manifest itself in a way that it is possible to name central/main events covered by the media in a given time period: in the case of Jüri Mosin, the main event was the trial, and in the case of Pekka Vennamo, the main event was the scandal. On the other hand, it is difficult to name one or even a few central events in the set of articles mentioning Mart Siimann, as he was mentioned in relation to a rather diverse set of events.

Naturally, if the set of articles is focused on a narrow set of events, there is less room for making mistakes, and even a minimal lexical similarity can indicate that both sentences are likely mentioning the same events. And if the set of articles is not focused, but describes a large set of events, it is likely that even more sophisticated coreference detection methods (compared to the simple lexical similarity) will be fallible, as there can be misleading similarities between event descriptions actually referring to different events.

We hypothesize that the media coverage of the events related to the person plays an important, albeit a latent role in narrowing down the set of events. Our dataset indicates that if the person is only mentioned in a few articles, the stories are likely built around a narrow set of events, while frequent mentions more likely indicate a large set of events. As discussed by Johan Galtung and Mari Holmboe Ruge (1965), various factors can explain these tendencies. One can note that events related to elite persons are reported more often, and events related to non-elite persons are reported infrequently and more likely if they are negative (factors that both contribute to the reports being more focused). Also, if the person has high media coverage, she/he is more likely mentioned in opinions and discussions, which are less concerned with reporting factual details about events, while low media coverage could indicate that the reports are more factual in nature. Considering this, we can speculate that it is easier to find coreferring event mentions if a person has low media coverage, as there is less room for mistakes (only events reported inside the fixed time period are likely the coreferring ones) and potentially more factual information.

Returning to our dataset and to the problem of whether the article set is focused on a narrow set of events, high lexical similarity seems to be one indicator of a narrow set of events. The sets of articles mentioning Jüri Mosin and Pekka Vennamo were both characterized by relatively high lexical similarity, compared to the set of articles mentioning Mart Siimann (Table 1). However, this indicator is affected by the length of the timeframe chosen for analysis, and one cannot rule out the possibility that two articles having relatively low lexical similarity are still discussing the same events, using different vocabularies.

In terms of fine-grained event coreference detection and evaluation, our results indicate that the simple lexical similarity method (lemma overlap between sentences) seems to be the most stable method for producing high-precision results. However, it is also the easiest method to judge: high lexical overlap between words leaves little room for doubt on whether two sentences refer to the same events or not. When we loosen the lexical similarity constraints and allow coreferring events to be detected only by words indicating the spatial and temporal context of events (the method 2), the task of judging event coreference becomes more difficult, and there will be more cases of uncertainty. The uncertainty indicates that there is some structural similarity between events mentioned in two sentences, so we cannot decide for sure that no coreference exists. Consider an example of an uncertain match from the set of articles mentioning Pekka Vennamo (returned by method 2):

- (5) Vennamo ise ütles eile Soome telekanalile MTV 3, et ei teinud midagi valesti ega ole andnud ka valeinfot Sonera aktsiatega sooritatud tehingute kohta. (ERC, Eesti Päevaleht 5.1.1999)
'Vennamo himself told the Finnish TV channel MTV 3 yesterday that he had not done anything wrong and he had not given any wrong information regarding the transactions with Sonera's shares.'
- (6) Varem Vennamot toetanud Soome sideminister Matti Aura loobus skandaali tõttu teda usaldamast ja astus eile ka ise tagasi. (ERC, Postimees 5.1.1999)
'A former supporter of Vennamo, Finnish Minister of Communications Matti Aura, stopped trusting Vennamo because of the scandal and also resigned yesterday from his position.'

The main event mentions of the sentences (i.e. 'told', 'had not done' and 'had not given' in sentence (5), and 'stopped' and 'resigned' in sentence (6)) are referring to events performed by different persons and thus cannot corefer. However, the event mention 'scandal' in sentence (6) refers to a rather general event (the event which was also the main reason why both mentioned individuals ended up in the spotlight of Estonian media in the given week), and arguably one can consider the events 'told', 'stopped trusting' and 'resignation' all as subevents of the event 'scandal', thus indicating a partial coreference according to Hovy et al. (2013).

Note that the uncertainty introduced when we loosen the lexical similarity constraints (moving from method 1 to method 2) can also be used as an indicator of the degree to which the set of articles is focused on a narrow set of events. In the set of articles mentioning Pekka Vennamo, switching from method 1 to method 2 introduces more uncertain pairs than incorrect pairs, while in the set of articles mentioning Mart Siimann, the switching introduces more incorrect pairs than uncertain pairs. If there are more uncertain than incorrect pairs, this indicates that the set of events is probably narrow, but the descriptions are revealing the structure of events (e.g. subevents and member events of the focused events are being described), so it is difficult to decide whether we have a coreference or not. However, if loosening the lexical similarity constraints introduces more incorrect than uncertain pairs, the set of articles is likely not focused enough on a narrow set of events.

In conclusion, our preliminary results confirm the findings of previous works that relatively high lexical similarity within a (sub)set of articles can provide an important

indicator that the set of articles is focused on a narrow set of events, and high-precision event coreference detection can be performed on the dataset. However, our main contribution is to outline the importance of media coverage of the events (related to the person searched for) as an underlying mechanism influencing the results. An interesting question is whether we can somehow formalize this notion and its characteristics influencing how factually (or non-factually) the events are discussed in media – which events or persons tend to evoke more factual reports. To some extent, the presence of the factors influencing the media coverage discussed by Galtung and Ruge (1965) can be detected, e.g. one can determine the negativity of the text by using automatic sentiment analysis.

References

- Allan, James; Carbonell, Jaime G.; Doddington, George; Yamron, Jonathan; Yang, Yiming 1998. Topic detection and tracking pilot study final report. – Proceedings of the DARPA Broadcast News Transcription and Understanding Workshop, February 8–11, 1998. <http://www.itl.nist.gov/iad/mig/publications/proceedings/darpa98/pdf/tdt2040.pdf> (28.2.2015).
- Baker, Collin F.; Fillmore, Charles J.; Lowe, John B. 1998. The Berkeley Framenet Project. – Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics, Vol. 1, ACL '98. Stroudsburg, PA: Association for Computational Linguistics, 86–90. <http://dx.doi.org/10.3115/980845.980860>
- Cybulska, Agata; Vossen, Piek 2013. Semantic relations between events and their time, locations and participants for event coreference resolution. – Recent Advances in Natural Language Processing. Proceedings, 156–163.
- Cybulska, Agata; Vossen, Piek 2014. Using a sledgehammer to crack a nut? Lexical diversity and event coreference resolution. – Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14), 4545–4552.
- Davidson, Donald 1969. The individuation of events. – N. Rescher et al. (Eds.). Essays in Honor of Carl G. Hempel. Synthese Library 24. Dordrecht: Reidel, 216–234. Reprinted in D. Davidson, Essays on Actions and Events. Oxford: Clarendon Press, 2001, 163–180. <http://dx.doi.org/10.1093/0199246270.003.0008>
- Davidson, Donald 1985. Reply to Quine on events. – E. LePore, B. McLaughlin (Eds.). Actions and Events: Perspectives on the Philosophy of Donald Davidson. Oxford–New York: Basil Blackwell, 172–176.
- Fokkens, Antske; Braake, Serge ter; Ockeloen, Niels; Vossen, Piek; Legêne, Susan; Schreiber, Guus 2014. BiographyNet: Methodological issues when NLP supports historical research. – Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14), 3728–3735.
- Galtung, Johan; Ruge, Mari Holmboe 1965. The structure of foreign news the presentation of the Congo, Cuba and Cyprus Crises in four Norwegian newspapers. – Journal of Peace Research, 2 (1), 64–90. <http://dx.doi.org/10.1177/002234336500200104>
- Glavaš, Goran; Šnajder, Jan 2013. Exploring coreference uncertainty of generically extracted event mentions. – Alexander Gelbukh (Ed.). Computational Linguistics and Intelligent Text Processing. 14th International Conference, CICLing 2013, Samos, Greece, March 24–30, 2013, Proceedings, Part 1. Berlin–Heidelberg: Springer, 408–422. http://dx.doi.org/10.1007/978-3-642-37247-6_33
- Hovy, Eduard; Mitamura, Teruko; Verdejo, Felisa; Araki, Jun; Philpot, Andrew 2013. Events are not simple: Identity, non-identity, and quasi-identity. – Proceedings of the

- 1st Workshop on EVENTS: Definition, Detection, Coreference, and Representation, Atlanta, Georgia, 14 June 2013. Association for Computational Linguistics, 21–28.
- Jaccard, Paul 1901. Etude comparative de la distribution florale dans une portion des Alpes et du Jura. – *Bulletin de la Société Vaudoise des Sciences Naturelles*, 37, 547–579.
- Kaalep, Heiki-Jaan; Muishchnek, Kadri; Uihoaed, Kristel; Veski, Kaarel 2010. The Estonian Reference Corpus: Its composition and morphology-aware user interface. – I. Skadina, A. Vasiljevs (Ed.). *Human Language Technologies – The Baltic Perspective. Proceedings of the Fourth International Conference Baltic HLT. Frontiers in Artificial Intelligence and Applications* 219. IOS Press, 143–146.
- Manning, Christopher D.; Raghavan, Prabhakar; Schütze, Hinrich 2008. *Introduction to Information Retrieval*. Vol. 1. Cambridge: Cambridge University Press. <http://dx.doi.org/10.1017/CBO9780511809071>
- Naughton, Martina 2009. *Sentence Level Event Detection and Coreference Resolution*. PhD dissertation. Dublin: National University of Ireland.
- Orasmaa, Siim 2012. *Automaatne ajaväljendite tuvastamine eestikeelsetes tekstides*. [Automatic recognition and normalization of temporal expressions in Estonian language texts.] – *Eesti Rakenduslingvistika Ühingu aastaraamat*, 8, 153–169. <http://dx.doi.org/10.5128/ERYa8.10>
- Pecina, Pavel 2010. Lexical association measures and collocation extraction. – *Language Resources and Evaluation*, 44 (1–2), 137–158. <http://dx.doi.org/10.1007/s10579-009-9101-4>
- Pustejovsky, James; Castaño, José M.; Ingria, Robert; Sauri, Roser; Gaizauskas, Robert J.; Setzer, Andrea; Katz, Graham; Radev, Dragomir R. 2003. TimeML: Robust specification of event and temporal expressions in text. – *New Directions in Question Answering*, 3, 28–34.
- Quine, Willard Van Orman 1985. Events and reification. – E. LePore, B. P. McLaughlin (Eds.). *Actions and Events: Perspectives on the Philosophy of Donald Davidson*. Oxford–New York: Basil Blackwell, 162–171.
- Schneider, Susan 2005. *Internet Encyclopedia of Philosophy: Events*. <http://www.iep.utm.edu/events/> (7.9.2014).
- Sprugnoli, Rachele; Lenci, Alessandro 2014. Crowdsourcing for the identification of event nominals: An experiment. – *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, 1949–1955.
- TimeML Working Group 2007. *Semantic Annotation: A TimeML Case Study*. http://semanticweb.kaist.ac.kr/research/tc37sc4/new_doc/ISO_TC37_SC4_N337_WG2_ISO-TimeML_Tilburg2007.pdf (9.2.2015).
- Vossen, Piek; Rigau, German; Serafini, Luciano; Stouten, Pim; Irving, Francis; Hage, Willem Van 2014. NewsReader: Recording history from daily news streams. – *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, 2000–2007.

Web references

- ACE. <https://www ldc.upenn.edu/collaborations/past-projects/ace> (4.9.2014).
- Estonian Reference Corpus. <http://www.cl.ut.ee/korpused/segakorpus/> (7.2.2015).

Siim Orasmaa (University of Tartu). Research interests: topics related to information extraction, with the specific focus on extraction of event and temporal information from natural language texts.
 J. Liivi Str 2, 50409 Tartu, Estonia
siim.orasmaa@ut.ee

SÜNDMUST VÄLJENDAVATE KEELENDITE SAMAVIITELISUSE TUVASTAMINE EESTI KEELE UUDISTEKSTIDES: ESIALGSED KATSETUSED

Siim Orasmaa

Tartu Ülikool

Artiklis uuritakse, kuidas automaatselt tuvastada samal ajaperioodil ilmunud ajaleheartiklitest samadele sündmustele viitavaid lauseid. Uurimisküsimusele keskendutakse eeskätt infootsingute kontekstis, kus samadele sündmustele viitavad laused peaksid kasutajale pakkuma esmast ülevaadet sündmuste meediakajastusest (st võimaldama erinevate viitamiskontekstide ja allikate kõrvutamist ning faktilise info esmast kontrollimist).

Antakse ülevaade lähenemisviisidest, mida senised tööd on uudistekstide sündmusanalüüsil kasutanud: eristatakse sündmusanalüüsi dokumendi, lause ja sõna/fraasi tasemel. Seejärel tutvustatakse probleemi teoreetilisi aluseid (osalise ja täieliku samaviitelisuse eristamist) ning tuuakse välja võimalikud lähte-eeldused (artiklitevaheline samaviitelisuse tuvastamine on lihtsam kui artiklisisene; kõrge artiklitevaheline leksikaalne sarnasus võimaldab ennustada sündmuste kattuvust), millele tuginedes võib katsetada ka lihtsamaid (peamiselt leksikaalsele sarnasusele tuginevaid) meetodeid samaviitelisuse tuvastamiseks.

Vaadeldakse infootsingu ülesannet, kus kolme päevalehe ühe nädala artiklite koguhulgast tuleb leida ühe konkreetse isikuga seotud sündmuse. Samale sündmusele viitavaid lauseid otsiti kõigest artiklitest, kus isikut mainiti nimepidi. Katsetati kahte Jaccard'i koefitsiendil põhinevat meetodit samaviitelisuse tuvastamiseks: lause sõnalemmade kokkulangevusel põhinev meetod ning aja- ja kohaväljendite kokkulangevusel põhinev meetod. Eksperimendis valiti välja kolm isikut, kellega seotud artiklite hulki iseloomustab erinev leksikaalse sarnasuse tase, ning uuriti meetodite täpsust nendel artiklihulkadel. Esialgsed tulemused kinnitasid varasemaid tähelepanekuid, et suure artiklitevahelise leksikaalse sarnasuse kontekstis võivad ka lihtsad meetodid anda kõrge täpsusega tulemusi. Samas paistab oluliseks tulemust mõjutavaks teguriks olevat just isikuga seotud meediakajastuse iseloom: kitsale sündmuste hulgale fokuseeritud kajastuste korral on ülesanne ilmselt lihtsam kui juhtudel, mil isikut mainitakse paljude erinevate sündmuste kontekstis. Mil määral meediakajastuse iseloomu saab formaliseerida ning rakendada sündmust väljendavate keelendite samaviitelisuse tuvastamisel (tuvastamise kvaliteedi ennustamisel) ning kuivõrd siinsete esialgsete katsete tulemused leiavad kinnitust ka suuremate andmehulkade analüüsil, jääb aga uurimiseks edaspidi.

Võtmesõnad: automaatne keeletöötlus, tekstianalüüs, eesti keel

USING MICRO-CONTEXTS TO DESCRIBE A WRITING PROCESS IN ESTONIAN AS A SECOND LANGUAGE ACROSS PROFICIENCY LEVELS

Olga Pastuhhova

Abstract. This paper aims to describe the writing process of native Russian-speaking students in Estonian as a second language. 34 participants were given the assignment of writing a text in the L2. The written texts were then rated as being at levels from A2 to C1 according to the Common European Framework of Reference for Languages (CEFR). The data were collected by computer keystroke logging and analysed based on the concept of a ‘micro-context’. Micro-contexts were analysed according to their frequency and duration and were compared across proficiency levels. The results show that writing in the L2 is not a smooth process. The longest transitions in micro-contexts reveal that the most cognitive effort is made between paragraphs and sentences and when deletions are involved. The growing number of consecutive deletions demonstrates that even with developing proficiency, the linear production text is subject to constant revision, correction and modification.

Keywords: keystroke, ScriptLog, Estonian, Russian

1. Introduction

Computer keystroke logging (Sullivan, Lindgren 2006, Van Waes et al. 2009, Van Waes et al. 2011, Leijten, Van Waes 2013) is a data collection method for recording writing in real time. It helps to reveal the cognitive processes underlying written text production. According to John R. Hayes (2012), the writing model is represented at three levels: the control level (e.g., motivation and goal setting), the process level (writing processes and task environment), and the resource level (e.g., attention, working memory). Writing processes at the process level include a proposer, translator, transcriber, and evaluator component. Mariëlle Leijten and Luuk Van Waes (2013) situate the methodological contribution of keystroke logging mainly

at the process level of Hayes' latest writing model (2012). Keystroke logging as a method is mainly aimed at observing translation (or formulation) and transcription processes, but is not limited to the process level. Since logging data provide information on pausing and revision behaviour, the method is also inferred to the resource level (Leijten, Van Waes 2013: 362).

Keystroke logging has been used in researching writing in the L1 and L2, in translation studies (Jakobsen 2006, Dimitrova 2006), as well as in researching writers with a language impairment (Asker-Árnason et al. 2010). Guido Nottbusch (2010) divides keystroke logging research into single word studies and typewritten composition. Evgeny Chukharev-Hudilainen (2014) links two strands of keystroke logging research to analyse keystroke patterns in spontaneous computer-mediated communication. In L1 and L2 writing research, keystroke logging has been used to investigate, for example, revision (Lindgren 2005, Stevenson et al. 2006), pauses (Spelman Miller 2000, 2006a, 2006b) and fluency (Uppstad, Solheim 2007, Palviainen et al. 2012).

The present study combines writing process and second language acquisition research. Namely, the writing process of native Russian-speaking students in Estonian as an L2 is under investigation. Writing is an important skill of second language proficiency. Previous research in the field of acquisition of Estonian as an L2 has largely focused on analysing the final product, while internationally it is the writing process itself that has become a subject to be explored. The shift to writing process research in L2 Estonian was noted by Olga Pastuhhova (2011) in an article which introduced computer keystroke logging as a relevant method for investigating the writing process, previous research based on the method and a case study of the writing process of a native Russian-speaking learner of Estonian.

The participants of the present study are Russian-speaking students at Tallinn University. The computer program ScriptLog was used to collect the data. The program keeps a detailed record of activity on the keyboard during a writing assignment (Strömquist, Karlsson 2002, Strömquist et al. 2006, Johansson 2009). The analysis is based on the concept of a micro-context, defined as 'the context around the certain transition between two keystrokes' (Wengelin 2006: 107). Although each transition between two keystrokes includes a short period of inactivity, it is not reasonable to interpret each transition as a pause. 'A working definition is that a pause is a transition time between two keystrokes which is longer than what can be expected to be necessary for the time needed to merely find the next key' (Wengelin 2006: 111). The criterion of two seconds in the current study is based on previous research that described inactivity between two keystrokes that was longer than two seconds as a pause (Wengelin 2006).

The present study aims to examine what an analysis of micro-contexts can tell us about text translation (or formulation) and transcription processes in terms of Hayes' latest writing model (2012) in general, and about L2 writing proficiency in particular. The study concerns L1 Russian writers of Estonian as an L2. Micro-contexts are analysed based on the frequency of their occurrence and the duration of inactivity between keystrokes. Since text production is investigated in L2 writing, the study aims to compare the findings across proficiency levels, from A2 to C1 according to the Common European Framework of Reference for Languages: Learning, Teaching Assessment (CEFR), to point out similarities and differences

in text production at the different levels. The data enable us to describe a writing process with the following components: text composition; text length; syntactic complexity; and deletions/mouse events (see section 3). The keywords of the data analysis in the present study are micro-contexts, frequency of their occurrence, duration, and proficiency levels.

2. Previous research

Micro-context is a key concept used by Åsa Wengelin (2006) in analysing the distribution of pauses in Swedish online text-writing by university students with no reading and writing difficulties and adults with reading and writing difficulties (dyslexia and deafness). The method was also applied to detect developmental change in the allocation of cognitive effort to different levels of text-writing by native speakers of Swedish across four age-specific groups: 4th, 7th and 10th graders, and university students. Their results indicate that at an early phase of text-writing development, more cognitive resources are allocated to low-level processes, such as spelling and punctuation. Later the focus is shifted to higher-level activities, such as the content, structure, and the text style (Wengelin, Strömquist 2004: 189–190). Pause frequency was observed in six micro-contexts: before word, within word, after word, before sentence, before delimiter¹, and after sentence. The pause frequency profiles of the four age-specific groups, the university students with no reading and writing difficulties, and the adults with reading and writing difficulties were found to be surprisingly similar (Wengelin 2006: 119–120). The 4th graders and the 7th graders differed from the 10th graders and university students in two main aspects only: the 4th graders made many more pauses before words in comparison with the others and both the 4th and the 7th graders made more pauses before major delimiters than before a sentence or after a delimiter. The pause frequency of the adults with reading and writing difficulties resembled a combination of the 4th and 7th graders. The main findings of pause behaviour of all groups of subjects were found to be as follows: ‘micro-contexts around sentence boundaries are more predictive of pauses than micro-contexts around words’, ‘micro-contexts before words are generally more predictive of pauses than micro-contexts within and after words’, ‘micro-contexts before sentences are more predictive of pauses than micro-contexts after sentences’, ‘micro-contexts just before a major delimiter change systematically with age or perhaps writing development’ (Wengelin 2006: 120–121).

Kristyan Spelman Miller (2006b: 141–142) points out that pausological research on writing has mainly focused on native speakers, although recently the situation has changed to include L2 writers as well. The author investigated the pausing of L1 and advanced L2 writers of English and found differences in pause duration and pause frequency between the groups. Pauses are longer and more frequent for L2 writers, especially at lower text (often spelling and morphological) unit levels.

The present study involves an analysis of micro-contexts in L2 writing by L1 Russian learners of Estonian.

¹ Major delimiters refer to a full stop, a question mark or an exclamation mark. Minor delimiters refer to a comma, a colon, a semicolon, a dash or a pair of brackets (Wengelin 2006: 114).

3. Data and method

The data of the study consist of texts written by 34 participants whose L1 is Russian and for whom Estonian is an L2. For the most part (31 out of the 34), the participants were bachelor students at Tallinn University; the other three were studying at the master's level. All participants attended an Estonian language course in different groups according to their proficiency level. Being students, the participants were regularly made to write, whether in the form of essays, examinations or assignments for language courses. The students were asked to participate in a study concerning the writing process in Estonian as an L2. 34 participants agreed to take part in the study and the author met everyone individually. All participants were asked to write a text on a topic related to education for the purpose of the study. The topic was given on the basis of participants' expected proficiency level. Since text type is in some sense a property of proficiency level, more advanced learners wrote an argumentative text on the topic 'Is higher education a requirement for everyone?', while less advanced learners wrote a narrative text on the topic 'What I have gained from university'. All texts were written on a computer using the program ScriptLog. The participants were not allowed to use dictionaries or Internet materials. All were given one hour to express their thoughts on the topic and finish the assignment, but no time pressure was established. Upon completion of the task, the students were asked to comment on their writing in general as well as the task specifically. These interviews were taped. Retrospective data remain outside the scope of the present study.

The written texts were rated by two experts on the basis of the illustrative scales provided for overall written production, creative writing, and reports and essays of the Common European Framework of Reference for Languages (CEFR: 61–62). The data were representative of four language proficiency levels: 4 texts corresponded to A2 (all texts narrative), 13 texts to B1 (12 texts narrative and 1 argumentative), 13 texts to B2 (12 texts argumentative and 1 narrative), and 4 texts to C1 (all texts argumentative).

The data were analysed based on the micro-contexts registered by ScriptLog. The micro-contexts allowed for the description of a text production process according to the following criteria: text composition; text length; syntactic complexity; and deletions/mouse events. Notations and descriptions of micro-contexts recordable by ScriptLog and used in the study are found in Appendix 1.

What is meant by **text composition** is the dividing of the text into paragraphs. It is tracked in terms of two micro-contexts: transition to a new paragraph ('.^c') and starting a new paragraph ('c^a'). As regards **text length**, ScriptLog provides a statistics output file stating the number of tokens in the linear as well as the final text. Although this is an easy and straightforward way to calculate the length, the data provided with micro-contexts will be discussed here (transitions inside words ('a^a'), between words ('a_+^a', 'a^_+a'), and between sentences ('a^.', '^_+a', '^_+^a')).

What is meant by **syntactic complexity** is the use of minor delimiters ('a^', '^_+a', '^_+^a'). **Deletions/mouse events** are registered in a number of different surroundings and they give an overview of deletion behaviour during text production: 'D^D' represents consecutive deletions; 'a^D', 'D^a' word level deletions;

‘a^_D’ the cases where a typer taps the spacebar upon completion of the previous word with the intention to continue producing the text but ends up with a deletion instead; ‘_^D’, ‘D^_’ deletions between words; and ‘D_+^a’ transition to continuing constructing the text after the previous text has been edited. It is worth noting that ScriptLog denotes both deletions and mouse events with the same symbol. For this reason the discussion of these micro-contexts concerns deletions, mouse events or both.

4. Results

The findings of the study are presented in the following order: first, the analysis of micro-contexts based on the frequency of their occurrence; second, the analysis of micro-contexts based on the duration of inactivity between the keystrokes. The frequency and duration of the micro-contexts registered are compared by proficiency level. Finally, a general overview of the text production process across the proficiency levels and potential pause locations is discussed. Since the number of participants varied per level and not all micro-contexts were present in every text, the mean numbers of micro-contexts were used and the mean of the mean transition time in every micro-context is analysed per text.

4.1. Micro-contexts by frequency

Table 1 below presents the most frequent micro-contexts across the proficiency levels.

Table 1. Micro-contexts with mean frequency of occurrence exceeding 10 times per text across the proficiency levels

A2 (N = 4)		B1 (N = 13)		B2 (N = 13)		C1 (N = 4)	
Micro-context	Mean N/text	Micro-context	Mean N/text	Micro-context	Mean N/text	Micro-context	Mean N/text
a^a	938	a^a	1132	a^a	1299	a^a	1214
a^_+a	180	D^D	333	D^D	340	D^D	391
a_+^a	180	a^_+a	212	a^_+a	220	a^_+a	208
D^D	156	a_+^a	212	a_+^a	220	a_+^a	208
D^a	43	D^a	56	D^a	69	D^a	59
a^D	41	a^D	52	a^D	62	a^D	56
a^.	27	_ ^D	23	,^_+a	31	,^_+a	25
a^,	22	a^,	22	,_+^a	31	,_+^a	25
,^_+a	26	,^_+a	20	_ ^D	27	a^,	25
,_+^a	26	,_+^a	20	a^,	25	_ ^D	18
_ ^D	16	a^.	19	a^_+D	20	a^_+D	14
		a^_+D	17	a^.	17	a^.	14
		D^_	13	_ ^_	22	D^_	13
				D^_	15	D_+^a	12
				D_+^a	11		

The most frequent micro-context across all proficiency levels reflects continuing to construct a word. However, the second most frequent micro-context at B1, B2 and C1 levels demonstrates that text production is not a smooth process and is characterised by a high level of consecutive deletions or mouse movements. Their number exceeds that of transitions between words across all the proficiency levels except for A2 level. The fifth and sixth positions in the frequency order across all levels involve deletions inside a word. The next micro-context at A2 level is associated with finishing a sentence followed by use of minor delimiters. At B1, B2 and C1 levels the opposite is true – the use of minor delimiters per text is more frequent micro-contexts related to sentence finishing. Other micro-contexts with mean frequency of occurrence exceeding 10 times per text include, again, transitions in several surroundings with deletion involved. The full list of micro-contexts across all levels together with their total number, mean number per text and mean of the mean transition time per text is to be found in the appendices (Appendix 2, 4, 6, 8).

The first six micro-contexts demonstrate that text is constructed across proficiency levels similarly. It is no surprise that a text is formulated by constructing and linking words. A tendency towards deletions across proficiency levels shows a requirement for additional research to understand the nature of cognitive difficulties in L2 Estonian writing.

4.2. Micro-contexts by duration

Table 2 below presents micro-contexts across proficiency levels with the mean of the mean transition time exceeding two seconds per text.

Table 2. Micro-contexts with the mean of the mean transition time per text exceeding two seconds across the proficiency levels

A2 (N = 4)		B1 (N = 13)		B2 (N = 13)		C1 (N = 4)	
Micro-context	Mean-mean transition time per text (s)	Micro-context	Mean-mean transition time per text (s)	Micro-context	Mean-mean transition time per text (s)	Micro-context	Mean-mean transition time per text (s)
.^c	7.69	.^c	8.19	c^a	11.45	.^c	10.52
_ ^D	6.16	_ ^D	6.13	_ +^a	8.28	_ ^D	7.03
_ +^a	3.90	c^a	5.84	.^c	7.61	.^ +a	4.94
D_ +^a	3.40	_ +^a	3.55	_ ^D	6.87	c^a	4.77
a^D	3.39	a^D	3.15	a^.	3.25	_ +^a	3.77
D^a	2.54	a^c	2.70	_ ^ _	3.19	a^D	2.48
a_ +^a	2.34	D^a	2.54	a^D	3.13	D_ +^a	2.47
a^,	2.14	a^,	2.47	.^ +a	2.87		
a^.	2.06	a^.	2.31	D^ _	2.04		
		.^ +a	2.27				

Long periods of inactivity between keystrokes across all proficiency levels take place at the end of the paragraph. The mean transition to a new paragraph exceeds seven seconds at each level and the mean transition time is the longest at A2, B1

and C1 levels. Using solely keystroke logging without any additional data collection methods does not provide information on what is going on during these periods of inactivity. It can be assumed that the writer is either rereading what he or she has previously written or planning the following utterance.

The most time-consuming micro-context at B2 level is associated with starting a new paragraph. Likewise, B1 and C1 levels are marked by inactivity preceding the writing of a new paragraph, although the duration of that inactivity is only about half as long as at B2 level. Considering that both transition to a new paragraph and starting a new paragraph cause keystroke inactivity, the duration of inactivity between paragraphs is even longer (14.03 s at B1 level, 19.06 s at B2 level, 15.29 s at C1 level).

A mean period of inactivity of over six seconds across all proficiency levels is inherent to a transition from a space to a deletion. Considering that the preceding word has been finished and the writer has moved on to continue producing the text but ended up with a deletion, it can be assumed that in the given context the writer is rereading what he or she has previously written, revising or planning corrections.

Starting a new sentence is characterised by a mean period of inactivity exceeding three seconds across all proficiency levels and reaching more than eight seconds at B2 level. If we include the transition from a major delimiter at the end of the preceding sentence to the space followed by the first letter of the next sentence, the inter-sentence period of inactivity will be even longer (5.82 s at B1 level, 11.15 s at B2 level and 8.71 s at C1 level). Thus the longest micro-contexts at B2 level demonstrate that the biggest cognitive effort is made at the beginning of a new paragraph, followed by the beginning of a new sentence. Likewise, ending a sentence is characterised by a maximum mean period of inactivity specifically at B2 level.

All proficiency levels are characterised by periods of inactivity within a word between a letter and a delete/backstroke, although the duration of this inactivity decreases with each subsequent proficiency level. Conversely, the transition exceeds two seconds only at A2 and B1 levels, although at B2 level it is very close to the pause criterion.

The continuation of the text composition process after the previously written text has been edited is preceded by a mean period of inactivity of 3.40 s at A2 level and 2.47 s at C1 level. Unlike the more proficient levels, A2 level is also characterised by a mean period of inactivity exceeding two seconds between words.

To sum up, the longest transitions occur between paragraphs, sentences and in the micro-contexts which involve deletions. The full lists of micro-contexts with mean numbers of the mean transition time per text, the mean of the minimum and maximum transition time per text and the mean number of the given micro-contexts per text can be found in the appendices (Appendix 3, 5, 7, 9). Since not all texts included all the contexts, the tables also contain the percentage of texts in which various micro-contexts occurred.

4.3. Text production process across proficiency levels: similarities and differences

The text production process is analysed according to the text composition, text length, syntactic complexity, and deletions and/or mouse events data; this is followed by a discussion of potential pause locations.

Table 3 below summarises the mean numbers per text of micro-contexts and the mean of mean transition times across all proficiency levels describing the text production on the basis of the available data. All numbers represent the linear text production process and not the final text.

Table 3. Text production based on micro-contexts data across proficiency levels: similarities and differences

Micro-context	A2 (N = 4)		B1 (N = 13)		B2 (N = 13)		C1 (N = 4)	
	Mean N/ text	Mean-mean transition time per text (s)	Mean N/ text	Mean-mean transition time per text (s)	Mean N/ text	Mean-mean transition time per text (s)	Mean N/ text	Mean-mean transition time per text (s)
Text composition								
.^c	3	7.69	1	8.19	2	7.61	4	10.52
c^a	0	0.00	2	5.84	3	11.45	5	4.77
Text length								
a^a	938	0.49	1132	0.49	1299	0.34	1214	0.30
a_^a	180	2.34	212	1.58	220	1.46	208	1.47
a^_+a	180	0.55	212	0.50	220	0.48	208	0.50
a^.	27	2.06	19	2.31	17	3.25	14	1.13
._+^a	7	3.90	7	3.55	10	8.28	8	3.77
.^_+a	7	1.96	7	2.27	10	2.87	8	4.94
Syntactic complexity								
a^,	22	2.14	22	2.47	25	1.94	25	1.60
,^_+a	26	0.42	20	0.63	31	0.57	25	0.38
._+^a	26	1.34	20	1.08	31	0.88	25	1.44
Deletions/mouse events								
D^D	156	1.12	333	0.51	340	0.97	391	0.49
a^D	41	3.39	52	3.15	62	3.13	56	2.48
D^a	43	2.54	56	2.54	69	1.99	59	1.45
a^_+D	10	1.87	17	0.68	20	0.72	14	0.74
._^D	16	6.16	23	6.13	27	6.87	18	7.03
D^_	10	0.97	13	1.90	15	2.04	13	1.93
D_^+a	6	3.40	8	1.67	11	1.97	12	2.47

If we compare the number of paragraphs across the proficiency levels, a tendency towards growth is noted along with the improvement in proficiency level. However, the given numbers describe only those cases where a paragraph is made with a transition from an enter keystroke to a letter. Several participants divided their texts into

paragraphs by other means, for example by pressing several spacebars after the enter key before a letter. Checking the texts confirms that paragraphs are used also at A2 level, but these cases were not registered by ScriptLog with the given micro-context.

As regards text length, linear texts tend to become longer from A2 level to B2 level and then slightly decrease in length at C1 level. The mean number of sentences based on micro-contexts associated with closing a sentence decreases along with growing proficiency, being almost equal at B1 level and B2 level, but twice as big at A2 level (27) as at C1 level (14). It can be concluded that the mean number of sentences is bigger at A2 level, but the sentences are shorter. At C1 level, conversely, the texts are shorter in terms of the number of sentences, but the sentences are longer. The mean numbers of micro-contexts associated with beginning and closing a sentence differ substantially. One reason for this is that this process is subject to constant modification. Also, several participants did not use spaces between sentences, so these cases were left outside this micro-context.

In the case of syntactic complexity, minor delimiters are used on each level and at a similar frequency, even at levels at the extremes: A2 and C1.

The writing process across all proficiency levels is characterised by deletions and mouse movements. Consecutive deletions grow from A2 level to C1 level, with twice as many at C1 level as at A2 level. Even with increasing language proficiency, linear texts are subject to constant revision. Even deletions inside and between words gradually increase from A2 level to B2 level. Although C1 level is characterised by the maximum number of consecutive deletions and mouse events, B2 level is the level with the most correction and modification inside and between words.

Finally, let us compare the means of mean transition time per text in order to reveal which of the micro-contexts cause potential pause locations. The findings are in line with Åsa Wengelin (2006), Veerle M. Baaijen, David Galbraith and Kees de Glopper (2012), and Kristyan Spelman Miller (2000), confirming that higher level text units are subject to longer pauses and pause length increases for higher-level locations.

Micro-contexts between paragraphs cause the longest inactivity and are subject to extensive pauses across all proficiency levels. So too do micro-contexts before sentences across all levels. Closing a sentence is accompanied with pauses from A2 level to B2 level. Pauses are more often made before sentences than at the end. Micro-contexts before words require more transition time than within and after words. That said, their mean transition times generally do not exceed two seconds, except at A2 level before words.

Pauses are made before minor delimiters at A2 level and B1 level. Pauses are longer before minor delimiters across all levels than pauses before major delimiters, except at B2 level. Transitions to minor delimiters are longer across all proficiency levels than transitions right after minor delimiters.

In the case of deletions and mouse events, the most time-consuming pause locations across all proficiency levels are between words before beginning editing after the preceding word has been finished and the spacebar pressed. The time is required to look back to a previously written text to plan a modification. The second time-consuming pause location is inside words before a deletion across all proficiency levels, and after a deletion at A2 and B1 levels. Transitions to a new word after a deletion and a spacebar cause pauses at A2 and C1 levels. Consecutive deletions are not connected with pauses.

5. Conclusion

The study presented in this article aimed to examine what an analysis of micro-contexts can tell us about text production in the L2 Estonian of L1 Russian students across proficiency levels from A2 to C1.

The writing process across all levels is not a smooth process and a lot of deletions are made during the writing session. The longest transitions in micro-contexts reveal that the most cognitive effort is made between paragraphs and sentences and when deletions are involved.

The study shows that along with proficiency development, the number of paragraphs in a text increases. Linear texts tend to become longer from A2 to B2 level. However, the mean number of sentences decreases along with increasing proficiency. The use of minor delimiters across proficiency levels is quite similar.

The growing amount of consecutive deletions demonstrates that even with developing proficiency, linear production text is subject to constant revision, correction and modification. Deletions inside and between words gradually increase from the A2 level to the B2 level. Although the C1 level is characterised by a maximum of consecutive deletions and mouse events, B2 is the level with the most correction and modification inside and between words.

However, this study involved a number of limitations. The cases in which participants did not follow the writing conventions were not included in the study. For example, some participants wrote partly without spaces between sentences or made a space before a minor or major delimiter. Still, using micro-contexts to describe text production provides valuable data which cannot be collected in any other way.

The results of the study do not lend themselves to generalizations but rather show tendencies in how text is formulated and transcribed. The study highlights the need for future research on writing in Estonian as an L2. One possible direction for future research concerns the allocation of time during writing. Since writing is characterised by making a lot of corrections, text deletions should be looked at more closely. The third option could be combining keystroke logging and retrospective methods to investigate what is occurring during inactivity in writing and what exactly makes the writing process complicated across proficiency levels.

Symbols

- ^ transition from a given keystroke to a subsequent keystroke
- a character in a word (either upper or lower case) or a digit
- spacebar
- c enter keystroke
- D delete/backspace keystroke or a mouse event
- . major delimiter, i.e. a full stop, a question mark or an exclamation mark
- ,
- min
- minor delimiter, i.e. a comma, a colon, a semicolon, a dash or brackets
- + one or more of the same character as the preceding one

References

- Asker-Árnason, Lena; Ibertsson, Tina; Wass, Malin; Wegelin, Åsa; Sahlén, Birgitta 2010. Picture-elicited written narratives, process and product, in 18 children with cochlear implants. – *Communication Disorders Quarterly*, 31 (4), 195–212. <http://dx.doi.org/10.1177/1525740109337734>

- Baaijen, Veerle M.; Galbraith, David; Gloppe, Kees de 2012. Keystroke analysis: Reflections on procedures and measures. – *Written Communication*, 29 (3), 246–277. <http://dx.doi.org/10.1177/0741088312451108>
- CEFR = Common European Framework of Reference for Languages: Learning, Teaching, Assessment. http://www.coe.int/t/dg4/linguistic/Source/Framework_EN.pdf (9.2.2015).
- Chukharev-Hudilainen, Evgeny 2014. Pauses in spontaneous written communication: A keystroke logging study. – *Journal of Writing Research*, 6 (1), 61–84.
- Dimitrova, Birgitta Englund 2006. Segmentation of the writing process in translation: Experts versus novices. – Kirk P. H. Sullivan, Eva Lindgren (Eds.). *Computer Keystroke Logging and Writing: Methods and Applications*. Oxford: Elsevier, 189–201.
- Hayes, John R. 2012. Modeling and remodeling writing. – *Written Communication*, 29 (3), 369–388. <http://dx.doi.org/10.1177/0741088312451260>
- Jakobsen, Arnt Lykke 2006. Research methods in translation – *Translog*. – Kirk P. H. Sullivan, Eva Lindgren (Eds.). *Computer Keystroke Logging and Writing: Methods and Applications*. Oxford: Elsevier, 95–105.
- Johansson, Victoria 2009. *Developmental Aspects of Text Production in Writing and Speech*. Travaux de l'institut de linguistique de Lund 48. Lund: Lund University.
- Leijten, Mariëlle; Waes, Luuk van 2013. Keystroke logging in writing research: Using inputlog to analyze and visualize writing processes. – *Written Communication*, 30 (3), 358–392. <http://dx.doi.org/10.1177/0741088313491692>
- Lindgren, Eva 2005. *Writing and Revising. Didactic and Methodological Implications of Keystroke Writing*. *Skrifter från moderna språk* 18. Umeå: Umeå universitet.
- Nottbusch, Guido 2010. Grammatical planning, execution, and control in written sentence production. – *Reading and Writing*, 23 (7), 777–801. <http://dx.doi.org/10.1007/s11145-009-9188-4>
- Palviainen, Åsa; Kalaja, Paula; Mäntylä, Katja 2012. Development of L2 writing: Fluency and proficiency. – L. Meriläinen, L. Kolehmainen, T. Nieminen (Eds.). *AFinLA-e Soveltavan Kielitieteen Tutkimuksia*, 4, 47–59.
- Pastuhhova, Olga 2011. Kirjaliku produtseerimisprotsessi uurimise võimalused programmi *ScriptLog* abil: juhtumiuuring. [Research on the writing process using *ScriptLog*: A case study.] – *Lähivõrdlusi. Lähivertailuja* 21, 185–212. <http://dx.doi.org/10.5128/LV21.08>
- Spelman Miller, Kristyan 2000. Academic writers on-line: Investigating pausing in the production of text. – *Language Teaching Research*, 4 (2), 123–148. <http://dx.doi.org/10.1177/136216880000400203>
- Spelman Miller, Kristyan 2006a. The pausological study of written language production. – Kirk P. H. Sullivan, Eva Lindgren (Eds.). *Computer Keystroke Logging and Writing: Methods and Applications*. Oxford: Elsevier, 11–30.
- Spelman Miller, Kristyan 2006b. Pausing, productivity and the processing of topic in online writing. – Kirk P. H. Sullivan, Eva Lindgren (Eds.). *Computer Keystroke Logging and Writing: Methods and Applications*. Oxford: Elsevier, 131–156.
- Stevenson, Marie; Schoonen, Rob; Gloppe, Kees de 2006. Revising in two languages: A multi-dimensional comparison of online writing revisions in L1 and FL. – *Journal of Second Language Writing*, 15 (3), 201–233. <http://dx.doi.org/10.1016/j.jslw.2006.06.002>
- Strömqvist, Sven; Karlsson, Henrik 2002. *ScriptLog for Windows. User's Manual*. University of Lund: Department of Linguistics, University College of Stavanger: Centre for Reading Research.
- Strömqvist, Sven; Holmqvist, Kenneth; Johansson, Victoria; Karlsson, Henrik; Wengelin Åsa 2006. What keystroke logging can reveal about writing. – Kirk P. H. Sullivan, Eva Lindgren (Eds.). *Computer Keystroke Logging and Writing: Methods and Applications*. Oxford: Elsevier, 45–71.

- Sullivan, Kirk P. H.; Lindgren, Eva (Eds.) 2006. Computer Keystroke Logging and Writing: Methods and Applications. Studies in Writing 18. Oxford: Elsevier.
- Uppstad, Per Henning; Solheim, Oddny Judith 2007. Aspects of fluency in writing. – Journal of Psycholinguistics Research, 36 (2), 79–87. <http://dx.doi.org/10.1007/s10936-006-9034-7>
- Waes, Luuk van; Leijten, Mariëlle; Weijen, Daphne van 2009. Keystroke logging in writing research. Observing writing processes with Inputlog. – GFL Journal, 2–3, 41–64.
- Waes, Luuk van; Leijten, Mariëlle; Wengelin, Åsa; Lindgren, Eva 2011. Logging tools to study digital writing processes. – Virginia Wise Berninger (Ed.). Past, Present, and Future Contributions of Cognitive Writing Research to Cognitive Psychology. New York: Psychology Press, 507–534.
- Wengelin, Åsa 2006. Examining pauses in writing: Theory, methods and empirical data. – Kirk P. H. Sullivan, Eva Lindgren (Eds.). Computer Keystroke Logging and Writing: Methods and Applications. Oxford: Elsevier, 107–130.
- Wengelin, Åsa; Strömqvist, Sven 2004. Text-writing development viewed through on-line pausing in Swedish. – R. Berman (Ed.). Language Development Across Childhood and Adolescence. Trends in Language Acquisition Research 3. Philadelphia: John Benjamins, 177–190. <http://dx.doi.org/10.1075/tilar.3.12wen>

Olga Pastuhhova's (Tallinn University) main research interests are acquisition of Estonian as second language and writing process.
 Narva mnt 25, 10120 Tallinn, Estonia
olgap@tlu.ee

Appendix 1. Types of micro-contexts recordable by ScriptLog and used in the study (Strömqvist, Karlsson 2002, Wengelin 2006: 114)

Notation	Description
a^a	Character, TRANSITION, character Inactivity between two subsequent letters (within a word)
a_+^a	Character, space, TRANSITION, character Inactivity before a letter and after a spacebar preceded by a letter
a^_+a	Character, TRANSITION, space, character Inactivity after the last letter of a word and before a spacebar followed by a letter
_+^a	Major delimiter, space, TRANSITION, character Inactivity before a letter and after a major delimiter followed by a spacebar
a^.	Character, TRANSITION, major delimiter Inactivity after the last letter of a word and before a major delimiter
.^c	Major delimiter, TRANSITION, enter Inactivity after a major delimiter and before an enter keystroke
c^a	Enter, TRANSITION, character Inactivity after an enter keystroke and before a letter
,^_+a	Minor delimiter, TRANSITION, space, character Inactivity after a minor delimiter and before a spacebar followed by a letter
,_+^a	Minor delimiter, space, TRANSITION, character Inactivity before a letter and after a minor delimiter followed by a spacebar
.^_+a	Major delimiter, TRANSITION, space, character Inactivity after a major delimiter and before a spacebar followed by a letter
D^D	Delete/mouse event, TRANSITION, delete/mouse event Inactivity between two delete/backstrokes
D^_	Delete/mouse event, TRANSITION, space Inactivity between a delete/backstroke and a spacebar
D^a	Delete/mouse event, TRANSITION, character Inactivity between a delete/backstroke and a letter
D_+^a	Delete/mouse event, space, TRANSITION, character Inactivity before a letter and after a delete/backstroke followed by a spacebar
_+^D	Space, TRANSITION, delete/mouse event Inactivity between a spacebar and a delete/backstroke
a^D	Character, TRANSITION, delete/mouse event Inactivity between a letter and a delete/backstroke
a^_+D	Character, TRANSITION, space, delete/mouse event Inactivity after a letter and before a spacebar followed by a delete/backstroke
a^,	Character, TRANSITION, minor delimiter Inactivity after a letter and before a minor delimiter
+^	Space, TRANSITION, space Inactivity between two spacebars
a^c	Character, TRANSITION, enter Inactivity after a letter and before an enter keystroke

Appendix 2. Full list of micro-contexts at A2 level according to the frequency of occurrence

Context	Number	Mean N/text	Mean-mean transition time per text (s)
a^a	3752	938	0.49
a^_+a	720	180	0.55
a_+^a	720	180	2.34
D^D	625	156	1.12
D^a	173	43	2.54
a^D	162	41	3.39
a^.	108	27	2.06
a^,	86	22	2.14
,^_+a	77	26	0.42
,_+^a	77	26	1.34
_ ^D	65	16	6.16
a^_+D	39	10	1.87
D^_	38	10	0.97
D_+^a	23	6	3.40
.^_+a	22	7	1.96
._+^a	22	7	3.90
_ ^_	12	6	1.28
.^c	6	3	7.69

Appendix 3. Full list of micro-contexts at A2 level according to the duration

Micro-context	Mean of the mean transition time per text (s)	Mean of the min transition per text (s)	Mean of the max transition per text (s)	% of texts in which micro-context occurred	Mean N/ text
.^c	7.69	6.05	8.81	50	3
_ ^D	6.16	0.96	36.61	100	16
._+^a	3.90	1.29	9.14	75	7
D_+^a	3.40	0.79	12.94	100	6
a^D	3.39	0.52	35.78	100	41
D^a	2.54	0.32	17.20	100	43
a_+^a	2.34	0.18	40.23	100	180
a^,	2.14	0.46	7.62	100	22
a^.	2.06	0.52	10.24	100	27
.^_+a	1.96	1.78	2.50	75	7
a^_+D	1.87	0.15	5.96	100	10
,_+^a	1.34	0.28	7.94	75	26
_ ^_	1.28	1.21	1.34	50	6
D^D	1.12	0.11	24.51	100	156
D^_	0.97	0.57	2.01	100	10
a^_+a	0.55	0.12	13.26	100	180
a^a	0.49	0.09	14.91	100	938
,^_+a	0.42	0.23	0.89	75	26

Appendix 4. Full list of micro-contexts at B1 level according to the frequency of occurrence

Context	Number	Mean N/text	Mean-mean transition time per text (s)
a^a	14713	1132	0.49
D^D	4324	333	0.51
a^_+a	2752	212	0.50
a_+^a	2752	212	1.58
D^a	727	56	2.54
a^D	677	52	3.15
_ ^D	303	23	6.13
a^,	281	22	2.47
a^.	240	19	2.31
a^_+D	219	17	0.68
,^_+a	215	20	0.63
,_+^a	215	20	1.08
D^_	169	13	1.90
D_+^a	101	8	1.67
.^_+a	72	7	2.27
._+^a	72	7	3.55
_ ^_	48	6	0.57
c^a	13	2	5.84
.^c	11	1	8.19
a^c	2	1	2.70

Appendix 5. Full list of micro-contexts at B1 level according to the duration

Micro-context	Mean of the mean transition time per text (s)	Mean of the median transition time per text (s)	Mean of the min transition per text (s)	Mean of the max transition per text (s)	% of texts in which micro-context occurred	Mean N/text
.^c	8.19	7.69	7.55	9.34	69	1
_ ^D	6.13	2.95	0.64	31.72	100	23
c^a	5.84	5.01	3.08	8.41	54	2
._+^a	3.55	1.31	0.92	10.22	77	7
a^D	3.15	1.32	0.41	31.07	100	52
a^c	2.70	2.70	2.70	2.70	15	1
D^a	2.54	0.82	0.19	20.92	100	56
a^,	2.47	1.35	0.64	14.94	100	22
a^.	2.31	1.54	0.53	8.94	100	19
.^_+a	2.27	0.84	0.69	5.42	77	7
D^_	1.90	0.67	0.29	10.86	100	13
D_+^a	1.67	0.49	0.30	6.44	100	8
a_+^a	1.58	0.69	0.16	24.74	100	212
,_+^a	1.08	0.56	0.22	7.79	85	20
a^_+D	0.68	0.39	0.16	3.35	100	17
,^_+a	0.63	0.44	0.29	1.96	85	20
_ ^_	0.57	0.60	0.44	0.72	62	6
D^D	0.51	0.17	0.05	21.24	100	333
a^_+a	0.50	0.26	0.09	10.75	100	212
a^a	0.49	0.33	0.05	8.40	100	1132

Appendix 6. Full list of micro-contexts at B2 level according to the frequency of occurrence

Context	Number	Mean N/text	Mean-mean transition time per text (s)
a^a	16887	1299	0.34
D^D	4415	340	0.97
a^_+a	2857	220	0.48
a_+^a	2857	220	1.46
D^a	899	69	1.99
a^D	807	62	3.13
,^_+a	397	31	0.57
,_+^a	397	31	0.88
_ ^D	356	27	6.87
a^,	326	25	1.94
a^_+D	258	20	0.72
a^.	224	17	3.25
_ ^_	202	22	3.19
D^_	196	15	2.04
D_+^a	145	11	1.97
,^_+a	128	10	2.87
,_+^a	128	10	8.28
c^a	29	3	11.45
,^c	23	2	7.61
a^c	1	1	0.78

Appendix 7. Full list of micro-contexts at B2 level according to the duration

Micro-context	Mean of the mean transition time per text (s)	Mean of the median transition time per text (s)	Mean of the min transition per text (s)	Mean of the max transition per text (s)	% of texts in which micro-context occurred	Mean N/text
c^a	11.45	4.91	1.06	27.96	68%	3
,_+^a	8.28	4.79	0.93	27.85	100%	10
,^c	7.61	5.99	4.22	11.45	92%	2
_ ^D	6.87	2.29	0.47	43.69	100%	27
a^.	3.25	0.94	0.28	17.47	100%	17
_ ^_	3.19	1.34	1.30	5.77	69%	22
a^D	3.13	1.15	0.39	33.70	100%	62
,^_+a	2.87	0.91	0.37	11.66	100%	10
D^_	2.04	0.73	0.24	15.75	100%	15
D^a	1.99	0.75	0.17	29.42	100%	69
D_+^a	1.97	0.53	0.20	10.51	100%	11
a^,	1.94	0.95	0.27	10.23	100%	25
a_+^a	1.46	0.52	0.10	37.48	100%	220
D^D	0.97	0.16	0.03	44.01	100%	340
,_+^a	0.88	0.40	0.15	7.01	100%	31
a^c	0.78	0.78	0.78	0.78	8%	1
a^_+D	0.72	0.24	0.11	7.95	100%	20
,^_+a	0.57	0.28	0.15	4.85	100%	31
a^_+a	0.48	0.20	0.07	14.26	100%	220
a^a	0.34	0.24	0.05	7.84	100%	1299

Appendix 8. Full list of micro-contexts at C1 level according to the frequency of occurrence

Context	Number	Mean N/text	Mean-mean transition time per text (s)
a^a	4854	1214	0.30
D^D	1565	391	0.49
a^_+a	833	208	0.50
a_+^a	833	208	1.47
D^a	237	59	1.45
a^D	225	56	2.48
,^_+a	101	25	0.38
,_+^a	101	25	1.44
a^,	98	25	1.60
_ ^D	73	18	7.03
a^_+D	57	14	0.74
a^.	55	14	1.13
D^_	52	13	1.93
D_+^a	47	12	2.47
.^_+a	30	8	4.94
._+^a	30	8	3.77
c^a	15	5	4.77
.^c	8	4	10.52

Appendix 9. Full list of micro-contexts at C1 level according to the duration

Micro-context	Mean of the mean transition time per text (s)	Mean of the median transition time per text (s)	Mean of the min transition per text (s)	Mean of the max transition per text (s)	% of texts in which micro-context occurred	Mean N/text
.^c	10.52	10.02	4.10	17.58	50	4
_ ^D	7.03	1.83	0.27	48.22	100	18
.^_+a	4.94	3.39	0.22	11.69	100	8
c^a	4.77	2.91	0.72	12.31	75	5
._+^a	3.77	0.95	0.52	17.43	100	8
a^D	2.48	0.94	0.27	35.61	100	56
D_+^a	2.47	0.67	0.18	15.71	100	12
D^_	1.93	0.65	0.27	10.19	100	13
a^,	1.60	0.64	0.18	7.79	100	25
a_+^a	1.47	0.42	0.12	55.02	100	208
D^a	1.45	0.47	0.12	27.39	100	59
,_+^a	1.44	0.45	0.18	15.48	100	25
a^.	1.13	0.47	0.15	6.83	100	14
a^_+D	0.74	0.34	0.10	3.47	100	14
a^_+a	0.50	0.20	0.06	14.47	100	208
D^D	0.49	0.13	0.03	33.29	100	391
,^_+a	0.38	0.22	0.14	2.45	100	25
a^a	0.30	0.23	0.05	4.08	100	1214

MIKROKONTEKSTID EESTI KEELE KUI TEISE KEELE KIRJUTAMISPROTSESSI KIRJELDAMISEL KEELEOSKUSTASEMETI

Olga Pastuhhova

Tallinna Ülikool

Artikkel kirjeldab venekeelsete eesti keele õppijate kirjutamisprotsessi mikrokonteksti mõiste abil. Uurimismaterjal on kogutud klaviatuuri klahvivajutuste salvestuse meetodil põhineva programmiga ScriptLog. Uuringus osales 34 Tallinna Ülikooli venekeelset üliõpilast, kellel paluti kirjutada arvutis loovkirjutis haridusteemal. Loovkirjutisi hindasid kaks eksperti, kelle hinnangul vastas osalejate eesti keele oskus antud ülesande põhjal “Euroopa keeleõppe raamdokumendi” neljale keeleoskustasemele järgnevalt: 4 teksti vastasid tasemele A2, 13 teksti tasemele B1, 13 teksti tasemele B2 ja 4 teksti tasemele C1.

Mikrokonteksti defineeritakse kui üleminekut ühelt klahvilt teisele teatud ümbruses (Wengelin 2006). Artiklis esitatakse mikrokontekstide analüüs nende esinemissageduse ja ülemineku aja põhjal. Seejärel antakse üldine ülevaade kirjutamisprotsessist ja potentsiaalsetest peatumiskohtadest. Mikrokontekstide andmeid on esitatud ja võrreldud keeleoskustasemeti.

Tulemused näitavad, et kirjutamisprotsess ei ole sujuv ja seda iseloomustab suur hulk parandusi. Pikimad klahvidevahelised üleminekid esinevad lõikude vahel, mikrokontekstides, mis on seotud uue lause algusega ja kustutustega. Järjestikuste kustutamiste ja/või hiireliigutuste arvu kasv näitab, et kirjutamisprotsessi iseloomustab ka keeleoskuse kasvades teksti pidev redigeerimine, parandamine, muutmine. Kustutamised sõna sees ja sõnade vahel suurenevad tasemest A2 kuni tasemeni B2. Kuigi uurimismaterjali põhjal saab järeldada, et tasemel C1 toimub kõige rohkem järjestikuseid kustutamisi, iseloomustab suurim teksti parandamine sõna sees ja sõnade vahel taset B2.

Võtmesõnad: klahvivajutus, ScriptLog, eesti keel, vene keel

THE LINGUISTIC CONSTRUCTION OF THE GIFTEDNESS DISCOURSE IN THE MEDIA TEXTS OF HISTORICAL AND DIGITAL TIMES

Halliki Põlda

Abstract. The aim of the study is to describe and explicate, using critical text analysis, how the socially weighty discourse of giftedness has been constructed historically and how it manifests in the media texts of the digital era. The diachronic analysis is based on the media texts of the 1890s–1990s stored in the Corpus of Standard Estonian, while the synchronic analysis applies to texts found in Delfi.ee. The results highlight the main media discourses dealing with giftedness, the relevant terms and expressions, and the social relations and meanings brought up in the media in connection with the topic. The study reveals that through history, the giftedness discourse has been subject to changes and, constructed with specific linguistic means, it plays an important role in modern social arrangements.

Keywords: critical discourse analysis, ideology, social interaction, media language, Estonian

1. Background and problem statement

The topic of giftedness as handled in speech and writing, and the words thereby used, are becoming ever more topical not only for research but also for a wider public aspect. Throughout history, giftedness has been regarded as a key sustainability factor for countries, peoples and communities (Simonton 2009: 911, Ziegler 2009: 1510, Shavinina 2013: 55). In the modern globalized world, giftedness is more and more emphasized in the context of small nation sustainability (EA 2012, Mazzoli Smith 2014).

Historically, giftedness has mostly been understood as human intellectual ability (Kaufman, Sternberg 2008: 72–73, Simonton 2009: 906). Nowadays giftedness is

regarded as a system of personal characteristics and external factors, in which an important role is played by processes of human development (Feldman 2000, Gagné 2004). Some authors (Mudrak 2011: 200, Hymer 2014: 8) understand giftedness as a social construct; according to James Borland (2005: 3) it was not before the 1910s that the term acquired its modern meaning, i.e. not before a combination of socio-cultural and socio-political factors rendered the construct beneficial to society. Today, the topic of giftedness is often associated with creativity, ethical norms and the importance of self-regulation (Sternberg, Davidson 2005). At a recent ECHA (European Council for High Ability) conference, giftedness was defined as high ability in one or more areas of life. Giftedness is usually found in an individual whose personality traits include perfectionism, achievement motivation and task commitment. Also, giftedness is often associated with cognitive processes and socio-cultural environment factors such as participation in special talent groups, peer relations and teacher support. (ECHA 2014)

In the latest studies, the environment of the gifted is considered important (e.g. Gagné 2004). However, little attention is given to the media, which is, after all, an essential component of this environment. Mutual relations of the gifted and the media have been investigated by Radford (1998), who focuses on the British press coverage of prodigies. A brief study of the media discussion on giftedness and gifted children has been done in Finland (Laine 2010).

The focus of the present study is on how and with what lexical means the discourse of giftedness has been constructed in Estonian media texts of historical times and of the modern digital era and what social meanings are thereby brought up.

Research questions are:

- 1) What were the lexical means used to construct the historical discourse of giftedness and what social relations and meanings were brought up in that connection in the media texts of the time?
- 2) What are the lexical means used to construct the historical discourse of giftedness in the modern digital era and what social relations and meanings are brought up in that connection in modern media texts?

In the study, two text sets have been used to analyse the meanings constructed through the press coverage of giftedness: the diachronic view is based on the media texts drawn from the Corpus of Standard Estonian¹, while the synchronic view is based on the media texts found with the help of Delfi.ee², which is the oldest and biggest web portal in Estonia.

2. Material and method

The historical texts analysed come from the Corpus of Standard Estonian, which covers media texts from the 1890s–1990s.³ The search words used were the strings *ande* and *anne*⁴, which transparently refer to the phenomenon studied (*anne* 'talent', *andekus* 'giftedness'), and the strings *genius*, *geniaal*, *intelligent*, *talent*, which are contained in the most frequent linguistic expressions referring to the research phenomenon in giftedness studies. As a result, 175 texts touching on giftedness were

¹ Corpus of Standard Estonian 1890s–1990s, <http://www.cl.ut.ee/korpused/baaskorpus/> (9.9.2014).

² Delfi.ee, <http://otsing.delfi.ee/i.php?lang=et&p=n&q=andekus> (30.8.2014).

³ As the referenced corpus lacks texts of the 1920s and 1940s part of the research material was found with the help of the analytical bibliographic database of Estonian press, <http://www2.kirmus.ee/biblioserver/index2.php> (6.8.2014).

⁴ The word *anne* 'talent, gift' was only used in the context of giftedness, not in the sense of donation (*annetus*) or edition (*väljaanne*).

found (Figure 1). Subsequently, qualitative analysis was used on the list of 175 to dismiss those texts whose original source was unknown or where the context of the research phenomenon was unclear. As a result, 146 historical texts were selected for close investigation.

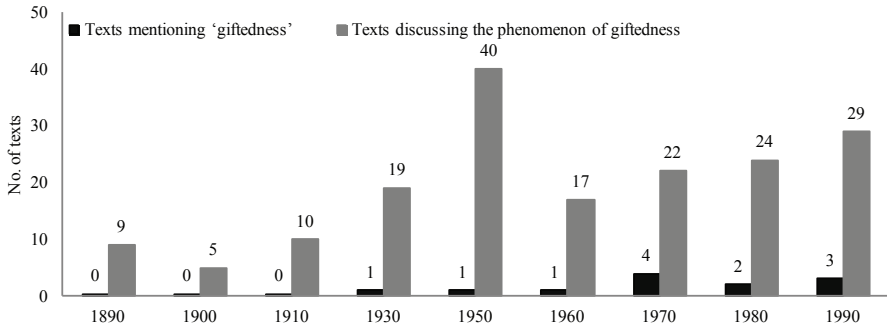


Figure 1. Historical texts (1890s–1990s) mentioning giftedness and discussing the phenomenon, stored in the Corpus of Standard Estonian

The modern digital era is understood as the current period since the beginning of the 21st century, characterized by an extensive advance of the Internet. The digital media texts analysed were found using the web portal Delfi.ee (Figure 2). One of the reasons behind the choice was that the portal offers the relevant texts for free. So it can be assumed that those articles will reach a wider audience than print media or online paid articles, thus making a major influence on public opinion, including the general ideas of giftedness (see McQuail 2003). Due to the great number of modern texts, synchronic analysis was confined to those directly referring to giftedness as a phenomenon, using the term *andekus* as reference. As a search word, the term gave 279 texts.

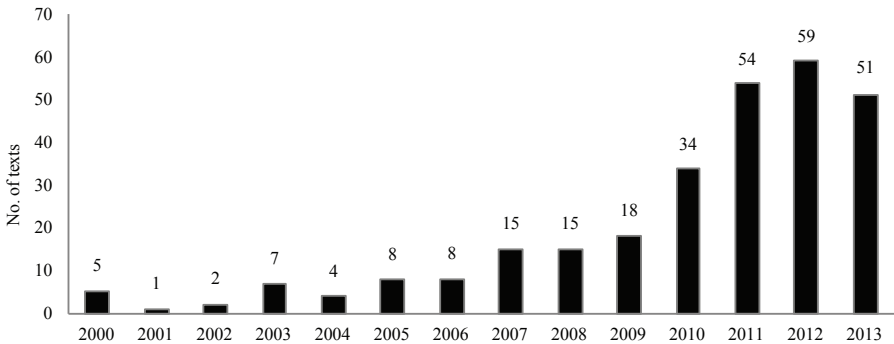


Figure 2. Digital media texts mentioning giftedness (portal Delfi.ee from 2000–2013)

As the aim of the present study is not to compare two eras, but to describe, using qualitative methods, the content of the concept of giftedness and the contexts characteristic of the coverage of the phenomenon the volumes of the texts representing the two periods need not be equal or even comparable.

There are various ways to study the attitudes and meanings concerning social phenomena, of which case studies and phenomenography are perhaps the most widespread approaches; some other authors have concentrated on the meanings propagated by the media (Cohen et al. 2007, cf. Richardson 2007). The phenomenon of giftedness has been analysed in a lot of case studies (e.g. Mudrak 2011, Mazzoli et al. 2012); the few studies available on the media coverage of the phenomenon are based on a qualitative content analysis of texts (Radford 1998, Laine 2010). In the present study, critical discourse analysis (CDA) was preferred as this is a method allowing to bring out social agreements and meanings manifested in linguistic choices.

CDA has been discussed by several theoreticians (e.g. Wodak 1998, Rogers 2004, Richardson 2007, Wodak, Meyer 2009). The method has mainly been criticized for lack of system and for insufficient transparency (Cohen et al. 2007: 391, Breeze 2011). Fairclough (2001: 92–93), however, has managed to provide CDA with a definite structure, the systematic nature of which attracted my preference. The approach addresses the linguistic meanings of texts, which are used to represent reality and to establish social relations, while the unit of analysis is a full text. This enables the analysis to move on from word and sentence levels to address relations between passages in order to see which topics are highlighted and which ones remain in the background.

Fairclough's (2001: 21–22) critical text analysis consists of three steps: 1) text description; 2) interpretation of the relationship between the text and act of communication, and 3) analysis of the relationship between the text and its social context. Here we can see one of the advantages of Fairclough's structured analysis. In the context of my own study, micro-level linguistic analysis can answer that part of my research question which concerns the lexical structuring of texts dealing with giftedness. This stage of analysis involves a description, in terms of vocabulary, grammar and text structure, of the linguistic choices made when writing on giftedness. Simultaneous interpretation addresses the meanings created by those choices in each particular text and the influence of those meanings on the available context. Social analysis, which addresses the macro level, helps to find out what meanings and social relations are constructed in the society when giftedness is at issue and how they are highlighted. In terms of the present study it means that wordings are tried to be explained from the ideological trends of the time.

Language use in the media is always ideological, reflecting hidden power relations (Fairclough 2001: 41, Foucault 2005: 11). Linguistic choices can reveal the attitudes and values that play a key role in constructing the understandings naturalised in different eras, thus becoming a benchmark of social and cultural evaluation (Dijk 2005: 94). One of the aims of my study is to highlight those power relations as manifested in the socially weighty discourse of giftedness, which has sometimes even been classified as a mainstream discourse (cf. Schultz 2005: 117). The principle of open markets, for example, which is one of the main features of the dominant neoliberal ideology, stands on the ideas of individualism, freedom and competition (Fairclough 2010: 11, Gilbert 2013: 9–10). The question whether this idea is spreading into texts on the gifted and giftedness and what linguistic means are used to construct the relevant relations is one of the interests of the present study.

3. Theoretical framework

The definition of giftedness was briefly covered in the Problem statement above. Sepp (2010) has given a longer Estonian overview of the theories of giftedness (which is whence my search words for the historical texts were drawn). According to Sepp, the current treatment of giftedness in the social and educational paradigm bears strongly on the theory launched by Gardner (1983) in his “Frames of Mind”⁵. He classified abilities and skills and pointed out different domains of giftedness. Later, the multiple intelligences suggested by Gardner have been added to by several theorists (e.g. Shavinina 2013). The idea of eugenics reflected in some historical treatments of giftedness, was introduced by Galton in his “Hereditary Genius” (Olari 2011, cf. Galton 1869: 336–340), who laid a foundation to the 19th-century studies of giftedness.

The main context of the present study is a critical approach to texts and society, which analyses the relations between the language use of individuals or institutions and power, concentrating on the roles of discourses in the reproduction of power relations (Richardson 2007: 1, Dijk 2005: 22). For the present study, critical discourse analysis (which is a rather widespread approach, e.g. Rogers 2004, Richardson 2007, Aava 2011, Hodge 2012, Holzscheiter 2014) works both as a method and a theoretical framework.

Etymologically the term ‘discourse’ originates in the Latin verb *discurrere*, which in a direct sense means ‘running back and forth’, but in a figurative sense it means ‘interaction of messages between two or more participants’ (Hodge 2012: 4). Hence, social interaction is one of the main characteristic features of discourse. Although many studies put an equal sign between *text* and *discourse*, interpreting discourse as a text in its context (Dijk 2005: 231), discourse means more than that (Wodak, Krzyzanowski 2008: 4–10 ff.). Foucault (1972, 2005: 11) associates discourse with power and ideology. Bringing together a text, the language use in the text, and the power represented by the latter Foucault defines discourse as a language use which simultaneously reflects and constructs the social world. Fairclough (2010: 3, 59), however, regards discourse as a system of which texts only make up a part, while the whole discourse is mainly about the creation of social reality and the construction of meanings. Holzscheiter (2014: 144) summarizes the ideas of modern discourse research, stating that discourse is a space where intersubjective meanings are created, held and changed, thereby constructing the reality around us (see also Richardson 2007: 21–24). Hence, the giftedness discourse is a space of thoughts and media, which manifests itself in texts; it is a communicative system where, on the one hand, the concept of giftedness is reflected in a way characteristic to the era, while on the other hand the same concept is being constructed by establishing its connections with a certain set of other discourses. This way, giftedness as a social construct that keeps reflecting the reality and simultaneously creating it.

Traditionally, discourse studies are based on the constructionist idea, arguing that the constructions of the social world depend, on the one hand, on language and our interpretations, and on the other hand, on an inter-subjective dimension containing definite models of reflection manifested in a discourse (cf. Holzscheiter 2014: 145). While usually inter-subjectivity is a term referring to a semantic space

⁵ In the first edition (1983) of his “Frames of Mind” Gardner mentions seven modules, while later (2006) he has written of nine different types of intelligence (Sepp 2010).

constructed in the context of human interaction, in media discourse it rather means a space of thinking formed by common ideological beliefs, which according to Richardson (2007: 19) results from an ongoing dialogue between language and society. The ideologies shimmering through media texts may be more or less well hidden (Kasik 2008: 18). Whether someone is called gifted in the media is usually a signal of community attitudes, shared by the author, towards giftedness and the gifted. This signal, in turn, has a formative influence on those attitudes.

4. Results

4.1. Construction of the giftedness discourse in historical media texts

As revealed by analysis, giftedness was an issue in Estonian media texts of as early as the 1890s. Four of the nine relevant texts found in the corpus as representative of that particular decade (Figure 1) refer to God as the source of the gift of talent⁶ (Examples 1, 2), while in four texts skills and abilities are pointed out as prerequisites for the manifestation of the gifts (1, 3).

- (1) *Imeilusal ja kaunil wõimsal helil tõi kunstnik kuus kunstlaulu kuuldawale, milles igasühes nii Loojalt kingitud talent kui hoolega haritud osawus selgesti nähtavale ilmus.* (AJA1890\vir0105)
 ‘With a voice of power and beauty the artist delivered six art-songs, each revealing both God-given talent and a carefully cultivated skill.’
- (2) *Tähendamata ei wõi meie ometi mitte jätta, mis imekallist armu ja annet Jumal selle Eesti wennale on taganud.* (AJA1890\pos1001)
 ‘We can hardly overlook the wondrous grace and talent lavished by God on this Estonian brother of ours.’
- (3) *Meil ei ole kaupmehe andeid ja meie oleme nii kaswatatud, et meie äri- asjadest ei mõista õiget lugu pidada.* (AJA1890\oleo301)
 ‘We lack the talents of a merchant, neither has business been an important part of our upbringing.’

Table 1 summarizes the discourses of giftedness, manifested in historical media texts and mainly differing by the lexical means used, as well as their construction.

Table 1. Major discourses of giftedness found in historical media texts

Discourse focus	Topics and metaphors
Divinity	Biblical message; talent as a gift of God
Heredity	Hereditary premises
Eugenics	‘Breeding’ by making the gifted marry their peers; untalented means inferior
Development	Industry and cultivation of innate abilities, environmental influences on development → environmental discourse
Environment	Environment surrounding gifted people, environmental factors

In the 1890s the divinity discourse seems to have dominated, while with the turn of the century the topics of heredity and eugenics rose to the fore (4, 5).

⁶ As in English the term *giftedness* refers to a gift of God, the Estonian *talent* originates in the biblical parable (Mt 25, 14–30).

- (4) *Kahe meie tuntud kirjaniku lapsed omavad silmapaistvalt teistest kõrgemaid ja oluliselt häid vaimuandeid. (PM 20.5.1934)*
 ‘The children of two prominent Estonian writers are remarkably more gifted than the rest.’
- (5) *Mitte ainult üksikud ei erine üksteisest omade vaimiliste omaduste ja andekuse poolest, vaid et selles suhtes erinevad ka oluliselt tõud ja rahvad .. Inglise rahvas on kõige paremate vaimiliste omadustega. (Ibid.)*
 ‘Differences in mental abilities and giftedness run not only between individuals but also between races and peoples .. The English have the best mental abilities.’

The line drawn between the two last mentioned discourses is rather arbitrary. Still, I used it in order to call attention to the idea of ‘breeding’ that began spreading in Estonia at the previous turn of century, while its credibility amounted to the establishment of a Chair of Eugenics at the University of Tartu in 1939 (Olari 2001). Such discourse construction shows that even at that time giftedness was a powerful ideological instrument to classify and select not only individuals but also whole nations and ethnicities (5).

The developmental discourse was constructed by speaking of natural prerequisites (6) and talent cultivation (1, 7). In the texts, mention was made of the development of individual talents (1 – *carefully cultivated skill*, 7 – *guidance appropriate to each separate actor*) as well as of a supportive environment necessary for talent development (7 – *good schooling and individual guidance*).

- (6) *.. loodus oli teda ilusate ja mitmekesiste annetega selle elukutse jaoks ehtinud. (AJA1910\pm0359)*
 ‘.. nature had adorned her with various beautiful talents necessary for the trade.’
- (7) *Ei wõi ütelda, et “Wanemuise” näitlejate seas anderikkaid jõudusid rohkem oleks .. Aga hea kool ja iga üksiku näitleja kohane juhatus on see, mis “Wanemuiselasi” “Estonialastest” hea tüki kõrgemale on suutnud tõsta. (AJA1910\wi0301)*
 ‘There are hardly more of talented actors at the “Wanemuine” theatre than .. Yet, it is good schooling and individual guidance offered to each single actor that has managed to raise this theatre a good deal higher than “Estonia”.’

Part of the developmental discourse can be classified under the environmental discourse where certain environmental factors supporting talent development are pointed out, such as school and education (7), family (8), money (8 – a talent cannot be cultivated without financial support) and society in general (3, 6). The instrumental approach manifested in the environmental discourse – a talent is someone to bring riches and fame to a private film studio (9) – was a reflection of the prevalent beliefs of the time.

- (8) *Kuna minu nooremad wennad omal ajal edasiõppimist ei saanud jatkata, sest et perekonna jõud minu toetamiseks ära kulus, siis oli soow, et nemad oma jõudu annete ja huwide kohaselt kasuliku töö pääle saaks kulutada. (AJA1910\pm0311)*

‘As my younger brothers had not been able to continue their studies in their time, because all of the family support was spent on myself, there was a wish that they could spend their energy on a useful job according to their talents and interests.’

- (9) *Metro-Goldwyn-Mayeri reshissöör Harry Raps oli saabunud New-Yorki otsima uusi talente ja awastas Billie. .. Studiote ei meeldinud Billie Cassini nimi. Billie muudeti ..* (AJA1930\paewa11)

‘Harry Raps, who was a director at MGM Studios had arrived in New York in search of new talents – and there he discovered Billie. .. The Studios disliked the name of Billie Cassin. Billie was changed into ..’

- (10) *Esmakordselt avanes [sotsialismi tingimustes] vaba tee rahva talentide õitsenguks ..* (AJA1950\rho268)

‘[Under socialist conditions,] for the first time the road was open for the talents of a free nation to flourish ..’

The discourses not included in Table 1 are those of gender (11 – *a woman genius*) and success (12 – *famous overnight*), which are associated with the modern neoliberal worldview, but were relatively less prevalent. Yet it is here that an interesting bridge to the modern discourse of giftedness can be observed (see Tables 2, 3).

- (11) *.. vaieldi selle üle, kas naisterahvas võib suurvaimuks saada.* (AJA1910\plo02)

‘The discussion was about whether a woman could ever become a genius.’

- (12) *Proua Bori sai üleöö kuulsaks oma muusikaliste löökpaladega.* (UUS\uus019)

‘Mrs. Bori became famous overnight with her musical hits.’

Noticeably, the historical discourses of giftedness were constructed not only in terms of the origin of the concept, but also using such elements as skills, abilities, or environmental factors, which are equally typical of the modern scientific discourse.

4.2. Construction of the giftedness discourse in the media texts of the digital era

The texts of the digital era use neoliberal rhetoric to emphasize the elite, winner’s, business and money discourses; also, neoliberal terms are found in the educational discourse (Table 2).

Table 2. Major discourses of giftedness in the media texts of the digital era

Discourse focus	Topics and metaphors
Elite	Being part of the social elite, elitist education for the gifted → educational discourse
Education	Giftedness as a criterion for admission to elite schools, giftedness as a social divisor, gifted = excellent student; giftedness as a special educational need
Winner	A winner-take-all model of the society, ⁷ i.e. achievement and success according to neoliberal ideology; the peak metaphor
Business	Free market economy with its rules – the gifted are bought and sold at market price like any other goods
Money	Poverty as an inhibitor of talent development; Riches achievable by the gifted

⁷ Robert Frank and Philip Cook have described a winner-take-all society, where only the first and best were worth something (Frank 2001). They also launched the term ‘winner’s discourse’.

For the media of the 2000s, giftedness was strongly associated with belonging to the social elite (13–17).

- (13) *Eliitkoolid on geniaalsed. Mida varem andekad lapsed üles leitakse ja kokku pannakse, seda parem.* (8.1.2014)⁸
'Elite schools are an ingenious invention. The earlier they find the gifted children and bring them together, the better.'
- (14) *Andekate õpilaste ja paremate õpetajate koondumine eliitkoolidesse suurendab veelgi hariduslikku lõhet.* (26.8.2013)
'The concentration of gifted students and better teachers in elite schools contributes even more to the educational divide.'
- (15) *Meil on prestiižsed hoopis nn eliitkoolid .. kus lapsevanem peab oma kukrut kergitama .. Parema hariduse eeldus on [eliitkoolis antav] kõrgem startipakk ja nii mängib rolli sünnipärane andekus.* (8.4.2011)
'In our country prestige rather belongs to the so-called elite schools .. , where parents have got to shell out .. A higher starting block [received from an elite school] is a prerequisite for better education [to follow] and this is where innate giftedness comes in.'
- (16) *Elitaarsetes õpivad ränkadel katsetel välja sõelutud oivikud.* (18.4.2011)
'The elite [schools] are filled with excellent students screened out by really tough tests.'
- (17) *25-aastase Adam Frayne'i andekus golfis on viinud teda Suurbritannia kuningannaga teed jooma.* (19.10.2005)
'The talent of the 25-year-old Adam Frayne in golf has taken him to a cup of tea with the Queen of Great Britain.'

The elite discourse consists of two kinds of texts. Part of them use the metaphor of an elite school to describe an educational institution for advanced studies (15 – *higher starting block for [further] education*), where parents are expected to pay (*to shell out*), while competition is emphasized as the ultimate value (16: the metaphoric expression *screened out by tests*). There are arguments for (13) as well as against the gifted to be attending elite schools (14). All this is evidence to the elite discourse being intertwined with the educational one. At the same time the elite discourse is concerned with the social elite at large, as its rhetoric is constructing giftedness as a privilege providing admission to the top of the society (17 – *a cup of tea with the Queen*).

A new rhetoric in the educational discourse refers to giftedness as a special need. Here, too, we meet pro- and contra-attitudes to such association (the rhetorical exclamation of (18) vs. (19) – *special need as a learning difficulty*).

- (18) *Lapsevanemad arvavad, et andekamad lapsed vajaksid rohkem tähelepanu. Andekus on ka ju erivajadus!* (20.6.2010)
'Parents believe that more gifted children should get more attention. Is giftedness also not a special need?'
- (19) *Erivajadust ollakse ühiskonnas harjunud tõlgendama kui ajutist või püsivat õpiraskust.* (15.11.2013)
'A special need is traditionally interpreted as a temporary or permanent learning difficulty.'

⁸ All examples in this chapter have been found in Delfi.ee and thus they are referenced by date only.

Example 19, according to which discourse 'special need' (*erivajadus*) refers to mental disability rather than to extraordinary ability, was corroborated by texts stating, on the one hand, that giftedness is indeed a special need, but on the other hand, when it came to creating the necessary learning conditions for children with special needs, the talk was confined to the facilities facilitating movement and psychiatric services, which are hardly necessary for most of the gifted (cf. 20).

(20) *Koolide kordategemisel arvestatakse asjaoluga, et erivajadustega õpilastele oleksid tagatud kõik vajalikud teenused ning sellele vastav ruumikeskkond.* (17.12.2009)

'When schools are fixed up, care will be taken to ensure all necessary conditions and the appropriate interior environment for pupils with special needs.'

Neoliberal rhetoric allows to differentiate between winner's, business and money discourses. The winner's discourse (21) reflects a winner-take-all model of society. The peak metaphor was the main means to emphasize the neoliberalist idea of winning as the ultimate value (22).

(21) *"Eesti otsib superstaari 2008" võitjaks on andekas Jana Kask, kes vaid 16-aastasena võitis tunnustava tiitli, plaadilepingu ja suure auhinna- raha.* (9.6.2008)

'The Estonian Idol 2008 is the talented Jana Kask, who at her early age of 16 has now won the title, a record deal and a solid prize money.'

(22) *Urmo Aava on tõusnud kiiresti maailma ralliässade sekka .. ta suudab tänu oma töökusele ja andekusele tõusta järgmise eestlasena ka WRC sarjas tipptegijate hulka.* (22.8.2007)

'Urmo Aava has made a quick rise to the top of world rally aces .. Owing to his diligence and talent he can well be the next Estonian to rise to the top performers in the WRC series as well.'

(23) *Andekuse kriteeriumiks pakkusid eliitkoolide juhid maakondlikel olümpiaadidel heade tulemuste saavutamise.* (3.8.2009)

'According to the directors of elite schools, good results at county olympiads should be used as a talent criterion.'

Similar lexical choices were characteristic of educational texts touching on giftedness. Good and winning results were the criterion for talent assessment and recognition (23).

It is hard to draw a clear line between business and money discourses. The business discourse is distinctive for using business vocabulary. However, business vocabulary is also used in that part of the money discourse where giftedness is referred to as a means to earning good money (25, 26).

(24) *2007. aastal leidis Vetteli toonane meeskond BMW, et noor sakslane pole kaugeltki nii andekas ja kiire .. ja lasi tal poole hooaja pealt minna Toro Rossosse. Juba 2009. aastaks oli Red Bull Vetteli endale otnud.* (27.11.2012)

'In 2007 the BMW, which was then Vettel's team, developed an impression that the young German was by far not talented or speedy enough ..

and they let him go, in the middle of a season, to Toro Rosso. By 2009 Red Bull had bought him over.’

Example (24) demonstrates how giftedness will become decisive in whether a person is still in the picture or suppressed to the background (*the BMW let him go*). The expression *bought over* refers to the talent as a good sold and bought at free market prices. The ‘price’ of the talented ‘good’ depends on its amount of giftedness. The texts have ceased to deal with prerequisites or abilities, because position in the business world has taken over.

In (25) irony is used (*giftedness depends on the time of day*) to make another reference to neoliberal values as the quoted message (of a cabinet minister) is about the situation where gifted physicians were leaving state health care for private practices in quest of higher income, which was a widespread problem of the time.

- (25) “*Arstid töötavad nii era- kui riigimediisins samaaegu, aga erasüsteemis tuleb nende anne paremini esile.*” [tsitaat tervishoiuministrilt] .. *Keskhaiglas sõltub andekus kellaajast. Näiteks dr Szirko on kolmapäeviti kella kaheteistkümnest kaheni andetum ja kella kahest poole kaheksani andekam.* (25.4.1997)⁹

“Physicians work parallelly in private and public medicine, yet the private system gives them more prominence.” [quote from the Minister of Public Health] .. In our Central Hospital giftedness depends on the time of day. On Wednesdays, for example, Dr. Szirko is less gifted from noon to 2 p.m. and more gifted from 2 to 7.30 p.m.’

The money discourse occurs in two contexts, on the one hand, treating giftedness as a means to making money (25, 26) and on the other hand, describing the gifted as a needy layer of society for their lack of financial support (27 *non-existent practising conditions*).

- (26) *Adam Frayne alustas karjääri koos selliste noorte golfikuulsustega nagu Justin Rose, kes täna teenib ligikaudu 22 miljonit Eesti krooni aastas.* (19.10.2005)

‘Adam Frayne started his career together with such famous young golfers as Justin Rose, whose annual income today equals nearly 22 million Estonian kroons.’

- (27) *Võttes arvesse koduseid, juba teist hooaega olematuid harjutamistingimusi, võib tulemused suures osas kirjutada poiste andekuse arvele.* (30.4.2008)

‘Considering the non-existent conditions for practising at home, continuing for a second season, the boys’ achievements can largely be accounted for by their own giftedness.’

Thus, the texts belonging to the money discourse emphasize not only the preconditions and achievements of the gifted, but also, if not mainly, the sums of money which can either be earned using giftedness or, vice versa, which are lacking from the amount necessary for best results.

⁹ Delfi.ee was founded in 1999, but the search also retrieved three texts representing 1997 and 1996.

By way of conclusion, the relevant texts of the digital era are rather diverse as, besides the discourses just described, the gender and success discourses continue to be active; in addition, some elements of the discourses of divinity, heredity and eugenics keep popping up here and there.

5. Conclusions and discussion

During my analysis of the relevant texts from the 1890s to the present day I pinpointed the ideologies, main domains and vessels of giftedness. Those markers enabled a periodization of the media coverage of giftedness (Table 3).

Table 3. Ideologies, main domains and vessels of giftedness as revealed in the construction of the giftedness discourse

Diachronic construction of giftedness					Construction of giftedness in the digital era
Decades	1890s–1910s	1930s	1950s–1980s	1990s	2000s
Ideology	Liberalism: talent or giftedness as a means to profit	Eugenics: talent is hereditary, important for the continuation of race	Socialism: Soviet system as the best environment for talent development	Early neo-liberalism: giftedness contributes to a person's value	Neoliberalism: giftedness as the property of a product, a basis for success
Vessel of talent	individual – adult	intellectuals, people, nation	collective, individual – head of state	people	individual – child; additional metonymical use – animals, home appliances etc.
Major domains	classical music, arts	arts, sciences	politics, industry	banking, enterprise	sports, politics, pop music

Reflection of the dominant ideologies in the printed press around the turn of the 20th century brought out liberal values, which saw giftedness as an object of profit (Example 9). The advance of neoliberalism added concrete profit numbers and the linguistic choices included economic and business rhetoric (Examples 22, 24, 26). Although the historical texts reflecting liberal ideology were not numerous, they create an essential bridge between the texts from the early periods of two centuries.

The idea of 'breeding' manifested in the media of the 1930s (Example 5) is rarely found today, mainly occurring in the modern elite discourse (Examples 13, 15). True, browsing the digital media, I noticed a discussion on how to improve the giftedness of peoples and nations ("*Hiiinas hakatakse lapsi geenide põhjal paika panema*" / 'In China children are beginning to be sorted by genes', 6.9.2009).

The socialism of the 1950s brought a rise in the coverage of the genius of the heads of state and, consequently, a considerable rise in the number of texts on giftedness (Figure 1). The ruling regime was described as the best possible environment for the life and development of the gifted (Example 10). None of the analysed texts of the digital era referred to the gifts of any modern head of state.

In the 1990s, which was the initial phase of neoliberalism the business discourse highlighted the individual and his values. Giftedness was also attributed to those who possessed the abilities and prerequisites necessary for entrepreneurship as soon as the Iron Curtain had been removed (hence the metaphor of a 'waffle sellers' generation' (*vahvliküsetajate põlvkond*) in Läck's "Mart Kalmu ..", EE 16.1.2002). In the media of the 2000s, however, those values have changed, notably, giftedness as a value of a person has come to signify, first and foremost, the value of a good for sale.

The list of the main walks of life associated with giftedness has also changed over time. Despite the relatively small amount of the relevant texts available for each decade a qualitative analysis enables the conclusion that different stages of history also differ in what domains are more prestigious than the rest and are thus made prominent by association with giftedness. A largely similar conclusion has been arrived at in another study of giftedness and media, stating that the people having certain gifts get more coverage than the vessels of some others (Laine 2010: 73). Speaking of multiple intelligences, students of giftedness emphasize only some walks of life, mainly seeing the gifted among those active in arts, music or sciences (Gardner 1983, Gagné 2004). Besides the last mentioned studies the present results on the business discourse and the domain of enterprise, can be compared with the speculations by Shavinina (2013: 55, 57), who pays considerable attention to giftedness and talent in enterprise or business. Focusing on creativity and innovation Shavinina emphasizes that while creativity in general is associated with generation of ideas that innovators try implementing, the gifted are the one group of society which can do both. This way the gifted are attributed with a very important social position. Thus, on the one hand, giftedness has a vital role in the sustainability of nations and societies, as pointed out in the introduction of the article (Simonton 2009, Ziegler 2009), on the other hand, our results suggest that in the conditions of free market economy a gifted person has come to be seen as a good with a definite price.

The other major discourse of the digital era is the elite discourse addressed by Simonton (2009: 906). He is one of the few who takes a critical approach to the available theories of giftedness, pointing out that giftedness meant elitism already for Galton. This confirms my own findings as well as the observations of some other authors to the effect that giftedness is a social construct signifying belonging to a certain social group (cf. Mudrak 2011).

In view of social values an important conclusion of the study reads that in the digital era, giftedness has become a vehicle of neoliberal ideology which emphasizes success as an ultimate value and victory as an ultimate goal. Success and achievement were also emphasized at the ECHA conference, stating that "the victories of the gifted are usually won at prestigious contests" (Fülöp 2014).

The study indicates that the tradition of using the gifted as an instrument of power is nothing new. It has even been seen so profitable for national economy that 'breeding' was discussed as a means to improve the people's genetic code. The study also revealed that the discourse of giftedness, which historically used to be constructed on the basis of the concept of 'adult', have undergone a change. Notably, in the modern digital era most of the persons referred to as gifted are children, while the talents and gifts of adults clearly remain in the background. Note that

most of the relevant research, both historical and modern, address giftedness in children (Kaufman, Sternberg 2008, ECHA 2014), whereas giftedness in adults seems to interest but a few (e.g. Shavinina 2013). According to Laine (2010), who has studied the media coverage of gifted children, the discussion of a gifted child in Finland could be broader and more comprehensive. The present study indicates that although Estonian media addresses most different aspects of gifted children, there is no discussion on the conception of giftedness as a whole.

In the light of the above conclusions, the discourse of giftedness has changed over time, concerning the gifted person (or, figuratively, even object), the relevant domains where giftedness is seen manifested, and the salient values. However, the phenomenon of giftedness has been important in maintaining and describing social agreements and relations both in the historical texts and in the texts of the modern digital era. The present study is a part of a comprehensive framework study with an aim to survey how the concept of giftedness is constructed in the print media of various cultures in order to find out what enlivens and what inhibits public media discussion on the essence of giftedness and also, to describe the relevant discourses constructed in different linguistic environments. At the same time it is necessary to extend the studying of the construction of the concept of giftedness here in Estonia, by supplementing the reported results on media texts with the experience and ideas of those who are in direct contact with the gifted, i.e. teachers, parents etc., as well as of those responsible for educational policies.

References

- Aava, Katrin 2011. Teadmuspõhise ühiskonna ja elukestva õppe diskursuste rekontekstualiseerimine Eestis. [On the recontextualization of discourses concerning a knowledge-based society and life-long learning in Estonia.] – Eesti Rakenduslingvistika Ühingu aastaraamat, 7, 5–18. <http://dx.doi.org/10.5128/ERYa7.01>
- Borland, James H. 2005. Gifted education without gifted children. The case for no conception of giftedness. – Robert Sternberg, Janet Davidson (Eds.). *Conception of Giftedness*. USA: Cambridge University Press.
- Breeze, Ruth 2011. Critical discourse analysis and its critics. – *Pragmatics*, 21 (4), 493–525. <http://dx.doi.org/10.1075/prag.21.4.01bre>
- Cohen, Louis; Manion, Lawrence; Morrison, Keith 2007. *Research Methods in Education*. 6th ed. New York: Routledge.
- Dijk, Teun A. van 2005. *Ideoloogia. Multidistsiplinaarne käsitus. [Ideology. A Multidisciplinary Approach.]* Tartu: Tartu Ülikooli kirjastus.
- ECHA 2014 = *Re:thinking Giftedness – Giftedness in the digital age*. European Council for High Ability International Conference. Sept, 17–20. Ljubljana.
- Eesti ajakirjanduse analüütilise bibliograafia andmebaas. <http://www.kirmus.ee/biblioserver/index2.php> (9.9.2014).
- EA 2012 = Eesti arengufond. Talendifoorumi “Kuidas hoida, püüda ja arendada talente” mõttealgatus: globaalne talendiralli ja mõju Eestile. [Global talent rally and its influence on Estonia. Initiative of the Talent Forum on how to keep, capture and cultivate talents.]
- Eesti kirjakeele korpus. <http://www.cl.ut.ee/korpused/baaskorpus/> (9.9.2014).
- Fairclough, Norman 2001. *Language and Power*. 2nd ed. New York: Longman.
- Fairclough, Norman 2010. *Critical Discourse Analysis: The Critical Study of Language*. 2nd ed. Harlow: Longman.
- Feldman, David H. 2000. The development of creativity. – Robert J. Sternberg (Ed.). *Handbook of Creativity*. UK: Cambridge University Press, 169–189.

- Foucault, Michael 1972. *The Archaeology of Knowledge & the Discourse of Language*. New York: Pantheon Books.
- Foucault, Michael 2005. Diskursuse kord. Collège de France'i inauguratsiooniloeng 2. detsembril 1970. [The Order of Discourse. Inauguration lecture of Collège de France.] Tallinn: Varrak.
- Frank, Robert 2001. Talent and the Winner-Take-All Society. <http://prospect.org/article/talent-and-winner-take-all-society> (12.9.2014).
- Fülöp, Márta 2014. Do contests and competitions enrich or ruin the life of the gifted? – Presentation in ECHA Conference. Sept, 17–20. Ljubljana.
- Gagné, François 2004. Transforming gifts into talents: The DMGT as a developmental theory. – *High Ability Studies*, 15 (2), 119–147. <http://dx.doi.org/10.1080/1359813042000314682>
- Galton, Francis 1869. *Hereditary Genius*. London: Macmillan. [Digitized document]. <http://www.mugu.com/galton/books/hereditary-genius/> (4.9.2014).
- Gardner, Howard 1983. *Frames of Mind: The Theory of Multiple Intelligences*. London: Fontana Press.
- Gilbert, Jeremy 2013. What kind of thing is 'neoliberalism'? – *New Formations*. Special Issue. *Neoliberal Culture*, 80–81, 7–22.
- Hodge, Bob 2012. Ideology, identity, interaction: Contradictions and challenges for critical discourse analysis. – *Critical Approaches to Discourse Analyses Across Disciplines*, 5 (2), 1–18.
- Holzschneider, Anna 2014. Between communicative interaction and structures of signification: Discourse theory and analysis in international relations. – *International Studies Perspectives*, 15 (2), 142–162. <http://dx.doi.org/10.1111/insp.12005>
- Hymmer, Barry 2014. Constructing and deconstructing giftedness: A reflective conversation between Tim Dracup and Barry Hymmer. – *Gifted Educational International*, 30 (1), 5–23. <http://dx.doi.org/10.1177/0261429413480418>
- Kasik, Reet 2008. Meediateksti analüüs: eesmärgid ja meetodika. [Analysis of media texts: Purposes and methods.] – Reet Kasik (Toim.). *Tekstid ja taustad V. Meediatekstide keelekasutus ja selle sotsiokultuurilised taustad*. Tartu: Tartu Ülikooli Kirjastus, 10–41.
- Kaufman, Scott; Sternberg, Robert J. 2008. Conceptions of giftedness. – Steven I. Pfeiffer (Ed.). *Handbook of Giftedness in Children. Psychoeducational Theory, Research, and Best Practices*. New York: Springer, 71–93. http://dx.doi.org/10.1007/978-0-387-74401-8_5
- Laine, Sonia 2010. The Finnish public discussion of giftedness and gifted children. – *High Ability Studies*, 21 (1), 63–76. <http://dx.doi.org/10.1080/13598139.2010.488092>
- Mazzoli Smith, Laura; Campbell, Jim 2012. *Families, Education and Giftedness. Case Studies in the Construction of High Achievement*. Rotterdam: Sense Publishers.
- Mazzoli Smith, Laura 2014. Giftedness and globalisation: The challenge of cultural diversity for gifted education programmes in a neoliberal educational marketplace. – *Gifted Education International*, 30 (3), 197–211. <http://dx.doi.org/10.1177/0261429413486572>
- McQuail, Denis 2003. *McQuaili massikommunikatsiooni teooria*. [McQuail's theory of mass communication.] Tartu: Tartu Ülikooli kirjastus.
- Mudrak, Jiri 2011. 'He was born that way': Parental construction of giftedness. – *High Ability Studies*, 22 (2), 199–217. <http://dx.doi.org/10.1080/13598139.2011.622941>
- Olari, Jaan 2001. *Eugeenika Eestis ja mujal*. [Eugenics in Estonia and beyond.] – *Terviseleht*, 8. http://www.terviseleht.ee/200108/8_olari.php?gallup=tulemused (30.9.2014).
- Radford, John 1998. Prodigies in the press. – *High Ability Studies*, 9 (2), 153–164. <http://dx.doi.org/10.1080/1359813980090202>
- Richardson, John E. 2007. *Analysing Newspapers. An Approach from Critical Discourse Analysis*. New York: Palgrave Macmillan.
- Rogers, Rebecca 2004. *An Introduction to Critical Discourse Analysis in Education*. Mahwah, New Jersey: Lawrence Erlbaum Associates.

- Schultz, Samantha 2005. The gifted: Identify construction through the practice of gifted education. – *International Education Journal*, 5 (5), 117–128.
- Sepp, Viire 2010. Andekusest ja andekatest lastest. [Giftedness and Gifted Children.] Tartu: Atlex.
- Shavinina, Larissa 2013. How to develop innovators? Innovation education for the gifted. – *Gifted Education International*, 29 (1), 54–68. <http://dx.doi.org/10.1177/0261429412440651>
- Simonton, Dean K. 2009. Gifts, talents and their societal repercussions. – Larissa Shavinina (Ed.). *International Handbook of Giftedness*. New York: Springer, 905–912.
- Sternberg, Robert, J.; Davidson, Janet E. 2005. *Conceptions of Giftedness*. 2nd ed. USA: Cambridge University Press.
- Ziegler, Albert 2009. Research on giftedness in the 21st century. – Larissa Shavinina (Ed.). *International Handbook on Giftedness*. New York: Springer, 1509–1524. http://dx.doi.org/10.1007/978-1-4020-6162-2_78
- Wodak, Ruth 1998. *Gender and Discourse*. London: Sage Publications.
- Wodak, Ruth; Krzyzanowski, Michal 2008. *Qualitative Discourse Analysis in the Social Sciences*. New York: Palgrave Macmillan.
- Wodak, Ruth; Meyer, Michael 2009. *Methods for Critical Discourse Analysis*. London: Sage Publications.

Halliki Põlda's (Tallinn University) academic research interests are qualitative analysis methodology and means of analysing and interpreting the construction of concepts and phenomena in public communication.

Narva mnt 25, 10120 Tallinn, Estonia
halliki.polda@gmail.com

ANDEKUSE DISKURSUSE KEELELINE KONSTRUEERIMINE AJALOOLISTES JA DIGIAJASTU MEEDIA TEKSTIDES

Halliki Põlda

Tallinna Ülikool

Uuringu eesmärk on kirjeldada ja selgitada, kuidas on ajalooliselt konstrueeritud andekuse diskursust ja millisena see avaldub tänase digiajastu meediatekstides. Kriitilise diskursuseanalüüsi abil selgitatakse, kuidas ja milliste leksikaalsete vahenditega andekust konstrueeritakse ning millised sotsiaalsed tähendused see läbi esile tõusevad.

Uuringus ilmnes, et teatud valdkonnad on läbi ajaloo olnud teistest oluliselt prestiižsemad ja andekuse mõiste kaudu esile tõstetud, ka võimu instrumentidena on andekaid läbi ajaloo ära kasutatud. Digiajastul on andekus muutunud uusliberalismile omase, edukust ülimalt väärtusena rõhutava, ideoloogia kandjaks. Selgus, et digiajastul kajastatakse andekatena ülekaalukalt lapsi, nii jääb täiskasvanu oma andekusega tagaplaanile. Kokkuvõttes näitab uuring, et andekuse diskursus on läbi ajaloo muutunud ja teema mängib kindla keelekasutuse najal konstrueerituna tänapäevastes ühiskondlikes kokkulepetes olulist rolli.

Võtmesõnad: kriitiline diskursuseanalüüs, ideoloogia, sotsiaalne interaktsioon, meediakeel, eesti keel

FACTORIAL STRUCTURE OF SILL REVISITED: ADAPTATION OF SILL FOR ESTONIAN EFL LEARNERS

Katrin Saks, Äli Leijen,
Kandela Õun, Karin Täht

Abstract. The present study aims to contribute to the understanding of the taxonomy of language learning strategies (Oxford 1990: 14–22) and its measuring instrument, the Strategy Inventory for Language Learning (SILL), through the adaptation process of the Estonian version. The translated version of the SILL was pilot-tested with a sample of 337 tertiary students. Cronbach's alpha (0.91) reveals an acceptable reliability of the instrument. The t-test that was conducted in addition to descriptive analysis, to check for gender differences, did not reveal any significant differences. The results of the exploratory factor analysis did not support the 2- or 6-factor structures but the 9-factor analysis provides a good comparison to similar studies published in 1995 (Oxford, Burry-Stock 1995). The ethnic and gender differences of the results are discussed from the cultural perspective.*

Keywords: exploratory factor analysis, questionnaire, reliability, t-test, English, Estonian

1. Introduction

Foreign language skills have become self-evident in today's globalized world. Although foreign language acquisition has not become substantially easier compared to decades ago, learning a foreign language is becoming faster and more efficient through the use of new methods (Ariza 2002, Jiamu 1997, Ellis 2013). The progress of language learning can be assessed in several ways. One of the most efficient is assessing the use of language learning strategies (LLSs) (Hsiao, Oxford 2002), which are believed to be connected with language proficiency and have much potential for enhancing learning. Language learning strategies are believed to play a vital role as they assist learners in mastering the forms and functions required for reception and

* The present research was supported by the European Social Fund.

production in the foreign language and thus affect achievement (see e.g. Bialystok 1979, Douglas 2001). The use of appropriate LLSs improve proficiency and achievement and, at the same time, enable students to take charge of their own learning by enhancing learner autonomy, independence and self-direction (Wong 2011).

Several researchers have produced different classifications of strategies (Rubin 1975, Stern 1975, O'Malley, Chamot 1990), giving their input to the six-strategy taxonomy designed by Oxford (1990: 14–22). The use of language learning strategies may be measured using different methods, from observation or interviews to diaries and think-aloud protocols. Student-completed, summative rating scales have been reported to be the most time-saving and cost-effective. Moreover, being self-scoring, they enable students to discover a great deal about themselves, giving valuable feedback about their learning habits (Oxford, Burry-Stock 1995). Even though self-report questionnaires have often been reported to suffer from serious validity problems (Veenman et al. 2014), we have chosen to utilize it because it provides a) information about learners' memories and interpretations of their actions and b) their explanations of cognitive and metacognitive processes researchers cannot observe (Turner 1995). The most widely used and analysed instrument in the domain of foreign language acquisition, the Strategy Inventory for Language Learning (SILL), has been translated into more than 20 languages and been validated by many researchers. Despite the wide use of the SILL, its factor analyses have still not provided acceptable results (Park 2011, Tragant et al. 2013). The poor fit to the data of factor analyses has been justified by high correlation among the constructs (Park 2011) as well as by national and cultural influences (Oxford, Burry-Stock 1995). Therefore, the authors of this article support Park's idea of undertaking more studies using participants from different cultures and learning contexts to clarify the most suitable factor model for the SILL (Park 2011). Also, research on strategy classification could help to understand the nature of LLSs and decide which classification system accounts best for them (Park 2011).

The current study was conducted in Estonia. English as a foreign language (EFL) has been taught in Estonian schools since the 1930s. During the Soviet period, the grammar-translation method was mainly used. In recent years, this has been replaced by a communicative approach, which involves changes in teaching methods, learning materials and learners' LLSs. Based on these facts and also that in Estonia the assessment of LLSs has not been purposefully dealt with, mainly because of the lack of a reliable LLS measuring instrument in the Estonian language, our goal is to contribute to the development of the taxonomies of LLSs, shedding light on the cultural distinctness of East-European EFL learners, as in the case of Estonia.

1.1. Language learning strategies and SILL

Strategies are the learner's toolkit for active, conscious, purposeful and attentive learning, and they pave the way towards greater proficiency, learner autonomy and self-regulation (Hsiao, Oxford 2002). According to Oxford (1990: 1-2), LLSs have the features of contributing to the main goal, allowing learners to become more self-directed, expanding the role of teachers, being problem-oriented, including

specific actions taken by the learner, involving many aspects of the learner, not just cognitive, supporting learning both directly and indirectly, being not always observable, being conscious, possible to be taught, flexible, and influenced by a variety of factors. According to Jones (1998), the system of LLSs developed by Oxford is believed to be more comprehensive and detailed than earlier classification models by her predecessors.

Language learning strategies research has gone through the development from simple lists of strategies to much more sophisticated investigations (Oxford, Crookall 1989). Oxford's SILL has been the most widely used instrument in the field of LLS research. The items of the SILL are divided into six strategy groups according to the results of early factor analysis (Oxford, Burry-Stock 1995). The groups are: memory (9 items), cognitive (14 items) and compensation (6 items) strategies, and metacognitive (9 items), affective (6 items) and social (6 items) strategies. The activities characteristic to certain strategies have been presented in Table 1. The first three subdivisions – memory, cognitive and compensation strategies, are direct strategies that involve direct learning and use of the subject matter, in this case a new language. The last three subdivisions – metacognitive, affective and social are indirect strategies that contribute indirectly but powerfully to learning (Oxford 1990: 14). The subdivisions designed to elicit students' use of metacognitive, affective and social strategies can also be used to assess the degree to which students report them having control of their own learning activities (Benson 2011: 98-99).

Table 1. Strategy groups (based on Oxford 1990: 16–17)

Strategy groups	Strategies	Activities
Direct strategies: directly involve the target language, such as reviewing and practising	Memory strategies: aid in entering information into long-term memory and retrieving information when needed for communication	<ul style="list-style-type: none"> • Creating mental images • Applying images and sounds • Reviewing well
	Cognitive strategies: used for forming and revising internal mental modes and receiving and producing messages in the target language	<ul style="list-style-type: none"> • Practising • Receiving and sending messages • Analysing and reasoning • Creating structure for input and output
	Compensation strategies: needed to overcome any gaps in knowledge of the language	<ul style="list-style-type: none"> • Guessing intelligently • Overcoming limitations in speaking and writing
Indirect strategies: provide indirect support for language learning, such as planning, cooperating and seeking opportunities	Metacognitive strategies: help learners exercise executive control in planning, arranging, focusing, and evaluating their own learning process	<ul style="list-style-type: none"> • Centring your learning • Arranging and planning your learning • Evaluating your learning
	Affective strategies: enable learners to control feelings, motivation and attitudes related to language learning	<ul style="list-style-type: none"> • Lowering your anxiety • Encouraging yourself • Taking your emotional temperature
	Social strategies: facilitate interaction with others, often in a discourse situation	<ul style="list-style-type: none"> • Asking questions • Cooperating with others • Empathizing with others

It has been stated that the factors of the SILL are correlated rather than orthogonal (Hsiao, Oxford 2002) whereas particular strategies could be viewed as related to more than one category (Oxford 1990: 17, 22) with the categories mutually supporting each other (Hsiao, Oxford 2002). Hsiao and Oxford (2002) believe that there will probably never be a strategy taxonomy in which intercorrelations among particular strategies are totally eliminated, because such a taxonomy would not reflect reality. However, this partial overlapping and strong intercorrelation are considered to be the main reason why the factors do not obtain clear outlines (Park 2011).

1.2. Former studies on validating SILL and their cultural distinctions

The SILL is considered to be a useful instrument as it has clearly indicated the relationship between strategy use and language performance, giving reasons for the belief that enhancing strategy use could improve language performance (Oxford, Burry-Stock 1995). The psychometric characteristics of the SILL have been widely explored and tested (Tragant et al. 2013, Ehrman, Oxford 1989, Fazeli 2012, Alhaisoni 2012, Wong 2011, Green, Oxford 1995, Griffiths 2003). Green and Oxford (1995) quote the reliability of the total instrument of the SILL using Cronbach's alpha for internal consistency as 0.93 to 0.98, and for subscales 0.68–0.80 (Tragant et al. 2013). Reliability of the SILL was slightly lower when not being administered in the native language of the respondents but in English (Oxford, Burry-Stock 1995). The lower reliability seems to be caused by measurement errors due to the language effect (Oxford, Burry-Stock 1995). This was also the reason why it was considered to be important to translate the questionnaire into Estonian for Estonian learners and not use the original English version.

To validate the SILL's underlying structure, confirmatory factor analysis (CFA) was used for the investigation of the hypothesized measurement structures of scales by Hsiao and Oxford (2002). Results showed that the six-factor strategy taxonomy was most consistent with learners' strategy use. However, the authors admitted that the model did not produce "a fully acceptable fit to the data" and that the format and structure of the whole instrument should be further revised (Hsiao, Oxford 2002).

However, several researchers still question the reliability of the instrument as no valid evidence has been found to support the six-fold classification of LLSs in the form of subclass reliabilities (Woodrow 2005, Park 2011, Rose 2012). According to Park (2011), the popularly used two-construct and six-construct classification systems of the SILL do not fit the data and the classification systems should be reinvestigated to understand better the structures of the SILL.

To explain the poor results of the factor analyses, Bedell and Oxford emphasize that nationality or ethnicity influences strategy use (1996: 47–49). This may also be the reason why the same SILL factor structure might not be appropriate for all people who are learning English as the second language (ESL) or as a foreign language (EFL). While European students are reported using LLSs more frequently than students of other nationalities (Griffith, Parr 2000), Taiwanese students seem to be aware of various LLSs but only a few of them report using these (Yang 1998, Griffiths 2004). Almost the same claims have been made by Politzer and McGroarty (1985), who compared Asian students' strategy use to that of Hispanic students.

Japanese students are typically regarded as passive learners and not very eager to use different LLSs (Usuki 2000, Griffiths 2004). Malaysian EFL learners can be considered quite frequent strategy users (Wong 2011), especially when compared to Saudi (Alhaisoni 2012) or Korean learners (Magno 2010). The way strategies are used may indicate a cultural habit. Lengkanawati (2004) points out that remembering difficult words is not as effective for Australian students as for Indonesians, who confess that they have a habit of rote learning behaviour. Bell (1995) and Porte (1988) admit, based on research as well as their own personal learning experience, that when learning a foreign language, the choice of strategies is also affected by those that the learner has used when learning a previous language or even their first language. Bell (1995) warns against the wholesale transfer of assumptions from first language literacy that can complicate the process of acquiring second language literacy. From this it may be concluded that, for example when learning English or Chinese, the same strategies do not work in the same way and the earlier language learning experience may complicate second language learning (Griffiths 2003).

National and cultural differences exist in the use of LLSs, even though an individual may not fully reflect the trends. As a solution, Oxford and Burry-Stock (1995) suggest creating country-by-country SILL norms around the world based on large-scale factor analyses.

Studies that have examined the relationship between gender and strategy use have come to mixed conclusions. There are several (Ehrman, Oxford 1989, Oxford, Nyikos 1989, Green, Oxford 1995) that have discovered distinct gender differences in strategy use. Females more often than males are claimed to use LLSs consciously, especially metacognitive, affective and social ones (Oxford, Ehrman 1995). Politzer (1983) found that social strategies are more popular with female learners and this may be attributed to their greater social orientation, which is also a sign of their need for social approval (Oxford et al. 1988). The studies of Ehrman and Oxford (1990) and Wong (2011), however, failed to discover any evidence of differing LLS use between the sexes. It might be concluded, perhaps, that although men and women do not always demonstrate differences in LLS use, where differences are found, women tend to use more LLSs than men (Griffiths 2004). Women's higher use of strategies has been associated with their deeper engagement in language related activities (Lynn, Mikk 2009). However, there are also studies that have shown that women use fewer LLSs than men (Tran 1988). A study on the usage of individual LLSs in Malaysia shows that males surpass females in the use of social, compensation and memory strategies (Wong 2011).

Despite the criticisms of the SILL's construct validity and poor factorial fit (Douglas 2001, Tragant et al. 2013), it is still the most widely used measure of LLSs at the moment. Construct validity is the degree to which a test measures what it claims, or purports, to be measuring (Cronbach, Meehl 1955), and that is why it is necessary to indicate it when estimating the quality of a measure. Moreover, thanks to its widespread use, in almost all continents and different cultures it provides good material for comparison, which may finally result in its reinvestigation and restructuring into an even more reliable instrument.

In this study we seek to investigate whether the SILL, translated and adapted for Estonian EFL learners, reflects two-, six- or nine-construct classification, and to compare the results with other similar studies conducted in different cultures. We

will also see if gender differences appear in the comparison of using LLSs. Proceeding from that the hypotheses that will be tested are: (H1) the structure of the SILL translated and adapted for Estonian learners reflects the two-factor structure of direct and indirect strategies; (H2) the structure of the SILL translated and adapted for Estonian learners reflects the six-factor structure of memory, cognitive, compensation, metacognitive, affective and social strategies; (H3) the structure of the SILL translated and adapted for Estonian learners reflects the nine-factor structure of LLSs comparable to the outcome of Oxford and Burry-Stock (1995).

2. Methodology

In Estonia, the issues of LLSs have not been studied regularly, partly because of the lack of a reliable means of measurement. This was one of the reasons for starting the adaptation process of the SILL. When translating a measuring instrument into another language, it is necessary to critically evaluate an instrument that has been designed and validated in one cultural context, in the context of another culture. Differences when interpreting certain terms determine the necessity to adapt them to make them understood in a different culture. The aim of the adaptation process is to achieve the highest possible unequivocalness and equivalence (Guillemin et al. 1993). Cross-cultural adaptation includes translating the original instrument while assuring the validity and reliability of the adapted version.

In the adaptation process of the SILL into Estonian, we used the adapted version of Guillemin's five-step methodology (Guillemin et al. 1993), which covered 1) translating the original instrument into Estonian by one translator, 2) back-translation by three independent translators to check if the translated version reflected the same content of terms used in the original version, 3) assessing, analysing and comparing all back-translations and the source text, 4) semantic editing and correcting the Estonian version, 5) linguistic editing by an Estonian language expert, 6) asking the respondents to assess the overall usability of the adapted instrument and the unequivocalness of terms while pre-testing.

2.1. Participants

To pilot-test the Estonian version of the SILL, data was collected from the students of the University of Tartu in March and April, 2013. The participants were majoring in different domains – economics, social work administration, teacher training, science, mathematics, law, psychology, etc. The students of philology were not included in the study. The sample was 374 students (the total number of students was 18,000), of which 337 sent back their questionnaires filled in completely. The 37 questionnaires that were not complete were not included in the study. Although the sample may not represent the population in general, it gives a good overview of all levels of tertiary students in Estonia's biggest university. The following diagram (Figure 1) presents the social-demographic traits of the sample.

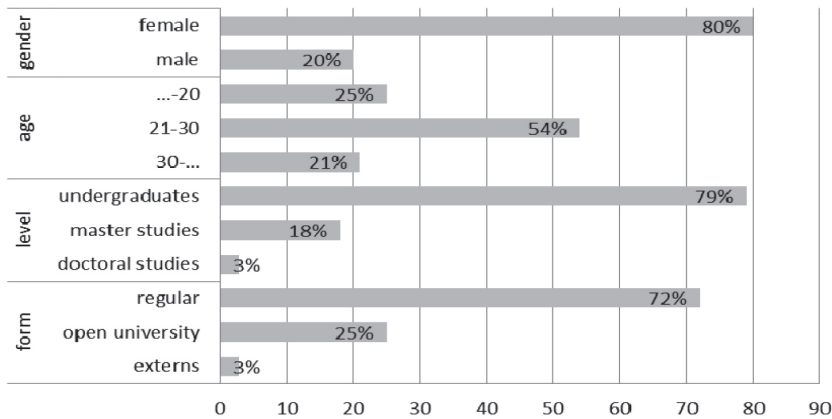


Figure 1. Sample division by gender, age, level and form of studies (numbers given in percentage)

2.2. Instrument

The self-report questionnaire SILL (see Appendix) translated and adapted for Estonian learners involves 50 items. These are statements that express learners' activities or learning strategies while learning a foreign language. The items are divided into six groups according to the division of strategies by Oxford (1990: 293–297): memory strategies – 9 items, cognitive strategies – 14 items, compensation strategies – 6 items, metacognitive strategies – 9 items, affective strategies – 6 items, and social strategies – 6 items. The students assess all items in the Likert-type scale from 1 to 5 where 1 stands for *Never or almost never true of me*, and 5 *Always or almost always true of me*. The questionnaire starts with a brief introduction where respondents are explained the idea of the instrument. They are also encouraged not to answer how they think they should be as there are no right or wrong answers to the statements.

2.3. Data collection

Data was collected on a voluntary basis, partly in the classroom and partly electronically. In the classroom setting participants completed the SILL individually with the test leader present who instructed the participants beforehand. For electronic questionnaire the environment LimeSurvey was used. The respondents were addressed and the URL of the questionnaire was forwarded via study departments. In the electronic form of the questionnaire the statements were presented in one page similarly to the print-out form. This made it possible for students to see all questions at a time, move backwards and forwards, and make corrections if they considered it necessary. Completion of the questionnaire took the students approximately 25–30 minutes.

2.4. Data analysis

Data analysis involved a reliability assessment using Cronbach's alpha for internal consistency, descriptive statistics, independent-samples t-test and exploratory factor analysis (EFA). Even though several previous studies have conducted confirmatory factor analysis to check the factor structure of the SILL, exploratory factor analysis was considered more appropriate in the current case as after translating the instrument into Estonian it was necessary to explore which the underlying factor structure in the case of this sample is. The analysis was conducted with SPSS 19.

Before the statistical analysis, the respondents' answers were standardized to search for outliers. One of the easiest standardization methods is to use two statistical parameters – empirical average and standard deviation – to find z-scores that will give the characteristic of normal distribution $N(0;1)$, which is centralized and standardized, and enables the comparison of characteristics with different content (Kreyszig 1979: 880). There were a couple of outliers in the database of the present study; one of these gave a reason for omitting one item. The original inventory of the SILL includes 50 items but, in the adapted Estonian version of the SILL, it was decided to omit item number 43 in the group of affective strategies – *I write down my feelings in a language learning diary*. The item was deleted based on feedback from participants in the pilot study referring to it as the one that could not be understood and responded to because they had never used a learning diary and could not even imagine what it was, and because of the outlier that emerged in the standardization process. The reason why Estonian students had difficulties assessing the item on learning diaries lies in Estonian pedagogical practice where using learning diaries is not widespread and this has left the students without the experience of self-reflection in the form of a learning diary.

3. Results

3.1. Reliability

The Cronbach's alpha coefficient of the total SILL is 0.91, supporting previous studies with a similar consistent result (Park 2011, Hsiao, Oxford 2002, Green, Oxford 1995). The Cronbach's alpha of the constructs of the SILL was investigated for measuring the internal consistency of the items within each construct. The results reveal that the only alpha coefficient that remained below the acceptable level of 0.60 was for memory strategies with 0.59; the other strategies were above it (Table 2). This finding indicates that the items within each construct of the SILL measure similar characteristics about LLSs.

Table 2. Cronbach's alpha coefficients of strategy groups

Strategies	Cronbach's α	Number of items in each strategy group
Memory	0.59	9
Cognitive	0.80	14
Compensation	0.65	6
Metacognitive	0.84	9
Affective	0.61	5
Social	0.77	6
Total	0.91	49

3.2. Descriptive statistics

Table 3 gives an overview of the means and standard deviations of all six strategy groups. The overall mean of the use of strategies as estimated with the Estonian version of SILL was 3.23, with a standard deviation of 1.27. Considering the ranges of low (less than 2.5), medium (2.5–3.5) and high (over 3.5) frequency, the overall use of LLSs in the case of Estonian EFL learners is medium. Social strategies as a group are used with the highest frequency with a mean of 3.54, and memory strategies with the lowest (2.85).

Table 3. The means of strategy groups

Strategies	Overall		Gender				Age groups					
			Female		Male		...-20		21-30		31-...	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Memory	2.85	1.32	2.88	1.32	2.68	1.31	2.79	1.28	2.87	1.34	2.86	1.29
Cognitive	3.33	1.23	3.35	1.23	3.22	1.22	3.17	1.19	3.43	1.22	3.25	1.27
Compensation	3.42	1.27	3.42	1.30	3.43	1.14	3.30	1.27	3.58	1.25	3.14	1.27
Metacognitive	3.34	1.19	3.37	1.19	3.21	1.20	3.33	1.19	3.39	1.19	3.24	1.18
Affective	2.90	1.30	2.92	1.30	2.80	1.31	2.86	1.29	2.97	1.32	2.75	1.25
Social	3.54	1.17	3.56	1.18	3.43	1.11	3.48	1.15	3.63	1.15	3.37	1.22
Overall means	3.23	1.27	3.25	1.27	3.13	1.25	3.16	1.24	3.31	1.27	3.10	1.27

When analysing the items separately, 20 items out of 49 (41%) have got a high range of use, 24 (49%) medium and 5 (10%) low. The highest scores go to using a similar word or phrase if a foreign language word cannot be thought of (4.26), asking someone to slow down or to say it again if something cannot be understood (4.12) and paying attention when someone is speaking in foreign language (4.08). The lowest scores go to physically acting out new foreign language words (1.52), using rhymes to remember new words (2.01) and using flashcards to remember new words (2.02).

The comparison of means based on gender gives rather predictable results, with females having slightly higher points than males. Similar results have been reported in several studies (Ehrman, Oxford 1989, Green, Oxford 1995, Hong-Nam, Leavell 2006). The only category where Estonian males slightly outperform females

is that of compensation strategies (Table 3). To exclude the possibility that women's predominance of strategy use has been caused by a bias of the sample to women, independent-samples t-test was conducted. However, the t-test did not reveal any significant differences ($p \geq .05$) and that is why they have not been reported here in greater detail.

The means of strategies in the comparison of three age groups (≤ 20 , 21–30, ≥ 31) are a little surprising. The EFL students aged from 21 to 30 outperform the other age groups in all strategy groups. The result may be considered surprising because the frequency of using, for example, cognitive and metacognitive strategies is usually expected to increase with age (Kolić-Vehovec et al. 2010, Khezrlou 2012) but these results do not support this. However, the period between 21 and 30 can still be considered the most active time of learning, when young people usually do not have any significant family or professional responsibilities as yet.

3.3. Exploratory factor analysis

For exploratory factor analysis, principal axis factoring, varimax rotation was conducted. According to the Kaiser rule of eigenvalues, up to a 14-factor analysis could have been conducted. The Kaiser rule is applied to drop the components with eigenvalues under 1.0 – this is the eigenvalue equal to the information accounted for by an average single item (Larsen, Warne 2010). We chose the 2-, 6- and 9-factor analysis (with eigenvalues of 3.48, 1.59 and 1.35 respectively) to test the hypotheses; the first two with the purpose of checking the possible factor structures according to the division of Oxford's classification (1990: 16), and the last one with the purpose of making it possible to compare the results with the analyses conducted by the researchers in Puerto Rico, Taiwan, China, Japan, Egypt and the United States (Oxford, Burry-Stock 1995).

To test the first hypothesis on two-factor structure of the SILL, exploratory factor analysis on two factors was conducted. Factor loadings greater than or equal to 0.4 accounted for over 25% of the variance that explains just a quarter of the strategy use being represented by the items in the SILL. In the case of the two-factor structure, it might have been assumed that two big strategy groups – direct and indirect ones – would form. Results of the factor analysis did not support this. The groups that formed were mixed, having 17 items of direct strategies and 7 items of indirect strategies in one factor, with the items for the other factor being 12 and 13 respectively. Therefore, it has to be admitted that the analysis on two factors did not give a satisfactory result and the first hypothesis is rejected.

To test the second hypothesis on six-factor structure of the SILL, exploratory factor analysis on six factors was conducted. Factor loadings greater than or equal to 0.4 accounted for over 46.3% of the variance, which explains a little less than half of the strategy use being represented by the items in the SILL. The factors that formed do not coincide with the taxonomy proposed by Oxford (1990: 16), however the result could be considered interesting.

Factor 1 characterizes the independent language learner with active language use mostly outside the classroom. Factor 2 comprises predominantly metacognitive strategies. Factor 3 covers strategies characterizing analytical activities like looking

for patterns and similarities between languages, dividing the words into parts and predicting the meaning. Factor 4 covers mostly social strategies. Factor 5 contains mostly cognitive strategies and Factor 6 forms the combination of memory, social and affective strategies. In brief, it has to be admitted that the six factors revealed in the current study do not coincide with the original taxonomy presented by Oxford (1990: 16) and therefore the second hypothesis on the six-factor structure also has to be rejected.

To test the third hypothesis on nine-factor structure of the SILL, exploratory factor analysis on nine factors was conducted. Factor loadings greater than or equal to 0.4 accounted for over 52% of the variance, which explains over half of the strategy use being represented by the items in the SILL. To make the factors formed in this analysis comparable with other similar studies (Oxford, Burry-Stock 1995), the following overview is presented, similar to the ones given for Puerto Rico, China, Japan, Egypt, Taiwan and the US (Table 4).

The Estonian study produced the following factors.

Factor 1 consisted basically of the strategies of active language use (similar to Puerto Rico, China, Japan and the US (Oxford, Burry-Stock 1995)) including reading for pleasure, watching TV and films in English, writing notes and messages and looking for opportunities to read as much as possible.

Factor 2 stood for metacognitive planning (similar to Puerto Rico, China and the US (Oxford, Burry-Stock 1995)), covering four strategies out of the nine in SILL Part D and comprising thinking about the progress in learning a foreign language, having clear goals for improving skills, being determined to become a better language learner and planning the schedule.

Factor 3 comprised the social strategies (partly similar to Puerto Rico, China and Egypt (Oxford, Burry-Stock 1995)) like asking for help from other foreign language speakers, asking others to correct errors while talking, asking others to slow down or say it again, asking questions in foreign language, practising it with other students and looking for people to talk to.

Factor 4 on analysis included finding patterns, dividing the word into parts to find its meaning, thinking of relationships between new and old and looking for words in one's own language resembling new words.

Factor 5 on cognitive and memory strategies included using foreign language words in different ways, summarizing information heard or read, using new words in a sentence to memorize the words better, remembering new words by using mental pictures, trying to talk like native speakers and practising the sounds of the foreign language.

Factor 6 (partly similar to Japan (Oxford, Burry-Stock 1995)) was mostly made up of the metacognitive, social and compensation strategies of paying attention when someone is speaking, asking the person to slow down or say it again if the sentence cannot be understood, noticing mistakes and trying to do better, and using synonyms if the right word cannot be thought of.

Factor 7, affective strategies, resembled the research results in Taiwan (Oxford, Burry-Stock 1995) and included noticing tension when using the foreign language, rewarding oneself in the case of success, talking to others about the feelings when the foreign language is learned and trying to relax when being afraid to use the foreign language.

Factor 8, covering the strategies of repetition and revision, included saying or writing new words several times and reviewing foreign language lessons.

Factor 9, sensory memory strategies, covered using rhymes and connecting the sound and image to remember new words, and physically acting out new words (similar to Egypt (Oxford, Burry-Stock 1995)).

As the table 4 reveals, the factors that evolved in the factor analysis of the Estonian version of the SILL are somewhat comparable with the results of other studies in different cultures. The grey cells in the table indicate the factors that coincide with the factors of these studies (Oxford, Burry-Stock 1995). Even though they do not present the best taxonomy for the LLSs, there are certain concurrencies with the outcomes of other similar studies. Therefore, it can be stated on the basis of the results of the exploratory factor analysis on nine-factor structure, the third hypothesis cannot totally be rejected.

4. Discussion

In this study we sought to investigate whether the SILL, translated and adapted for Estonian EFL learners, reflected two-, six- or nine-construct classification, and compare the results with other similar studies conducted in different cultures. Proceeding from that three hypotheses were tested: (H1) the structure of the SILL translated and adapted for Estonian learners reflects the two-factor structure of direct and indirect strategies; (H2) the structure of the SILL translated and adapted for Estonian learners reflects the six-factor structure of memory, cognitive, compensation, metacognitive, affective and social strategies; (H3) the structure of the SILL translated and adapted for Estonian learners reflects the nine-factor structure of LLSs comparable to the outcome of Oxford and Burry-Stock (1995).

The results revealed that the exploratory factor analysis used to test Oxford's two-construct and six-construct taxonomy of the SILL did not provide a fully acceptable fit to the data, and therefore H1 and H2 were rejected. This has been explained with high correlations among the constructs (Park 2011). In the case of the current study, the sample ($n = 337$) could be considered sufficient according to Zhao (2009). The nine-construct analysis provided nine factors relevant for comparison with the results of the study conducted by Oxford and Burry-Stock (1995) on Puerto Rico, Taiwan, China, Japan, Egypt and the US, and therefore H3 was not rejected. The factor structure of Estonian research has most overlappings with Puerto Rico and PR China; the least with Taiwan and Japan (Table 4).

When comparing the data on gender differences collected within the study, it can be said that similarly to the results of many other studies (Wong 2011, Green, Oxford 1995), Estonian females demonstrated a slightly more frequent use of LLSs compared to males. Although many researchers have explained the gender differences within cultural or ethnical contexts (Oxford, Nyikos 1989), few have given a substantial explanation for these distinctions. When looking for reasons why Estonian females use more LLSs than males, we should look at the statistics describing education in Estonia. Estonian women seem to be valuing education and higher qualifications more than men, and are probably have more aspirations

Table 4. The comparison of results of 9-factor analysis between Estonia and six countries compared by Oxford and Burry-Stock (1995)

Factor/ location	Estonia	Puerto Rico	Taiwan	PR China	Japan	Egypt	Combined US
1	Active language use	Active language use	Metacognitive planning	Active language use	Active language use	Metacognitive planning	Active language use
2	Metacognitive planning	Metacognitive planning	Active language use	Metacognitive planning	Sensory memory strategies	Sensory memory strategies	Metacognitive planning
3	Social	Affective and social	Memory and analysis	Affective and social	Metacognitive/social/affective	Affective and social	Affective strategies
4	Analysis	Reflection (analysis and anxiety)	Formal oral practice	Sensory memory strategies	Compensation and analysis	Active language use	Sensory memory strategies
5	Cognitive and memory	Sensory memory strategies	Social strategies	Compensation in reading	Formal oral practice	Request and repetition	Social strategies
6	Metacognitive/social/compensation	Social/cognitive conversation	Compensation in reading	Metacognitive and affective	Affective strategies	Sensory memory and anxiety	Compensation and analysis
7	Affective strategies	Sensory (visual) memory	Affective strategies	Sensory (visual) memory	Compensation in speaking	Compensation in reading and listening	Metacognitive planning
8	Repetition and revision	Cognitive and relaxation	Compensation in speaking	Attention to key details	Attention to key details	General memory strategies	General memory strategies
9	Sensory memory strategies	General compensation	General memory strategies	General memory strategies	Reflection (analysis and anxiety)	Sensory memory strategies	Compensation and nonanalytic

when acquiring education. This has been confirmed at the level of secondary education already, and in tertiary education even more (Leoma 2011). According to the data presented by Eurostat (2013), the gender gap in education in Estonia is one of the biggest in European countries. The percentage of 30–34-year-old males having successfully completed tertiary education is 28.1%, whereas for the same age range in females it is 50.4%. The same tendency has been noted in almost all European countries (Eurostat 2013). Moreover, this explanation is supported by the sample division of the present research as well having only 20% of the sample male and 80% female. Males' relatively small presence in tertiary education does not determine the success of the country, but shows what may be the educational balance of males and females in the country in the future (Leoma 2011). The greater use of LLSs by women in Estonia could also be related to their bigger commitment to acquiring education.

At the same time, it is interesting to see how language teaching methodology has changed in Estonia over time. During the Soviet period, the grammar-translation method was mainly used. Due to its behaviouristic theoretical base – habit formation via repetition and reinforcement – it supported using mostly memory and cognitive strategies. Developing analytical skills was not encouraged, which led to metacognitive strategies being underexploited. As the grammar-translation method does not support developing active language use, people felt tense and nervous when they had to communicate in English. At present, the situation outside schools has changed a lot – borders are open, there is a tremendous information explosion and people have many opportunities to use the language – and that is one of the reasons why young people are highly motivated to learn English. This has caused the change in the use of LLSs – the role of social and compensation strategies has grown and metacognitive strategies are accruing gradually.

5. Limitations of the study and recommendations for further research

The first limitation of the current study is the uneven sample because not all of the respondents were active EFL learners at the moment of questioning – some students had actively participated in EFL courses some time ago; some were currently learning. That may have affected the reliability of students' reporting on their learning strategies to a small extent, but certainly not so much that it influenced the overall results as, when studying at the tertiary level, students are still expected to work with English study materials and articles. This, and the fact that a large amount of data was collected electronically, will lead us to the second limitation, which is the lack of opportunity to check the validity of the instrument related to language proficiency. Comparing the results according to language proficiency would have given a better picture in the comparison with other similar studies. Thirdly, as measuring language proficiency was not included in the study, it was not possible to assess the efficiency of strategy use but only the frequency of strategy use.

Despite these limitations, the study has contributed to clarification of the factor model of LLSs. Having many overlaps with the nine-factor model described by Oxford and Burry-Stock (1995), it provides good material for comparison with

similar studies conducted before. But, although we consider it reasonable to continue gathering similar data from different countries as it would shed light on cultural and regional features, it has also turned out to be necessary to reinvestigate and restructure the existing taxonomy as many studies have reported its unacceptable fit (Park 2011).

References

- Alhaisoni, Eid 2012. Language learning strategy use of Saudi EFL students in an intensive English learning context. – *Asian Social Science*, 8 (13), 115–128. <http://dx.doi.org/10.5539/ass.v8n13p115>
- Ariza, Eileen N. 2002. Resurrecting “old” language learning methods to reduce anxiety for new language learners: Community language learning to the rescue. – *Bilingual Research Journal*, 26 (3), 717–728. <http://dx.doi.org/10.1080/15235882.2002.10162586>
- Bedell, David; Oxford, Rebecca 1996. Cross-cultural comparisons of language learning strategies in the People’s Republic of China and other countries. – Rebecca Oxford (Ed.). *Language Learning Strategies Around the World: Cross-cultural perspectives*. Honolulu: University of Hawai’i, Second Language Teaching and Curriculum Center, 47–60.
- Bell, Jill Sinclair 1995. The relationship between L1 and L2 literacy: Some complicating factors. – *TESOL Quarterly*, 29 (4), 687–704. <http://dx.doi.org/10.2307/3588170>
- Benson, Phil 2011. *Teaching and Researching Autonomy*. Great Britain: Pearson Education Ltd.
- Bialystok, Ellen 1979. The role of conscious strategies in second language proficiency. – *Modern Language Journal*, 65 (1), 24–35. <http://dx.doi.org/10.1111/j.1540-4781.1981.tb00949.x>
- Cronbach, Lee; Meehl, Paul 1955. Construct validity in psychological tests. – *Psychological Bulletin*, 52 (4), 281–302. <http://dx.doi.org/10.1037/h0040957>
- Douglas, Dan 2001. Performance consistency in second language acquisition and language testing research: A conceptual gap. – *Second Language Research*, 17 (4), 442–456. <http://dx.doi.org/10.1191/026765801681495903>
- Ehrman, Madeline; Oxford, Rebecca 1989. Effects of sex differences, career choice and psychological type on adult language learning strategies. – *The Modern Language Journal*, 73 (1), 1–13. <http://dx.doi.org/10.1111/j.1540-4781.1989.tb05302.x>
- Ehrman, Madeline; Oxford, Rebecca 1990. Adult language learning styles and strategies in an intensive training setting. – *The Modern Language Journal*, 74 (3), 311–327. <http://dx.doi.org/10.1111/j.1540-4781.1990.tb01069.x>
- Ellis, Rod 2013. Changing trends in language teaching research. – *Language Teaching Research*, 17 (2), 141–143. <http://dx.doi.org/10.1177/1362168812460807>
- Eurostat News Release. Europe 2020 headline indicators on education. http://epp.eurostat.ec.europa.eu/cache/ITY_PUBLIC/3-11042013-BP/EN/3-11042013-BP-EN.PDF (5.7.2013).
- Fazeli, Seyed Hossein 2012. The psychometric analysis of the Persian version of the Strategy Inventory for Language Learning of Rebecca L. Oxford. – *Indian Journal of Science and Technology*, 5 (4), 2638–2644.
- Frankfort-Nachmias, Chava; Leon-Guerrero, Anna 2011. *Social Statistics for a Diverse Society*. Sage Publications Ltd.
- Green, John M.; Oxford, Rebecca 1995. A closer look at learning strategies, L2 proficiency, and gender. – *TESOL Quarterly*, 29 (2), 261–297. <http://dx.doi.org/10.2307/3587625>
- Griffiths, Carol 2003. Language learning strategy use and proficiency: The relationship between patterns of reported LLS use by speakers of other languages and proficiency with implications for the teaching/learning situation. – University of Auckland. <https://>

researchspace.auckland.ac.nz/bitstream/handle/2292/9/02whole.pdf?sequence=6
(26.6.2013).

- Griffiths, Carol 2004. Language learning strategies: Theory and research. http://www.crie.org.nz/research-papers/c_griffiths_op1.pdf (22.6.2013).
- Griffiths, Carol; Parr, Judy 2000. Language learning strategies, nationality, independence and proficiency. – *Independence*, 28 (1), 7–10.
- Guillemin, Francis; Bombardier, Claire; Beaton, Dorcas 1993. Cross-cultural adaptation of health-related quality of life measures: Literature review and proposed guidelines. – *Journal of Clinical Epidemiology*, 46 (12), 1417–1432. [http://dx.doi.org/10.1016/0895-4356\(93\)90142-N](http://dx.doi.org/10.1016/0895-4356(93)90142-N)
- Hong-Nam, Kyungsim; Leavell, Alexandra 2006. Language learning strategy use of ESL students in an intensive English learning context. – *System*, 34 (3), 399–415. <http://dx.doi.org/10.1016/j.system.2006.02.002>
- Hsiao, Tsung-Yuan; Oxford, Rebecca 2002. Comparing theories of language learning strategies: A confirmatory factor analysis. – *The Modern Language Journal*, 86 (3), 368–383. <http://dx.doi.org/10.1111/1540-4781.00155>
- Jiamu, Chen 1997. The right methods for adults: On English acquisition/ learning. – *Análise Psicológica*, 15 (4), 587–593.
- Jones, Sabine 1998. Learning styles and learning strategies: Towards learner independence. – *Forum for Modern Language Studies*, 34 (2), 114–129. <http://dx.doi.org/10.1093/fmls/xxxiv.2.114>
- Khezrlou, Sima 2012. The relationship between cognitive and metacognitive strategies, age, and level of education. – *Reading Matrix: An International Online Journal*, 12 (1), 50–61.
- Kolić-Vehovec, Svjetlana; Bajšanski, Igor; Zubković, Barbara 2010. Metacognition and reading comprehension: Age and gender differences. – Efkliides Anastasia, Plousia Misailidi (Eds.). *Trends and Prospects in Metacognition Research*. New York: Springer Science + Business Media, 327–344.
- Kreyszig, Erwin 1979. *Applied Mathematics*. Wiley Press.
- Larsen, Ross; Warne, Russell 2010. Estimating confidence intervals for eigenvalues in exploratory factor analysis. – *Behavior Research Methods*, 42 (3), 871–876. <http://dx.doi.org/10.3758/BRM.42.3.871>
- Lengkanawati, Nenden Sri 2004. How learners from different cultural backgrounds learn a foreign language. – *Asian EFL Journal*, 6 (1), 1–8.
- Leoma, Rain 2011. In education women spurt, men jog? – Kairit Pöder (Ed.). *Man's home is the world, Woman's world is her home?* Tallinn: Statistikaamet, 100–110.
- Lynn, Richard; Mikk, Jaan 2009. Sex differences in reading achievement. – *Trames*, 13 (1), 3–13. <http://dx.doi.org/10.3176/tr.2009.1.01>
- Magno, Carlo 2010. Korean students' language learning strategies and years of studying English as predictors of proficiency in English. – *TESOL International Journal*, 2, 39–61.
- O'Malley, J. Michael; Chamot, Anna Uhl 1990. *Learning Strategies in Second Language Acquisition*. Cambridge: Cambridge University Press.
- Oxford, Rebecca 1990. *Language Learning Strategies: What Every Teacher Should Know*. New York: Newbury House/Harper and Row.
- Oxford, Rebecca 1996. *Language Learning Strategies Around the World: Cross-Cultural Perspectives*. Manoa: University of Hawai'i Press.
- Oxford, Rebecca; Burry-Stock, Judith A. 1995. Assessing the use of language learning strategies worldwide with the ESL/EFL version of the Strategy Inventory of Language Learning (SILL). – *System*, 23 (1), 1–23. [http://dx.doi.org/10.1016/0346-251X\(94\)00047-A](http://dx.doi.org/10.1016/0346-251X(94)00047-A)
- Oxford, Rebecca; Crookall, David 1989. Research on language learning strategies: methods, findings, and instructional issues. – *The Modern Language Journal*, 73 (4), 404–419. <http://dx.doi.org/10.1111/j.1540-4781.1989.tb05321.x>

- Oxford, Rebecca; Ehrman, Madeline 1995. Adults' language learning strategies in an intensive foreign language program in the United States. – *System*, 23 (3), 359–386. [http://dx.doi.org/10.1016/0346-251X\(95\)00023-D](http://dx.doi.org/10.1016/0346-251X(95)00023-D)
- Oxford, Rebecca; Nyikos, Martha 1989. Variables affecting choice of language learning strategies by university students. – *Modern Language Journal*, 73 (3), 291–300. <http://dx.doi.org/10.1111/j.1540-4781.1989.tb06367.x>
- Oxford, Rebecca; Nyikos, Martha; Ehrman, Madeline 1988. Vive la difference? Reflections on sex differences in use of language learning strategies. – *Foreign Language Annals*, 21 (4), 321–329. <http://dx.doi.org/10.1111/j.1944-9720.1988.tb01076.x>
- Park, Gi-Pyo 2011. The validation process of the SILL: A confirmatory factor analysis. – *English Language Teaching*, 4 (4), 21–27. <http://dx.doi.org/10.5539/elt.v4n4p21>
- Politzer, Robert L. 1983. An exploratory study of self-reported language learning behaviors and their relation to achievement. – *Studies in Second Language Acquisition*, 6 (1), 54–65. <http://dx.doi.org/10.1017/S0272263100000292>
- Politzer, Robert L.; McGroarty, Mary 1985. An exploratory study of learning behaviours and their relationship to gains in linguistic and communicative competence. – *TESOL Quarterly*, 19 (1), 104–124. <http://dx.doi.org/10.2307/3586774>
- Porte, Graeme 1988. Poor language learners and their strategies for dealing with new vocabulary. – *ELT Journal*, 42 (3), 167–171. <http://dx.doi.org/10.1093/elt/42.3.167>
- Rose, Heath 2012. Language learning strategy research: Where do we go from here? – *Studies in Self-Access Learning Journal*, 3 (2), 137–148.
- Rubin, Joan 1975. What the “good language learner” can teach us. – *TESOL Quarterly*, 9 (1), 41–51. <http://dx.doi.org/10.2307/3586011>
- Stern, H. H. 1975. What can we learn from the good language learner? – *Canadian Modern Language Review*, 31, 304–318.
- Tragant, Elsa; Thompson, Marilyn; Victori, Mia 2013. Understanding foreign language learning strategies: A validation study. – *System*, 41 (1), 95–108. <http://dx.doi.org/10.1016/j.system.2013.01.007>
- Tran, Thanh 1988. Sex differences in English language acculturation and learning strategies among Vietnamese adults aged 40 and over in the United States. – *Sex Roles*, 19 (11–12), 747–758. <http://dx.doi.org/10.1007/BF00288990>
- Turner, Julianne 1995. The influence of classroom contexts on young children's motivation for literacy. – *Reading Research Quarterly*, 30 (3), 410–441. <http://dx.doi.org/10.2307/747624>
- Usuki, M. 2000. A new understanding of Japanese students views on classroom learning. – *Independence*, 27, 2–6.
- Veenman, Marcel; Hesselink, Rob; Sleeuwaegeen, Shannon; Liem, Sophie; Haaren, Marieke 2014. Assessing developmental differences in metacognitive skills with computer logfiles: Gender by age interactions. – *Psychological Topics*, 23 (1), 99–113.
- Wong, Mary Siew-Lian 2011. Language learning strategy use: A study of pre-service teachers in Malaysia. <http://www.eric.ed.gov/PDFS/ED521415.pdf> (13.6.2013).
- Woodrow, Lindy 2005. The challenge of measuring language learning strategies. – *Foreign Language Annals*, 38 (1), 90–98. <http://dx.doi.org/10.1111/j.1944-9720.2005.tb02456.x>
- Yang, N-D. 1998. An interviewing study of college students' English learning strategy use. – *Studies in English Language and Literature*, 4, 1–11.
- Zhao, Nathan 2009. The minimum sample size in factor analysis. <http://www.encorewiki.org/display/~nzhao/The+Minimum+Sample+Size+in+Factor+Analysis> (24.7.2013).

Appendix. Strategy Inventory for Language Learning (SILL)

This form of the strategy inventory for language learning (SILL) is for students of a foreign language (FL). Please read each statement and circle the score of the response (1, 2, 3, 4, or 5) that tells HOW TRUE THE STATEMENT IS.

1. Never or almost never true of me
2. Usually not true of me
3. Somewhat true of me
4. Usually true of me
5. Always or almost always true of me

Answer in terms of how well the statement describes you. Do not answer how you think you should be, or what other people do. **There are no right or wrong answers** to these statements.

Part A		Score				
1	I think of relationships between what I already know and new things I learn in the SL	5	4	3	2	1
2	I use new SL words in a sentence so I can remember them	5	4	3	2	1
3	I connect the sound of a new SL word and an image or picture of the word to help me remember the word	5	4	3	2	1
4	I remember a new SL word by making a mental picture of a situation in which the word might be used	5	4	3	2	1
5	I use rhymes to remember new SL words	5	4	3	2	1
6	I use flashcards to remember new SL words	5	4	3	2	1
7	I physically act out new SL words	5	4	3	2	1
8	I review SL lessons often	5	4	3	2	1
9	I remember new SL words or phrases by remembering their location on the page, on the board, or on a street sign	5	4	3	2	1
Part B		Score				
10	I say or write new SL words several times	5	4	3	2	1
11	I try to talk like native SL speakers	5	4	3	2	1
12	I practice the sounds of SL	5	4	3	2	1
13	I use the SL words I know in different ways	5	4	3	2	1
14	I start conversations in the SL	5	4	3	2	1
15	I watch SL language TV shows spoken in SL or go to movies spoken in SL	5	4	3	2	1
16	I read for pleasure in the SL	5	4	3	2	1
17	I write notes, messages, letters, or reports in the SL	5	4	3	2	1
18	I first skim an SL passage (read over the passage quickly) then go back and read carefully	5	4	3	2	1
19	I look for words in my own language that are similar to new words in the SL	5	4	3	2	1
20	I try to find patterns in the SL	5	4	3	2	1
21	I find the meaning of an SL word by dividing it into parts that I understand	5	4	3	2	1

22	I try not to translate word for word	5	4	3	2	1
23	I make summaries of information that I hear or read in the SL	5	4	3	2	1
Part C		Score				
24	To understand unfamiliar SL words, I make guesses	5	4	3	2	1
25	When I can't think of a word during a conversation in the SL, I use gestures	5	4	3	2	1
26	I make up new words if I do not know the right ones in the SL	5	4	3	2	1
27	I read SL without looking up every new word	5	4	3	2	1
28	I try to guess what the other person will say next in the SL	5	4	3	2	1
29	If I can't think of an SL word, I use a word or phrase that means the same thing	5	4	3	2	1
Part D		Score				
30	I try to find as many ways as I can to use my SL	5	4	3	2	1
31	I notice my SL mistakes and use that information to help me do better	5	4	3	2	1
32	I pay attention when someone is speaking SL	5	4	3	2	1
33	I try to find out how to be a better learner of SL	5	4	3	2	1
34	I plan my schedule so I will have enough time to study SL	5	4	3	2	1
35	I look for people I can talk to in SL	5	4	3	2	1
36	I look for opportunities to read as much as possible in SL	5	4	3	2	1
37	I have clear goals for improving my SL skills	5	4	3	2	1
38	I think about my progress in learning SL	5	4	3	2	1
Part E		Score				
39	I try to relax whenever I feel afraid of using SL	5	4	3	2	1
40	I encourage myself to speak SL even when I am afraid of making a mistake	5	4	3	2	1
41	I give myself a reward or treat when I do well in SL	5	4	3	2	1
42	I notice if I am tense or nervous when I am studying or using SL	5	4	3	2	1
43	I write down my feelings in a language learning dairy	5	4	3	2	1
44	I talk to someone else about how I feel when I am learning SL	5	4	3	2	1
Part F		Score				
45	If I do not understand something in SL, I ask the other person to slow down or say it again	5	4	3	2	1
46	I ask SL speakers to correct me when I talk	5	4	3	2	1
47	I practice SL with other students	5	4	3	2	1
48	I ask for help from SL speakers	5	4	3	2	1
49	I ask questions in SL	5	4	3	2	1
50	I try to learn about the culture of SL speakers	5	4	3	2	1

Katrin Saks (University of Tartu) is an English philologist, she defended the Master's thesis in Education. In her doctoral research she studies the ways of supporting learners' language learning strategies and self-regulation in technologically enhanced learning environments.
University of Tartu Pärnu College, Ringi 35, 80010 Pärnu, Estonia
katrin.saks@ut.ee

Äli Leijen (University of Tartu) defended her Master's thesis and Doctoral dissertation in Education. Her research topics focus on reflection and developing student-teachers' professional identity.
University of Tartu, Institute of Education, Salme 1a, r. 29, 50103 Tartu, Estonia
ali.leijen@ut.ee

Kandela Õun (University of Tartu) defended her Master's thesis in applied statistics. Her research topics are connected with compiling constructs and their measurement.
University of Tartu Pärnu College, Ringi 35, 80010 Pärnu, Estonia
kandela.oun@ut.ee

Karin Täht (University of Tartu) defended her Master's thesis in mathematics and Doctoral dissertation in psychology. Her research topics are connected with students' learning motivation in different cultures.
University of Tartu, Department of Individual and Social Psychology, Näituse 2, 50409 Tartu, Estonia
karin.taht@ut.ee

KEELEÕPPESTRATEEGIATE MÕÕTEVAHENDI (SILL) FAKTORSTUKTUURI UURING: SILL-I ADAPTEERIMINE EESTI INGLISE KEELE ÕPPIJATE JAKS, KEELEÕPPESTRATEEGIATE KASUTAMISE KULTUURILISED ISEÄRASUSED

**Katrin Saks, Äli Leijen,
Kandela Öun, Karin Täht**

Tartu Ülikool

Keeleõpe on kaasaegses globaliseerivas maailmas muutunud iseenesestmõistetavaks sõltumata õppijate taustast, asukohast või elualast. Keeleõppe edukust saab hinnata erinevalt, kuid üheks enamlevinud viisiks on teha seda õppijate keeleõppestrateegiatega hindamise kaudu (Hsiao, Oxford 2002). SILL on maailmas enim kasutatud enesekohane küsimustik, millega mõõdetakse inglise keele kui võõrkeele õppimisel kasutatavaid õpistrateegiaid. See põhineb Oxfordi (1990) loodud taksonoomial, mis eristab kaht suuremat strateegiatega gruppi (otsesed ja kaudsed) ning 6 alagruppi: mälu-, kognitiivsed, kompensatsiooni, metakognitiivsed, afektiivsed ja sotsiaalsed strateegiad. Ometi on mõned uuringud näidanud 2- ja 6-faktorilise struktuuri sobimatust keeleõppestrateegiatega mõõtmiseks (Park 2011). Käesoleva uuringu eesmärgiks on panustada keeleõppestrateegiatega taksonoomia ja selle mõõtevahendi SILL paremasse mõistmisse selle eestikeelse versiooni kohandamisprotsessi kaudu.

Eesti keelde tõlgitud ja kohandatud küsimustikku hinnati 337 üliõpilasest koosneva valimiga. Viidi läbi uuriv faktoranalüüs, mille tulemused ei toetanud ei 2- ega 6-faktorilist struktuuri, mida eeldanuks originaalstruktuur. 9-faktorilise analüüsi tulemused pakuvad huvitavaid võrdlusmomente sarnase uuringuga, mis viidi läbi 1995. aastal (Oxford, Burry-Stock 1995). Lisaks Oxfordi pakutud kuuele faktorile (1990), eristusid käesolevas uuringus aktiivse keelekasutuse strateegiad, mis on omased kommunikatiivsele keeleõppele, samuti analüüsiioskust eeldavad ja uute struktuuride loomisel varasemate teadmiste aktiveerimist eeldavad strateegiad. Mälustrateegiad jagunesid kahte eraldiseisvasse faktorisse: sensoorsed mälustrateegiad ja kordamisstrateegiad.

Võtmesõnad: uuriv faktoranalüüs, küsimustik, reliaablus, t-test, inglise keel, eesti keel

EVIDENTSAALSUSE JA EPISTEMILISE MODAALSUSE SUHETEST EESTI LASTEKEELES KÄITUMISKATSETE PÕHJAL

Anne Tamm, Reili Argus, Airi Kapanen,
Andra Kütt, Kadri Suurmäe

Ülevaade. Evidentsiaalsuse ja episteemilise modaalsuse sarnasused ja erinevused on viimasel aastakümneil olnud mitme uuringu aluseks. Eesti *vat*-morfeemi (*ole-vat*) kohta on arvatud, et see keeleüksus võib kanda kaht tähendust: nii evidentsiaalset kui ka episteemilis-modaalset. Uurisime selles artiklis, kuivõrd tõlgendati *vat*-morfeemi episteemilise modaalina. Täpsemalt selgitasime eesti laste käitumiskatsetega, mis on evidentsiaalide olemuses sellist, et nad on seotud episteemilise modaalsusega. Katsetest selgus, et evidentsiaalse markerina omandatakse *vat*-morfeem enne üheksandat eluaastat. Episteemilise modaalina tõlgendamine üheksandaks eluaastaks vaid kasvab, seega omandatakse eesti *vat*-morfeemi tähendustest kõigepealt evidentsiaalsus ja alles siis episteemiline modaalsus.*

Võtmesõnad: lastekeel, semantika, pragmaatika, morfoloogia, meeleteooria, eksperimentaalsed meetodid, eesti keel

1. Sissejuhatus

Artikkel uurib episteemilise modaalsuse ja evidentsiaalsuse seoseid eesti keele *vat*-morfeemi tähenduses ja kasutuses. Evidentsiaalsus on keeleliselt vormistatud viide info allikale ning evidentsiaamid seega grammatilised markerid, mis näitavad, kas öeldu kohta on olemas tõendus ja milline see on (Aikhenvald 2004: 3). Episteemilise modaalsuse kategooria hõlmab lausungiga väljendatud propositsiooni tõenäosushinnangut (Palmer 1986: 10–11).

Kuna eesti keeles on mitme kategooria seostumine pigem reegel kui erand (nt aspekt ja sihtis, käändelised verbivormid ning aja- ja aspektikategooria, vt nt Tamm 2012, 2011a), siis on episteemilise modaalsuse ja evidentsiaalsuse seos ootuspärane

* Täname lapsi, lapsevanemaid ja personali Tallinna lasteaedadest Karikakar, Kiikhobu, Kullerkupu, Mikumanni, Männi ja Piriita-Kose, Viljandi lasteaiaist Mesimumm, Värska lasteaiaist, Saku valla lasteaedadest Päikesekild ja Terake, Rocca al Mare koolist ja Tallinna Reaalkoolist. Täname Kesk-Euroopa Ülikooli uurimistoetuse CEU RSS F-1010-3.1 "Evidentsiaalsus ja geneerilise sotsiaalkognitiivse arengu seisukohast" (2013) eest. Autorite panused: teoreetiline raam R. Argus ja A. Tamm; katsete koostaja A. Tamm; katsete tegemine A. Kapanen, A. Kütt ja K. Suurmäe; analüüs R. Argus, A. Tamm, A. Kütt ja K. Suurmäe. Täname ka katsete läbiviimise eest Sirli Zuppingut ja varustuse eest Gergely Csibrat.

nähtus, mille täpseks uurimiseks on vaja leida metoodika. Evidentsiaalsuse ja episteemilise modaalsuse seoseid on katseliselt vähe uuritud. Eesti keel avab harukordse võimaluse uurida neid nähtusi korraga.

Vaatlesime kõigepealt põimunud tähenduste olemust korpuses (interneti juhuotsingu abil) ja koostasime seejärel kaks katset, kus võrdlesime laste ütlusi info vahendatuse kohta ja laste valikuid, mis peegeldavad nende tõenäosushinnanguid. Hüpoteesiks on, et kaudsus on *vat*-i¹ kasutamise eeltingimus ja kasutus ennast tingib vajadus edastada kuulajale, et propositsiooni tõenäosus ei ole piisav. Oletame, et evidentsiaalsus ja episteemiline modaalsus on erinevad kategooriad, ent nende tähendused ja kasutus on põimunud.

2. Evidentsiaalsus ja modaalsus

Evidentsiaalsuse varasemad käsitlused, mis tuginevad põhiliselt Euroopa keeltele (nt Palmer 1996: 66–67), on paigutanud evidentsiaalsuse episteemilise modaalsuse alla. Alates Aikhenvaldist (2004), kes on esitanud näiteid vähetuntud keelte kohta, lahutatakse need kaks kategooriat teineteisest. Nii saab kategooriaid jagada nende keelte andmete alusel, kus evidentsiaalsus ja episteemiline modaalsus on grammatiseerunud eraldi morfeemidena.

Artiklis eristame mõistmise hõlbustamiseks kaht liiki kategooriaid. Keelepetsiifiliseks grammatiliseks kategooriaks peame keeles mingi tähenduse väljendamiseks kasutatavat ja selgelt eristatavat vormielementi. Eesti keeles on selline kategooria näiteks partitiiv, mis erineb oma semantikalt ja kasutuselt saami või soome partitiivist. Selles artiklis on keelepetsiifiline kategooria morfeem *-vat* oma evidentsiaalses ja episteemilis-modaalses funktsioonis. Teine kategooria liik on keeleteadlaste konstrueeritud tajukategooriatele tuginevad mõisteväljad ehk mõistekategooriad. Osa-terviku tajukategooriale tuginev mõisteväli ei pruugi kattuda keelepetsiifilise kategooriaga, näiteks eesti keele partitiiv ei väljenda üksnes osa-terviku mõistelisi suhteid, vaid ka aspekti ja muid mõisteid (vt näiteid kahte tüüpi kategooriate kohta eesti keeles Tamm 2014). Mõistekategooriad on niisiis abstraktsed ja ei pruugi keeles esinevate grammatiliste kategooriatega üks-üheselt kattuda. Üks keelepetsiifiline grammatiline kategooria võib väljendada mitut mõistekategooriat, nagu väljendab eesti grammatiline partitiiv osa-terviku ja aspekti mõistekategooriaid. Mõni keelepetsiifiline grammatiline kategooria võib põhiliselt ka ainult üht mõistekategooriat väljendada, nagu väljendab *ei* standardeitust. Mõni teeb sedagi vaid osaliselt, näiteks abessiiv väljendab eitust, aga nihkega ning vaid osaliselt (vt täpsemalt Tamm 2011b). Mõistetüüpi kategooriateks on siinses artiklis evidentsiaalsus ja episteemiline modaalsus. Küsimus on niisiis selles, kuidas suhestuvad mõistekategooriad *evidentsiaalsus* ja *episteemiline modaalsus* ning eesti keele grammatikakategooria *-vat*.

Evidentsiaalsuse esinemist eesti keeles ei ole enam tarvis tõestada. Nähtusele on pühendatud viimasel aastakümnel artikleid ning raamatute, magistri- ja doktori-tööde olulisi osi (Erelt 2002, Metslang, Paljusalu 2002, Aikhenvald 2004, Sepper 2005, Erelt jt 2006, Kehayov 2008). Eesti keel on üks väheseid Euroopa keeli, kus on olemas grammatiline evidentsiaalsus – keeleliselt vormistuv viide info allikale (Erelt 2013: 121).

¹ Morfeemi *-vat* all mõistame ka sellega väljendatud keelepetsiifilist kategooriat.

Evidentsiaalsus võib olla otsene, näiteks kõneleja nägi või kuulis ise, st oli sündmuse vahetu tunnistaja, või kaudne, st kõneleja vahendab kelleltki kuulnud või järeldab – tema teadmine põhineb mingitel tundemärkidel (vt Aikhenvald 2004: 3, Öztürk, Papafragou 2007: 256, Erelt 2013: 121). Eesti keeles ei ole otsese evidentsiaalsuse grammatikavahendeid, kuid kaudse evidentsiaalsuse väljendami- seks saab kasutada mitut leksikaalset ja grammatilist võimalust (Erelt 2013: 121, Kehayov 2004: 829). Seda on võimalik väljendada kaudse kõneviisi vormiga (*Ta olevat kodus*), aga ka *da*-infinitiivi vormiga (moodustatud eelkõige *olema*-verbist, nt *Seal olla üht bokserit nähtud*) ning *pidama*-verbi ja *ma*-infinitiivi konstrukt- siooniga (*Ta pidi päris arukas olema*) (vt ka Erelt 2013: 122). Kõige sagedasem on kaudse kõneviisi vorm (vt kirjaliku keele kohta Sepper 2005: 85). Võib oletada, et suulises kõnes esineb ka samasuguse sagedusega *pidama*-verbi konstruktsiooni ning *olema*-verbi *da*-infinitiivi vormi.

Keelekasutust uurivaid, pragmaatika ja sotsiolingvistikaga tegelevaid keeletead- lasi ei rahulda käsitlused, mis evidentsiaalsuse ja episteemilise modaalsuse rangelt lahus hoiavad. Formaalse keeleteaduse probleemiks on aga hägusad kirjeldused, mis need kaks nähtust põhimõtteliselt koos hoiavad. On vaja seletusi ja tõendeid, kas semantiliselt on tegemist kahe kategooriaga, kui keelekasutuses ilmneb mõnel kas evidentsiaaliks või episteemiliseks modaaliiks peetud morfeemil kaks tahku korraga: evidentsiaalsus ja episteemiline modaalsus. De Haan (1999, 2005) selgitab seost tõenduse kaudu: evidentsiaaliga esitatakse tõendus ja episteemilise modaaliiga antakse sellele tõenäosushinnang. Uurides traditsioonilise grammatikakirjelduse ääremaale kuuluvate keelte kategooriate tähenduseseid, täheldab de Haan, et evidentsiaalsus sarnaneb kõneleja ja kuulaja ning sündmuse suhete markeerimise poolest pigem deiksisekategooriaga kui modaalsusega.

Eesti keele evidentsiaali ajalooline areng kõrvallauses esinevast partitiivvormist toetab oletust, et tegu on eri tähendusi ühendava morfeemiga. *vat*-i kohta on varemgi väidetud, et ta väljendab nii evidentsiaalsust kui ka episteemilist modaalsust (nt Erelt 2013). Eesti grammatikates esinev kvotatiivi kategooria ehk kaudne kõneviis vormistab nimelt kaudset teatelaadi, kus kõneleja on kolmandalt isikult kuulnud info vahendaja ning kaudse kõneviisi kasutamisega väljendab ta, et kahtleb öeldu sisus ega võta vastutust selle tõelevastavuse eest. Teate vahendatusega võib kaas- neda tõenäosushinnang, kõneleja võib tegevussituatsiooni käsitada nii reaalse kui ka irrealsena (EKG I: 82). Lisaks infoallikale saab evidentsiaalsussüsteemi vahenditega väljendada ka suhtumist informatsiooni (vt nt Kehayov 2004: 813). *vat*-morfeem väljendab nii kaudsust (seda, et info pärineb kelleltki teiselt) kui ka modaalsust (seda, et rääkija ei ole väites täielikult kindel) (vt nt Erelt 2013: 126, tõestavat testi Tamm 2009: 376–379).

Tähenduste põimumine võib (aga ei pruugi) tuleneda keelekontaktidest selliste keeltega, kus kaks kategooriat on tähenduslikult või keelekasutuse kaudu põimunud. Kontaktid võivad tugevdada kognitiivselt niigi lähedasi kategooriaid ja rikastada morfeemi semantikat. Intuiitiivselt tuleneb kategooriatevaheline seos ka kõne- situatsioonidest, suhtluse iseärasustest, mida on kirjeldanud Rätsep (1971), Erelt, Metslang ja Pajusalu (2006) ning Tamm (2008) formaalsemast aspektist. Kaudne evidentsiaalsus on tihedas seoses sotsiaalkognitiivsete võimete arenguga. Kaudsuse mõistmiseks on vaja meelelist esitust (representatsiooni), et vestluspartner on osalenud ka teises kõnesituatsioonis ja vahendab teadet. Ainult siis on võimalik

teha järeldus evidentsiaali tähenduse kohta. Evidentsiaalsuse mõistmiseks peab niisiis olema arenenud meeleteooria (ingl *Theory of Mind*, vt Premack, Woodruff 1978), st võime kujutada kellegi teise esitusi (ingl *representations*). Meeleteooria areng võimaldab esitada propositsioone teise inimese esitustes. Kuna modaalsuse puhul ei ole meeleteooriat niivõrd vaja, sest esitama peab vaid propositsioonide tõenäosuse ja mitte nende eksisteerimise teiste inimeste esitustes, võib arvata, et modaalsuse kategooriat on lastel lihtsam omandada. Samas saab oletada, et kui meeleteooria (tajus) ja evidentsiaalsus (keeles) on omandatud, siis saab sellele toetada ka modaalseid järeldusi. Modaalseks järelduseks võib olla näiteks see, et öeldu on vähem tõenäoline, kui tegemist on vahendatud infoga.

Enne laste grammatikamõistmise katsete arutelu vaatleme täiskasvanute kirjutatud ajakirjandustekstidest, kas kaudsus on pigem evidentsiaali kasutamise eeltingimus. Oletame, et kasutust tingib vajadus edastada kuulajale, et info tõenäosus on ebapiisav. Kui oletus osutub tõeseks, siis peab edasi olema võimalik katsetega selgitada, kas kaks kategooriat on koos või lahus, ja kindlaks teha, kuidas nad omavahel suhestuvad. Siinjuures on oluline uurida, kuidas kategooriate kooseksisteerimine keele omandamise käigus välja kujuneb ja mida keeles esineva kategooriaga seoses tajutakse.

3. Kasutust tingib tõenäosushinnang

Millele mõtleme eesti keele *vat*-i puhul evidentsiaalsuse ja episteemilise modaalsuse seoste all? Grammatikud on üldiselt omaks võetnud, et *-vat* on evidentsiaal, kuid sellegipoolest on selle morfeemi seosed modaalsusega tugevad. Kui ütleme, et *Kuu tiirleb ümber Maa*, siis ei toetu see teadmine kindlasti meie isiklikule tähelepanekule, vaid pigem kelleltki kuuldule või koolis õpitule. Vaatamata teabe vahendatusele ei väljenda eesti keele kasutaja seda *vat*-iga ega ütle *Kuu tiirlevat ümber Maa*. Kuigi *vat*-iga lause pole keeleliselt väär, jääb eesti keelt emakeelena kõnelevale inimesele siiski mulje, nagu kõneleja tegelikult kahtleks, et *Kuu tiirleb ümber Maa*. Peaaegu kogu meie maailma kohta käiv teave põhineb vahendatud ütlustel. Sellegipoolest väljendatakse seda eesti keeles enamjaolt kindlas kõneviisis ja ilma *vat*-morfeemita.

Tihti võib leida *vat*-i lausungitest, mis on kõneleja seisukohast väikese tõenäosusega. Analüüsime paari ajakirjandusest leitud näidet. Näites (1) leiduv *-vat* (*väited, mis tõestavat*) ei ole mõeldud selleks, et lugejat informeerida, et teooriad Estonia huku kohta on kelleltki kuuldud, vaid pigem on soovitud edasi anda, et ajakirjaniku hinnangul ei ole Estonia huku teooriatel erilist usaldusväarsust.

- (1) Ametlikest raportitest erinevad teooriad Estonia huku kohta ja väited, mis tõestavat, et valitsused varjavad tõde, elavad rahvasuus oma elu ning on äärmiselt visad kaduma isegi siis, kui leidub fakte, mis need ümber lükkavad. (<http://www.postimees.ee/2934453/10-vandenoueteoriat-estoniast-usk-toe-varjamise-pole-kadunud> 27.9.2014)

Sellised näited annavad põhjust kahelda, kas *vat*-i kasutatakse eelkõige selle märkimiseks, et info on vahendatud. Vaadeldes lauset (2) kontekstis, mis kirjeldab kaudselt saadud teavet, on näha, et *-vat* esineb lauses, kus a) teave pärineb kaudselt

allikast (FSB avalike suhete keskuselt), aga b) selles väljendub ka hinnang ütluse tõenäosusele.

- (2) Reede õhtul väitis Venemaa Föderaalne Julgeolekuteenistuse (FSB) avalike suhete keskus, et Eesti kaitsepolitsei ametnik peeti kinni Pihkva oblastis, kus ta **olevat** viinud läbi salajast operatsiooni. (<http://www.postimees.ee/2912215/konsulile-pole-endiselt-voimaldatud-kohtumist-eston-koh-veriga> 7.9.2014)

Näidetes (1) ja (2) väljendub ka huvitav kontrast. Täheldada võib, et *vat*-morfeemi kasutamist ei määra info kaudsus – sest kaudsus ei kutsu automaatselt esile tunnuse kasutamist lause ülejäänud verbivormides –, vaid pigem rõhutatult vähene tõenäosus. Kaudset teavet, et *Eesti kaitsepolitsei ametnik peeti kinni* (2) või et *väited* [---] *elavad rahvasuus oma elu* (1), evidentsiaalivormiga ei markeerita.

Kui kaudsuse edastamisel *vat*-morfeemi kohustuslikult ei rakendata, siis info väidete tõestamise ja salajase operatsiooni läbiviimise kohta on kõneleja seisukohast väiksema tõenäosusega – ajaleheartikli kirjutaja märgib seda *vat*-morfeemiga näidetes (1) ja (2).

Suhtluses, kus keegi vahendab kellelegi infot, järeldeb kuulaja ka väiksemat tõenäosust, sest teadmise allikas pole kõneleja ise. Seetõttu võiks oletada, et kaudsuse markeritel (nt *da*-infinitiiv, *-vat*) on alati episteemilise modaalsuse semantika. Vaatleme tähenduste kombineerumist teistes evidentsiaalinäidetes, nagu lauses (3).

- (3) Putin **olla** Porošenkole **öelnud**: “Kui ma tahaksin, siis võiksid Vene väed olla kahe päeva jooksul mitte ainult Kiievis, vaid ka Riias, Vilniuses, Tallinnas, Varssavis või Bukarestis.” (<http://www.postimees.ee/2925035/euroopa-liit-ei-kinnitanud-putini-vaidetavat-ahvardust> 18.9.2014)

Näide (3) (*Putin olla Porošenkole öelnud...*) näitab, et evidentsiaalsuse väljendamise grammatilised vahendid kannavad ka muud tüüpi hinnanguid. Evidentsiaalsus, mis on väljendatud *da*-infinitiiviga, on seotud negatiivse hoiakuga vahendatava, kõneleja tehtu või ütluse sisu kohta. Domineerib varjatud pahakspanu või iroonia, hinnang on seega vaid moraalne ja erinevalt *vat*-morfeemist lähtutakse tõenäosuse hindamisel ainult kontekstist. See viitab episteemilise modaalsuse elemendi puudumisele *da*-evidentsiaali semantikas. Sellest järeldeb, et kaudne evidentsiaalsus ei pruugi automaatselt põimuda episteemilise modaalsuse tähendustega. Evidentsiaalitähendusega ei seostu iseenesest episteemilis-modaalne hinnang, pigem on see hinnang üks võimalikest järeldest.

Nüüsiis ei saa kinnitust oletus, et kaudsuse väljendamise vahendid vähendavad tõenäosust. Väljendatuna *vat*-morfeemiga on täiskasvanute keelekasutuses evidentsiaalsus ja episteemiline modaalsus põimunud. Väljendatuna *da*-infinitiiviga on evidentsiaalsus seevastu vaid kontekstist sõltuvate järeldeste kaudu seotud episteemilise modaalsuse mõisteväljaga, sest tõenäosuse kohta hinnang puudub.

4. Evidentsiaalsuse ja modaalsuse omandamisest

Modaalsuse omandamise uuringutes on leitud, et lapsed omandavad deontilise modaalsuse varakult, episteemilise hilja. Deontilise modaalsuse varase omandamise taga on nähtud lapsekesksetes kõnesituatsioonides domineerivaid direktiivseid kõneakte (Stephany 1993: 39).

Kaudse evidentsiaalsuse hilise omandamise põhjaks on selle kategooria olemuslik ja kognitiivne keerukus (Öztürk, Papafragou 2007). Keeleline ja kognitiivne areng on tihedalt seotud. Kuigi väidetavalt rajavad lapsed omandatavad keelelised kategooriad kognitiivsetele (Clark 2004: 472), toimub mõju ka teises suunas, keeleline areng aitab omakorda kaasa kognitiivsele arengule (Papafragou jt 2007: 254). On arvatud, et evidentsiaalsus omandatakse hilja seetõttu, et selle kategooria omandamiseks peavad olema arenenud teatud kognitiivsed võimed (nt meeleteooria). Et saada aru, kust info tegelikult pärineb, peab laps olema võimeline eristama info allikat ja suutma siduda kaht eri ajal toimunud sündmust. Kaudse evidentsiaali puhul ei ole lapsel võimalik pelgalt kõnesituatsioonist järeldada, et kunagi varem on olnud veel üks info edastamise situatsioon, milles info edastati tegelasele, kes osaleb praegu selles kõnesituatsioonis, milles ka laps ise. Seega, kuna evidentsiaalsust väljendav marker sisaldab enne kõnehetke toimunud teabe edastamist ühelt isikult teisele – vahendatust –, on see lapse jaoks kognitiivselt keerukas.

Kognitiivse keerukuse kõrval mõjutavad kaudse evidentsiaali hilist omandamist ka pragmaatilised tegurid. Lapsekesksetes kõnesituatsioonides pole kaudse evidentsiaali vorm sage, kuna olukorrad, kus vahendatakse mujalt tulnud ja vähese tõenäosusega infot, ei ole väikese lapsega kõneledes tavalised. Ka eesti laste spontaanse kõne materjal ei paku evidentsiaali omandamise uurimiseks sobivaid keeleandmeid (vt Argus jt 2014).

Samas, evidentsiaalsust tähendust võivad lapsed hakata varakult väljendama ka keeltes, kus morfoloogilist evidentsiaali ei ole (nt inglise keele kohta vt Rett jt 2013: 11). Siiski paistab evidentsiaalide omandamine sõltuvat mingil määral sellest, kas vastav marker on keeles kohustuslik ja sage või mitte. Keeles, kus tuleb markeerida nii otsest kui ka kaudset evidentsiaalsust (türgi), omandatakse vastavad markerid varakult (Choi 1995, Aksu-Koç 1988). Kuigi neis keelis, kus evidentsiaalsust väljendavad markerid on kohustuslikud ja lapsed suudavad neid kasutada varakult, nt türgi ja korea keeles 2–3-aastaselt (Aksu-Koç 1988, Papafragou jt 2007), on nendegi keelte kohta väidetud, et täielik arusaam evidentsiaalidest tekib lastel alles palju hiljem (Öztürk, Papafragou 2007: 2) ja nii võib nende markerite täielik mõistmine tekitada probleeme veel 6-aastastelgi (Zufferey 2010: 88).

Ka eri tüüpi evidentsiaalid paistavad lapsed omandavat eri ajal: otsesed evidentsiaalid omandatakse varem (kui keeles on selline marker) kui kaudsed (Öztürk, Papafragou 2007, de Villiers jt 2009, Rett jt 2013). Türki järelduslik evidentsiaal omandatakse varakult, 1;6–2;0 aasta vanuselt, ja nn *hearsay* ehk eesti kaudsele kõneviisile vastav evidentsiaal vanuses 2;0–3;0 või isegi hiljem, umbes nelja aasta vanuselt (Slobin, Aksu 1982, Aksu-Koç, Slobin 1986, Fitneva 2001).

Kuna eesti evidentsiaalsussüsteemis on esindatud ainult eelneva kaudne evidentsiaal ning seegi on valikuline ja lapsele suunatud kõnes harv, on eesti evidentsiaali hiline omandamine igati ootuspärane. Esialgsed omandamisandmed seda ka kinnitavad. Esimestes eesti evidentsiaali omandamise uuringutes (Argus jt 2014, Tamm jt 2014)

on vaadeldud, kuidas lapsed evidentsiaaliga esitatud teabele reageerivad. Tulemused näitasid, et 4-aastaste laste hulgas tekitas kuulnud evidentsiaal huvi uurida uute esemete omadusi või uue grammatilise vormi tähendust, samas kui 6-aastastele *-vat* suuremat huvi enam ei tekitanud. Seega võib väita, et 4-aastased reageerivad kaudse evidentsiaaliga vahendatud teabele teisiti kui otsesele pigem vormilistel põhjustel. Võib oletada, et 4-aastased lapsed suhtuvad neile tundmatusse grammatikaelementi lihtsalt kui millessegi, mis on neile tundmatu, uus ja huvitav ning see paneb neid eseme omadusi enam üldistama ja neid katsetama (Tamm jt 2014). Situatsioonis, kus lapsega tegeldakse, tõlgendab väiksem laps ütlust või pedagoogilist demonstratsiooni kui midagi olulist, mis tuleb ära õppida ja mille alusel saab üldistusi teha (Gelman 2010, Csibra, Gergely 2011, Butler, Markman 2012).

Loomulikult tuleb ka arvestada, et eri meetodiga võib grammatika omandamise kohta saada erinevaid tulemusi. Näiteks on laste keelekasutust uurinud Kadri Suurmäe eksperimentaalse meetodikaga leidnud, et kuigi juba enne 2-aastaseks saamist kasutavad lapsed spontaanses kõnes kausatiivi liidet, ei suuda nad veel 6-aastaseltki tajuda, et kausatiiviliitel on oma tähendus ja sellega saab uusi sõnu luua (Suurmäe 2014). Seega võivad pikiuuringu ja eksperimentaalse meetodi tulemused olla üsna erinevad.

5. Katsete eesmärk, kirjeldus ning üldised tulemused

Siinses uurimuses koostatud katsete eesmärk oli selgitada välja evidentsiaalsuse ja modaalsuse seosed. Katsetega vaadeldi nende kategooriate omandamist eri vanuserühmades. Tabelis 1 on esitatud teave valimi kohta.

Tabel 1. Valimi jaotus

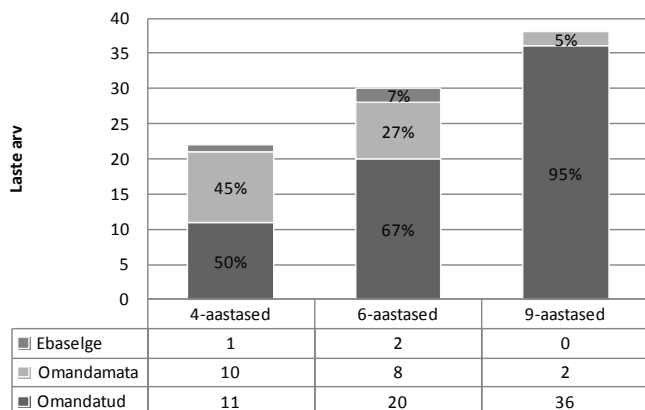
Katse tüüp	Laste vanus	Laste arv	Kokku
Evidentsiaalsuse omandamine	4-aastased	22	90
	6-aastased	30	
	9-aastased	38	
<i>vat</i> -i tõlgendamine episteemilise modaalina	4-aastased	22	90
	6-aastased	30	
	9-aastased	38	

5.1. Katse 1. Evidentsiaalsuse omandamine

Katse tehti 91 eesti keelt emakeelena kõneleva lapsega, kellele esitati kahe tegelase vahendusel *vat*-tunnusega lause ning küsiti, kas seda vormi kasutanud tegelane on ise info valdaja või on see vahendatud info. Katse tehti 4- ja 6-aastaste lasteaialastega ning 9-aastaste koolilastega. Tulemuste analüüsis määrati, mis vanuses eesti lapsed morfeemi *-vat* tähendust evidentsiaalina mõistavad. Katse oli 2013. aastal tehtud suurema eesti evidentsiaalsuse kasutamise ja omandamise uurimuse osa (täpse kirjelduse ning tulemused leiab Arguse jt juba ilmunud artiklist (2014)). Siinkohal piirdume katse põhipunktide esiletoomisega, et võrdlus episteemilise modaalsuse tähenduse omandamisega lugejale parema ettekujutuse annaks.

Katses esitati lapsele kahe tegelase vestlus käpiknukkude Alfi ja Pärteliga, üks neist kasutas info edastamiseks evidentsiaali. Katse tegija näitas lapsele kaht nukku, öeldes, et need on Alf ja Pärtel, keda tuleb hoolikalt kuulata. Seejärel esitati lapsele kahe nuku dialoog. Dialoogis esitati uus, muu kontekstita info evidentsiaaliga: *Minu ema töö juures olevat punaste plikititega² arvutimäng*. Hiljem küsis katse tegija, kas lausungi esitaja oli mängu ise näinud või sellest kusagilt kuulnud. Pärast küsimusele vastamist võis laps soovi korral oma vastust ka kommenteerida. Katse videosalvestati ja tulemused kodeeriti. Selle käigus jagati laste vastused kolme rühma: a) õiged vastused: laps ütles, et Alf on selle mängu kohta kuulnud kelleltki teiselt, b) vale vastused: laps ütles, et Alf on mängu ise näinud, ning c) muud vastused: laps ütles, et ta ei tea, kas Alf on mängu ise näinud või selle kohta kelleltki kuulnud.

Katse tulemused erinesid kolmes vanuserühmas selgelt ja vanuse kasvades tõusis õigete vastuste osakaal märkimisväärselt. Kui 4-aastased ei saanud veel kaudse kõneviisi vormist aru ja õigete vastuste arv oli sama suur kui valede vastuste arv, siis 6-aastaste õigete vastuste osakaal oli juba kaks kolmandikku kõigist vastustest. Seega võib oletada, et just selles vanuses hakkavad lapsed kaudse kõneviisi vormist evidentsiaali tähenduses aru saama. 9-aastastel oli õigete vastuste osakaal juba 95%, seega võib öelda, et nad mõistavad *vat*-i tähendust kaudse evidentsiaalina (vt joonis 1).



Joonis 1. Evidentsiaalsuse omandamine

Katses osalenud laste kommentaaride (vt ka Argus jt 2014) analüüsimisel saime kinnitust, et evidentsiaalne tähendus omandatakse enne üheksandat eluaastat. Enamik 9-aastasi lapsi selgitas oma vastust evidentsiaalina, näiteks:

- (4) Noo, kui seal on **olevat**, siis talle pigem öeldi.
- (5) Minu arust ta ei ole ise näinud, sellepärast et ta ütles **olevat**.
- (6) Ma arvan, et talle on keegi öelnud, sest **olevat**.

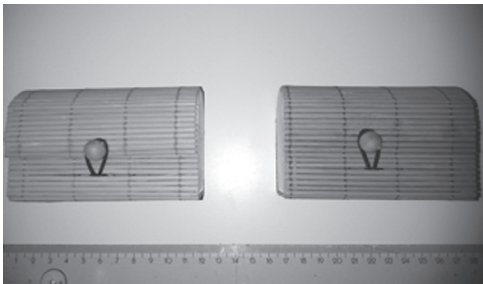
² Katses kasutati väljamõeldud sõna *plikit*, mis tähistas magnetiliste omadustega eset, millega laps oli tutvunud katse esimeses pooles.

5.2. Katse 2. vat-i tõlgendamine episteemilise modaalina

Episteemilise modaalsuse eksperimentaalse uurimisega on tegelnud Audin, Aksu-Koç (2015), kelle türgi keele episteemiliste markerite omandamise katsete stsenaariumi kohandades oleme koostanud siinse eesti keele *vat*-morfeemi katse. Katse tehti samade 4-, 6- ja 9-aastaste lastega, kes osalesid evidentsiaalitähenduse omandamise katses (vt 5.1). Tulemuste analüüsis püüti määrata, mis vanuses eesti lapsed *vat*-i tähendust episteemilise modaalsusena mõistavad.

Episteemilise modaalsuse tähenduselemendi väljaselgitamiseks testiti laste hinnangut ütluse tõenäosusele. Selleks lasti 4-, 6- ja 9-aastastel lastel otsida peidetud eset teabe põhjal, mida anti kolmel moel: a) jaatav indikatiivilause (*Komm on sinises karbis*), b) eitav indikatiivilause (*Komm ei ole roosas karbis*) või c) jaatav *vat*-iga lause (*Komm olevat roosas karbis*). Lähtusime mõttekäigust, et kui *-vat* väljendab tõenäosushinnangut, siis väljendub see valikutes nii, et *vat*-iga väljendatu jääb episteemilise hinnangu skaalal “tõene-väär” tõese ja väära vahele, st jaatava indikatiivilause ja eitava indikatiivilause vahele. Täpsemalt oletasime, et kui *-vat* väljendab laste jaoks tõenäosushinnangut, siis eelistavad nad eseme peidukoha hüpoteesi loomisel *vat*-morfeemiga lauset eitavale lausele. Sama mõttekäigu järgi oleks ootuspärane eelistada eseme peidukoha hüpoteesi loomisel jaatavat lauset *vat*-iga lausele.

Nii karpide värvusi, tegelasi kui ka tegelaste antud infot juhuslikustati, st värve ja tegelasi kombineeriti eri vormis lausetega. Eesmärk oli, et keelevormide valik ei tuleneks laste lemmikvärvist või eelistatud tegelasest. Vältisime olukorda, kus nukud Alf ja Pärtel oleksid pöördunud teabe edastamisel otse lapse poole, ning imiteerisime suhtlust, kus nukud andsid teineteisele infot karpide sisu kohta. Lapse roll oli olla nukude kõneluse kuulaja. Lastele esitati kaks karpi (üht katsevahendina kasutatud karbipaari näeb joonisel 2).



Joonis 2. Karbipaari episteemilise modaalsuse mõistmiskatses

Katse tegija näitas lapsele kaht nukku ning ütles, et need on Alf ja Pärtel, keda tuleb hoolikalt kuulata. Katse tegija asetaskorral kolm korda kaks eri värvi karpi korraga. Iga paari korral said lapsed vastukäivat infot kahe nuku dialoogi kaudu. Vastukäiv info esitati erineva grammatikaga lausetega, ent kõikidele lastele ühes järjekorras. Karpide värvus varieerus (varieeruv element on allpoolsetes dialooginäidetes 1–3 alla joonitud). Nukud pakkusid oma teadmisi iga kolme karbipaari kohta. Nad teatasid, millises karbis on komm, kasutades üksteise järel kolme katseolukorda, kus esitasid erineva infoga dialoogid:

Dialoog 1

Alf: "Rohelises karbis **ei ole** kommi."

Pärtel: "Sinises karbis **on** komm."

Dialoog 2

Alf: "Oranžis karbis **olevat** komm."

Pärtel: "Roosas karbis **on** komm."

Dialoog 3

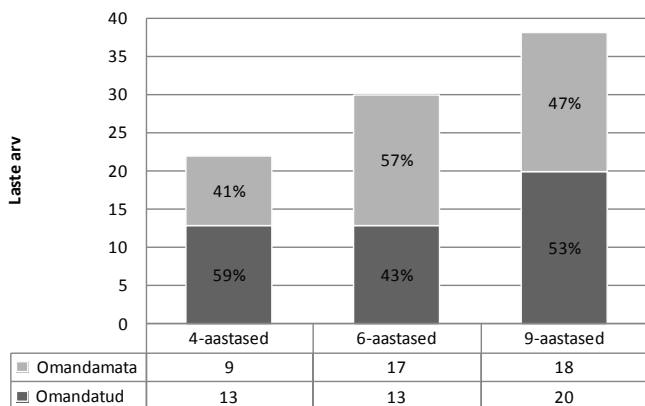
Alf: "Punases karbis **olevat** komm."

Pärtel: "Kollases karbis **ei ole** kommi."

Iga kord pärast kahe karbi tutvustamist sai laps näidata, kummas karbis tema arvates komm on. Esimesel korral avati karbid kohe pärast lapse valikut, näitamaks lapsele, et mõlemad nukud ütlevad kindlalt just seda, mida ise teavad, ja et kui hoolega kuulata, siis ühest karbist leiabki ütluste järgi kommi. Kahe järgneva karbipaari puhul pandi lapse valitud karp kõigepealt kõrvale. Need avati alles katse lõpus ja laps sai järele vaadata, kas ta oli nukkude vestlusest õigesti aru saanud: kui laps oli teavet õigesti tõlgendanud, leidis ta karbist kommi. Mängimine filmiti ja salvestisi kasutati, et fikseerida lapse valitud karbikeste järjekord ning analüüsida, kas lapsed mõistavad kaudse evidentsiaali modaaltähendust, hinnates propositsiooni tõenäosust.

Mõõtmiseks kasutasime vaid neid katse osi, kus vastandati evidentsiaaliga lauset eitava või jaatava lausega. Episteemilise modaalsuse mõistmise tunnusooneks kodeerisime sellised vastused, kus eitavale infole eelistati evidentsiaaliga esitatud infot ja evidentsiaaliga esitatud infole jaatavas oleviku indikatiivilises esitatud infot.

Tulemuseks saime, et episteemilise modaali tähendusest sai aru 46 last 90 lapsest (51,1%). Omandamise dünaamikat näitab see, et 4- ja 6-aastaste õigesti arvanute arv on võrdne, ent nooremad olid modaaltähenduse mõistatamises isegi suhteliselt paremad: 13 nelja-aastast (59%) ja 13 kuueaastast (43%) tajus episteemilise modaalsuse mõttelisel skaalal evidentsiaalitähendust kindla ja võimatu vahel. 4- ja 6-aastaste vastuste vahel oli palju varieeruvust ja oletusi. 9-aastaste seas hakkas modaalsest tähendusest aru saavate laste arv jälle suurenema: neid oli 20 (53%, vt joonis 3).

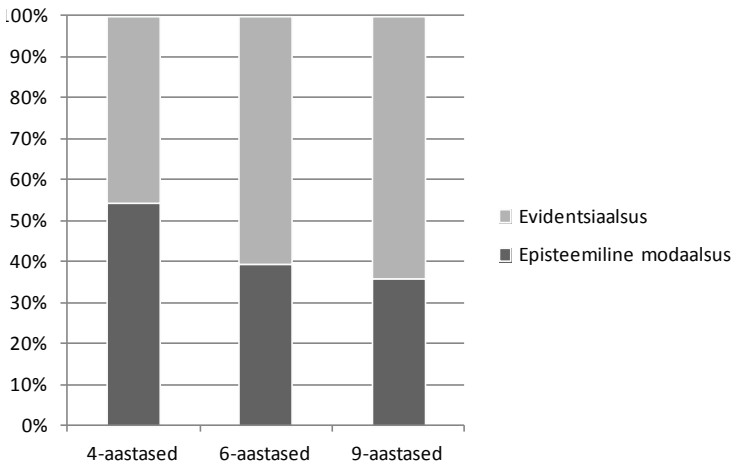


Joonis 3. Episteemilise modaalsuse omandamine

Ühe lapse kommentaar vähesele tõenäosushinnangule (*Ma ei tea, kas ta valetas või siis läks ta seda mängu mängima*) kinnitab fakti, et 9-aastastel on seos episteemilise modaalsusega tekkinud.

5.3. Kahe mõistevälja omandamise dünaamika

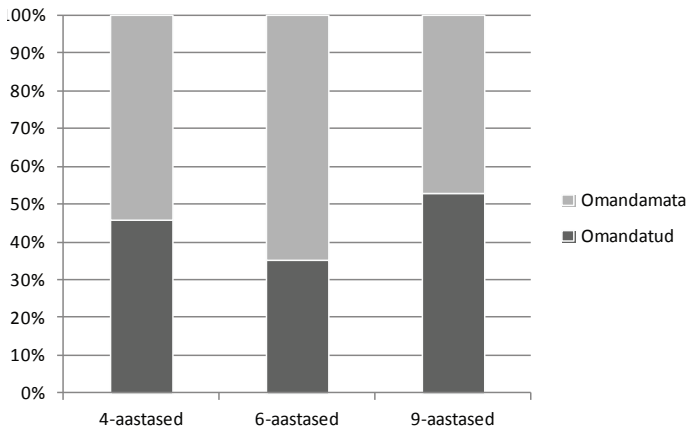
Järgnevalt vaatlesime kahe mõistevälja omandamise dünaamikat. Joonis 4 kujutab kahes katses (vt 5.1 ja 5.2) antud õigete vastuste suhet. Episteemilise tähenduse omandamise osakaalu tähistab tulpade tumehall ja evidentsiaalitähenduse omandamise osakaalu helehall osa.



Joonis 4. Evidentsiaalsuse ja episteemilise modaalsuse mõistmise areng

Kõigepealt antakse morfeemile modaalitähendus (54%), ent alates 6. eluaastast suureneb *vat*-morfeemi tõlgendamine evidentsiaalina (61%, üheksa-aastastest 64%). Mõned on omandanud ainult evidentsiaalitähenduse, teised ainult modaalitähenduse ning kolmandad on õppinud ära mõlemad. Järgnevalt vaatlesime, kas konkreetsed lapsed õpivad tähendused etapiti ja eraldi või omandavad nad mõlemad tähendused koos. Uurisime individuaalseid erinevusi, et selgitada kahe tähenduse põimumise arengut. Võrdlesime, a) kui palju lapsi neist, kes valivad õigesti – “keegi teine ütles” –, valivad jaatava indikatiivilausega vastanduse puhul n-ö jaatava karbi ja eitava indikatiivilausega vastanduse puhul *vat*-iga esitatud karbi; b) ja vastupidi, kui palju neist, kes valivad õigesti – jaatava indikatiivilausega vastanduse puhul n-ö jaatava karbi ja eitava indikatiivilausega vastanduse puhul *vat*-iga esitatud karbi, – valivad ka õigesti “keegi teine ütles”.

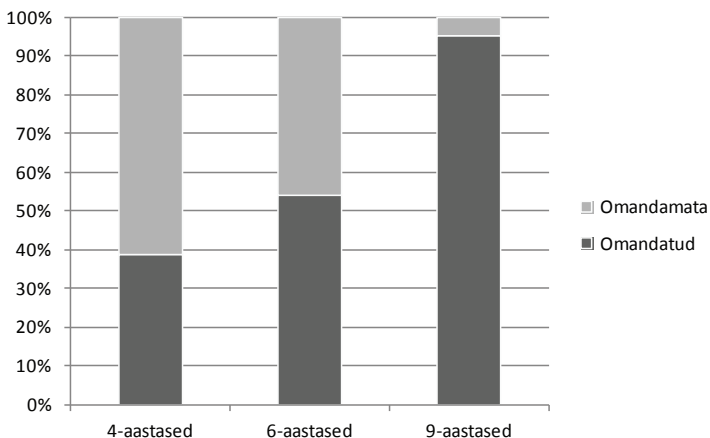
Kõigepealt uurisime, kui palju lapsi mõistis evidentsiaalitähendust. Tulemuseks saime 67 last (11 nelja-aastast, 20 kuueaastast, 36 üheksa-aastast). Nendest sai episteemilisest tähendusest aru 31 last (5 nelja-aastast, 7 kuueaastast ja 19 üheksa-aastast). Joonisel 5 kajastab tulpade tumehall osa nende laste osakaalu, kes lisaks evidentsiaalitähendusele on omandanud ka episteemilise tähenduse.



Joonis 5. Episteemilise modaalsuse omandamine omandatud evidentsiaalsuse puhul

6-aastaste modaalikatsetes langeb edukate laste osakaal 45%-lt 35%-le. 9-aastaste edukate laste osakaal jälle kasvab (52,8%). Evidentsiaalitähenduse omandanud lastest umbes pooled ei olnud niisiis modaalikatses edukad.

Tegime ka vastupidise arvestuse. Modaalitähendusest sai aru 46 last (13 nelja-aastast, 13 kuueaastast, 20 üheksa-aastast), nendest sai evidentsiaalsusest aru 31 last: 5 nelja-aastast (38,5%), 7 kuueaastast (53,8%) ja 19 üheksa-aastast (95%). Joonisel 6 tähistab tulpade tumehall osa evidentsiaalitähenduse omandanute osakaalu modaalitähendust mõistvate laste seas.



Joonis 6. Evidentsiaalsuse omandamine omandatud episteemilise modaalsuse puhul

Kuigi võib arvata, et katsed on kognitiivselt erineva raskusega, on suhtarvud siiski kõnekad: episteemilist modaalsust mõistvatest lastest 95% olid lõpuks edukad ka evidentsiaalikatses (joonis 6).

6. Arutlus

Korpust analüüsid leidsime, et kõik kaudsuse väljendamise vahendid ei pruugi tõenäosust vähendada. Teabe kaudsus ei tingi kaudse evidentsiaali kasutamist. Näitasime, et väljendatuna *da*-infinitiivi vormiga on evidentsiaalsus vaid marginaalselt seotud episteemilise modaalsusega, sest domineerib moraalne hinnang või iroonia ja tõenäosuse kohta hinnang puudub. Korpusest saime uurida, kuidas märgib kaudsust morfeem *-vat* siis, kui seda tingib väiksem tõenäosus. Seetõttu saab juba korpuseuuringu meetodikaga näidata, et evidentsiaalsus ja episteemiline modaalsus võivad eesti keeles esineda eraldi mõistekategooriatena, ent pole välistatud ka kooseksisteerimine.

Lastega tehtud katsete abil otsisime kinnitust evidentsiaalsuse ja episteemilise modaalsuse kategooria eristatavuse kohta, samas proovides teha nähtavaks nende kooseksisteerimise tingimused eri vanuserühmades. Evidentsiaalsuse omandamist näitas lapse õige arvamus info allika kohta (arvamus, et kõnelejast erinev isik on allikas). Morfeemi tähenduse asetamist tõenäosusskaala keskosas aga pidasime täiskasvanute keelekasutusega sarnanevaks episteemilise modaalsuse tõlgenduseks (vrd katseid Kütt jt 2014, Argus jt 2013).

Kategooriad õpiti esialgu eraldi (paralleelselt) ja hiljem koos (üksteisest sõltuvalt). Evidentsiaalsuse omandamine toimus vanuserühmades ühtlaselt ja tõusujoones, omandanute protsent vanuserühmades oli vastavalt 50%, 67% ja 95%. Meeleteooria arengu seaduspärasustele toetudes ei ole see ootamatu tulemus, sest kaudne evidentsiaalsus on tihedas seoses sotsiaalkognitiivsete võimete arenguga. Oletasime, et modaalsuse mõistekategooriat on lastel lihtsam omandada, ent mõistmisel ilmnesid vastakad modaali-tõlgendused (vrd Tamm jt 2014, Argus jt 2014). Nelja-aastastest lastest tajus 59% ja kuueaastastest ainult 43% *vat*-morfeemi modaalsust tähendust. Kuna modaali-katsetes edukate osakaal langeb kuueaastaste laste hulgas ja jälle kasvab üheksa-aastaste seas (53%), võib arvata, et tähenduste põimumisel toimivad erineva iseloomuga nähtused.

Seda kinnitab, et üheksa-aastased said mõistmiskatsel evidentsiaali-tähendusest täielikult aru, ent episteemilise modaalsuse omast mitte. Samas mõistsid selles vanuserühmas evidentsiaali-tähendust 95% lastest, kes olid omandanud modaali-tähenduse, aga mitte vastupidi: modaali-tähendust mõistvaid lapsi oli vaid 52,8% evidentsiaali-tähendust mõistvatest lastest.

Üheksa-aastaste laste selgitused oma tõlgendusotsuste kohta andsid enamjaolt tunnistust evidentsiaali-tõlgendusest, ent leidis ka modaali-tõlgendusi. Vastupidist suundumust ei ilmnenu, st sõnaliste tõlgenduste seas ei leidu andmeid, et episteemilise modaalsuse tähendusest saaksid lapsed paremini aru enne evidentsiaali-tähendust.

Leidis ka nelja-aastaseid lapsi, kes – veel mõistmata kaudse evidentsiaali tähendust – taipasid, et *-vat* seondub tõenäosusega: juhatab sisse kas jaatavast lausest väiksema või eitavast lausest suurema tõenäosusega propositsiooni (või isegi mõlemat). Nooremate laste hüpotees *vat*-i tähenduse kohta vastab esialgu evidentsiaalsusele üksnes osaliselt, kuna meeleteooria ei ole ilmselt piisavalt välja arenenud. Niipea kui omandatakse evidentsiaali-tähendus, saab hakata kinnistuma kaudsuse mõistmisest tulenevate järelduste modaalne semantika. Tabel 2 esitab kokkuvõtte saadud tulemustest.

Tabel 2. Põimunud kategooriate mõistmise dünaamika

Kategooria	Vanus aastates		
	4 aastat	6 aastat	9 aastat
Episteemiline modaalsus	Episteemiline hüpotees uue vormielemendi kohta episteemilisel skaalal	Evidentsiaalsusest sõltumatu episteemiline tõlgendus taandub	Suureneb info vahendatusest järeldatud modaalne tõlgendus
Evidentsiaalsus	Evidentsiaalsuse mõistmiseks vajaliku meeleteooria väljakujunemine, evidentsiaaltähendus omandamata	Algab evidentsiaaltähenduse omandamine	Evidentsiaaltähendus omandatud

Eksperimentaalsete meetoditega saame üsna veenvalt kinnitada keeleteaduses välja pakutud, ent intuiitivseks jäänud mõtet: morfeemi *-vat* semantika hõlmab nii evidentsiaalsust kui ka episteemilist modaalsust. Kuna keelelise arengu käigus kujunevad *vat*-morfeemi kaks tähendust etapiti, on ka keelesiseste grammatikakategooriate arengu algetapil tegu eri kategooriatega. Episteemilise modaalsuse tähendus ehitatakse seejuures pigem evidentsiaaltähendusele kui vastupidi.

7. Kokkuvõtteks

Meie korpuseanalüüsi ja katsete tulemuste kõrvutaval vaatlemisel selgus, et eesti keele *vat*-morfeemiga väljendatud evidentsiaalsuse ja episteemilise modaalsuse mõistekategooriad eristuvad eesti laste keelelises arengus. Morfeemi tähendustest omandati evidentsiaalsus enne episteemilist modaalsust. Mõlemad kategooriad on keeleomandamises aga suhteliselt hilised ja omandatakse alles pärast 6. eluaastat. Samuti kinnitasid tulemused, et *vat*-morfeem on kas kahtetähenduslik või on kahe kategooria semantika ja pragmaatika tihedasti seotud.

Töötasime välja meetodika ja koostasime katsed, millega on võimalik võrrelda informatsiooni edastamise grammatikavalikuga kaasnevaid episteemilisi mõjureid ning eristada evidentsiaalsus ja episteemiline modaalsus. Tulevikus on nüüd võimalik selgitada, kas kaudsete evidentsiaalide kasutamisel tehakse universaalseid järeldusi teabe vähese tõenäosuse kohta või on selliste järelduste tegemine hoopis keespetsiifiline. Samuti on nüüdsest peale võimalik võrrelda grammatiseerunud evidentsiaalsusega keeli selliste keeltega, kus evidentsiaalsus on väljendatud leksikaalselt, näiteks adverbidega, ning selgitada, mis mõju on evidentsiaalidel teabe mõistmisele, esitusele ja kasutamisele suhtlussituatsioonides.

Viidatud kirjandus

- Aikhenvald, Alexandra Y. 2004. Evidentiality. Oxford: Oxford University Press.
- Aksu-Koç, Ayhan 1988. The Acquisition of Aspect and Modality: The Case of Past Reference in Turkish. Cambridge: Cambridge University Press. <http://dx.doi.org/10.1017/CBO9780511554353>
- Aksu-Koç, Ayhan; Slobin, Dan I. 1986. A psychological account of the development and use of evidentials in Turkish. – Wallace Chafe, Johanna Nichols (Eds.). Evidentiality: The Linguistic Coding of Epistemology. Hillsdale: Ablex, 159–167.

- Argus, Reili; Kapanen, Airi; Kütt, Andra; Parm, Sirli; Suurmäe, Kadri; Tamm, Anne 2013. Evidentsiaalidel on täiskasvanute ja laste seas erinev funktsioon. – Ettekanne konverentsil “Teoreetiline keeleteadus Eestis”, Tartu, 16.–17. detsember 2013.
- Argus, Reili; Suurmäe, Kadri; Kütt, Andra; Tamm, Anne 2014. Eesti kaudse evidentsiaalsuse morfoloogilise markeri omandamisest: mõistmiskatse tulemused. [The acquisition of the morphological marker of evidentiality in Estonian: The results of a comprehension experiment.] – Krista Kerge (Toim.). Uurimusi keele omandamisest, õppimisest ja korpustest. Tallinna Ülikooli eesti keele ja kultuuri instituudi toimetised 16. Tallinn: Tallinna Ülikooli Kirjastus, 83–102.
- Aydın, Çağla; Aksu-Koç, Ayhan 2015. Children’s knowledge acquisition: Comprehension of the epistemic function of the marker –DIR in Turkish. – Poster Session A. Budapest CEU Conference on Cognitive Development, January 8–10, 2015, Budapest.
- Butler, Lucas P.; Markman, Ellen M. 2012. Preschoolers use intentional and pedagogical cues to guide inductive inferences and exploration. – *Child Development*, 83 (4), 1416–1428. <http://dx.doi.org/10.1111/j.1467-8624.2012.01775.x>
- Choi, Soonja 1995. The development of epistemic sentence-ending modal forms and functions in Korean children. – Joan Bybee, Susanne Fleischman (Eds.). *Modality in Grammar and Discourse. Typological Studies in Language* 32. Amsterdam: John Benjamins Publishing Company, 165–204.
- Clark, Eve 2004. How language acquisition builds on cognitive development. – *Trends in Cognitive Sciences*, 8 (10), 472–478. <http://dx.doi.org/10.1016/j.tics.2004.08.012>
- Csibra, Gergely; György, Gergely 2011. Natural pedagogy as evolutionary adaptation. – *Philosophical Transactions of the Royal Society, B* 366, 1149–1157. <http://dx.doi.org/10.1098/rstb.2010.0319>
- de Haan, Ferdinand 1999. Evidentiality and epistemic modality: Setting boundaries. – *South-west Journal of Linguistics*, 18, 83–101.
- de Haan, Ferdinand 2005. Encoding speaker perspective: Evidentials. – Zygmunt Frajzyngier, Adam Hodges, David S. Rood (Eds.). *Linguistic Diversity and Language Theories. Studies in Language Companion Series* 72. Amsterdam: Benjamins. <http://dx.doi.org/10.1075/slcs.72.i8haa>
- de Villiers, Jill; Garfield, Jay; Gernet-Girard, Harper; Roeper, Tom; Speas, Margaret 2009. Evidentials in Tibetan: Acquisition, semantics, and cognitive development. – Stanka A. Fitneva, Tomoko Matsui (Eds.). *Evidentiality: A Window into Language and Cognitive Development. New Directions for Child and Adolescent Development* 125. San Francisco: Jossey-Bass, 29–47.
- EKG I = Erelt, Mati; Kasik, Reet; Metslang, Helle; Rajandi, Henno; Ross, Kristiina; Saari, Henn; Tael, Kaja; Vare, Silvi 1995. *Eesti Keele Grammatika I. Morfoloogia. Sõnamoodustus. [Estonian Grammar I. Morphology. Word-formation.]* Tallinn: Eesti Teaduste Akadeemia Eesti Keele Instituut.
- Erelt, Mati 2002. Evidentiality in Estonian and some other languages. Introductory remarks. – *Linguistica Uralica*, 38 (2), 91–97.
- Erelt, Mati 2013. Eesti keele lauseõpetus. Sissejuhatus. Õeldis. [Estonian Syntax. Introduction.] Tartu ülikooli eesti keele osakonna preprintid 4. Tartu. <http://hdl.handle.net/10062/34069>
- Erelt, Mati; Metslang, Helle; Pajusalu, Karl 2006. Tense and evidentiality in Estonian. – Bert Cornillie, Nicole Delbecque (Eds.). *Topics in Subjectification and Modalization. Belgian Journal of Linguistics* 20. Amsterdam: John Benjamins, 125–136. <http://dx.doi.org/10.1075/bjl.20.09ere>
- Fitneva, Stanka A. 2001. Epistemic marking and reliability judgments: Evidence from Bulgarian. – *Journal of Pragmatics*, 33 (3), 401–420. [http://dx.doi.org/10.1016/S0378-2166\(00\)00010-2](http://dx.doi.org/10.1016/S0378-2166(00)00010-2)

- Gelman, Susan A. 2010. Generics as a window onto young children's concepts. – F. J. Pelletier (Ed.). *Kinds, Things, and Stuff: The Cognitive Side of Generics and Mass Terms*. New Directions in Cognitive Science 12. New York: Oxford University Press.
- Kehayov, Petar 2004. Eesti keele evidentsiaalsussüsteem mõne teise keele taustal. Morfo-süntaks ja distributsioon. [The Estonian evidential system in contrast with some other grammatical evidential systems. Morphosyntax and distribution.] – *Keel ja Kirjandus*, 11, 812–829.
- Kehayov, Petar 2008. *An Areal-Typological Perspective to Evidentiality: The Cases of Balkan and Baltic Linguistic Areas*. *Dissertationes linguisticae Universitatis Tartuensis* 10. Tartu: Tartu Ülikooli Kirjastus.
- Kütt, Andra; Tamm, Anne; Argus, Reili 2014. Evidentsiaalsus ja kognitiivne dissonants: ühe esmase katse tulemused. [Evidentiality and cognitive dissonance.] – *Oma Keel*, 1, 51–58.
- Metslang, Helle; Pajusalu, Karl 2002. Evidentiality in South Estonian. – *Linguistica Uralica*, 38 (2), 98–109.
- Palmer, Frank Robert 1986. *Mood and Modality*. Cambridge: Cambridge University Press.
- Palmer, Frank Robert 1996. *Semantics*. Cambridge: Cambridge University Press.
- Papafraçou, Anna; Li, Peggy; Choi, Youngon; Han, Chung-hye 2007. Evidentiality in language and cognition. – *Cognition*, 103 (2), 253–299. <http://dx.doi.org/10.1016/j.cognition.2006.04.001>
- Postimees. <http://www.postimees.ee/>.
- Premack, David; Woodruff, Guy 1978. Does the chimpanzee have a theory of mind? – *Behavioral and Brain Sciences*, 1 (4), 515–526. <http://dx.doi.org/10.1017/S0140525X00076512>
- Rett, Jessica; Hyams, Nina; Winans, Lauren 2013. *The Effects of Syntax on the Acquisition of Evidentiality*. BUCLD 37 Proceedings. Cascadilla Press.
- Rätsep, Huno 1971. Kas kaudne kõneviis on kõneviis? Verbivormide situatsioonianalüüsi. [Is the oblique mood a mood?] – *Keel ja struktuur* 5. Töid struktuuralse ja matemaatilise lingvistika alalt. Tartu Riiklik Ülikool, eesti keele kateeder. Tartu, 43–67.
- Sepper, Maria-Maren 2005. *Vahendatuse morfosüntaktiline väljendamine eesti kirjakeeles* 19. sajandi lõpust 20. sajandi lõpuni. [Indirect evidentiality in Estonian journalistic and fiction texts of the late 19th and the 20th centuries.] *Magistritöö*. Tallinn: Tallinna Ülikool.
- Slobin, Dan; Aksu, Ayhan 1982. Tense, aspect and modality in the use of the Turkish evidential. – Paul J. Hopper (Ed.). *Tense-Aspect: Between Semantics and Pragmatics*. *Typological Studies in Language* 1. Philadelphia: John Benjamins Publications, 185–200.
- Stephany, Ursula 1993. Modality in first language acquisition: The state of the art. – Norbert Dittmar, Astrid Reich (Eds.). *Modality in Language Acquisition*. *Sociolinguistics and Language Contact* 6. Berlin–New York: De Gruyter, 133–144.
- Zufferey, Sandrine 2010. *Lexical Pragmatics and Theory of Mind: The Acquisition of Connectives*. *Pragmatics & Beyond New Series* 201. Amsterdam: John Benjamins Publications. <http://dx.doi.org/10.1075/pbns.201>
- Tamm, Anne 2008. Partitive morphosemantics across Estonian grammatical categories, and case variation with equi and raising. – Miriam Butt, Tracy Holloway King (Eds.). *Proceedings of the LFG08 Conference*. Stanford, CA: CSLI Publications, 473–493. <http://web.stanford.edu/group/cslipublications/cslipublications/LFG/13/papers/lfg08tamm.pdf> (1.10.2014).
- Tamm, Anne 2009. The Estonian partitive evidential. Some notes on the semantic parallels between the aspect and evidential categories. – Lotte Hogeweg, Helen de Hoop, Andrei Malchukov (Eds.). *Cross-Linguistic Semantics of Tense, Aspect, and Modality*. *Linguistics Today* 148. Amsterdam: John Benjamins, 365–402. <http://dx.doi.org/10.1075/la.148.15tam>

- Tamm, Anne 2011a. Cross-categorial spatial case in the Finnic non-finitive system: Focus on the absentive TAM semantics and pragmatics of the Estonian inessive m-formative and non-finites. – *Linguistics: An Interdisciplinary Journal of the Language Sciences*, 49 (4), 835–944. <http://dx.doi.org/10.1515/ling.2011.025>
- Tamm, Anne 2011b. The case of Nincstelenek ('the Destitue'): Nounless and verbless in the Uralic grammar. – Sándor Csúcs, Nóra Falk, Viktória Tóth, Cábör Zaicz (Eds.). *Proceedings of Congressus XI Internationalis Fenno-Ugristarum: Piliscsaba, Hungary, August 9–14, 2010*. Piliscsaba: Reguly Társaság, 270–276.
- Tamm, Anne 2014. The partive concept versus linguistic partitives: From abstract concepts to evidentiality in the Uralic languages. – Silvia Luraghi, Tuomas Huumo (Eds.). *Partitive Cases and Related Categories. Empirical Approaches to Language Typology* 54. Berlin: Mouton de Gruyter, 89–151. <http://dx.doi.org/10.1515/9783110346060.89>
- Tamm, Anne; Argus, Reili; Suurmäe, Kadri 2014. Evidentsiaalid: kas uskuda või mitte? Teabe vormi mõju eesti koolieelikute käitumisele. [Evidential sentences guiding Estonian preschoolers' exploration of novel objects and their properties.] – *Emakeele Seltsi aastaraamat*, 59 (2013), 244–261. <http://dx.doi.org/10.3176/esa59.13>
- Öztürk, Özge; Papafragou, Anna 2007. Children's acquisition of evidentiality. – *BUCLD 31 Proceedings Supplement*. <http://papafragou.psych.udel.edu/papers/BU-Ozge.pdf> (1.10.2014).

Anne Tamm (Gáspár Károli nimeline Ungari Protestantlik Ülikool / Johns Hopkinsi Ülikool / Kesk-Euroopa Ülikool) uurib keele ja taju seoseid. 243 Krieger Hall, 3400 N. Charles St. Baltimore, MD 21218, USA
anne.tamm@jhu.edu

Reili Argus (Tallinna Ülikool) uurib esimese keele omandamist. Tallinna Ülikooli eesti keele ja kultuuri instituut, Narva mnt 29, 10120 Tallinn, Estonia
reili.argus@tlu.ee

Airi Kapaneni (Tallinna Ülikool) uurimisvaldkond on esimese keele omandamine. Tallinna Ülikooli eesti keele ja kultuuri instituut, Narva mnt 29, 10120 Tallinn, Estonia
airikapanen@gmail.com

Andra Kütt (Tallinna Ülikool) uurib sotsiolingvistilise keskkonna mõju laste keelelisele arengule. Tallinna Ülikooli eesti keele ja kultuuri instituut, Narva mnt 29, 10120 Tallinn, Estonia
andra.kytt@gmail.com

Kadri Suurmäe (Tallinna Ülikool) on uurinud kausatiivide omandamist. Tallinna Ülikooli eesti keele ja kultuuri instituut, Narva mnt 29, 10120 Tallinn, Estonia
kadri.suurmae@gmail.com

EVIDENTIALITY AND EPISTEMIC MODALITY COMBINE IN THE ESTONIAN MORPHEME -VAT BUT ARE DIFFERENT CATEGORIES IN ACQUISITION: EXPERIMENTAL EVIDENCE

Anne Tamm

Central European University, Johns Hopkins University,
Károli Gáspár University of the Hungarian Reformed Church

**Reili Argus, Airi Kapanen,
Andra Kütt, Kadri Suurmäe**

Tallinn University

Since Aikhenvald 2004, evidentiality has been severed from epistemic modality. We studied the evidential and epistemic modal aspects of the Estonian evidential *-vat* experimentally and on the basis of a corpus study. We designed two experiments to separate the two categories. Four-, six-, and nine-year-old children were first asked to identify the source of the information provided with the morpheme. Then they performed a forced choice task guided by grammar: affirmative indicative, negative, and evidential sentences. Four-year-olds did not understand the evidential and modal *-vat*. Six year-olds showed an increased understanding of the evidential meaning but not the modal meaning or use. Nine-year-olds demonstrated full understanding of the evidential and an increased understanding of the epistemic modal aspect of the evidential. Estonian children thus develop the correct understanding of the morpheme *-vat* as an epistemic modal only after learning its evidential meaning.

Keywords: language acquisition, semantics, pragmatics, morphology, Theory of Mind, experimental methods, Estonian

KEELETEADUSLIKE ANDMETE RUUMILISI VISUALISEERIMISVÕIMALUSI

Kristel Uiboed, Aki-Juhani Kyröläinen

Ülevaade. Keeleteadusliku uurimismaterjali ja -tulemuste geograafiline visualiseerimine on dialektoloogias olnud alati kesksel kohal, kuid see pole omane ainult murdeuurimisele, vaid on oluline osa mis tahes lingvistilisest tööst, kuhu on kaasatud ruumiline komponent. Tänapäeval on olemas arvukalt programme ja kaardiressursse, mis võimaldavad neid ülesandeid täita. Sageli on aga nende kasutamise eelduseks üsna põhjalikud teadmised geoinfosüsteemidest või mõnest konkreetsest programmist. Käesoleva artikli eesmärk on pakkuda üks lihtne ja kiire võimalus geograafiliste andmete esitamiseks, eeldamata kasutajalt suuri tehnilisi oskusi. Pakutud lahendus ja materjalid on vabavarana saadaval ning iga soovija võib neid oma vajadustele vastavalt kohandada või selle baasilt välja töötada enda jaoks sobivad vahendid. Artiklis kasutame andmestikuna eesti murrete korpust, sh vadja keele alamkorpust, kuid võimalused eri andmete esitamiseks ei ole ühe konkreetse andmestikuga piiratud.*

Võtmesõnad: dialektoloogia, murdesüntaks, geolingvistika, korpuslingvistika, eesti keel

1. Sissejuhatus

Lingvistiliste nähtuste visualiseerimine kaartidel on alati olnud oluline osa murdeuurimisest. Traditsiooniliselt on selleks kasutatud olemasolevaid kaarte ning lingvistiline informatsioon on kaardile kantud enamasti käsitsi. Arvutiajastu on andmete esitusvõimalusi märkimisväärselt avardanud. Sellega on aga tehnoloogia keerulisemaks muutunud ja see eeldab lingvistidelt tunduvalt mitmekülgsemaid teadmisi ja tehnilisi oskusi. Selliste oskuste omandamine klassikalise keeleteadusliku hariduse juurde praegu ei kuulu. Artikli eesmärk ongi esitada üks võimalik viis geograafiliste andmete visualiseerimiseks, mis ei nõua spetsiifilisi tehnilisi oskusi ega pikemaajalist süvenemist eri programmidesse.

* Artikli valmimist on toetanud Eesti Teadusagentuuri projekt PUT90. Aitäh anonüümsetele retsensentidele parandusettepanekute eest ja Maarja-Liisa Pilvikule kommentaaride ja tagasiside eest.

Enamasti kasutatakse geograafiliste andmete visualiseerimiseks mitmesuguseid GIS-tarkvara (ingl *geographic information system*) võimalusi, mille rakendamise eelduseks on aga üsna põhjalikud teadmised GIS-süsteemidest ja geograafilisest terminoloogiast ning kindlad tehnilised oskused. GIS-süsteemid on sageli tasulised, mis seab nende kasutusele muidki piiranguid. Meie välja töötatud võimalus on lihtne viis andmete visualiseerimiseks. Kombineerime nii vabavaralist statistika-programmi R (R Development CoreTeam 2013) kui ka olemasolevaid vabavaralisi kaardiressursse, näiteks Google'i või Regio kaarte (2014). Siinne lähenemine ei eelda kasutajalt sügavamaid teadmisi geoinfosüsteemidest ega ka põhjalikke programmeerimisoskusi R-is (mõningad teadmised tulevad muidugi kasuks). Oleme artikli esitanud justkui kasutusjuhendi kõigi vajalike sammudega ning igal uurijal on juhiseid järgides võimalik oma andmestiku põhjal esitada soovitud infot. Eelduseks on vaid statistikaprogrammi R oma arvutisse paigaldamine. Käesoleva töö tulemusena valminud kaardiressursid, näidetes kasutatud failid ja R-i skriptid on kättesaadavad veebis ning kõik võivad neid vabalt kasutada ja vastavalt oma vajadustele muuta.¹ Viitamisjuhendid leiab veebilehelt ja failidest. Programmikoodi me artiklis põhjalikumalt ei kommenteeri, detailsem info on esitatud artikliga kaasnevas skriptifailis. Programmi R keeleteaduses rakendamise kohta on mitmeid väga häid õpikuid, mille abil on võimalik omandada sügavamaid teadmisi (Baayen 2008, Gries 2009, Johnson 2008). Ruumiliste andmete töötlemisest R-is on põhjaliku monograafia kirjutanud Bivand jt (2013).

Artikli alustuseks kirjeldame töö tausta ja põhjusi, miks me sellise lahenduseni oleme jõudnud. Kolmandas peatükis demonstreerime võimalusi, kuidas saada endale sobiv kaart, mida oleks võimalik kasutada edasiseks töötlemiseks ja oma uurimisandmete ja -tulemuste esitamiseks. Neljandas peatükis kirjeldame uurimisandmete, sagedusinfo ja statistilise analüüsi tulemuste kaartidel esitamist.

2. Taust

Praeguseks on olemas erinevaid visualiseerimis- ja analüüsirakendusi, mis on spetsiaalselt välja töötatud keeleteadlastele (nt Goebel 2006, Nerbonne jt 2011) lingvistiliste andmete analüüsimiseks ja uurimistulemuste esitamiseks. Need rakendused on hästi dokumenteeritud, nende kasutamine on tehtud lihtsaks ja mugavaks ka uurijatele, kes pole varem sarnaste töövahenditega kokku puutunud. Selliste rakenduste väljatöötamine ja olemasolu on väga tänuväärsed ning praeguseks võimaldavad need rakendused teha üsnagi keerulisi statistilisi analüüse. Nagu igasuguste valmistoodete puhul paratamatu, seavad nad kasutajale ka teatavaid piiranguid nii kaartide kasutamisel kui ka statistiliste meetodite valikul. Seega kui kasutaja soovib rakendada oma andmete analüüsimiseks meetodit, mida programmeerijad pole rakendustesse kaasanud, pole kasutajal võimalik seda teha. Selleks tuleb otsida alternatiivseid lahendusi.

Tänapäeval on hulgaliselt kaardiressursse, mida saab kasutada ruumiliste andmetega seotud mis tahes töös, kaasa arvatud lingvistikas. Sellised kaardiressursid põhinevad sageli praegusel või ajaloolisel haldusterritoriaalsel jaotusel. Murdepiiride esitamiseks on selliste ressursside kasutamine aga mõnevõrra

¹ <https://github.com/kristel-/Keeleteaduslike-andmete-ruumilisi-visualiseerimisvoimalusi> (Uiboed, Kyröläinen 2015).

problemaatiline. Traditsioonilised murdepiirid ei järgi enamasti haldusterritoriaalseid piire ja keealad võivad ka riigipiire ületada. Selline probleematika ei ole levinud ainult dialektoloogias. Näiteks on keeruline kasutada tänapäeva haldusterritoriaalseid kaarte ka juhul, kui uuritakse keelt, mis pole riigikeel ning pole seotud ühegi konkreetse tänapäevase või ajaloolise haldusüksusega. Näiteid pole vaja kaugelt otsida: probleem on tuttav kõigile väiksematele soome-ugri keelte uurijatele, rääkimata murdeuurijatest. Järgnevalt esitamegi ühe väga lihtsa võimaluse, kuidas selliseid kaarte ise joonistada, ning seejärel näitame, kuidas on võimalik kasutada oma andmete esitamiseks juba olemasolevaid elektroonilisi kaarte.

3. Kaardiandmestik

3.1. Poolkäsitsi joonistatud kaardid

Selles alapeatükis demonstreerime, kuidas üsna lihtsa vaevaga n-ö joonistada kaart, mida on võimalik automaatselt töödelda. Iga geograafilise andmestiku esitamise eeldus on koordinaatsüsteemil põhineva info olemasolu. Seega on meil alustamiseks vaja koordinaatide andmeid piirkonnast, mida soovime visualiseerida. Selleks on võimalik kasutada GIS-süsteemi ressursse. Suure tõenäosusega aga ei ole GIS-süsteemides kaarte, mis esitaksid nt liivi keele alad või eesti murdejaotuse. Niisiis on üks lahendus võtta mõni olemasolev kaart ja sellele toetudes n-ö joonistada uus kaart meid huvitavate piirkondadega. Keskne mõiste geograafias ja kartograafias on polügoon (ehk hulknurk), mis siinses kontekstis tähendab ühte kaardile kantud ala. Selleks on vaja teatavaid punkte, mida oleks võimalik ühendada, nii et neist moodustuks üks piirkond. Meie kasutasime Regio (2014) kaardisüsteemi, mis esitab pikkus- ja laiuskraadid. Sama eesmärgi täidab mis tahes sama infot esitav kaart. Eri- nevalt Google'ist pakub Regio infot ka kihelkondade kohta ning sobivad murdealad saamegi välja joonistada just kihelkonna piiride abil. Polügooni joonistamiseks piisab ainult mõnest punktist, kuid mida rohkem punkte, seda täpsem on kaart.²

Tabel 1. Fragment koordinaatandmestikust (failist *murdealadeKoordinaadid.csv*)

Murre	Latituud	Longituud
kirde	58.99	27.73
kirde	59.23	27.89
kirde	59.3	28.04
kirde	59.28	28.13
kirde
kirde	NA	NA

Tabel 1 esitab fragmendi failist *murdealadeKoordinaadid.csv* (Uiboaed, Kyröläinen 2015), mis sisaldab infot kõigi murdealade kohta ning nende alade joonistamiseks vajalikku infot. Iga murdeala (või ühe polügooni, näiteks saarte murdeala koosneb mitmest polügoonist) lõpetab rida, kus pikkus- ja laiuskraadide lahtrites on väärtus NA (ingl *not available*), mis on vajalik selleks, et polügooni joonistamine n-ö lõpetada või katkestada, et punkte ei ühendataks üle mitme polügooni. Oluline on

² Täpsemalt geograafilise terminoloogia ja vastava teoreetilise aluse kohta vaata Bibiko (2012) ja Krikmann (2005).

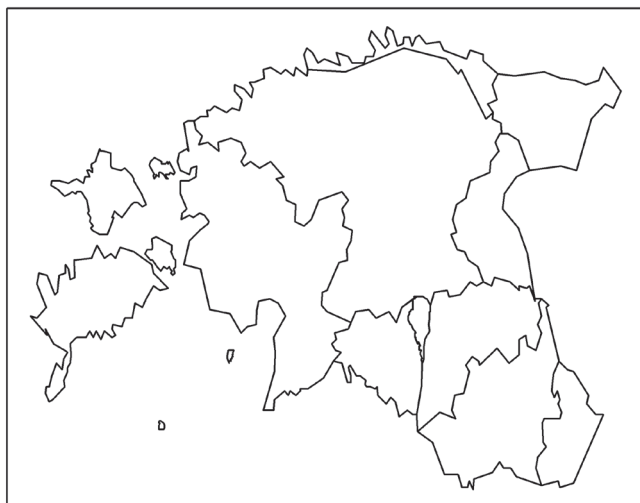
jälgida, et geograafilise andmestiku murdenimed vastaksid täpselt, ka kuju poolest, murdenimedele, mida kasutatakse lingvistilises andmestikus (täpsemalt vt allpool). Kõik tabelis esitatud koordinaadid on sisestatud käsitsi ehk kogutud punkti punkti haaval aluseks olevalt kaardilt (Regio 2014). Nagu eelnevalt mainitud, võib sama protsessi korrata mis tahes muul kaardiandmestikul, mis vajaliku info esitab.

Kui punktid on kogutud ja sobivas formaadis tabelisse kantud, on võimalik nende abil programmis R (R Development CoreTeam 2013) juba üsna lihtsalt kaart joonistada. Selleks tuleb andmestik esmalt R-i sisestada.

```
# loeme sisse kaardiandmestiku
koordinaadid <- read.csv("murdealadeKoordinaadid.csv", header=T,
sep=";")
```

Kui andmestik on programmi sisse loetud, piisab sellest lihtsakoelise kaardi joonistamiseks (tulemust vt kaardilt 1).

```
plot(koordinaadid$lon, koordinaadid$lat, type="l")
```



Kaart 1. Poolkäsitsi joonistatud eesti murdealade kaart

Järgnevas kirjeldame, kuidas sellele kaardile andmeid kanda ja seda näiteks sageduste visualiseerimiseks kasutada. Siin on oluline märkida, et parema tulemuse saab GIS-tarkvara³ kasutades georefererimise teel, mis on tehniliselt pisut nõudlikum viis, kuid tagab detailsema ja visuaalselt kvaliteetsema tulemuse. Täpsemalt vt selle kohta Kyröläinen ja Uiboed (ilmumas).

³ Vabavaraline GIS-tarkvara on näiteks Quantum GIS (QGIS Development Team 2014).

3.2. ggmap ja ggplot2 programmis R ning kasutatud kaardiressursid

Kui oma kaardi joonistamine pole hädavajalik ning sobivad olemasolevad kaardiressursid, on olukord veelgi lihtsam. R pakub selleks väga lihtsaid võimalusi, näiteks on muude võimaluste hulgas ka Google'i kaarte toetav pakett *ggmap* (Kahle, Wickham 2013), mida on üsna lihtne kasutada koos laialt levinud graafikapaketiga *ggplot2* (Wickham 2009). Alloleva skriptiosa käivitamisel saab edasiseks töötluks alla laadida Eesti kaardi värvilise või mustvalgena (värvilise kaardi jaoks esitame vaid koodi). Tulemus on esitatud kaardil 2.

```
# vaikimisi värviline kaart
qmap("Eesti", zoom=7)

# mustvalge kaart
qmap("Eesti", zoom=7, color="bw")
```



Kaart 2. Paketiga *ggmap* saadud mustvalge Eesti kaart

Kaardi edasiseks töötluks tuleks see salvestada järgneval viisil.

```
eesti <- qmap("Eesti", zoom=7, color = "bw")
```

Arvukate lisavõimaluste kohta (näiteks lisaks Google'i kaartidele on võimalik kasutada muidki kaarte) vaata paketi dokumentatsiooni (Kahle, Wickham 2013).

4. Keeleteaduslikud (meta)andmed ja nende esitus kaartidel

4.1. Sagedusandmete esitusvõimalusi

Eespool esitatud kaartidele on lihtne kanda sagedusandmestikku. Kasutame näitena finiiitse verbi *saama* ja *tud*-partitsiibi konstruktsioonide andmestikku (Uiboaed 2013). Konstruktsiooniga saab eesti keeles väljendada nii passiivi kui ka impersonaali, selle kasutus varieerub murrete lõikes märkimisväärselt (Uiboaed 2013, Uiboaed jt 2013). Eespool esitatud kaardil võime visualiseerida selliste ühendite (mitte konstruktsioonide, st ühendi tähendus ei ole eristatud) sagedust niimoodi, et tumedamad alad tähistaksid kõrgemat sagedust ja heledamad vastupidi, madalamat sagedust ehk kasutame n-ö üleminekukaarti (ingl *gradiance map*).⁴ Sisendiks on tabel, kus on esitatud vaadeldava ühendi sagedus eesti murretes.

Tabel 2. Finiitse *saama*-verbi ja *tud*-partitsiibi ühendite sagedusandmestik (fail *tudSaama.csv*, Uiboaed, Kyröläinen 2015)

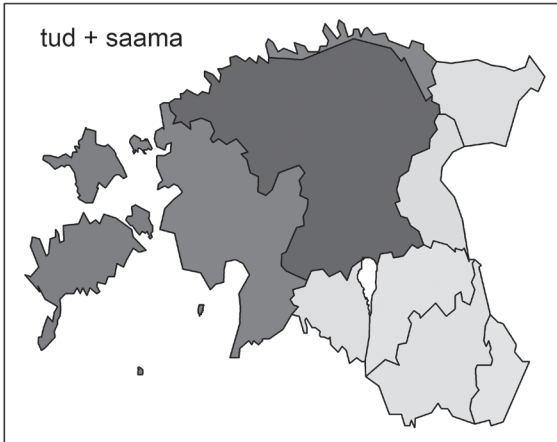
Murre	tudSaama
kirde	19
ida	29
tartu	13
võru	17
lääne	173
mulgi	21
seto	9
kesk	199
saarte	180
ranna	178

```
# kasutame juba tuttavat koordinaatandmestikku
koordinaadid <- read.csv("murdealadeKoordinaadid.csv", header=T,
sep=";")

# loeme sisse verbiühendite sagedusandmestiku
tudSaama <- read.csv("tudSaama.csv", header=T, sep=";")

# ja seejärel illustreerime neid sagedusi heledamate ja tumedamate
toonidega
plot(lat~lon, type="n", data=koordinaadid, axes=F, ylab="", xlab="")
for (i in 1: length(unique(koordinaadid$murre))) {
  temp=koordinaadid$murre==unique(koordinaadid$murre)[i]
  polygon(x=koordinaadid$lon[temp], y=koordinaadid$lat[temp],
col=gray.colors(max(tudSaama$tudSaama))
[ max(tudSaama$tudSaama)-tudSaama$tudSaama[tudSaama$murre==
unique(koordinaadid$murre)[i]]+1]
)
}
box()
```

```
text(x=min(koordinaadid$lon,na.rm=T), y=max(koordinaadid$lat,na.
rm=T)-0.1,labels="tud + saama",pos=4, cex=2.5)
```



Kaart 3. Finiitse *saama*-verbi ja *tud*-partitsiibi ühendite kasutussagedus eesti murretes

Kaart 3 esitab finiidse *saama*-verbi ja *tud*-partitsiibi sageduste üleminekukaardi. Kaart illustreerib, kuidas selliste ühendite kasutussagedus esitatud andmestiku põhjal on suurim keskmurdes ja väiksem kõigis idapoolsetes murretes.

Samal viisil võime visualiseerida konstruktsioonide sagedust eraldi. Sagedusandmestik peab selle valiku korral olema jällegi eraldi tabelis (kasutame tabelit failis *tudKonstr.csv*, Uiboaed, Kyröläinen 2015).

```
# kasutame juba tuttavat koordinaatandmestikku (fail murdealadeKoor-
dinaadid.csv)
koordinaadid <- read.csv("murdealadeKoordinaadid.csv", header=T,
sep=";")

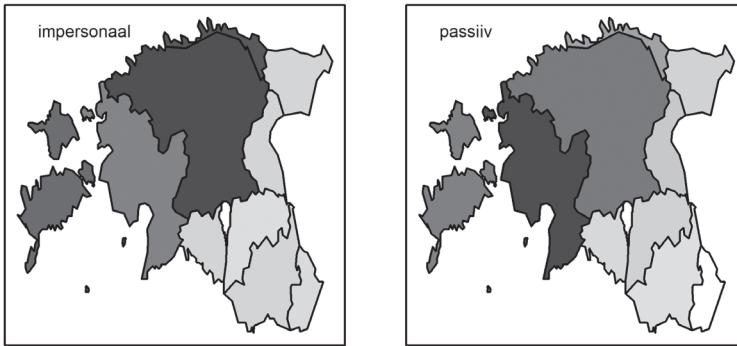
# loeme sisse konstruktsioonide andmestiku (fail tudKonstr.csv)
tudKonstr <- read.csv("tudKonstr.csv", header=T, sep=";")

# joonistame mõlema konstruktsiooni sageduste illustreerimiseks kaar-
did
plot(lat~lon, type="n", data=koordinaadid, axes=F, ylab="", xlab="")
for (i in 1: length(unique(koordinaadid$murre))) {
  temp=koordinaadid$murre==unique(koordinaadid$murre)[i]
  polygon(x=koordinaadid$lon[temp], y=koordinaadid$lat[temp],
col=gray.colors(max(tudKonstr$imp))[max(tudKonstr$imp)-
tudKonstr$imp[tempKonstr$murre==unique(koordinaadid$murre)
[i]]+1]
)
}
box()
text(x=min(koordinaadid$lon,na.rm=T), y=max(koordinaadid$lat,na.
rm=T)-0.1,labels="impersonaal",pos=4, cex=2.5)
```

```

plot(lat~lon, type="n", data=koordinaadid, axes=F, ylab="", xlab="")
for (i in 1: length(unique(koordinaadid$murre))) {
  temp=koordinaadid$murre==unique(koordinaadid$murre)[i]
  polygon(x=koordinaadid$lon[temp], y=koordinaadid$lat[temp],
    col=gray.colors(max(tudKonstr$pass))[max(tudKonstr$pass)-
tudKonstr$pass[tudKonstr$murre==unique(koordinaadid$murre)
[i]]+1]
  )
}
box()
text(x=min(koordinaadid$lon,na.rm=T), y=max(koordinaadid$lat,na.
rm=T)-0.1,labels="passiiv",pos=4, cex=2.5)

```



Kaart 4. Eesti murrete finiiitse *saama*-verbi ja *tud*-partitsiibiga moodustatud impersonaali- ja passiivikonstruktsioonide sageduste üleminekukaardid

Kaart 4 esitab finiiitse *saama*-verbi ja *tud*-partitsiibi abil moodustatud impersonaali- ja passiivikonstruktsioonide sagedused eesti murrete korpuse andmestiku põhjal, kus tumedamad alad tähistavad kõrgemat ja heledamad madalamat sagedust. Kaardilt 4 võib näha, et mõlema konstruktsiooni kasutussagedus on suurem läänepoolsetes murretes. Impersonaali konstruktsioonid on levinumad keskmurdes ning passiivi kasutussagedus on suurim läänemurdes.

4.2. Geograafiliste punktide ja piirkondade esitamine

Sageli on vaja esitada ainult geograafilisi punkte, kust uurimismaterjali on kogutud, või märkida kaardil, millisest piirkonnast on mingit vormi, konstruktsiooni vms leitud. Selleks kasutatakse sageli nn sümbolkaarte, kus soovitud piirkonna võib kaardil tähistada lihtsalt punktiga. Eesti murrete kontekstis on klassikaliseks näiteks Andrus Saareste (1955) murdeatlase kaardid, mida oleks samuti võimalik esitada elektrooniliselt või kombineerida neid uute andmetega. Neid võimalusi me käesolevas artiklis ei esita, selle kohta vt täpsemalt Kyröläinen ja Uiboaed (ilmumas).

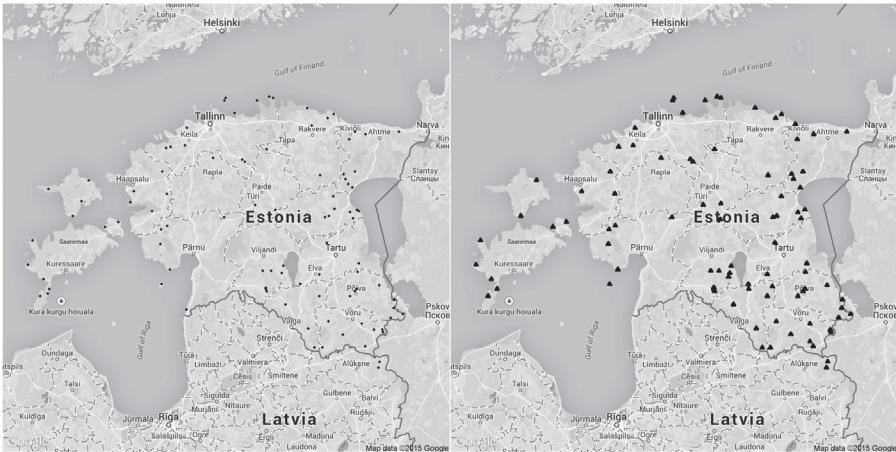
Sümbolkaartide esitamiseks on taas vaja tähistatavate piirkondade koordinaatandmestikku, mis on esitatud eraldi tabelina. Põhjana võib kasutada eespool esitatud kaarte.

Alolev kaart 5 illustreerib materjali ühest eesti murrete analüütilisi ja sünteetilisi kvotatiivikonstruktsioone käsitlevast uurimusest (Pilvik, Uiboaed 2014). Iga punkt kaardil tähistab küla, kust on töösse kaasatud uurimismaterjali kogutud. Näiteandmestikud on failides *kvotAn.csv* ja *kvotSyn.csv* (Uiboaed, Kyröläinen 2015).

```
# loeme sisse andmestikud failidest kvotAn.csv ja kvotSyn.csv, mis esitavad külad, kust tuvastati analüütilisi ja sünteetilisi vorme
kvotAn <- read.csv("kvotAn.csv", header=T, sep=";")
kvotSyn <- read.csv("kvotSyn.csv", header=T, sep=";")
# põhjana kasutame eelnevalt esitatud Google'i kaarti
# siin anname ette kaardi keskpunkti koordinaatidega, mitte kohanimega
eesti <- get_map(c(lon = 25.01361, lat = 58.59527), zoom=7, color = "bw")
```

Viimati nimetatud sammuga salvestame Eesti kaardi (objekt *eesti*), millele kanname andmestiku külade kohta, kust vaadeldavaid vorme tuvastati.

```
eesti_pohi <- ggmap(eesti)
eesti_pohi + geom_point(data = kvotAn, aes(x = lon, y = lat), size=1)
eesti_pohi + geom_point(data = kvotSyn, aes(x = lon, y = lat), size=1)
```



Kaart 5. Analüütiliste (vasakul) ja sünteetiliste (paremal) kvotatiivikonstruktsioonide tuvastamispunktid eesti murrete korpuse andmestikus (Pilvik, Uiboaed 2014)

Iseenesestmõistetavalt ei piirdu esitatud andmete visualiseerimisvõimalused ainult eesti murdealadega. Sarnase kaardi võime visualiseerida ka vadja korpusedmestikuga. Eesti murrete korpuse (EMK 2014) üks osa sisaldab morfoloogiliselt märgendatud vadja keele tekste. Visuaalselt võime esitada sellise andmestiku analoogselt eelnevatega: tähistame kaardil geograafilised punktid, kust on korpusedmestiku tekstid pärit. Kuna Vadjamaad kui ametlikku haldusüksust ei eksisteeri, siis võib keskpunktiks võtta mis tahes sobiva piirkonna, olenevalt sellest, millist kaardikuju eelistada. Järgnevas näites kaardil 6 oleme selleks n-ö keskpunktiks valinud Habolovo järve.

```

# vadja keele alamkorpuse sagedusandmestik on failis vadja.csv, mille
# loeme järgneva käsuga R-i
vad <- read.csv("vadja.csv", header=T, sep=";", dec=",")

# laeme vajaliku kaardi
vadja <- get_map("Habolovo", zoom=9, color = "bw")
vadja_pohi <- ggmap(vadja,)

# esitame külad
vadja_pohi + geom_point(data = vad, aes(x = lon, y = lat), size=1)

```



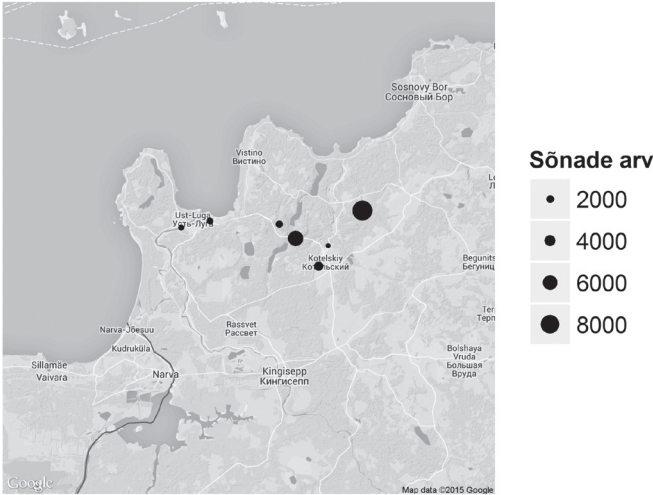
Kaart 6. Murdekorpuse vadja keele alamkorpuse materjali kogumispunktid

Samale kaardile võime lisada rohkem visuaalset informatsiooni, näiteks muuta punktide suurust vastavalt sellele, kui palju materjali korpuses mingist külast sisaldub: punktide suurus on seotud sõnade arvuga konkreetse küla korpusedmestikus.

```

vad <- read.csv("vadja.csv", header=T, sep=";", dec=",")
vadja <- get_map("Habolovo", zoom=9, color = "bw")
vadja_pohi <- ggmap(vadja)
vadja_pohi + geom_point(data = vad, aes(x = lon, y = lat, size=sonu))
+
scale_size_continuous(range = c(1, 4), name="Sõnade arv")

```



Kaart 7. Eesti murrete korpuse vadja keele alamkorpuse materjali kogumispunktid ja sõnade arv visuaalselt

Kaardil 7 on musta punktiga tähistatud koht, kust materjali on kogutud ning punkti suurus on viidud vastavusse sõnade arvuga igast punktist. Mida suurem punkt, seda rohkem on sellest külast korpuses sõnu ja vastupidi.

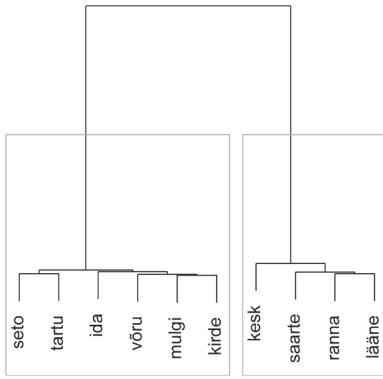
4.3. Statistilise analüüsi tulemuste esitamine

Järgnevalt näitame, kuidas samu vahendeid kasutades esitada kaardil klasteranalüüsi tulemusi. Klasteranalüüs on üks klassifitseerimismeetoditest, mis sarnaste tunnuste alusel grupeerib arvutuslikel põhimõtetel uuritavaid nähtusi (Everitt jt 2011). Siin me ei peatu meetodi üksikasjadel, vaid esitame lihtsalt ühe siinkirjutaja varasema uurimuse (Uiboed 2013) klasteranalüüsi tulemused. Kasutame juba tuttavat *tud*-partitsiibi ja finiiitse *saama*-verbi ühendite andmestikku. Klasteranalüüsiga saame dendrogrammi, mis on esitatud alloleval joonisel 8.

```
# analüüsiks loeme taas sisse andmestiku
klaster <- read.csv("tudSaama.csv", header=T, sep=";", row.names=1)

# arvutame kaugusmaatriksi, vajalik klasteranalüüsi sisendiks
andmed.kaugus <- dist(klaster, method="euclidean", upper=TRUE,
diag=TRUE)
klasteranalüüs <- hclust(andmed.kaugus, method="ward.D")

# graafik
plot(klasteranalüüs)
rect.hclust(klasteranalüüs)
```



Joonis 8. Klasteranalüüsi dendrogramm *tud*-partiisiibi ja finiiitse *saama*-verbi ühendite sageduse põhjal eesti murretes

Joonise 8 kohta võib lihtsustatult öelda, et sarnased objektid grupeeritakse samasse rühma ning moodustub kaks homogeenet ja teineteisest eristuvat rühma, mis on joonisel tähistatud heledamate joontega. Vasakpoolsesse kuuluvad Seto, Tartu, ida-, Võru, Mulgi ja kirdemurre ning teise, parempoolsesse rühma kesk-, saarte, ranna- ja läänemurre. Üsna sarnane pilt oli hoomatav ka sageduspõhistel üleminekukaartidel.

Järgnevalt kombineerime eelnevalt esitatud võimalusi ja esitame klasteranalüüsi tulemused ka geograafiliselt. Selleks kasutame faili *klasterTulemus.csv* (Uiboed, Kyröläinen 2015), kus on esitatud polügoonide joonistamise järjekord, koordinaadid (sama andmestik, mis on failis *koordinaadid.csv*), murded ning klasteranalüüsi tulemusena saadud grupitunnus ja põhjana kasutame Google'i kaarti. Praegu visualiseerime vaid kaht analüüsis selgemalt eristunud gruppi.

```
klusterVis <- read.csv("klasterTulemus.csv", header=T, sep=";")
qmap("Eesti", zoom=7, color = "bw") +
geom_polygon(aes(x = lon, y = lat, group = murre, fill = klaster,
data=klusterVis))
```



Kaart 9. Klasteranalüüsi tulemused geograafiliselt

Kaart 9 esitab klasteranalüüsi tulemused kaardil, mis on kombineeritud ülal esitatud ise joonistatud kaardist ning Google'i kaardist. Sellisel viisil on kaks eelist. Ühelt poolt saame kasutada murdepiire, mis ei ühti haldusterritoriaalsete piiridega, teisalt muudame kaardi informatiivsemaks, lisades ühe lisakihina tänapäevase kaardi, mis annab võimaluse esitada uurimismaterjali ja -tulemusi kahes kontekstis.

5. Kokkuvõte

Artiklis esitasime ühe lihtsa lingvistiliste andmete kaartidel esitamise võimaluse, rakendades selleks statistikaprogrammi R ja olemasolevaid vabavaralisi kaardiressursse. Näitasime, kuidas on võimalik ise poolautomaatselt kaarti joonistada, kui olemasolevad kaardid pole piisavad, ning kuidas seda kaarti elektrooniliselt töödelda. Pakkusime välja võimalusi sagedusandmestiku esitamiseks kaartidel ning näitasime, kuidas võib kaartidel visualiseerida statistilise analüüsi, täpsemalt klasteranalüüsi tulemusi.

Artikli põhieesmärk oli pakkuda võimalikult lihtne lahendus kirjeldatud ülesannete täitmiseks, eeldamata kasutajalt sügavaid teadmisi geoinformaatikast ja eri programmidest. Välja töötatud ja rakendatud vahendid on vabalt kasutatavad (vt Uiboaed, Kyröläinen 2015) ning nende eeskujul on võimalik lihtsa vaevaga luua endale sobivad kaardid. Pakutud lahendus muudab lingvistiliste andmete ruumilise esitamise tehniliselt väga lihtsaks ja kiireks ning sobib suurepäraselt esmaseks analüüsi ja andmete visualiseerimiseks.

Viidatud kirjandus

- Baayen, Harald R. 2008. *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R*. Cambridge: Cambridge University Press.
- Bibiko, Hans-Jörg 2012. Visualization and online presentation of linguistic data. – Frank Seifart, Geoffrey Haig, Nikolaus P. Himmelmann, Dagmar Jung, Anna Margetts, Paul Trilsbeek (Eds.). *Potentials of Language Documentation: Methods, Analyses, and Utilization*. Honolulu: University of Hawai'i Press, 96–104.
- Bivand, Roger S.; Pebesma, Edzer J.; Gómez-Rubio, Virgilio 2013. *Applied Spatial Data Analysis with R*. New York–Heidelberg–Dordrecht–London: Springer.
- EMK 2014. Eesti murrete korpus. <http://www.murre.ut.ee/mkweb/> (1.12.2013).
- Everitt, Brian S.; Landau, Sabine; Leese, Morven; Stahl, Daniel 2011. *Cluster Analysis*. 5th revised edition. Wiley Series in Probability and Statistics. Chichester: Wiley-Blackwell.
- Goebel, Hans 2006. Recent Advances in Salzburg Dialectometry. – *Literary and Linguistic Computing: Journal of the Association for Literary and Linguistic Computing*, 21 (4), 411–435. <http://dx.doi.org/10.1093/lc/fql042>
- Google Maps 2014.
- Gries, Stefan Th. 2009. *Statistics for Linguistics with R: A Practical Introduction*. Berlin: Mouton de Gruyter. <http://dx.doi.org/10.1515/9783110216042>
- Johnson, Keith 2008. *Quantitative Methods in Linguistics*. Malden: Blackwell Publishing Ltd.
- Kahle, David; Wickham, Hadley 2013. ggmap: A package for spatial visualization with Google Maps and OpenStreetMap. R package version 2.3. <http://CRAN.R-project.org/package=ggmap> (15.2.2015).
- Krikmann, Arvo 2005. Mart R Emmeli “geograafilised figuurid”. [“Geographical figures” by Mart R Emmel.] – Joel Sang (Toim.). *Endspiel*. Kummardus Mart R Emmelile. Tallinn: Eesti Keele Sihtasutus, 91–112.

- Kyröläinen, Aki-Juhani; Uiboaed, Kristel (ilmumas). Visualization Techniques for Spatial Linguistic Data Using ggplot2 and ggmap.
- Nerbonne, John; Colen, Rinke; Gooskens, Charlotte; Kleiweg, Peter; Leinonen, Therese 2011. Gabmap – A Web Application for Dialectology. – *Dialectologia*, Special Issue II, 65–89.
- Pilvik, Maarja-Liisa; Uiboaed, Kristel 2014. Grammatical evidentiality in Estonian: Areal evidence of variation. – *Ettekanne konverentsil Syntax of the World's Languages VI (SWL6)*, Pavia.
- QGIS Development Team 2014. QGIS Geographic Information System. Open Source Geospatial Foundation. <http://qgis.osgeo.org>.
- R Development CoreTeam 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing. <http://www.R-project.org/>.
- Regio 2014. Eesti kihelkondade kaart. Regio AS. <http://kaart.delfi.ee/>.
- Saareste, Andrus 1955. *Petit atlas des parlers estoniens. Väike eesti murdeatlas*. Uppsala: Almqvist & Wiksell.
- Uiboaed, Kristel 2013. *Verbiühendid eesti murretes*. [Verb Constructions in Estonian Dialects.] *Dissertationes philologiae estonicae Universitatis Tartuensis* 34. Tartu: Tartu Ülikooli Kirjastus. <http://hdl.handle.net/10062/34499>
- Uiboaed, Kristel; Hasselblatt, Cornelius; Lindström, Liina; Muischnek, Kadri; Nerbonne, John 2013. Variation of verbal constructions in Estonian dialects. – *Literary & Linguistic Computing*, 28 (1), 42–62.
- Uiboaed, Kristel; Kyröläinen, Aki-Juhani 2015. *Keeleteaduslike-andmete-ruumilisi-visualiseerimisvoimalusi*. [Techniques for spatial data visualization in linguistics.] *AndmestikeFailid*. <https://github.com/kristel-/Keeleteaduslike-andmete-ruumilisi-visualiseerimisvoimalusi> (22.2.2015).
- Wickham, Hadley 2009. *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer.

Kristel Uiboaed (Tartu Ülikool) on eesti murrete teadur. Peamised uurimisvaldkonnad on keeleteaduslike andmete kvantitatiivne analüüs (eriti varieerumise uurimisel), korpuslingvistika, eesti murrete süntaktiline varieerumine, geograafiliste andmete esitusvoimalused. Tartu Ülikool, eesti ja üldkeeleteaduse instituut, Jakobi 2, 51014 Tartu, Estonia kristel.uiboaed@ut.ee

Aki-Juhani Kyröläinen (Turu Ülikool, Tartu Ülikool) on järeldoktor. Peamised uurimisvaldkonnad on statistilised ja katselised meetodid keeleteaduses ning uurimistulemuste ja -andmete visualiseerimine. Tartu Ülikool, eesti ja üldkeeleteaduse instituut, Jakobi 2, 51014 Tartu, Estonia akkyro@gmail.com

TECHNIQUES FOR SPATIAL DATA VISUALIZATION IN LINGUISTICS

Kristel Uiboaed¹, Aki-Juhani Kyröläinen^{2,1}

University of Tartu¹, University of Turku²

Data visualization is an integral part of scientific inquiry in order to represent data and communicate findings. Recent developments such as the rise of large-scale corpora show that techniques to relate linguistically informed analysis and spatial data visualization have become increasingly important for quantitative analysis. Although spatial data visualization has gained momentum, these techniques may not be readily available for small or understudied languages. Here, we give an introduction to spatial data visualization using publicly available resources.

We use case studies on Estonian and Votic data to illustrate certain basic tasks in quantitative dialectology. We give solutions to create spatial maps based on either self-extracted coordinates or Google Maps. These maps can be used as a base layer and additional information, such as metadata and frequency distributions, can be represented on top of them. This approach can also be applied to statistical analysis. We illustrate this by carrying out a cluster analysis and its visualization using Google Maps. Thus, a toolkit is provided for quantitative analysis and spatial visualization in dialectology.

Keywords: dialectology, dialect syntax, geolinguistics, corpus linguistics, Estonian

THE TREATMENT OF LEXICAL COLLOCATIONS IN EFL COURSEBOOKS IN THE ESTONIAN SECONDARY SCHOOL CONTEXT

Liina Vassiljev, Liljana Skopinskaja,
Suliko Liiv

Abstract. The article investigates lexical collocations encountered in English as a foreign language (EFL) instruction in Estonian upper secondary schools. This is achieved through a statistical analysis of collocations featuring in three coursebooks where the collocations found are analysed in terms of their type, frequency and usefulness index by studying them through an online language corpus (Collins Wordbanks Online). The coursebooks are systematically compared and contrasted relying upon the data gathered. The results of the study reveal that the frequency and range of lexical collocations in a language corpus have not been regarded as an essential criterion for their selection and practice by any of the coursebook authors under discussion.

Keywords: coursebook evaluation, corpus analysis, lexical collocation, EFL instruction, Estonian secondary school level

1. Introduction

Effective communication in a foreign language (FL) requires, among other aspects, a vast repertoire of lexical knowledge which does not refer to the broad knowledge of single words but rather to the company these words keep, as proposed by Firth (1968: 179): “You shall know a word by the company it keeps.” Collocations, as a type of frequently co-occurring, pre-fabricated items, help learners to achieve native-like command and fluency in a FL, and their significance as a useful input material should be acknowledged in the classroom in terms of teaching materials’ design and employment.

1.1. Key concepts used in the research

The availability of corpus analyses of spoken and written texts has provided us with the evidence of lexical patterning in language output (Moon 1998: 81, O’Keefe et al. 2007: 60–61), and Bengt Altenberg (1998: 102) has even suggested that over 80% of “words in the corpus form part of recurrent word-combinations in one form or the other.”

Over 50 different terms have been offered to describe formulaicity, such as chunks, collocations, fixed expressions, formulae, multi-word items, ready-made expressions, and semi-preconstructed phrases that constitute single choices (Wray 2005: 9). Despite the diversity of formulaic sequences, some basic criteria, such as institutionalisation, fixedness, non-compositionality, and frequency of occurrence (in a corpus) may be indicated (Schmitt, Carter 2004: 3). According to Nadja Nesselhauf (2005: 1), collocations, as arbitrarily restricted lexeme combinations, form one type of recurrent word combinations. From an instructional point of view, which is the focus of the current study, it makes sense to regard collocations as items which frequently occur together and have some degree of semantic unpredictability (Nation 2001: 317).

The notion of collocation was first taken up by Harold E. Palmer (1933: 4), who defined it as a succession of two or more words whose meaning is not obvious from its component parts, and must thus be learnt “as an integral whole or independent entity.” Notwithstanding the importance of Palmer’s contribution, theoretical prominence to the notion was given by J. R. Firth (1957: 194–195) who brought forward the technical term “meaning by collocation”, calling attention to the environment of a word in determining its meaning, and claiming that the co-occurring words determine in part the meaning of a particular word. The decades following Firth’s work up to the present have witnessed a great number of descriptive studies examining the nature of collocations in linguistics, lexicography, corpus linguistics, pedagogy, and translation studies (Krishnamurthy 2006: 598) which, however, has resulted in a lack of precision while defining the concept (Bahns 1993: 57, van den Meer 1998: 313, Nesselhauf 2005: 11).

Despite the variety of interpretations, two main perspectives on clarifying the notion may be outlined: a purely statistical approach and a more linguistically motivated approach (Seretan 2011: 10–17). The advocates of the former view the notion as a statistical phenomenon of word co-occurrence, describing it in terms of corpus analysis and frequency counts (Halliday 1994, Halliday 2002, Sinclair 1991, Sinclair 1997). Based upon a large-scale frequency count into the study of lexis and collocation (The OST Report 1970), Sinclair (1991: 110) maintains that two diametrically opposed models of interpretation are needed to understand the relationship between meaning and language text: the open-choice principle and the idiom principle. The former conveys the view of the connection between a language text and a broad range of choices regulated by grammar, while the latter acknowledges the importance of semi-preconstructed phrases as single choices. Contrary to the purely statistical approach, the proponents of a linguistically motivated approach do not exclusively view collocation as a frequent association of words, but rather as a syntactically bound combination (Halliday, Hasan 1976, Howarth 1998, Mel’chuk 1998, Nesselhauf 2005). For the purposes of the present research, the statistical

approach has been adopted, where collocation is regarded as an example of the idiom principle, that is, as an instance where “words appear to be chosen in pairs or groups and these are not necessarily adjacent” (Sinclair 1997: 115). Collocation is thus viewed as “the occurrence of two or more words within a short space of each other in a text” (Sinclair 1997: 170).

1.2. Pedagogical implications for learning/teaching collocations in EFL instruction

Considering the role of collocational knowledge in EFL instruction, a range of arguments have been put forward to justify the incorporation of collocations in teaching materials (Hill 2000: 53–56, Nation 2001: 318–328).

- Collocational knowledge is the essence of language knowledge, as language knowledge and use are based on associations between sequentially observed language items.
- Fluent language use requires collocational knowledge, since naturally occurring text is to a large extent made up of prefabricated language.
- The use of collocations exemplifies the non-arbitrary nature and predictable patterning of the lexicon.
- Collocational knowledge facilitates learners’ thinking, memorisation, and expression of complex ideas.

This proves that collocational knowledge is a vital tool for EFL learners to achieve native-like fluency and confidence while communicating meaning. However, the results of a considerable amount of studies on non-native speakers’ command of collocations in written discourse (Howarth 1998, Siyanova, Schmitt 2008, Laufer, Waldman 2011) point to major challenges students experience, especially in the use of common and uncommon collocations, as well as high-frequency and medium-frequency collocations (Siyanova, Schmitt 2008). Apart from Mark Koprowski’s (2005) study of lexical phrases as featured in the three EFL coursebooks conducted on the basis of large-scale corpora, not much research has been done into the treatment of lexical collocations in teaching materials in terms of their amount and usefulness.

In view of the lack of studies on this topic in Estonia, there is a need for research into the treatment of collocations in EFL coursebooks currently utilised in Estonian upper-secondary school education in order to achieve a fundamental change in the methodological position of collocations, giving them “the same status as the other aspects of language – pronunciation, intonation, stress and grammar” (Hill 2000: 59).

Based on Morton Benson et al. (1997: ix) distinction between grammatical collocations and lexical collocations, the current study focuses exclusively on the latter, which are viewed as having the following structures: verb + noun, adjective + noun, noun + verb, noun + noun, adverb + adjective, verb + adverb, and verb + adjective.

In addition, the research relies upon Ute Roemer’s (2008: 113–118) classification of pedagogical corpus applications into direct and indirect corpus applications, where direct corpus applications focus on teacher/learner-based analysis of corpora for data-driven learning purposes and indirect corpus applications provide

descriptive information about a language for the evaluation and design of teaching materials and syllabi. It is the latter type of pedagogical applications which is employed here.

The present study aims to establish the relevance of lexical collocations in EFL coursebooks currently utilised at upper secondary school level in Estonia in terms of the amount and usefulness of lexical collocations featuring in collocation exercises and other vocabulary-related tasks.

2. Research material and methods

Prior to coursebook evaluation, a small-scale online survey was undertaken among 40 upper-secondary school level teachers in Estonia to find out the most frequently used teaching materials in Form 12. The participants were required to rate the amount of collocation exercises as well as the selection of lexical collocations (in terms of frequency and range of their occurrence) in the exercises of the coursebooks employed on the scale of 1–5, where 1 stands for a negligible amount/random selection, and 5 for a considerable amount/careful selection of the collocations. The three most frequently used coursebooks were selected for the purposes of the study: *Upstream Advanced Student's Book* (2003), *Advanced Expert CAE Coursebook* (2005), and *Mission 1 Coursebook* (1996). These titles are henceforth abbreviated as *UA*, *AE*, and *M1*.

In order to examine the amount and usefulness of lexical collocations found in the aforementioned coursebooks, the following steps were followed.

First, the number of overall vocabulary-related exercises (VRE) was counted and compared to that of collocation exercises (CE). VRE in Vocabulary/Language Focus sections, English in Use, Reading, Listening and Writing sections of the coursebooks were explored to determine whether the exercises contained any collocational input via word lists, phrases, etc. in, for example, matching tasks. An exercise was qualified as a CE only if it used the word “collocation” in its instruction or headline. The comparison between the total number of VRE and that of CE exemplifies the coursebook designers’ general attitude towards the notion – whether lexical collocations are deemed as an essential or a relatively neglected aspect of vocabulary practice.

Second, the total amount of lexical items in VRE and CE was compared to that of lexical collocations in CE, and to the number of lexical collocations in the other VRE. This comparison gives some insight into the authors’ overall approach towards collocations: if a VRE includes, for example, a multiple-choice selection of vocabulary items, then all the items have been counted and checked, albeit disregarding their definitions.

In order to ascertain whether a lexical item in VRE could be regarded as a lexical collocation, its occurrence was checked in each exercise by means of the *Oxford Collocation Dictionary for Students of English (OCD)* (CD-ROM version), which is based on the 100 million word British National Corpus (McIntosh et al. 2009: viii). If both the word under scrutiny (i.e. node according to Sinclair (1991: 115–116)) and the other word that occurs in the specified environment (i.e. collocate according to Sinclair (1991: 115–116)) have been included in the *OCD*, then the collocation undergoes statistical analysis.

The usefulness of lexical collocations in CE and VRE of the three coursebooks was analysed according to the adapted research model of Mark Koprowski (2005: 323). Koprowski suggests implementing frequency and range as the measures of the usefulness of vocabulary, which is consistent with the views of other scholars (McCarthy 1990: 69, Verghese 2007: 86, Boers, Lindstromberg 2008: 10, Granger 2009: 134). The usefulness score is derived from both frequency (i.e. the number of times an item occurs per million words in a computerised corpus) and range (i.e. the analysis of the five most common sub-corpora where the item has been found (Koprowski 2005: 324)). The usefulness score is the arithmetic mean of the frequency scores of the lexical item in the five sub-corpora.

The usefulness of lexical collocations derived from CE and other VRE was calculated according to t-scores of the collocations found in the Collins Wordbanks Online corpus, which was chosen from the variety of available language and learner corpora (Kitsnik 2006, Eslon, Metslang 2007) for two reasons: its adherence to the ideas of the Lexical Approach (Lewis 2000), which stresses the role of lexical, including collocational, input in EFL teaching, and its free accessibility at the time of the research.

Since range has been regarded as another crucial element for the selection of language input for EFL learners, four sub-corpora (with a total of 308,351,602 tokens) within the Collins Wordbanks Online corpus were created:

- Books (fiction and non-fiction) with a total of 152,400,529 tokens.
- News with a total of 83,928,217 tokens.
- Spoken language with a total of 61,527,354 tokens.
- Magazines, ephemera (i.e., any transitory written/printed matter not meant to be preserved, e.g. posters, etc) with a total of 40,495,502 token.

The minimum frequency of the collocation in the corpus is set to 5 and the minimum frequency in a given range is 3. In other words, a collocation under investigation must occur at least 5 times in the total corpus and at least 3 times in the chosen sub-corpus. Using this measure of statistical significance, t-scores were computed from the four sub-corpora for each lexical collocation found in the CE and VRE of the three coursebooks, and the arithmetic means of the t-scores were calculated to record a usefulness score for each lexical collocation.

The usefulness score, i.e. the average t-score for the four sub-corpora, can range from 0 to over 40, where 0 stands for a particular lexical collocation occurring less than 3 times in the given sub-corpus and less than 5 times in the total corpora. The usefulness scores of lexical collocations were compared and analysed according to the type of lexical collocations (as classified in Benson et al. 1997) as well as among the three coursebooks to give a better insight into the coursebook authors' procedures for selecting collocational input.

3. Results and discussion

In order to investigate the amount and usefulness of lexical collocations featuring in *UA*, *AE*, and *M1*, two studies were conducted. First, an analysis of the amount and type of collocations found in CE and VRE of the selected coursebooks was undertaken, and second, the usefulness scores (in terms of frequency and range) of the employed lexical collocations in the three coursebooks were calculated, compared and contrasted.

3.1. Analysis of the amount and type of lexical collocations

The counting of the total amount of VRE (including collocation exercises) in the Vocabulary/Language Focus, English in Use, Reading, Listening, Speaking and Writing sections reveals that quite different emphasis has generally been placed on the vocabulary input in the three coursebooks. *UA* and *M1* contain almost twice as many VRE (188 in both) as *AE* (99). It may be argued that it is the number of CE, as compared to that of VRE, that offers the most comprehensive overview of the coursebook authors' attitude towards the notion. However, only 8% (15) of the VRE in *UA*, 6% (6) in *AE*, and 0% (0) in *M1* comprise specific CE, clearly indicating that collocational practice forms a minor part in the overall vocabulary training provided in these coursebooks. Moreover, the considerable difference between the amount of CE in the three coursebooks gives some insight into the overall treatment of collocations. Thus, five units out of the total of ten in *UA* contain one CE, and five units out of ten include two CE. In *AE*, four units out of ten modules have one CE, one module comprises two CE, and five modules do not employ any collocational focus. In the case of *M1*, which lacks any specific CE, it is the responsibility of a teacher to highlight the concept for the learners, as "collocation is mostly a matter of noticing and recording" (Woolard 2000: 35). However, *M1* provides a greater amount of collocational input in the VRE than the other two coursebooks (see Figure 1).

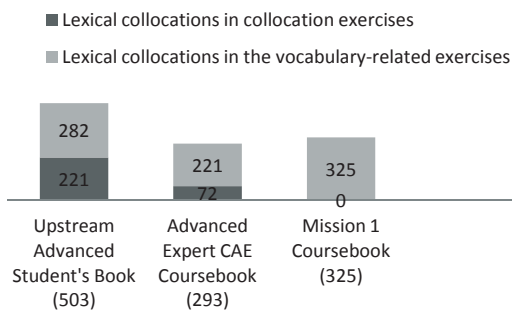


Figure 1. The comparison of the overall amount of lexical collocations in CE and in the other VRE

For a more detailed overview of the importance attached to lexical collocations, the total amount of lexical items in the VRE (including CE) has been compared with the number of lexical collocations in the CE and VRE (see Figure 2).

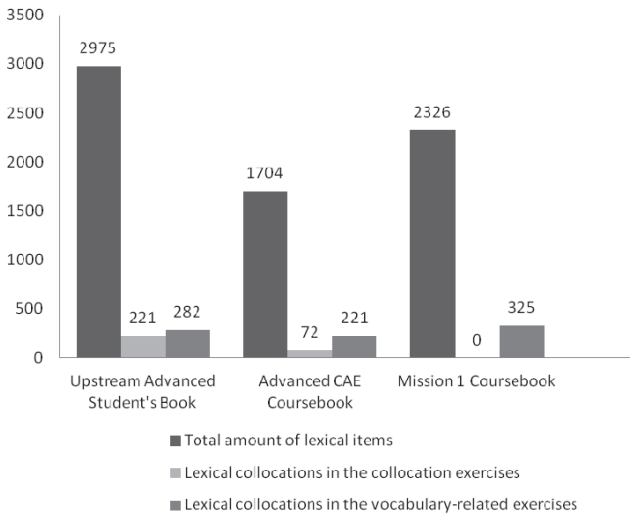


Figure 2. The comparison of the coursebooks based on the total amount of lexical items in the VRE (inc. CE), the numbers of lexical collocations in the CE and the other VRE

The results show that *UA* contains the greatest number of lexical items in the VRE (2975), followed by *M1* (2326), while the smallest amount is present in *AE* (1704). The comparison between the three coursebooks indicates that collocational practice does not form a substantial language learning element “from lesson one” as suggested by Hill (2000: 60), this meaning that a vast proportion of lexical items presented in the coursebooks are being introduced as single words following the traditional concept of the word as the unit of meaning (Sinclair 1991).

As *M1* does not contain any CE, the amounts of lexical collocations per type in CE have been counted and compared for *UA* and *AE* only (see Figure 3).

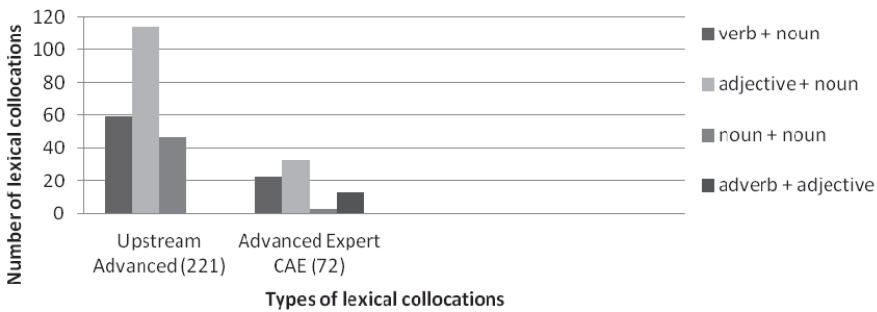


Figure 3. The amount of lexical collocations per type in collocation exercises in *UA* and *AE*

Out of the total of 221 lexical collocations found in the CE of *UA*, 114 are of the type adjective + noun, 59 of verb + noun, 47 of noun + noun, and 1 of adverb + adjective. Out of the total of 72 lexical collocations encountered in the CE of *AE*, 33 are of the type of adjective + noun, 23 of verb + noun, 13 of adverb + adjective, and 3 of noun + noun.

Figure 3 shows, therefore, that *UA* contains almost three times more lexical collocations (221) in the CE than does *AE* (72), which is consistent with the overall number of collocational practice afforded in these coursebooks. There is almost three times less collocational training included in *AE* (6 CE) than in *UA* (15 CE). In terms of the types of lexical collocations, the type of adjective + noun prevails in both CE and VRE in *UA*. This type has been presented twice as often as that of verb + noun in the CE of this book.

The amounts of lexical collocations per type in the VRE (see Figure 4) indicate that the most common type in *UA* is that of adjective + noun (111), followed by verb + noun (102) and noun + noun (32). The largest number of collocations in the VRE of *AE* is that of verb + noun (80), followed by adjective + noun (61) and adverb + adjective (50).

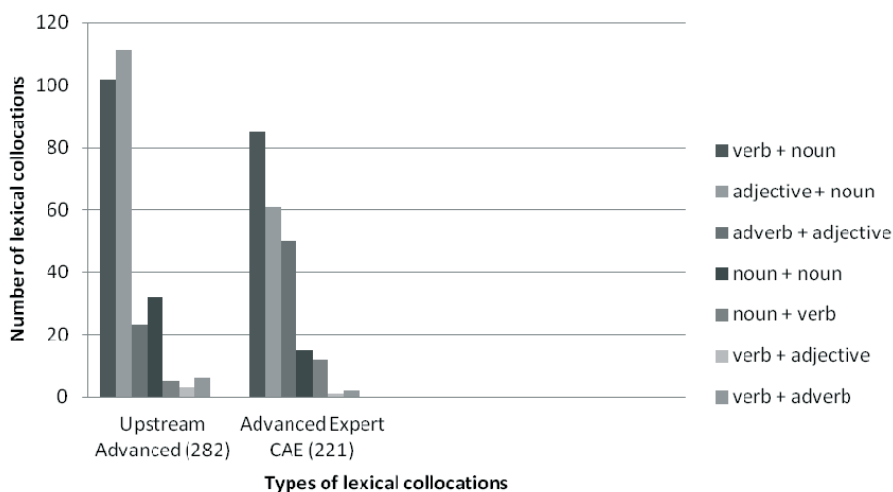


Figure 4. The amount of lexical collocations per type in the other VRE of *UA* and *AE*

As to the overall number of lexical collocations found in both CE and VRE (see Figure 5), the total amount of lexical collocations in *UA* is 503, compared to 293 in *AE* and 325 in *M1*. The most typical types of collocations employed in the exercises of *UA* are those of adjective + noun (225), verb + noun (161), and noun + noun (79). In *AE*, the most widespread types are verb + noun (103), adjective + noun (94), and adverb + adjective (63). The types of verb + noun (170), adjective + noun (88) and noun + noun (43) are the most recurrent ones in *M1*.

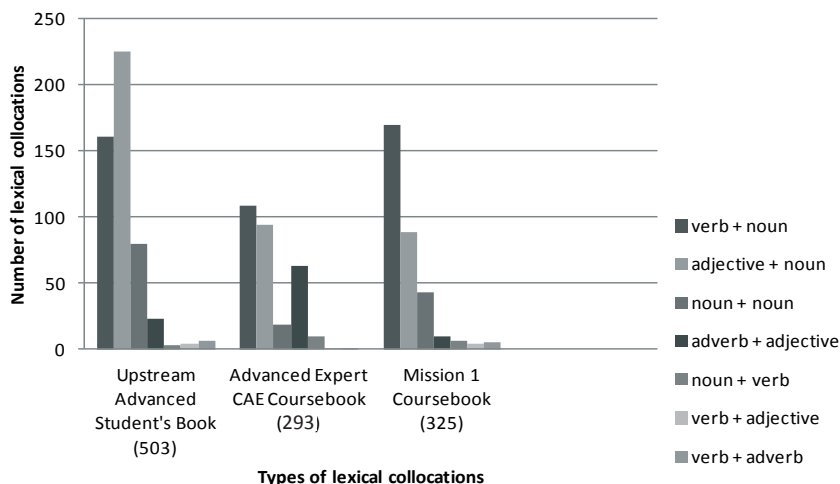


Figure 5 . The total amount of lexical collocations per type in both CE and the other VRE of *UA*, *AE* and *M1*

As can be seen from the analysis of the amount and types of lexical collocations in the three coursebooks, their authors have made different decisions about the variety of collocation types to be incorporated in the study material. *UA* has the highest amount of collocation exercises among the selected coursebooks. Moreover, most of the information provided regarding different types of collocations is also found in the exercises of *UA*. Much less data of this type is to be found in *AE* and *M1*. Ideally, a coursebook should naturally offer varied collocational practice in the EFL classroom.

3.2. Analysis of the usefulness scores of lexical collocations

The amount of lexical collocations in the aforementioned coursebooks indicates the quantity of their input, yet the quality of their selection and treatment is assessed by means of usefulness scores using the Collins Wordbanks Online corpus.

When comparing the proportion of different usefulness scores for the lexical collocations in the coursebooks (Figure 6), it can be seen that *AE* includes the greatest proportion of lexical collocations (11%) with the lowest usefulness score (0), which means that these collocations occur extremely rarely in the four sub-corpora created for the purposes of the current study. 11% of the lexical collocations in the CE and VRE of *AE*, 8% of *UA* and 5% of *M1* score that low on the usefulness measurement. In terms of the actual amount of lexical collocations with the lowest score of 0, the numbers for *UA*, *AE*, and *M1* are 40, 32 and 16 respectively.

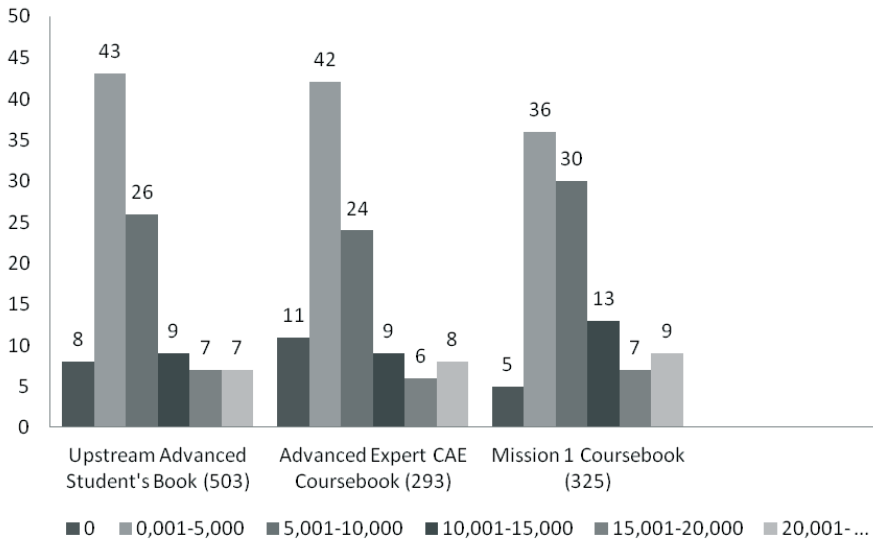


Figure 6. The comparison of lexical collocations found in CE and the other VRE of *UA*, *AE* and *M1* according to their usefulness score values in percentages

Introducing lexical collocations with the lowest usefulness score, such as *a vicious satire/headache*, *a nondescript suburban house/grey suit*, *an acknowledged comedy/story* found in the CE of *AE*, or *a delightful/quaint building*, *a pristine/littered coastline* (of the type of adjective + noun) encountered in the CE of *UA*, or *to utter a name/sound*, *to contract malaria/hepatitis* (of the type of verb + noun) taught in *AE* can be rather unhelpful for the learner. The fact that such lexical collocations have been included in the CE, i.e. in the exercises that should provide excellent input about lexical collocations in terms of their usefulness and range, raises questions about the general criteria used in selecting the teaching materials. One cannot but agree with Sylviane Granger (2009: 134) that too much valuable teaching time “is wasted on words and phrases that are not even worth bringing to learners’ attention for receptive purposes, let alone for productive purposes”.

When examining the proportion of lexical collocations with usefulness scores from 0.001–5.000, the CE and VRE of *UA* and *AE* have slightly more of such cases than does *M1*. However, collocations of such a low usefulness score (e.g. *government endorsement*) are rather infrequent in terms of their occurrences in large corpora, and their inclusion in the study materials at upper secondary level does not seem reasonable.

As regards the percentage of lexical collocations with usefulness scores ranging from 5.001 to over 20 (which may be regarded as a good score for the selected collocational material), *M1* leads the way at 59%, followed by *UA* at 49% and *AE* at 47%. Some examples are: *basic facts* (5.001–10.000), *harmful effects* (10.001–15.000), *a major concern* (15.000–20.000), *face a problem* (20.001...).

M1 seems to be the best in its selection of the collocational input, having the lowest amount of items with low usefulness scores (5%), and the highest number of collocations with higher usefulness scores (59%). Yet *M1* has no collocation-focused

exercises, which is a serious drawback concerning its overall attitude to the notion of collocations.

AE offers the poorest coverage of lexical collocations, containing the lowest amount of lexical collocations in the CE and VRE, with only six activities focusing on collocations (CE) and the highest proportion of infrequent lexical collocations with the usefulness score of 0 (11%).

Since *UA* has the greatest amount of lexical collocations in the CE and VRE, containing one or two activities in each unit, it may be argued that its users obtain more information about lexical collocations than those of *AE* and *M1*.

In summary, the analysis of the usefulness scores of the lexical collocations encountered in the three coursebooks reveals that, in general, the notion of collocations has not received close attention by any of the coursebook authors under discussion. However, this is merely an analysis of the treatment of lexical collocations in three coursebooks, in other words, an observation into how often the collocational input is provided, how useful it is, and how often a lexical item has been described by “the company it keeps” (Firth 1968: 179). It does not give any insight into the actual processes of teaching/learning in the classroom.

4. Conclusion

The present study has focused on the treatment of lexical collocations, i.e. collocations consisting of noun, adjective, verb and adverb, in the three EFL coursebooks currently employed in Form 12 of Estonian upper-secondary schools: *Upstream Advanced Student’s Book*, *Advanced Expert CAE Coursebook*, and *Mission 1 Coursebook*.

The treatment of lexical collocations refers here to the study of the amount and usefulness of lexical collocations that have been selected from the aforementioned coursebooks. In order to analyse the amount of lexical collocations, all their instances have been counted, in collocation exercises as well as in the other vocabulary-related exercises, and then compared among the coursebooks. For determining whether a lexical combination may be regarded as a lexical collocation, *Oxford Collocation Dictionary for Learners of English* has been consulted.

The usefulness of lexical collocations has been calculated according to t-scores of the collocations in the four sub-corpora of Collins WordBanks Online corpus. The usefulness scores of lexical collocations have been compared and analysed by type among the three coursebooks to give a better insight into the general selection policies of the collocational input by the coursebook authors.

The selection reveals the coursebook designers’ general attitude towards the notion of collocation – whether collocational knowledge is deemed as rendering substantial assistance for the achievement of native-like fluency and command of the target language, or whether collocations are regarded as a peripheral aspect of vocabulary practice.

The research has shown that lexical collocations are neglected rather than central in these textbooks, and their selection in terms of frequency and usefulness value is quite random. Although *Upstream Advanced* contains a considerably higher number of collocation exercises (15) than *Advanced Expert* (6) and *Mission*

1 (0), as well as that of lexical collocations, in general (503, 293, 325 for *Upstream Advanced*, *Advanced Express*, and *Mission 1* respectively), it may be argued that it is the word rather than the lexical collocation that is deemed the unit of meaning by the authors of all three coursebooks.

It has to be admitted, however, that the present study concerns only the exercises, and not the texts featuring in these coursebooks. Hence, further research is needed in this area.

Since these coursebooks are in current use in the EFL classroom in Estonia, the findings, though conflicting, should prove useful for both teachers and learners employing the same textbooks, but most of all, for coursebook designers.

References

- Altenberg, Bengt 1998. On the phraseology of spoken English: The evidence of recurrent word-combinations. – Anthony Paul Cowie (Ed.). *Phraseology: Theory, Analysis, and Applications*. Oxford: Oxford University Press, 101–122.
- Bahns, Jens 1993. Lexical collocations: A contrastive view. – *ELT Journal*, 47 (1), 56–63. <http://dx.doi.org/10.1093/elt/47.1.56>
- Bell, Jon; Gower, Roger; Hyde, David 2005. *Advanced Expert CAE Coursebook*. Harlow: Pearson.
- Benson, Morton; Benson, Evelyn; Ilson, Robert F. (Eds.) 1997. *The BBI Dictionary of English Word Combinations*. Rev. ed. Amsterdam: John Benjamins. [http://dx.doi.org/10.1075/z.bbi1\(2nd\)](http://dx.doi.org/10.1075/z.bbi1(2nd))
- Boers, Frank; Lindstromberg, Seth 2008. How cognitive linguistics can further vocabulary teaching. – Frank Boers, Seth Lindstromberg (Eds.). *Cognitive Approaches to Teaching Vocabulary and Phraseology*. Berlin: Mouton de Gruyter, 1–61.
- Collins Wordbanks Online. <http://www.collins.co.uk/page/Wordbanks+Online> (2.5.2013).
- Eslon, Pille; Metslang, Helena 2007. Öppijakeel ja eesti vahekeele korpus. [Learner language and Estonian interlanguage corpus.] – *Eesti Rakenduslingvistika Ühingu aastaraamat*, 3, 99–116. <http://dx.doi.org/10.5128/ERYa3.07>
- Evans, Virginia; Dooley, Jenny 1996. *Mission 1*. Berkshire: Express Publishing.
- Evans, Virginia; Edwards, Lynda 2003. *Upstream Advanced: Student's Book*. Berkshire: Express Publishing.
- Firth, John Rupert 1957. *Papers in Linguistics 1934–1951*. London: Oxford University Press.
- Firth, John Rupert 1968. A synopsis of a linguistic theory. – Frank Robert Palmer (Ed.). *Selected Papers of J. R. Firth*. Oxford: Oxford University Press, 179.
- Granger, Sylviane 2009. From phraseology to pedagogy: Challenges and prospects. – Thomas Herbst, Susen Faulhaber, Peter Uhrig (Eds.). *The Phraseological View of Language: A Tribute to John Sinclair*. New York: Walter de Gruyter, 123–146.
- Halliday, M.A.K. 1994. *An Introduction to Functional Grammar*. Rev. ed. London: Edward Arnold.
- Halliday, M.A.K. 2002. Lexis as a linguistic level. – Jonathan J. Webster (Ed.). *Collected Works of M.A.K.Halliday*. Vol. 1. London: Continuum, 158–172.
- Halliday, M.A.K.; Hasan, Ruqaiya 1976. *Cohesion in English*. New York: Longman.
- Hill, Jimmie 2000. Revising priorities: From grammatical failures to collocational success. – Michael Lewis (Ed.). *Teaching Collocation: Further Developments in the Lexical Approach*. Boston: Thomason/Heinle, 47–67.
- Howarth, Peter 1998. The phraseology of learners' academic writing. – Anthony Paul Cowie (Ed.). *Phraseology: Theory, Analysis, and Applications*. Oxford: Oxford University Press, 161–186.

- Kitsnik, Mare 2006. Keelekorpused ja võõrkeeļõpe. [Language corpora and foreign language teaching.] – Eesti Rakenduslingvistika Ūhingu aastaraamat, 2, 93–107. <http://dx.doi.org/10.5128/ERYa2.07>
- Koprowski, Mark 2005. Investigating the usefulness of lexical phrases in contemporary coursebooks. – *ELT Journal*, 59 (4), 322–332. <http://dx.doi.org/10.1093/elt/ccio61>
- Krishnamurthy, Ramesh 2006. Collocations. – Keith Brown (Ed.). *Encyclopedia of Language and Linguistics*. 2nd. ed. Vancouver: Elsevier, 598.
- Laufer, Batia; Waldman, Tina 2011. Verb-noun collocations in second language writing: A corpus analysis of learners' English. – *Language Learning*, 61 (2), 647–672. <http://dx.doi.org/10.1111/j.1467-9922.2010.00621.x>
- Lewis, Michael (Ed.) 2000. *Teaching Collocation: Further Developments in the Lexical Approach*. Bosten: Thomson/Heinle.
- McCarthy, Michael 1990. *Vocabulary*. Oxford: Oxford University Press.
- McIntosh, Colin; Francis, Ben; Poole, Richard (Eds.) 2009. *Oxford Collocations Dictionary for Students of English: A Corpus-Based Dictionary with CD-ROM Which Shows the Most Frequently Used Word Combinations in British and American English*. 2nd revised ed. Oxford: Oxford University Press.
- Mel'cuk, Igor 1998. Collocations and lexical functions. – Anthony Paul Cowie (Ed.). *Phraseology: Theory, Analysis, and Applications*. Oxford: Oxford University Press, 23–53.
- Moon, Rosamunde 1998. Frequencies and forms of phrasal lexemes in English. – Anthony Paul Cowie (Ed.). *Phraseology: Theory, Analysis, and Applications*. Oxford: Oxford University Press, 79–100.
- Nation, I.S.P. 2001. *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press. <http://dx.doi.org/10.1017/CBO9781139524759>
- Nesselhauf, Nadja 2005. *Collocations in a Learner Corpus*. Amsterdam: John Benjamins. <http://dx.doi.org/10.1075/sc1.14>
- O'Keefe, Anne; McCarthy, Michael; Carter, Ronald 2007. *From Corpus to Classroom: Language Use and Language Teaching*. Cambridge: Cambridge University Press. <http://dx.doi.org/10.1017/CBO9780511497650>
- Palmer, Harold Edward 1933. *Second Interim Report on English Collocations*. Tokyo: Kaitakusha.
- Roemer, Ute 2008. Corpora and language teaching. – Anke Lüdelin, Kytö Meja (Eds.). *Corpus Linguistics: An International Handbook*. Vol. 1. New York: Walter de Gruyter, 112–131.
- Schmitt, Norbert; Carter, Ronald 2004. Formulaic sequences in action: An introduction. – Norbert Schmitt (Ed.). *Formulaic Sequences: Acquisition, Processing and Use*. Amsterdam: John Benjamins, 1–22.
- Seretan, Violeta 2011. *Syntax-Based Collocation Extraction*. Dordrecht: Springer. <http://dx.doi.org/10.1007/978-94-007-0134-2>
- Sinclair, John Mchardy 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Sinclair, John McHardy 1997. Corpus evidence in language description. – Anne Wichmann (Ed.). *Teaching and Language Corpora*. London: Longman, 27–39.
- Siyanova, Anna; Schmitt, Norbert 2008. L2 learner production and processing of collocation: A multi-study perspective. – *The Canadian Modern Language Review*, 64 (3), 429–458. http://www.victoria.ac.nz/lals/about/staff/publications/Siyanova-and-Schmitt_CMLR2008.pdf (8.1.2013).
- van der Meer, Geert 1998. Collocations as one particular type of conventional word combinations: Their definition and character. – *Proceedings of the 8th Euralex Conference*, 4–8 August, 1998. http://www.euralex.org/elx_proceedings/Euralex1998_1/Geert/VAN/DERMEER/Collocations/as/one/particular/type/of/conventional/word/combinations/Their/de.pdf (19.2.2013).

- Verghese, Paul 2007. *Teaching English as a Second Language*. New Dehli: Sterling.
- Woolard, George 2000. Collocation – encouraging learner independence. – Michael Lewis (Ed.). *Teaching Collocation: Further Developments in the Lexical Approach*. Boston: Heinle, 28–46.
- Wray, Alison 2005. *Formulaic Language and Lexicon*. Cambridge: Cambridge University Press.

Liina Vassiljev is a teacher of English at upper-secondary school level and her research interests include coursebook evaluation.

Narva mnt 29, 10120 Tallinn, Estonia
liina.vassiljev@yahoo.com

Liljana Skopinskaja (Tallinn University) primary research interests are intercultural communication and its assessment, coursebook evaluation, foreign language teaching methodology, and language teacher education.

Narva mnt 29, 10120 Tallinn, Estonia
liljana.skopinskaja@tlu.ee

Suliko Liiv (Tallinn University) research interests are contrastive linguistics, intercultural communication, language policy, testing, methods of foreign language teaching.

Narva mnt 29, 10120 Tallinn, Estonia
liiv@tlu.ee

LESIKAALSETE KOLLOKATSIOONIDE KÄSITLEMINE INGLISE KEELE KUI VÕORKEELE ÕPIKUTES EESTI KESKKOOLI KONTEKSTIS

**Liina Vassiljev, Liljana Skopinskaja,
Suliko Liiv**

Tallinna Ülikool

Käesolev artikkel uurib leksikaalsete kollokatsioonide käsitlust kolmes 12. klassi inglise keele õpikus: “Upstream Advanced Student’s Book” (2003), “Advanced Expert CAE Coursebook” (2005) ja “Mission 1 Coursebook” (1996), mis valiti välja 12. klassi õpetajate ankeetküsitluse alusel kui kõige kasutatavamad õpikud selles kooliastmes. Lisaks uuriti ka sõnavara harjutustes esinevate kollokatsioonide hulka ja kasulikkust. Kollokatsioonide kasulikkus mõõdeti veebipõhise inglise keele korpuse (Collins Wordbanks Online) statistilise arvutusega ning kollokatsioonide esinemist analüüsiti neljas alamkorpuses, mis esindavad erinevaid tekstitüüpe. Kollokatsioonide üldarvu ja kasulikkuse uurimine õpikutes näitas, et kollokatsioone ei käsitleta kuigi põhjalikult ning õpiku autorite poolne kollokatsioonide valik on pigem juhuslik kui läbimõeldud. Õpikud erinevad nii üldise sõnavaraharjutuste arvu kui ka konkreetsete kollokatsioonide harjutuste poolest.

Võtmesõnad: õpikuanalüüs, korpusanalüüs, leksikaalne kollokatsioon, inglise keel kui võõrkeel, keskkooliaste

Trükitud 150 eksemplari