

Eesti Rakenduslingvistika Ühing

Eesti Rakenduslingvistika Ühingu aastaraamat 10

Estonian Papers in Applied Linguistics 10

Toimetanud
Helle Metslang, Margit Langemets
ja Maria-Maren Sepper

Tallinn 2014

EESTI RAKENDUSLINGVISTIKA ÜHINGU AASTARAAMAT 10 **Estonian Papers in Applied Linguistics 10**

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

Toimetuskolleegium / Advisory board:

Martin Ehala (Tallinn), Auli Hakulinen (Helsinki), Birute Klaas (Tartu), Eino Koponen (Helsinki), Irina Külmoja (Tartu), Maisa Martin (Jyväskylä), Jaan Mikk (Tartu), Hille Pajupuu (Tallinn), János 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)
European Reference Index for the Humanities (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 10. numbri väljaandmist on toetanud Haridus- ja Teadusministeeriumi riiklik programm "Eesti keel ja kultuurimälu II" ja Eesti Keele Instituut.

Autoriõigus / Copyright: autorid ja Eesti Rakenduslingvistika Ühing

Kujundaja / Design: Piia Ruber

ISSN 1736-2563 (trükis)

ISSN 2228-0677 (võrguväljaanne)

doi:10.5128/ERYa.1736-2563

SISUKORD

Assessment literacy of national examination interviewers and raters – experience with the CEFR <i>Riigieksami intervjuerijate ja hindajate hindamisalane pädevus – keeleõppe raamdokument ja selle tundmine</i> Ene Alas, Suliko Liiv	7
Dialoogide pragmaatilise analüüsi tarkvara <i>Software for pragmatic analysis of dialogues</i> Sven Aller, Olga Gerassimenko, Tiit Hennoste, Riina Kasterpalu, Mare Koit, Krista Mihkels, Kirsi Laanesoo, Andriela Rääbis	23
Sõnaliigid eesti lapse kõnes ja lapsele suunatud kõnes <i>Word classes in the child's speech and in the child-directed speech</i> Reili Argus, Helen Kõrgesaar	37 53
Morfosüntaktilise ja leksikaalse varieerumise piiridest: ilukirjandus- ja õppijakeele kasutusmuutrite võrdlus <i>Constraints on morphosyntactic and lexical variability</i> Pille Eslon	55
Käändevormide kasutamise oskus eakohase arenguga vene-eesti kakskeelsetel ning spetsiifilise kõnearengu puudega ükskeelsetel lastel <i>Estonian case morphology in second language acquisition and specific language impairment</i> Merit Hallap, Marika Padrik, Signe Raudik	73
Switching between Estonian Sign Language and spoken Estonian in a school environment <i>Eesti viipekeele ja suulise eesti keele vaheline koodimuutus koolikeskkonnas</i> Kadri Hein	91
Tõhususe tahud sõjandusterminoloogia näitel: ülediferentsimine või erinev tunnetusväärtus? <i>Aspects of specialised language efficiency: Over-differentiation or cognitive motivation?</i> Reet Hendrikson	107

The translation of idioms in children's cartoons: A comparative analysis of English dialogues and Lithuanian subtitles <i>Idioomide tõlkimine laste multifilmides: ingliskeelsete dialoogide ja leedukeelsete subtiitrite võrdlev analüüs</i> Ligita Judickaitė-Pašvenskienė	125
Sissevaade varasesse eesti keele süntaksi omandamisse <i>Insight into the early acquisition of syntax in Estonian</i> Airi Kapanen	139
Teismee loovkirjutiste sõnavara ja selle hindamine <i>Teenage vocabulary and its assessment in creative writing</i> Krista Kerge, Anne Uusen, Halliki Põlda	157
Grammatikakategooriate sekundaarsest kasutusest: vene ja eesti ajavormid <i>On secondary use of grammatical categories: Russian and Estonian tense forms</i> Sirje Kupp-Sazonov, Irina Külmoja	177
Küsimused eesti lapsele suunatud kõnes leedu ja vene keele taustal <i>Interrogatives in Estonian child-directed speech</i> Helen Kõrgesaar	193
<i>Põhimõtteliselt või praktiliselt Paides</i> – kahe sarnase funktsiooniga sõna käekäigust <i>Põhimõtteliselt or praktiliselt in Paide</i> – two words with similar functions Annika Küngas	209
Ühendverbid eesti keele pindsüntaktilises analüüsis <i>Particle verbs in Estonian shallow syntactic analysis</i> Kadri Muischnek, Kaili Müürisepp, Tiina Puolakainen	227
Palved eesti, soome ja vene keeles: grammatika pragmaatika teenistuses <i>Requests in Estonian, Finnish and Russian: Grammar and pragmatics</i> Renate Pajusalu	241
Authorship verification of opinion pieces in Estonian <i>Autorituvastus eestikeelsetes arvamuskirjeldustes</i> Timo Petmanson	259
Investigating cultural variability in rater judgements of oral proficiency interviews <i>Kultuuriliste erinevuste uuring suulise keelepädevustesti hindaja käitumises</i> Irina Stassenko, Liljana Skopinskaja, Suliko Liiv	269

Ilukirjanduskeelest uue pilguga <i>On belletristic language from a new perspective</i> Jekaterina Trainis, Kais Allkivi	283
Adjektiivide polüseemia korpuses ja sõnaraamatus <i>Adjectival polysemy in corpora and dictionaries</i> Maria Tuulik	307
Sõnasemantika ja teksti mõistmine põhikooli esimeses ja teises kooliastmes <i>Word semantics and text comprehension in the first and second stage of the primary school</i> Krista Uibu, Maile Timm	319
Kelle emotsioonile anda teksti ette lugedes hääli? Vaatus mitmest perspektiivist <i>Who's emotions to express while reading out loud?</i> <i>A multi-perspective approach</i> Ene Vainik	335

ASSESSMENT LITERACY OF NATIONAL EXAMINATION INTERVIEWERS AND RATERS – EXPERIENCE WITH THE CEFR

Ene Alas, Suliko Liiv

Abstract. The article investigates the training needs for the English language national examination interviewers and raters in Estonia in light of the new national examination that is set to measure students' proficiency at level B of the Common European Framework of Reference for Languages (henceforth CEFR) scale. A questionnaire study was designed to explore CEFR-related assessment literacy – the extent to which the target population employed the CEFR in their daily professional life, how accessible they felt the CEFR was for their professional needs and if they thought they could accurately place students on the CEFR levels. Additionally, the respondents assigned CEFR can do statements to the levels deemed appropriate. The results revealed ambiguity about the CEFR levels among the respondents, which lead the authors to propose a training sequence designed to empower the interviewers and raters to function more efficiently within the new national examination speaking test in Estonia.*

Keywords: assessment literacy, interviewer, rater, oral proficiency interview, CEFR, training

1. Introduction. Background to the study

The national examination in the English language in Estonia up until 2013 attempted developed measure to what extent the National Curriculum requirements concerning the learning outcomes in the English language have been met with a view that by the end of gymnasium students should reach B2 level on the CEFR scale as stipulated in the national curriculum of 2002. The exam developed was a 100 point exam, functioning as a B2 level achievement test. 'However, no research has been carried out to determine whether this is really the case. Anecdotal evidence suggests that, at least in the case of the English examination, the language competence

* This research has been supported by the Estonian Science Foundation Grant No 9037 "Assessing Variables: Creating Benchmarks in High-Stakes Proficiency Tests".

of those students who have gained at least 80-85 points out of the possible 100 (about 25-30 per cent of the test-taking population) is at level C1.' (Türk, Tender 2013: 233) which was also corroborated by SurveyLang findings reported in 2012 (European Commission 2012: 212–213). Like many other national examinations, the current exam has also acted as a proficiency measurement tool generating data that serve gate-keeping purposes in Estonia. Most universities in Estonia have set a national examination standard that has to be met in order to be admitted to a university programme.

In 2011, a new national curriculum was adopted by the Estonian Ministry of Education and Research (cf. *Gümnaasiumi riiklik õppekava*), which stipulates that students finishing Gymnasium should have (CEFR) B level proficiency in two foreign languages. Moreover, the new curriculum maintains that 'ainevaldkonda kuuluvate võõrkeelte õppe kirjeldus on üles ehitatud, lähtudes keeleoskustasemetest kirjeldustest Euroopa keeleõppe raamdokumendis [---]. Kõikide võõrkeelte [---] õpitulemused on raamdokumendile toetudes kirjeldatud ühtsetel alustel' (ibid.: Appendix 2, p. 1) ['description of the learning of foreign languages belonging to the subject area has been developed proceeding from the level descriptors of the Common European Framework of Languages [---]. The learning outcomes of all foreign languages have been described in a unitary fashion relying on the CEFR'.] Thus a conscious decision has been made to align foreign language instruction and evaluation to the guidelines of the CEFR, implying that also language proficiency testing via national examinations would have to follow suit.

As the national curriculum does not specify clearly which of the two Independent User B levels – B1 or B2 – is to be achieved, and the students may in actual fact reach either, the new national examination in the English language has been developed to allow the student to be placed at either level. The new 2014 exam is designed to be a 100 point exam like its predecessor, comprising 4 tests: writing, listening, reading and speaking. A foreign language national examination will be a compulsory graduation requirement as of next year. Although English has always been the most popular foreign language in terms of how many students choose it as one of their school-leaving examinations (the number varying between 7500 and 9500 students annually), the amount of test-takers may increase as a result of the new regulation, creating added challenges for test administration and marking.

The process of developing a new test has meant overcoming a number of challenges in terms of overall approach (a bi-level examination vs. a single level examination), validity and practicality issues like the number and nature of tasks (how many texts/ tasks/ items would be sufficient to be able to draw informed conclusions about a person's language level without completely exhausting the test-taker/ interviewer/ rater), new marking scales for writing and speaking, benchmarking, score reporting, etc.

As a new format was proposed for the speaking test, additional challenges emerged concerning interviewer/ rater awareness and behaviour during the interview, i.e. interviewer/ rater training to tackle reliability issues. For the interviewers, the novelty is predominantly procedural: familiarising themselves with the new interviewer script, adhering to the interviewer frames, understanding the nature of tasks and managing them effortlessly. Relying on earlier research on the subject (cf. Alas 2010, Reemant et al. 2013), a substantial amount of awareness-building would

also have to be spent on the issue of time management during the conduct of interviews. For those who rate the student performances during the speaking test, the challenge lies mostly in learning to implement the new marking scale appropriately and consistently, which presupposes an ability to recognise CEFR criterial levels of performance, distinguishing between all CEFR levels but more notably between levels A2/B1/B2/C1. As can be seen from the above, interviewer and rater training has assumed notable significance for the speaking test to be conducted consistently and marked validly and reliably. The task is made ever more daunting by the fact that the number of people involved in the speaking test as interviewers / raters in Estonia is 998 and they all require re-qualification before the implementation of the new examination in May 2014.

Before embarking on planning a re-training programme, information was sought as to the interviewers'/ raters' degree of familiarity with the level descriptors of the CEFR as well as their attitude to level descriptors' usefulness in their day-to-day teaching and assessment practices, i.e. the usefulness of the CEFR as an assessment/ testing instrument. The current research is thus attempting to study assessment literacy with respect to the CEFR.

The term 'assessment literacy' was coined by Stiggins (1991, 1997) and has been defined by Malone as 'stakeholders' (often with a focus on instructors') familiarity with measurement practices and the application of this knowledge to classroom practices in general and specifically to issues of assessing language' (Malone 2013: 330). There seems to be agreement (Brindley 2001, Malone 2008, Popham 2009, Fulcher 2012) that many test users – and here we specifically include teachers who prepare students for tests, develop tests and act as interviewers and raters during high stakes tests like national examinations – have limited knowledge about language testing, sometimes irrespective of the fact that language testing and assessment have been part of their teacher education. Research into assessment literacy has focused on the kinds of textbook that are available and their relevance for particular groups in language education population like teachers, testers and researchers (Davies 2008, Fulcher 2012), suggesting that more accessible testing/ assessment materials be developed for educators who are not necessarily testing experts but need to make informed assessment related decisions in their daily professional life. Another area of research are the characteristic features of language testing courses (Brindley 2001, Brown, Bailey 2008, Inbar-Lourie 2008, Malone 2008) where the findings point to 'the presence of a stable knowledge base [among test users] that is evolving and expanding, rather than shifting radically' (Brown, Bailey 2008: 371), still emphasising that a framework of core competences of language assessment be established for language testing courses (Inbar-Lourie 2008: 396–398) with the competences involving three areas: language testing/ assessment skills, language testing/ assessment knowledge and language testing/ assessment principles (Davies 2008: 328). Malone reminds the reader of the fact that language testing textbooks are not the only way to convey language testing/ assessment related expertise (Malone 2013: 332). In fact, other modern approaches, such as 'traditional as well as face-to-face workshops, online and downloadable tutorials, materials produced by professional language testing associations, reference frameworks, such as the CEFR, video projects, pre-conference workshops and series of narrative accounts about developing assessment literacy' (ibid.) probably outweigh the importance of textbooks in developing assessment literacy

among test users. With the CEFR and Manual for Language Test Development and Examining related to it being undoubtedly the most ‘influential publications [---] in the field of language learning and language testing in Europe’ (Figueras et al. 2005: 261), with research literature abounding in articles discussing the process of relating examinations to the CEFR (Figueras et al. 2005, O’Sullivan 2010, Kecker, Eckes 2010, Martyniuk 2010, Noijons et al. 2011) and how the CEFR can serve as a guide to develop assessment instruments (e.g. Little 2005), there is relatively little research into the how CEFR literate test users are. The current study aims to bridge some of that gap.

2. Method

The questionnaire study, developed for the purposes of discovering how extensively teachers employ the CEFR in their professional life and how accurate they are in describing particular CEFR levels of performance, invited the participants – teachers training to become national examination speaking test interviewers/ raters – to place 6 statements (I use CEFR for test development/ marking student work/ textbook evaluation/ curriculum design/ course evaluation, I can successfully distinguish between different CEFR levels) on a Likert scale, indicating their use of the CEFR. This closed-response section was supplemented by 5 open-ended questions, asking about the usefulness of the CEFR for the participant as a language teacher, about a particular section of the CEFR that was most useful, about the ease of use of the CEFR, about having obtained any CEFR-related training and commenting on the usefulness of the latter, and about any training that the respondents thought they might need as national examination interviewers/ raters. In each case, the respondents were also requested to account for their opinion. In the third part of the questionnaire, the respondents had to place 20 speaking sub-skills (can do statements) drawn from the CEFR on the correct level. The statements for the questionnaire were chosen from the CEFR levels A2, B1, B2 and C1 as it is these levels that the students will potentially be demonstrating during the national examination speaking test.

The questionnaire was completed by 64 respondents who were starting their interviewer/ examiner initial or re-training. The respondents fell into two groups with regard to the amount of experience they had had as national examination interviewers and trainers: novice interviewers/ trainers who had had no interviewing/ rating experience, comprising 27 people; and the experienced interviewers/ raters – 37 people who had had at least 1 year of respective practice, i.e. they had been interviewing/ rating the national examination speaking test during at least one examination session.

3. Analysis and results

This section presents the findings of the current study along with some interpretation. Although the number of participants is fairly small, the results are given in percentage value to allow a more generalised view of the outcome. An attempt was also made to see if the results of both groups differed significantly from each other

using the chi-square test. The interviewers'/ assessors' confidence level about their ability to distinguish between CEFR levels is reflected in Figure 1 below.

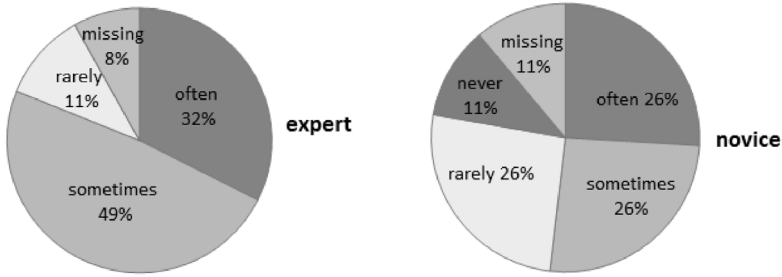


Figure 1. Interviewers'/ raters' confidence of their ability to distinguish between CEFR levels

A higher confidence level can be detected among the expert group as expected, as about a third (32%) believe themselves to be able to successfully distinguish between CEFR levels and a further 49% claim to be able to do that sometimes ($p \leq 0.05$). There is nobody in this group who said they could never do it but there is a fairly noticeable group of 8% who have not responded to this question, which may be an indication of lack of ability to separate levels. The novice group falls roughly in half with 52% of the respondents claiming to be able to identify levels successfully (26% often and 26% sometimes) and 48% needing more practice with level identification, as they can either rarely (26%) or never (11%) identify the level with a further 11% not responding to the question.

It was assumed that the interviewers and raters involved in the study who in their daily life are gymnasium level English teachers would be familiar with the CEFR and use the document in their professional life for various purposes. Five potential uses of the CEFR were proposed in the questionnaire with the request to rate the frequency of those uses. Figures 2 and 3 below reflect the results of this query for expert and novice interviewers/ raters respectively.

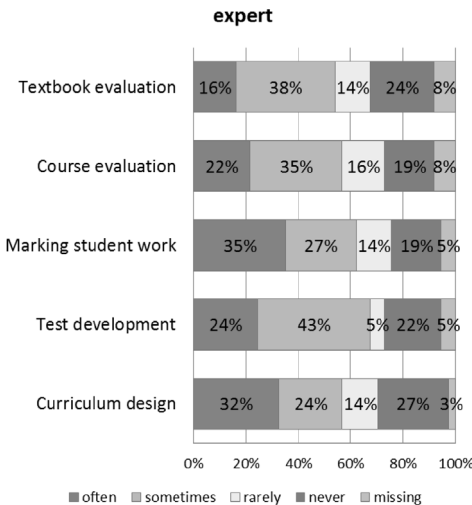


Figure 2. Using the CEFR in their professional life. Experts

It appears that the most frequent uses of the CEFR among expert interviewers/ raters are for test development (24% often and 43% sometimes) and marking student work (35% often and 27% sometimes) which is to be expected, as the CEFR is probably considered by the majority of teachers as a testing document. It is surprising, however, that more than about a third of the respondents admit either consulting the CEFR rarely or never, including the 5% who have not responded to the question. Other aspects of evaluation pertaining teaching – textbook evaluation (16% often, 38% sometimes), course evaluation (22% often, 35% sometimes) and curriculum design (32% often, 24% sometimes) are represented slightly less compared to test development and marking student work but still to an almost equal degree. Here, the group of rare or non-users is over 40% in all categories.

The comparison of the above data to that of the novice group gives a slightly different picture (cf. Figure 3 below).

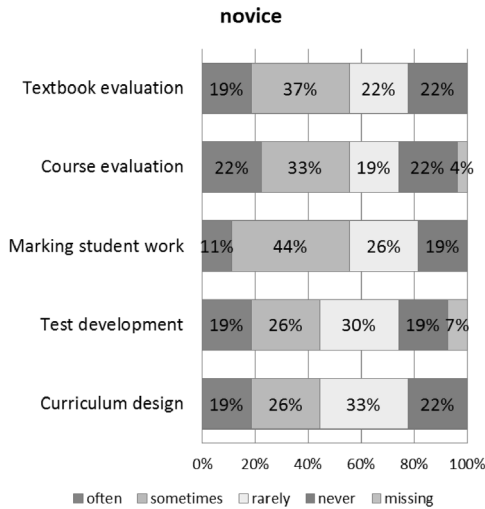


Figure 3. Using the CEFR in their professional life. Novices

The most frequently noted uses in this group are for marking student work (11% often, 44% sometimes), textbook evaluation (19% often, 37% sometimes) and course evaluation (22% often, 33% sometimes), whereas it can be noticed that the group who claims to use the CEFR frequently for textbook evaluation is larger than that of the expert group. Overall, the number of rare or non-users noticeably exceeds that of the expert group ranging from 44% for textbook evaluation to 56% in test development. Being novice national examination interviewers/ raters usually means that those respondents have also had somewhat less teaching experience, and thus they may have had less exposure to such activities as designing a course, planning a curriculum/ syllabus or evaluating the appropriacy of a textbook.

The CEFR is one of the key documents in managing (foreign) language instruction and evaluation in Europe on the governmental level. The questionnaire enquired in one of its open-ended questions how useful the respondents considered it as language teachers. The results are displayed in Figure 4 below.

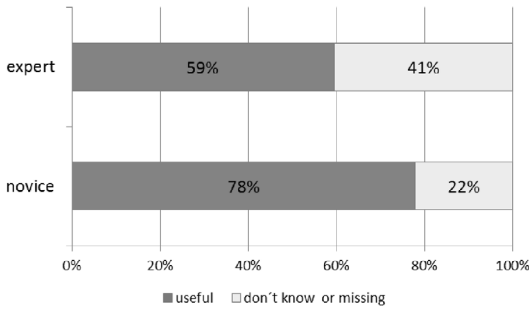


Figure 4. Usefulness of the CEFR

As can be seen, novice interviewers/ rater consider it useful more often than experts with as many as 78% of the novice respondents saying so, compared to 59% of the expert group ($p \leq 0.03$). It was speculated at this point of the analysis that the result would have something to do with the recency of the novice group's teacher education, being more familiar with the CEFR through teacher training, knowing its role and impact in planning language instruction/ evaluation. In their responses, the novice examiners/ rater pointed out CEFR's usefulness in terms of assessing the level of students, setting long-term goals, evaluating text-books and preparing for the exams; the respondents also valued CEFR because 'it is adaptable to every language situation'. Expert interviewers/ rater pointed out a larger variety of CEFR uses compared to novices: it is a guide for different kinds of evaluation, it allows the teacher to select/ customise resources, it serves as a common framework for assessment and feedback; it helps to find criteria, distinguish between proficiency levels and establish a uniform grading system; it explains different aspects of language testing. The respondents in this group valued CEFR in their development of placement tests and preparation for exams.

Although it was expected that the interviewers/ rater would consider CEFR useful for their professional purposes, the scope and the complexity of the document was predicted to cause problems for the respondent. The open-ended question concerning the ease of use of the document yielded the responses featured in Figures 5 and 6.

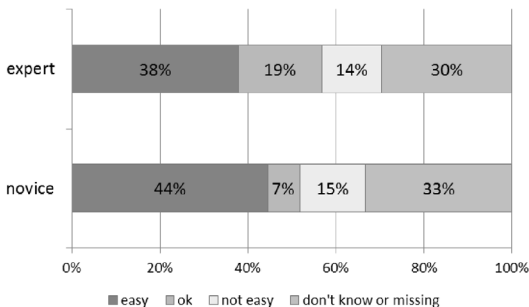


Figure 5. Ease of use of the CEFR

novices and experts together

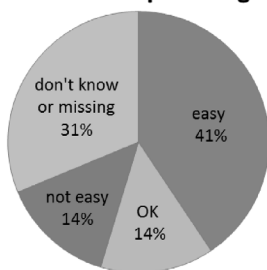


Figure 6. Ease of use of the CEFR. All respondents together

Comparing the groups that maintain that using the CEFR is either easy or mostly easy (OK) shows a slightly higher confidence level in the expert group (57% vs. 51%), the significance level of $p \leq 0.34$ showing no real difference between the groups though. While asked what made the CEFR easy to use the experts pointed out that the CEFR was well structured, logical, that the levels were clearly described but also the fact that they had gained experience with it (e.g. 'I have got used to it', 'I have studied it properly and introduced it to my teaching practice step by step'). Novices mentioned fewer reasons for the ease of use: the precision of level descriptors, clear tables and the fact that the skills were separately described. It is interesting, though, that the number of those who think it easy is larger among the novices (44% vs. 38%). This finding may be compatible with the results to the previous question: being more familiar with the CEFR through training, the respondents would be more confident using it. The fact that in both groups almost half of the respondents either considered using the CEFR hard to use (experts 14%, novices 15%) or refrained from answering the question, possibly because they had no experience using the CEFR (30% vs. 33%), points to the need of training with respect to the CEFR levels and their potential uses. Those experts who marked the CEFR not easy to use supported this by saying that language level evaluation was never easy, that the levels could be interpreted in various ways, that the evaluation was made more complicated by the difference in tasks and their own lack of practice in using the CEFR scales in their day-to-day practice. Novices mentioned similarity of level descriptors that caused confusion and their lack of experience with the scales.

All respondent were invited to share the amount and nature of their CEFR-related training so far. Figures 7 and 8 below indicate that, contrary to the assumptions made on the basis of responses to the earlier questions regarding CEFR-related training, the number of people in the expert group having gone through CEFR-related training noticeably surpasses those in the novice group (41% vs. 19%), $p \leq 0.051$.

The expert respondents who noted having had CEFR-related training indicated that although the training was useful, it had taken place quite some time ago and needed refreshing. Only a few respondents in both groups commented on the contents of the course and described it fairly ambiguously as 'working with levels'. It is noteworthy from the point of view of further training that a third of all the respondents claimed to have had no CEFR related training and 34% did not respond to the question which might be a further sign of no CEFR related training received.

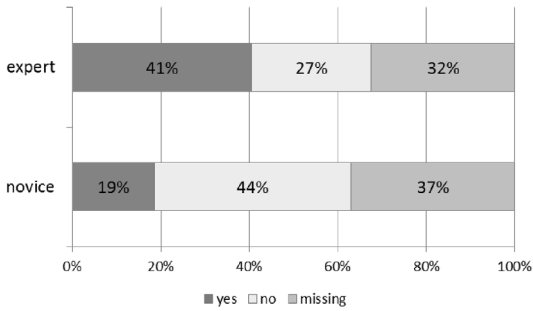


Figure 7. CEFR-related training so far

novices and experts together

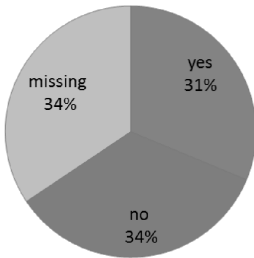


Figure 8. CEFR-related training so far. All respondents together

The final survey question concerning CEFR related training needs for the respondents as national examination interviewers/ raters provided further support for the research assumption that the target group was all but familiar with the CEFR and its potential. The responses to the question have been summarised in Figure 9 below.

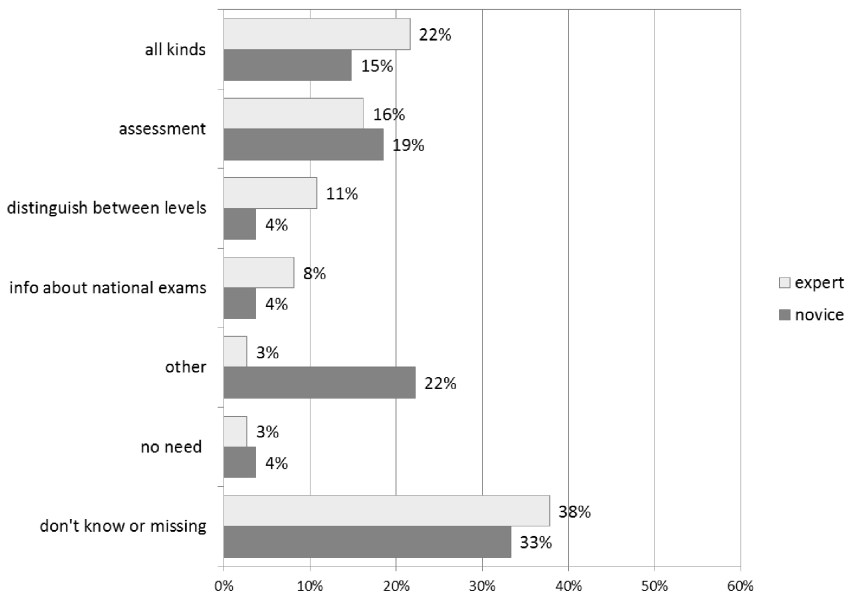


Figure 9. Further CEFR-related training needs

There is a large proportion of respondents (38% of experts, 33% of novices) who have either indicated that either they did not know the kind of training that was necessary or did not respond to the question. The second largest group consisted of those whose response seems diametrically the opposite – ‘all kinds of training’ (experts 22%, novices 15%). The answer, however, is from the training point unhelpful, as it does not point to any problem areas that training might focus on or the interviewers/ assessors are aware of or admit. The response differs from the previous group’s answer, though, in that it seems to indicate the respondents’ interest in the suggested area of development. Also, not knowing, at the time of the survey, what the new national examination would be like, and to what extent working with the new marking scale would require familiarity with the CEFR levels, it is no surprise that the prevailing responses are somewhat vague. The more focused CEFR related training needs mention assessment of student work and being able to distinguish between CEFR levels, especially when rating speaking. The more specific training considered necessary are connected to the information about national examinations (‘I need all the information connected with the new national exam as an interviewer and an assessor’), which seems more relevant to the respondents as language teachers, preparing their students for national examinations, rather than interviewers and raters. With the given group of respondents, the CEFR as a training focus is not mentioned. In this connection, it is perhaps also noteworthy that 3% of experts and 4% of novices indicated that there was no need for CEFR related training, without further specification, however.

In the final part of the survey, the respondents were asked to place 20 can do statements involving speaking derived from the CEFR levels A2, B1, B2 and C1 on the correct level. The success rate of the exercise is represented in Figures 10 and 11. The former (Figure 10) represents the overall result of assigning descriptors to the levels, whereas Figure 11 illustrates the success rate of the group of respondents who maintained that they could often distinguish between CEFR levels successfully. The figures show if the descriptor has been placed on a correct CEFR level (exact), or on a higher level (overestimated) or on a lower level than it belongs to (underestimated). As the respondents only had four levels to choose from, it was impossible for them to place A2 descriptors lower and C1 descriptors higher than they belonged to – hence the missing category of underestimated in section A2 and overestimated category in section C1.

As can be seen, the overall success rate is only marginally higher with the expert group compared to the results of all the respondents together. The level descriptors are assigned correctly on less than a half of the occasions with all the levels except C1 where the percentage of correct placement is 62 overall and 58 with the experts (sic!). It is also interesting that the descriptors are more frequently placed on a higher level than they actually belong. It may be that the respondents view the descriptors as illustrating perfect performances on that level, without allowing that although a student on A2 level would be able to ‘describe plans and arrangements, habits and routines, past activities and personal experiences’ (CEFR: 59) or a B2 level speaker would be able to ‘give detailed descriptions and presentations on a wide range of subjects related to his/ her field of interest, expanding and supporting ideas with subsidiary points and relevant examples’ (CEFR: 58) both students would still be struggling with language, A2 student with grammar as well as vocabulary and B2 student probably not so much with grammar but still with the range and appropriacy of vocabulary.

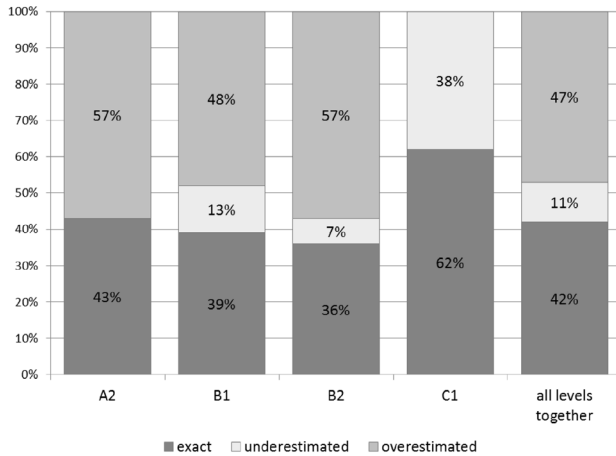


Figure 10. Assigning can do statements to appropriate CEFR levels. Overall

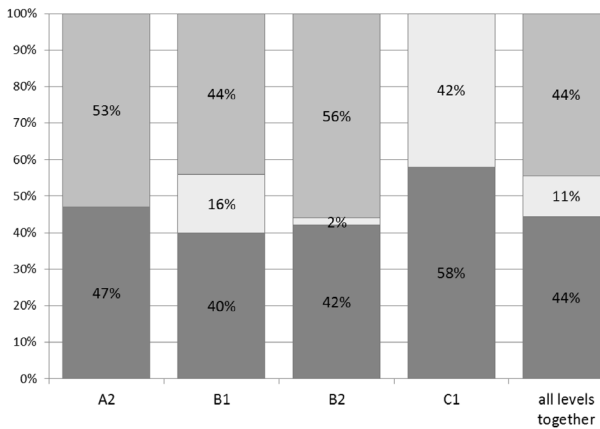


Figure 11. Assigning can do statements to appropriate CEFR levels. Experts

Having the above expectations might result in rating the students in the speaking test lower than they belong on a CEFR scale. Another explanation might be the one offered by Papageorgiou: ‘judges tend to have their own preconception of the CEFR levels based on what examinees they know can do with the language and such understanding [can] have an effect on [their judgements]’ (Papageorgiou 2010: 276).

4. Discussion and intended training

The above study was designed to investigate the interviewers’/ raters’ degree of familiarity with the level descriptors of the CEFR as well as their attitude to its usefulness in their day-to-day teaching and assessment practices. The results show that, although there is an expectation that the national examination interviewers/ assessors start implementing a marking scale that is closely connected to CEFR

level descriptors, the degree of CEFR-related literacy varies considerably among the respondents. There is a fairly large group who seems to have had very little or no contact with the document whatsoever. And even those interviewers/ raters who admit having had CEFR related training, refer to the necessity of updating the knowledge about it and skills of working with it.

Although the degree of familiarity with the CEFR is generally low, many respondents expressed confidence with respect to being able to place students on appropriate CEFR level, the confidence being high both in the expert and novice group. The practical task of the survey of identifying particular descriptors with CEFR levels demonstrated a low success rate even among expert raters. Allowing that the CEFR descriptors are not without problems themselves (cf. Fulcher et al. 2011: 8), such a low success rate among the respondents calls for training that would aim at a more standardised understanding of what a student on a particular level can do, and how students on adjacent levels differ from each other in terms of their skills.

The need for CEFR related training having been established, the precise nature of the training is harder to be determined judged solely by the results of the questionnaire. The vast majority of the respondents refrain from making suggestions or respond with a fairly ambiguous request for ‘any kind’ of training. These findings are somewhat at odds with Fulcher who found that teachers who also act as testers want concrete help with ‘a text that is not light on theory but explains concepts clearly [---], a practical “how-to” guidance [---], a balance between classroom and large-scale testing’ (2012: 122). Though the current respondents’ suggestions to incorporate assessment and distinguishing between levels into the training programme are helpful to some extent, the actual content of the training course is left to the discretion of the test developer with the hope that information would be obtained first and foremost about the national examination and also about the CEFR, if the latter has a bearing to the exam.

The interviewer/ rater training, bearing in mind the new examination format as well as the results of the given survey, has been planned relying on the recommendations given in *Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR). A Manual*, chapter 5 (henceforth the Manual, Noijons et al. 2011), which has been adapted to the current needs. The training is intended to be divided into 4 stages:

- **Phase 1: Familiarisation.** This is a pre-training phase and consists in the participants working independently with the CEFR descriptors and completing task-sheets that focus the participants’ attention on the differences between levels A2 to C1. Before coming to the training site, the participants should watch the sample interview of the speaking test and familiarise themselves with the marking scale developed for the national examination rating of speaking. The documents, tasks with keys and the sample interview will be available to the participant on-line, on the training web-page.
- **Phase 2: Illustration.** In this phase sample interviews will be watched in a group to identify salient aspects of interviewer behaviour on the one hand and sample performances of different CEFR levels, on the other. Here it will be important to view and discuss samples of both ‘flat’ and ‘uneven’ profiles (Manual: 42), benchmarked to the CEFR levels (Takala 2004).

- Phase 3: **Practice**. Here, the participants proceed to rate samples of different levels individually and in a group, to gain further confidence. In the process, lenient and strict raters will be found, followed by further discussion and adjustment of scores.
- Phase 4: **Individual assessment**. To complete the training, the raters will have to individually rate a set of speaking examination performances, complete rating grids and provide commentary if they feel the performance was extraordinary in some respect. Completed forms will then be handed in to the trainers. Individual feedback would be provided with regard to the success rate.

5. Conclusion

The article investigated CEFR-related assessment literacy among teachers who act as interviewers/ raters during the national examination in the English language in Estonia and studied their ability to successfully distinguish between CEFR levels. The findings show a fairly low level of familiarity among the teachers with the opportunities for the potential use of the CEFR for language learning purposes and confirm what research into the relevant literature had found earlier that ‘respondents tend to say that everything presented to them is important, resulting in little variation’ (Fulcher 2012: 118). The responses also indicated the need for a more substantial attention to be paid to CEFR-related pre- and in-service teacher training.

The study was conducted within one cultural context (Estonia) and involved a fairly small amount of respondents, still displaying significant differences on occasion between the two groups investigated. For more generalised conclusions, teachers assuming similar roles (interviewer/ tester working with the CEFR-related scales) could be involved in other cultural contexts. Further research should also focus on the success rate of the intended training to see if, as a result of a more focused CEFR-related interviewer/ rater training, more level-awareness will develop among the raters.

References

- Alas, Ene 2010. The English Language National Examination Validity Defined by Its Oral Proficiency Interview Interlocutor Behaviour. Tallinn University Dissertations on Humanities 22. Tallinn: Tallinn University Press.
- Brindley, Geoff 2001. Language assessment and professional development. – C. Elder, A. Brown, E. Grove, K. Hall, N. Iwashita, T. Lumley, T. McNamara, K. O’Loughlin (Eds.). *Experimenting with Uncertainty. Essays in Honour of Alan Davies*. Studies in Language Testing. Cambridge: Cambridge University Press, 126–136.
- Brown, James Dean; Bailey, Kathleen M. 2008. Language testing courses: What are they in 2007? – *Language Testing*, 25 (3), 349–385. <http://dx.doi.org/10.1177/0265532208090157>
- CEFR = Council of Europe 2001. *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Cambridge: Cambridge University Press.
- Davies, Alan 2008. Textbook trends in teaching language testing. – *Language Testing*, 25 (3), 327–349. <http://dx.doi.org/10.1177/0265532208090156>

- European Commission 2012. First European Survey on Language Competences: Final Report. http://ec.europa.eu/languages/policy/strategic-framework/documents/language-survey-final-report_en.pdf (11.2.2014).
- Figueras, Neus; North, Brian; Takala, Sauli; Verhelst, Norman; van Avermaet, Piet 2005. Relating examinations to the Common European Framework: A manual. – *Language Testing*, 22 (3), 261–279. <http://dx.doi.org/10.1191/0265532205lt3080a>
- Fulcher, Glenn 2012. Assessment literacy for the language classroom. – *Language Assessment Quarterly*, 9 (2), 113–132. <http://dx.doi.org/10.1080/15434303.2011.642041>
- Fulcher, Glenn; Davidson, Fred; Kemp, Jenny 2011. Effective rating scale development for speaking tests: Performance decision trees. – *Language Testing*, 28 (1), 5–29. <http://dx.doi.org/10.1177/0265532209359514>
- Gümnaasiumi riiklik õppekava. [National curriculum for upper secondary schools.] RT I, 28.8.2013, 8. <https://www.riigiteataja.ee/akt/128082013008> (14.2.2014).
- Inbar-Lourie, Ofra 2008. Constructing a language assessment knowledge base. A focus on language assessment courses. – *Language Testing*, 25 (3), 385–403. <http://dx.doi.org/10.1177/0265532208090158>
- Kecker, G.; Eckes, T. 2010. Putting the manual to the test: The test-Daf – CEFR Linking Project. – W. Martyniuk (Ed.). *Relating Language Examinations to the Common European Framework of Reference for Languages: Case studies and reflections on the use of the Council of Europe's Draft Manual*. Cambridge: Cambridge University Press, 50–79.
- Little, David 2005. The Common European Framework and the European Language Portfolio: Involving learners and their judgements in the assessment process. – *Language Testing*, 22 (3), 321–336. <http://dx.doi.org/10.1191/0265532205lt3110a>
- Malone, Margaret E. 2008. Training in language assessment. – E. Shohamy, N. Hornberger (Eds.). *Encyclopaedia of Language and Education*, Vol. 7: Language Testing and Assessment. 2nd ed. New York: Springer Science and Business Media, 225–233.
- Malone, Margaret E. 2013. The essentials of assessment literacy: Contrast between testers and users. – *Language Testing*, 30 (3), 329–344. <http://dx.doi.org/10.1177/0265532213480129>
- Manual for Language Test Development and Examining. For use with the CEFR. Produced by ALTE on behalf of the Language Policy Division, Council of Europe 2011. Strasbourg: Council of Europe
- Martyniuk, Waldemar (Ed.) 2010. *Aligning Tests with the CEFR*. Studies in Language Testing 33. Cambridge: Cambridge University Press.
- Noijons, José; Bérešová, Jana; Breton, Gilles; Szabó, Gábor (Eds.) 2011. *Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR). Highlights from the Manual*. Strasbourg: Council of Europe Publishing.
- O'Sullivan, B. 2010. The City and Guilds Communicator Examination Linking Project: A brief overview with reflections on the process. – W. Martyniuk (Ed.). *Relating Language Examinations to the Common European Framework of Reference for Languages: Case studies and reflections on the use of the Council of Europe's Draft Manual*. Cambridge: Cambridge University Press, 33–49.
- Papageorgiou, Spiros 2010. Investigating the decision-making process of standard setting participants. – *Language Testing*, 27 (2), 261–282. <http://dx.doi.org/10.1177/0265532209349472>
- Popham, William James 2009. Assessment of literacy of teachers: Faddish or fundamental? – *Theory into Practice*, 48 (1), 4–11. <http://dx.doi.org/10.1080/00405840802577536>
- Reemann, Edith; Alas, Ene; Liiv, Suliko 2013. Interviewer behaviour during oral proficiency interviews: A gender perspective. – *Estonian Papers in Applied Linguistics*, 9, 209–227. <http://dx.doi.org/10.5128/ERYa9.14>

- Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR). A Manual. Strasbourg: Council of Europe, 2009.
- Stiggins, Rick 1991. Assessment literacy. – The Phi Delta Kappan, 72, 534–539.
- Stiggins, Rick 1997. Student-Centered Classroom Assessment. Upper Saddle River: Prentice Hall.
- Takala, Sauli (Ed.) 2004. Reference Supplement to the Preliminary Version of the Manual for Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, Teaching and Assessment. Strasbourg: Council of Europe.
- Türk, Ülle; Tender, Tõnu 2013. Re-designing the school-leaving foreign language examinations in Estonia. – Jozef Colpaert, Mathea Simons, Ann Aerts, Margret Oberhofer (Eds.). Proceedings, “Language Testing in Europe: Time for a New Framework?”. Antwerp: University of Antwerp.

Ene Alase (Tallinna Ülikool) teadushuvid on keeletestimine, testide koostamine ja nende kvaliteedi hindamine, õpetajakoolitus ja õppekirjanduse hindamine.
Narva mnt 25, 10120 Tallinn, Estonia
enealas@tlu.ee

Suliko Liivi (Tallinna Ülikool) uurimisvaldkonnad on kontrastiivuuringud, kultuuridevaheline suhtluspädevus, keelepoliitika, võõrkeelte õpetamise meetodika, testimine.
Narva mnt 25, 10120 Tallinn, Estonia
liiv@tlu.ee

RIIGIEKSAMI INTERVJUEERIJATE JA HINDAJATE HINDAMISALANE PÄDEVUS – KEELEÕPPE RAAMDOKUMENT JA SELLE TUNDMINE

Ene Alas, Suliko Liiv

Tallinna Ülikool

Artiklis vaadeldakse inglise keele riigieksami intervjuerijate ja hindajate väljaõppe vajadusi tulenevalt uue inglise keele riigieksami kehtima hakkamisest. Uus eksam on koostatud nii, et see mõõdaks Euroopa keeleõppe raamdokumendi iseseisva keeleoskuse B alatasemeid B1 ja B2 ühe eksami kaudu. Kuna inglise keele õpetamine ja selle hindamine on riiklikus õppekavas otseselt seotud Euroopa keeleõppe raamdokumendi tasemetega, püüdis käesolev uurimus välja selgitada, kui pädevad on inglise keele riigieksami eksamineerijad ja hindajad raamdokumendi kasutamisel: mil määral raamdokumenti igapäevases koolitöös kasutatakse, kui lihtne on dokumenti kasutada ning kuivõrd täpselt vastajad enda arvates õpilaste teadmisi raamdokumendi skaaladele toetudes hinnata oskavad. Lisaks paluti vastajatel raamdokumendi tasemekirjeldusi õigete tasemetega ühendada. Uurimusest selgus, et raamdokumenti kasutatakse peamiselt õpilaste ja õppekirjanduse hindamisega seotud tegevustes ning tasemekirjeldustes kiputakse kirjet õigest tasemest kõrgemale asetama. Tulemustest lähtuvalt on autorid välja pakkunud intervjuerijate ja hindajate koolituskava, mis aitaks Euroopa keeleõppe raamdokumendi tasemeid paremini tundma õppida ja Eesti inglise keele riigieksami suulist osa usaldusväärsemalt hinnata.

Võtmesõnad: hindamisalane pädevus, intervjuerija, hindaja, suuline keeleeksam, Euroopa keeleõppe raamdokument, väljaõpe

DIALOOGIDE PRAGMAATILISE ANALÜÜSI TARKVARA

Sven Aller, Olga Gerassimenko, Tiit Hennoste, Riina Kasterpalu, Mare Koit, Krista Mihkels, Kirsi Laanesoo, Andriela Rääbis

Ülevaade. Meie uurimisobjektiks on eestikeelsed dialoogid (inimestevaheliste suuliste dialoogide transkriptsioonid ja inim-masindialoogide logifailid). Eesmärgiks on määrata, missuguseid dialoogiakte ja missuguseid suhtlusstrateegiaid kasutavad dialoogis osalejad ning missugustest struktuuriosadest dialoogid koosnevad, s.o läbi viia pragmaatilist analüüsi. Selleks, et lihtsustada dialoogide pragmaatilist analüüsi, arendame tarkvara, mis võimaldab tuvastada ja märgendada dialoogi tekstis dialoogiakte, suhtlusstrateegiaid ja dialoogi struktuuriosi (sh alamdialooge). Dialoogiaktide tuvastamiseks rakendame andmepõhist meetodit, samas suhtlusstrateegiate ja dialoogi struktuuri määramine põhineb reeglitel. Tarkvara kasutavad lingvistid dialoogide uurimisel, mille kaugem eesmärk on dialoogsüsteemi arendamine, mis suhtleks kasutajaga loomulikus keeles, järgides inimestevahelise suhtluse norme.*

Võtmesõnad: dialoog, dialoogiakt, suhtlusstrateegia, struktuur, eesti keel

1. Sissejuhatus

Traditsioonilise käsitluse kohaselt eelnevad seotud teksti pragmaatilisele analüüsile teksti koosseisus olevate lausete morfoloogiline, süntaktiline ja semantiline analüüs, kus iga järgmine analüüsietapp kasutab eelmiste etappide tulemusi. Meie püüame läbi viia pragmaatilist analüüsi nii, et see ei eeldaks eelmisi analüüsietappe (morfoloogiline jne), sest nende läbiviimiseks ei tarvitse veel leiduda (piisavalt töökindlat) tarkvara. Seega on pragmaatilise analüüsi sisendiks puhas tekst, st eeltöötlemata tekstifail. Esmalt määrame selles dialoogiaktid (DA), kasutades statistilist meetodit. Seejärel toimub dialoogistrateegiate ja dialoogi struktuuri tuvastamine, kasutades selleks reegleid, mis põhinevad DA-del.

* Artikli valmimist on toetanud Euroopa Regionaalarengufond Eesti Arvutiteaduse Tippkeskuse kaudu, Eesti Teadusagentuur (projektid SF0180078s08 "Loomulike keelte arvutitöötuse formalismide ja efektiivsete algoritmide väljatöötamine ning eesti keelele rakendamine", ETF9124 "Suhtlusagendi modelleerimine ja Eesti dialoogikorpus", ETF8558 "Eestikeelse spontaanse dialoogi struktuuri loomise keelised vahendid") ning Haridus- ja Teadusministeerium (projekt EKT5 "Eestikeelse dialoogi pragmaatika analüsaator").

Meie eesmärgiks on luua tarkvara, mida saaksid kasutada keeleteadlased dialoogide märgendamisel, nende uurimisel ja struktuuri võrdlemisel.

Maailmas on välja töötatud mitmeid erinevaid DA-de tüpoloogiaid (nt Sinclair, Coulthard 1975, Stenström 1994, Allwood jt 2001). Praegu on kujunenud standardiks DAMSL (Allen, Core 1997), mis võimaldab dialoogides märgendada lausungeid erinevatel tasemetel ja ühtlasi määrata kõneaktide vahelisi seoseid. Tartu Ülikoolis on välja töötatud vestlusanalüüsil (Hutchby, Wooffitt 1998) põhinev tüpologia, mida kasutame oma dialoogikorpuses DA-de märgendamisel.

DA-de automaatseks tuvastamiseks dialoogides on kasutatud mitmesuguseid erinevaid andmepõhiseid meetodeid: n-gramme, Markovi peitmudeleid, Bayesi klassifikaatoreid, tehisnärvivõrke, otsustuspuid, transformatsioonipõhist masinõpet, mälu põhised õpet jne (Reithinger, Maier 1995, Wright jt 1999, Keizer jt 2002, Grau jt 2004, Levin jt 2003, Samuel jt 1998, Fernandez jt 2005, vt ka Fishel 2007a). Meie kasutame Naiivse Bayesi klassifikaatorit.

Suhtlusstrateegiaid on märgendatud ja analüüsitud infodialoogides (Jokinen 1996, Eskor 2006, 2005, 2004) ja argumenteerimisdialoogides (Eskor 2007, Georgila jt 2011). Viimasel juhul on kasutatud hüvitusega masinõpet. Meie lähtume oma tarkvara arendamisel Jokineni konstruktiivses dialoogimudelil (*Constructive Dialogue Model*, CDM) toodud suhtlusstrateegia mõistest (Jokinen 2009, 1996). Lausungitele suhtlusstrateegiate omistamisel kasutame reegleid, mis põhinevad DA-del. Dialoogi struktuuri tuvastamine toimub samuti reeglipõhiselt.

Artikli ülesehitus on järgmine. Osas 2 tutvustame oma empiirilist materjali: Eesti dialoogikorpust ja dialoogiaktide tüpoloogiat. Osad 3 kuni 5 kirjeldavad meie tarkvara võimalusi: poolautomaatne DA-de tuvastamine, suhtlusstrateegiate ja dialoogi struktuuri automaatne tuvastamine. Osas 6 teeme kokkuvõtteid.

2. Korpus ja dialoogiaktide tüpologia

Eesti dialoogikorpus sisaldab kolme liiki dialooge. Korpuse esimese osa moodustavad inimestevahelised suulised vestlused, mis on salvestatud autentsetes tingimustes ja transkribeeritud, kasutades vestlusanalüüsi (VA) transkriptsiooni. Selliste dialoogide koguarv on üle 1000. Dialoogide salvestamise põhieesmärk on olnud uurida inimestevahelist suhtlust. Seetõttu on korpuses erinevat tüüpi dialooge: telefonikõnesid (helistamine infotelefonile, reisibüroosse, bussijaama, polikliinikusse) ja silmast-silma vestlusi (poes, teeninduses, reisibüroos, teejuhatamised tänaval jne). Selline mitmekesisus teeb dialoogide automaatse analüüsi muidugi raskemaks, aga on vajalik inimsuhtluse igakülgse uurimisel. Korpus on avatud ja kasvav, järjest lisanduvad uued dialoogid. Enamuse korpusest moodustavad siiski institutsionaalsed infotelefonikõned (Hennoste jt 2009).

Korpuse teine osa on kogutud võlur Ozi meetodil, kus arvuti rolli mängib kasutaja eest varjatult teine inimene (Kullasaar 2001, Pärkson 2011). Eksperimentideks kasutasime veebipõhist tarkvara (Treumuth 2011, Käbin 2011). Kasutaja sisestab oma teksti (infosoovi) klaviatuurilt ja loeb võluri vastuseid ekraanilt. Selliste dialoogide arv on umbes 100.

Kolmanda osa korpusest moodustavad dialoogid kahe veebis kasutatava dialoogsüsteemiga. Neist üks annab infot kinodes linastuvate filmide kohta ja teine

hambaraviinfot. Nagu võlur Ozi eksperimentides, sisestab ka siin kasutaja oma eestikeelse teksti klaviatuurilt ja näeb “arvuti” vastuseid ekraanilt (Treumuth 2011). Selliste dialoogide arv on praegu samuti umbes 100.

Tarkvara arendamiseks on kasutatud ühte osa korpuses leiduvatest suulistest infodialoogidest. Samas on tarkvara kavandatud kogu järjest täieneva korpuse automaatseks analüüsiks.

Dialoogiaktide märgendamiseks erinevates korpustes on varem välja töötatud mitmeid erinevaid tüpoloogiaid, millest tuntuim on DAMSL (Allen, Core 1997). Selle tüpoloogia põhieesmärk on esitada iga lausungi kõikvõimalikud funktsioonid, mida lausung saab kanda, ja seosed erinevate kõneaktide vahel.

Kuna meie tegeleme põhiliselt inimestevahelise suhtluse uurimisega, siis oleme aastatel 2001–2004 välja töötanud omaenese DA-de tüpoloogia, mis põhineb VA printsiipidel ja mida me kasutame DA-de märgendamiseks oma korpuses. Meie tüpoloogias jagatakse kõik DA-d kahte rühma: naabruspaariaktid, kus esiliige ootab teatud kindlat järelliiget (nagu küsimus – vastus), ja üksikaktid, mis ei oota reaktsiooni (nt vastuvõtuteade). Teisalt on kõik DA-d jagatud infoaktideks (nt erinevad küsimuste tüübid) ja suhtluse juhtimise aktideks (nt partneri algatatud parandused). Iga DA nimi koosneb kahest osast, mis on teineteisest eraldatud kooloniga (nt KY: suletud kas, VR: vastuvõtuteade): esimene osa näitab aktiklassi nime (nt KY – küsimused, VR – vabatahtlikud reaktsioonid) ja teine osa on akti pärisnimi (nt suletud kas, vastuvõtuteade). Aktide koguarv on meie tüpoloogias 126. Võrreldes mõne muu tüpoloogiaga (nt Allwood jt 2001) on meie omas rohkem tagasisideakte, sh nii üksikaktid (nagu vabatahtlikud reaktsioonid) kui ka naabruspaariaktid (nt partneri algatatud parandused). DA-de täieliku nimekirja võib leida nt teoses Hennoste, Rääbis 2004.

Näide (1) kujutab endast suulise dialoogi transkriptsiooni, kus on märgendatud DA-d meie tüpoloogia kohaselt (H – helistaja ehk klient, V – vastaja ehk ametnik). DA-de nimed on püstkriipsude vahel. Mõnel lausungil on mitu märgendit, st need lausungid kannavad korraga mitut funktsiooni. Näites on kasutatud VA transkriptsiooni.

(1)
((numbri valimine, kutsung)) | RIE: KUTSUNG |
V: info | RIJ: KUTSUNGI VASTUVÕTMINE | | RY: TUTVUSTUS |
tere | RIE: TERVITUS |
H: .hh tere | RIJ: VASTUTERVITUS |
ma paluks (.) rahandusülikon- õõ rahandusosakonnast=õ (.) `Siili telefoni-
numbrit. | DIE: SOOV |
(...)
V: Anne `Siil on see.= | PPE: ÜMBERSÕNASTAMINE | | KYE: VASTUST
PAKKUV |
H: =jah.= | PPJ: LÄBIVIIMINE | | KYJ: JAH |
V: = `kolm seitse viis üks kolm `kolm. | DIJ: INFO ANDMINE |
(0.5)
H: kolm seitse `viis üks kolm `kolm. | VR: NEUTRAALNE VASTUVÕTU-
TEADE |
aitäh? | RIE: TÄNAN |
(.) nägemist | RIE: HÜVASTIJÄTT |

Siiani oleme järginud sellist lähenemisviisi, kus kaks lingvistit märgendavad teineteisest sõltumatult DA-d käsitsi, kasutades tarkvara, mis lihtsustab dialoogide valikut korpusest ja DA-de valikut nimestikust (Vutt 2001), ja seejärel ühtlustab kolmas isik märgendused. Automaatne märgendamine teeks selle töö tunduvalt hõlpsamaks.

3. Dialoogiaktide tuvastamine

Nagu eespool mainitud, on DA-de tuvastamiseks dialoogitekstides kasutatud mitmesuguseid andmepõhiseid meetodeid. Ka meie otsustasime valida mõne andmepõhise meetodi, eelistades seda reeglite koostamisele, sest tuvastatavaid DA-sid on meie tüpoloogias väga palju.

Eestikeelsetes dialoogides DA-de klassifitseerimiseks meie tüpoloogia alusel on testitud mitmeid erinevaid meetodeid: (mitmekihilised) tajurid, otsustuspuud, sufiksipuud, Naiivse Bayesi klassifikaator (Fishel 2007a, Kikas 2007, Fišel, Kikas 2006). Kahjuks ei osutunud ükski meetod piisavalt heaks, et seda rakendada DA-de täiesti automaatseks tuvastamiseks. See on tingitud vähemalt kahest asjaolust: DA-de tüpoloogia keerukus ja meie (suhteliselt väikese) korpuse mitmekesisus, mis teeb raskeks meetodite treenimise. Seetõttu otsustasime oma tarkvaras realiseerida DA-de poolautomaatse märgendamise. Realiseerimiseks valisime kõige tõrkekindlama ja lihtsama meetodi: Naiivse Bayesi.

Programm tükeldatakse dialoogi teksti lausungiteks ja määrab igale lausungile kuni viis tõenäolisemat DA märgendit. Seejärel võib inimene-märgendaja parandada vigu ja vajadusel korrata automaatset märgendamist. Sisendiks on .txt fail: dialoogi tekst, milles voorud (aga mitte lausungid) asuvad igaüks erinevas reas. Väljund on .txt fail, kus voorud on tükeldatakse lausungiteks, mis on paigutatud erinevatesse ridadesse, ja lausungitele on määratud DA-d Bayesi klassifikaatoriga. Katsetuste teel on klassifikaatoris valitud kasutamiseks järgmised tunnused, mis andsid parimaid tulemusi: sõnade trigrammide tõenäosus, lausungi pikkus ja DA märgendite tõenäosuste geomeetriline keskmine (vt Fishel 2007b).

Märgendaja töötab kahes faasis: treenimine ja märgendamine. Treenimisele võib eelneda ristvalideerimine. Treeningu alguses initsieeritakse uus sessioon, luuakse mudel ja kasutatakse seda uute andmete märgendamisel. Klassifikaator on realiseeritud programmeerimiskeeles Perl. Treenimiseks valiti eesti dialoogikorpusest 800 infodialoogi ja kasutati 10-kordset ristvalideerimist. Klassifikaatori saagis on 64,7% ja täpsus 33,0%. Need tulemused tunduvad olevat kehvad, kuid tuleb arvesse võtta, et arvutused tehti iga lausungi jaoks kõige tõenäolisemat märgendit kasutades, samas kui programm pakub kuni viis märgendit tõenäosuste kahane-mise järjekorras. Inimene-märgendaja ei pea enam otsima sobivat märgendit kogu DA-de nimestikust, vaid enamasti leiab selle nende viie hulgast, mida klassifikaator pakkus.

Näita nõuandeid

Lausungi tekst	Tulemuse eelvaade	Tõenäosuslikud DAd	Kõik DAd
((427a1 ülikooli infotelefon))			
((numbri valimine, kutsung))	RIE: KUTSUNG	<ul style="list-style-type: none"> RIE: KUTSUNG YA: PRAAK YA: MUU RD: KUTSUNGI VASTUVÖTMINE RIE: TERVITUS 	
V: info tere	RD: KUTSUNGI VASTUVÖTMINE	<ul style="list-style-type: none"> RD: KUTSUNGI VASTUVÖTMINE RY: TUTVUSTUS RIE: TERVITUS YA: PRAAK YA: MUU 	
H: .hh tere ma paluks (.) rahanduslikon- õs rahandusosakonnast=õ (.) "Siili telef oninumbrit.	DIE: SOOV	<ul style="list-style-type: none"> DIE: SOOV RIE: TERVITUS RY: TUTVUSTUS KKE: ALGATUS 	
(...)			
V: Anne "Siil on see. =	KYE: VASTUST PAKKUV	<ul style="list-style-type: none"> KYE: VASTUST PAKKUV PPE: ÜMBERSONASTAMINE DIJ: INFO ANDMINE PPE: ÜLEKUSIMINE VTE: VASTUSE TINGIMUSTE TÄPSUSTAMI 	
H: =jah. =	KYJ: JAH	<ul style="list-style-type: none"> KYJ: JAH PPJ: LÄBIVIIMINE VTJ: VASTUSE TINGIMUSTE TÄPSUSTAMI KYE: VASTUST PAKKUV 	

Joonis 1. Lõik märgendatavast dialoogist märgendamislehel (vt ka näide 1)

Veebiliides käivitab kasutajaga suheldes serveris Perli skripte. Liides koosneb viiest lehest: 1) kasutajate haldus, 2) kasutajafailide haldus, 3) sisendfaili toimetamine, 4) märgendamine ja 5) väljundfailide kuvamine. Liides on realiseeritud keeles PHP, kasutajamugavuse ja efektiivsuse tõstmiseks on lisatud mõned skriptid keeles JavaScript (joonis 1, vt ka Aller 2012).

4. Suhtlusstrateegiate tuvastamine

Teadaolevalt on suhtlusstrateegiaid märgendatud ja uuritud infodialoogides ja kokkuleppimisdioloogides. Viimastes on tuvastatud argumenteerimisstrateegiaid ja kasutatud selleks hüvitusega masinõpet (Georgila jt 2011). Meie lähtume infodialoogide märgendamisel suhtlusstrateegia mõistest, mille on sisse toonud Kristiina Jokinen oma konstruktiivses dialoogimudelil (Jokinen 1996). Suhtluses osaleja kasutab suhtlusstrateegiat oma järgmise lausungi ülesehitamisel reaktsioonina partneri lausungile. Mudel arvestab suhtlusstrateegia valikul nelja binaarsete väärtustega kontekstifaktorit:

- 1) ootused – kas partneri lausung on ootuspärane (1) või mitte (0),
- 2) teema – kas partneri lausung jätkab sama teemat (1) või mitte (0),
- 3) initsiatiiv – kas kõnelejal on initsiatiiv (1) või mitte (0),
- 4) eesmärgid – kas kõneleja eesmärgid on täidetud (1) või mitte (0).

Seega on võimalikke suhtlusstrateegiaid 16 (tabel 1, vt ka Eskor 2005).

Tabel 1. Suhtlusstrateegiad konstruktiivses dialoogimudelil vastavalt neljale kontekstifaktorile (strateegiate nimetused on ingliskeelsed ja paksus kirjas, antud on ka eestikeelsed tõlked)

Ootused	Teema	Eesmärk	Initsiatiiv kõnelejal	Initsiatiiv kuulajal
oodatud	seotud	täitmata	backto tagasi eelmise juurde	follow-up-old jätku eelmisega
		täidetud	finish/start eelmine lõpetatud, alusta uuega	follow-up-new jätku uuega
	mitteseotud	täitmata	repeat-new korda uut	new question uus küsimus
		täidetud	specify täpsusta eelmist	new request uus soov
mitteoodatud	seotud	täitmata	subquestion,X lisaküsimus	continue jätku
		täidetud	new dialogue uus dialoog	somethingelse muu
	mitteseotud	täitmata	object,X vaidle vastu	notrelated teemaga mitteseotud
		täidetud	specify-new paku uus	new st-request uus väide

Igale strateegiale võib vastavusse seada kontekstifaktorite väärtustest moodustatud vektori (tabel 2).

Tabel 2. Suhtlusstrateegiatele vastavad vektorid (kontekstifaktorite väärtused: oodatud – seotud – kõneleja initsiatiiv – eesmärk)

Suhtlusstrateegia	Vektor
<i>notrelated</i>	0000
<i>new-st-request</i>	0001
<i>object,X</i>	0010
<i>specify-new</i>	0011
<i>continue</i>	0100
<i>somethingelse</i>	0101
<i>subquestion,X</i>	0110
<i>new-dialogue</i>	0111
<i>new question</i>	1000
<i>new-request</i>	1001
<i>repeat-new</i>	1010
<i>specify</i>	1011
<i>follow-up-old</i>	1100
<i>follow-up-new</i>	1101
<i>backto</i>	1110
<i>finish/start</i>	1111

Märgendasime käsitsi suhtlusstrateegiad 50 infodialoogis, mis olid juhuslikult valitud eesti dialoogikorpusest ja märgendatud DA-dega. Nende dialoogide uurimise käigus sõnastasime reeglid, mis kasutavad suhtlusstrateegia määramiseks DA-sid ja osalejatunnuseid. Selle tulemuseks oli järgmine algoritm (silumise käigus on seda hiljem täpsustatud).¹

Ei märgenda

1. Rituaalseid akte üldiselt ei märgenda, v.a
RIE: lõpusignaali – *specify-new*.
RIJ: lõpetamise vastuvõtmine – *follow-up-new*.
RIJ: lõpetamise tagasilükkamine – *somethingelse*.
2. Praaki (PRAAK) ei märgenda.

Naabruspaariaktid

Direktiivid, küsimused, arvamused

Esiliikmed

3. Esimene H: DIE/KYE on *finish/start*.
4. Hilisem H: DIE/KYE (ainsa märgendiga)
 - a. Kui ei eelne V: DIJ/KYJ, siis *backto*.
 - b. Kui eelneb V: info puudumine või teemavahetus, siis (algab uus teema) *specify-new*.
 - c. Kui eelneb V: info andmine, siis oli saadud info puudulik, jätkub sama teema: *new-dialogue*.
5. V: KYE: alternatiiv/jutustav kas või DIE: pakkumine on *new-dialogue*.
6. H: DIE+TVE on *specify-new*.
7. H: DIE+PA on *specify-new* (vahetab eneseparandusega teemat).
8. SEE: arvamus on *new-dialogue*.

Järelliikmed

9. DIJ/KYJ: info andmine/nõustumine on *follow-up-old*.
10. DIJ/KYJ: info puudumine/mittenõustumine on *continue*.
11. DIJ: nõustuv ei on *continue* (nagu info puudumine).
12. SEJ: muu on *continue* (nagu mittenõustumine).

Topeltmärgenditega (järel- ja esiliige):

13. KYJ+KYE on *new-question*.
14. DIJ+DIE on *new-st-request*.

Kontakti kontroll

15. KKE on *specify-new*.
16. KKJ on *follow-up-new*.

Alamdialoogid

17. VTE ja PPE on *subquestion,X*.
18. VTJ ja PPJ on *follow-up-old*.

Üksikaktid

Infolisad

19. H: IL: täpsustamine on (tema DIE/KYE täpsustamine) *backto*.

¹ Kasutame siin suhtlusstrateegiate ingliskeelseid nimetusi, nagu need esinevad konstruktiivses dialoogimudelis.

20. V: IL: täpsustamine/põhjendamine/ülerõhutamine sama osaleja info andmise järel on *follow-up-new*.

Vabatahtlikud reaktsioonid

21. H: VR: neutraalne/hinnanguline jätkaja on *continue*.

22. Kõik ülejäänud VR on *somethingelse*, v.a

a. VR: neutraalne/hinnanguline piiritleja.

i. Kui järgneb info andmine, siis ta on *object,X*.

ii. Kui ei järgne info andmist (dialoogi lõpus), siis *specify*.

b. VR: neutraalne/hinnanguline info osutamine uueks on *repeat-new*.

Üksikaktid

23. YA: info andmine on *follow-up-old*.

24. YA: muu on *somethingelse*.

Automaatse märgendaja sisendiks on dialoogi tekst (.txt), milles on märgendatud DA-d, ja väljundiks tekstifail, kus on märgendatud suhtlusstrateegiad ning lisatud neile vastavad vektorid, mis esitavad nelja kontekstifaktori väärtusi (0 või 1). Mõned lausungid jäävad märgendita, sest strateegiad on seotud eeskätt info soovimise ja info andmisega (näide 2 – programmi väljund; vrd ka näide 1). Programmeerimiskeeleks on PHP.

(2)

((numbri valimine, kutsung)) | RIE: KUTSUNG

V: info | RIJ: KUTSUNGI VASTUVÕTMINE | RY: TUTVUSTUS

tere | RIE: TERVITUS

H:.hh tere | RIJ: VASTUTERVITUS

ma paluks (.) rahandusülikon- õõ rahandusosakonnast=õ (.) `Siili telefoninumbrit. | DIE: SOOV [**finish/start**] [**1111**]

V: Anne `Siil on see.= | PPE: ÜMBERSÕNASTAMINE | KYE: VASTUST PAKKUV [**subquestion,X**] [**0110**]

H:=jah.= | PPJ: LÄBIVIIMINE | KYJ: JAH [**follow-up-old**] [**1100**]

V: = `kolm seitse viis üks kolm `kolm. | DIJ: INFO ANDMINE [**follow-up-old**] [**1100**]

H:kolm seitse `viis üks kolm `kolm. | VR: NEUTRAALNE VASTUVÕTUTEADE [**somethingelse**] [**0101**]

aitäh? | RIE: TÄNAN

(.) nägemist | RIE: HÜVASTIJÄTT

5. Infodialoogi struktuuri tuvastamine

Tüüpiline kõne infotelefonile koosneb kolmest osast: 1) rituaalne algus, 2) põhiosa, kus esitatakse infosoov ja saadakse vastus, ja 3) rituaalne lõpp.

```
[RITUAALNE ALGUS]
H: ((kutsung)) RIE: Kutsung
V: RIJ: Kutsungi vastuvõtmine RY: Tutvustus [V esitleb teenusepakkujat]
  (RY: Tutvustus [V esitleb ennast])
  (RIE: Tervitus )
H: RIJ: Vastutervitus
[PÕHIOSA]
H: DIE: Soov / KYE: Avatud/ Jutustav kas
  ( [INFOJAGAMISDIALOG]
--> V: VTE: Vastuse tingimuste täpsustamine
<-- H: VTJ: Vastuse tingimuste täpsustamine
  )
  ([PARTNERI ALGATATUD PARANDUS]
--> V/H: PPE: Ümbersõnastamine/ Üleküsimine/Mitterõistmine
H/V: PPJ: Läbiviimine
<-- (V/H: VR: Paranduse hindamine )
  )
V: ( VR: Neutraalne vastuvõtuteade DIJ/KYJ: Keeldumine ) DIJ/KYJ Info andmine
  ([PARTNERI ALGATATUD PARANDUS]
--> H/V: PPE: Ümbersõnastamine/ Üleküsimine/Mitterõistmine
V/H: PPJ: Läbiviimine
<-- (H/V: VR: Paranduse hindamine )
  )
  (H: VR: Neutraalne vastuvõtuteade / Neutraalne piiritleja/ Neutraalne info osutamine uueks
  )
[RITUAALNE LÕPP]
H: RIE: Tänan ( RIE: Hüvastijätt )
(V: RIJ: Palun RIJ: Vastuhüvastijätt )
```

Joonis 2. Infodialoogi osad: rituaalne algus, põhiosa, rituaalne lõpp. Alamdialoogi algust tähistab --> ja lõppu <--. Ümarsulgudes on DA-d(e) järjendid, mis võivad puududa (vt ka Koit 2012). Suhtlejad on H (helistaja) ja V (vastaja)

Põhiosa tuumaks on naabruspaar direktiiv – direktiivi täitmine või küsimus – vastus. Põhiosa koosseisus võib esineda alamdialooge: helistaja infosoovile võib järgneda vastaja täpsustav küsimus ja helistaja vastus sellele või algatab üks osalejatest paranduse, mille viib läbi partner.

Korpuse analüüsi tulemusel leidsime, et dialoogi struktuursete osade ja alamdialoogide tuvastamiseks saab kasutada DA-de naabruspaare kui põhilisi märguandeid.

Rituaalse alguse ja lõpu saab tuvastada rituaalsete naabruspaariaktide ja üksikakti RY: Tutvustus alusel dialoogi alguses või vastavalt lõpus. Põhiosa algab kohe pärast rituaalset algust soovi või küsimusega ja kestab kuni rituaalse lõpu alguseni. Põhiosas sisalduvaid alamdialooge saab tuvastada topeltmärgendite abil: täpsustav küsimus kannab alati lisaks küsimuse märgendile ka märgendit VTE: Vastuse tingimuste täpsustamine ja partneri algatatud parandus kas märgendit PPE: Mitterõistmine, PPE: Üleküsimine või PPE: Ümbersõnastamine. Naabruspaaride

järelliikmed algavad siis vastavalt märgenditega VTJ: Vastuse tingimuste täpsustamine ja PPJ: Parandus (vt joonis 2).

Näites (3) (programmi väljund) on märgendatud dialoogi osad (RI – rituaalne algus või lõpp, PRIM – põhiosa, RP – parandus). Põhiosas sisaldub alamdialoog: kliendi H algatatud parandus, mille viib läbi ametnik V (vrd ka näide 1).

(3) /Rituaalne algus (RI)/
RI ((numbri valimine, kutsung) | RIE: KUTSUNG |
RI V: info | RIJ: KUTSUNGI VASTUVÕTMINE | |RY: TUTVUSTUS|
RI tere | RIE: TERVITUS |
RI H: .hh tere | RIJ: VASTUTERVITUS |
/Põhiosa (PRIM)/
PRIM ma paluks (.) rahandusülikon- õõ rahandusosakonnast=õ (.) `Siili
telefoninumbrit. | DIE: SOOV |
/Paranduse alamdialoog (RP) (algatab V)/
RP V: Anne `Siil on see.= | PPE: ÜMBERSÕNASTAMINE | |KYE:
VASTUST PAKKUV|
RP H: =jah.= | PPJ: LÄBIVIIMINE | |KYJ: JAH|
/Põhiosa (PRIM)/
PRIM V: = `kolm seitse viis üks kolm `kolm. | DIJ: INFO ANDMINE |
PRIM H: kolm seitse `viis üks kolm `kolm. | VR: NEUTRAALNE VASTU-
VÕTUTEADE |
/Rituaalne lõpp (RI)/
RI aitäh? | RIE: TÄNAN |
RI (.) nägemist | RIE: HÜVASTIJÄTT |

Dialoogi struktuuri märgendaja saab sisendiks dialoogi teksti, milles on märgendatud DA-d (.txt fail), ja kasutab dialoogi osade äratundmiseks reegleid. Väljund antakse kahes formaadis: .txt ja .xml. Programmeerimiskeeleks on PHP.

6. Kokkuvõte ja edasine töö

Artiklis esitleti pragmaatilise analüüsi tarkvara, mis on kavandatud eestikeelsete dialoogide uurimise abivahendiks. Tarkvara võimaldab tuvastada ja märgendada dialoogiakte, suhtlusstrateegiaid ja dialoogide struktuuri. Esialgu kavandasime täiesti automaatset DA-de märgendamist meie tüpoloogia kohaselt ja testisime mitut erinevat andmepõhist meetodit. Kahjuks ei osutunud ükski testitud meetod piisavalt töökindlaks, sest tüpoloogia on keeruline (sisaldab 126 DA-d, mille tuvastamisel pole ka lingvistid alati üksmeelel) ja korpus ei ole homogeenne, sisaldades erinevat liiki dialooge, mida soovitakse kasutada inimestevahelise ning inimese ja arvuti vahelise suhtluse analüüsimisel ja võrdlemisel. Seetõttu on meie tarkvaras realiseeritud DA-de poolautomaatne märgendamine: programm jagab dialoogi vord lausungiteks ning omistab igale lausungile kuni viis DA märgendit, kasutades selleks Naiivse Bayesi klassifikaatorit. Seejärel võimaldab kasutajaliides lingvistil vajadusel parandada vigu. Kui DA-d on märgendatud, siis võib lasta dialoogis

määrata suhtlusstrateegiad ja ühtlasi märgendada dialoogi struktuuri, leides alguse, põhiosa ja lõpu ning põhiosas sisalduda võivad alamdialoogid.

Pragmaatilise analüüsi tarkvara on kasutatav internetis. Kasutaja saab valida korpusest märgendamata dialoogi ja lasta selles märgendada DA-d (kus tuleb seejärel võib-olla parandada märgendusvigu) või valida dialoogi, kus on juba märgendatud DA-d ja lasta märgendada dialoogi struktuuri ja/või suhtlusstrateegiad.

Olgugi et nii suhtlusstrateegiate kui ka dialoogi struktuuri tuvastamine toimub praegu eeldusel, et sisendiks on dialoog, milles on märgendatud DA-d meie tüpoloogias kohaselt, usume, et meie tarkvara aluseks olevaid ideid võib rakendada ka mõne muu tüpologia puhul (näiteks eelnevalt seades teise tüpologia igale DA-le vastavusse DA meie tüpoloogias). Poolautomaatset DA-de tuvastajat võib treenida ka mõnes muus, eesti keelest erinevas keeles üleskirjutatud dialoogidel ja sel viisil märgendada DA-sid meie tüpologia järgi.

Meie tulevane töö seisneb eestikeelsete vestluste analüüsimises kirjeldatud tarkvara abil. Ühtlasi täiendame tarkvara, võttes arvesse kasutajatelt saadud tagasisidet. Kaugem eesmärk on arendada dialoogsüsteeme, mis suhtlevad kasutajaga eesti keeles, järgides inimestevahelise suhtluse norme ja reegleid.

Viidatud kirjandus

- Allen, James; Core, Mark 1997. Draft of DAMSL: Dialog Act Markup in Several Layers. <http://www.cs.rochester.edu/research/speech/damsl/RevisedManual/> (14.2.2014).
- Aller, Sven 2012. Dialoogiaktide märgendamine Eesti dialoogikorpuses: ülevaade ressursisidest ja tarkvaraarendus. [Recognition of Dialogue Acts in the Estonian Dialogue Corpus: Overview of Resources and Software Development.] Magistritöö. Tartu Ülikool.
- Allwood, J.; Ahlsen, E.; Björnberg, M.; Nivre, J. 2001. Social activity and communication act-related coding. – J. Allwood (ed.). *Dialog Coding – Function and Grammar*. Göteborg Coding Schemas. Gothenburg Papers in Theoretical Linguistics, 85, 1–28.
- Daelemans, W.; Zavrel, J.; van der Sloot, K.; van den Bosch, A. 2004. *TiMBL: Tilburg Memory-Based Learner Reference Guide*. Technical Report ILK 04-02. Tilburg University and University of Antwerp.
- Eskor, Liina 2004. Dialoogiaktid ja suhtlusstrateegiad: eesti dialoogikorpuse analüüs. [Dialogue acts and communicative strategies: analysis of Estonian dialogue corpus.] Magistritöö. Tartu Ülikool.
- Eskor, Liina 2005. Dialoogiaktid ja suhtlusstrateegiad: eesti dialoogikorpuse analüüs. [Dialogue acts and communicative strategies: analysis of Estonian dialogue corpus.] – *Keel ja Kirjandus*, 9, 711–727.
- Eskor, Liina 2006. Suhtlusstrateegiad infodialoogides. [Communicative strategies in information dialogues.] – Mare Koit, Renate Pajusalu, Haldur Õim (toim.). *Keel ja arvuti*. [Language and Computer.] Tartu: Tartu Ülikooli Kirjastus, 183–195.
- Eskor, Liina 2007. Suhtlusstrateegiad ja -taktikad müügivestlustes. [Communicative strategies and tactics in telemarketing calls.] – *Eesti Rakenduslingvistika Ühingu aastaraamat*, 3, 83–97. <http://dx.doi.org/10.5128/ERYa3.06>
- Fernandez, R.; Ginzburg, J.; Lappin, S. 2005. Using machine learning for non-sentential utterance classification. – *Proceedings of the 6th SIGdial Workshop on Discourse and Dialogue*. Lisbon, Portugal, 77–86.
- Fishel, Mark 2007a. Machine learning techniques in dialogue act recognition. – *Eesti Rakenduslingvistika Ühingu aastaraamat*, 3, 117–134. <http://dx.doi.org/10.5128/ERYa3.08>
- Fishel, Mark 2007b. Complex taxonomy dialogue act recognition with a Bayesian classifier. – *Proceedings: DECALOG'2007 Workshop on the Semantics and Pragmatics of Dialogue*. Rovereto, Italy, 161–162.

- Fišel, Mark; Kikas, Taavet 2006. Dialoogiaktide automaatne tuvastamine. [Automatic recognition of dialogue acts.] – Mare Koit, Renate Pajusalu, Haldur Õim (toim.). Keel ja arvuti. [Language and Computer.] Tartu: Tartu Ülikooli Kirjastus, 233–245.
- Georgila, K.; Artstein, R.; Nazarian, A.; Rushforth, M.; Traum, D. R.; Sycara, K. 2011. An annotation scheme for cross-cultural argumentation and persuasion dialogues. – 12th Annual SIGdial Meeting on Discourse and Dialogue. Portland, Oregon, USA, 272–278.
- Grau, S.; Sanchis, E.; Castro, M. J.; Vilar, D. 2004. Dialogue act classification using a Bayesian approach. – Proceedings of the 9th International Conference Speech and Computer, 495–499.
- 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>
- Hennoste, Tiit; Rääbis, Andriela 2004. Dialoogiaktid eesti infodialoogides: tüpoloogia ja analüüs. [Dialogue acts in Estonian information dialogues: a typology and analysis.] Tartu: TÜ Kirjastus. <http://dSPACE.utlib.ee/dSPACE/handle/10062/18995> (10.12.2013).
- Hutchby, Ian; Wooffitt, Robin 1998. Conversation Analysis. Principles, Practices and Applications. Cambridge, UK: Polity Press.
- Jokinen, Kristiina 1996. Cooperative response planning in CDM: Reasoning about communicative strategies. – S. LuperFoy, A. Nijholt, G. Veldhuijzen van Zanten (Eds.). TWLT11. Dialogue Management in Natural Language Systems. Enschede: Universiteit Twente, 159–168.
- Jokinen, Kristiina 2009. Constructive Dialogue Modelling: Speech Interaction and Rational Agents. John Wiley & Sons Ltd. <http://dx.doi.org/10.1002/9780470511275>
- Keizer, S.; op den Akker, R.; Nijholt, A. 2002. Dialogue Act Recognition with Bayesian Networks for Dutch Dialogues. – Proceedings of the 3rd SIGdial Workshop on Discourse and Dialogue. Philadelphia, USA, 88–94. <http://dx.doi.org/10.3115/1118121.1118134>
- Kikas, Taavet 2007. Dialoogiaktide tuvastamine eestikeelsetes dialoogides sufiksipuude abil. [Recognition of dialogue acts using suffix trees.] Magistritöö. Tartu Ülikool. <http://dSPACE.utlib.ee/dSPACE/handle/10062/2755> (10.12.2013).
- Koit, Mare 2012. Towards automatic recognition of the structure of Estonian directory inquiries. – A. Tavast, K. Muischnek, M. Koit (Eds.). Proceedings of 5th International Conference on Human Language Technologies: the Baltic Perspective, Baltic HLT 2012. IOS Press, 120–128.
- Kullasaar, Maret 2001. Eestikeelse dialoogikorpuse arendamine “võlur Ozi” tehnikaga. [Developing the Estonian dialogue korpus by „Wizard of Oz“ technique.] Magistritöö. Tartu Ülikool.
- Käbin, Tiit 2011. Eestikeelsete dialoogide kogumise veebirakendus. [Estonian dialogues capture web application.] Bakalaureusetöö. Tartu Ülikool. http://comserv.cs.ut.ee/forms/ati_report/index.php?year=2011 (10.12.2013).
- Levin, L.; Langley, C.; Lavie, A.; Gates, D.; Wallace, D.; Peterson, K. 2003. Domain specific speech acts for spoken language translation. – Proceedings of the 4th SIGdial Workshop on Discourse and Dialogue, Sapporo, Japan.
- Pärkson, Siiri 2011. Võlur Ozi eksperimentide kogumine ja partneri algatatud paranduste analüüs. [Wizard of Oz experiments and analysis of collected dialogues.] – Eesti Rakenduslingvistika Ühingu aastaraamat, 7, 197–214. <http://dx.doi.org/10.5128/ERYa7.12>
- Reithinger, Norbert; Maier, Elisabeth 1995. Utilizing statistical dialogue act processing in VERBMOBIL. – Proceedings of the 33rd Annual Meeting of the Association for Computational Linguistics. Cambridge, MA, 116–121. <http://dx.doi.org/10.3115/981658.981674>

- Samuel, Ken; Carberry, Sandra; Shanker-Vijay K. 1998. Dialogue act tagging with transformation-based learning. – Proceedings of the 17th International Conference on Computational linguistics. Vol. 2. Montreal, Quebec, 1150–1156. <http://dx.doi.org/10.3115/980432.980757>
- Sinclair, J.; Coulthard, R. M. 1975. *Toward an Analysis of Discourse*. Oxford: Oxford University Press.
- Stenström, A.-B. 1994. *An Introduction to Spoken Interaction*. London, New York: Longman.
- Treumuth, Margus 2011. *A framework for asynchronous dialogue systems: concepts, issues and design aspects*. *Dissertationes Mathematicae Universitatis Tartuensis* 72. Tartu: Tartu University Press. http://dspace.utlib.ee/dspace/bitstream/handle/10062/17522/treumuth_margus.pdf?sequence=1 (10.12.2013).
- Vutt, Evelyn 2001. *Eestikeelse dialoogikorpuse märgendamistarkvara*. [Software for annotation of Estonian dialogue korpus.] Magistritöö. Tartu Ülikool.
- Wright, H.; Poesio, M.; Isard, S. 1999. Using high level dialogue information for dialogue act recognition using prosodic features. – Proceedings of an ESCA Tutorial and Research Workshop on Dialogue and Prosody. Eindhoven, 139–143.

Sven Aller (Tartu Ülikool), magistrikraad informaatikas, on programmeerinud keeletötlustarkvara. Liivi 2, 50409 Tartu, Estonia
sven.aller@ut.ee

Olga Gerassimenko (Tartu Ülikool), magistrikraad eesti keele alal, on uurinud tagasisidepartikleid eesti ja vene keeles. Jakobi 2, 51014 Tartu, Estonia
olga.gerassimenko@ut.ee

Tiit Hennoste (Tartu Ülikool), filosoofiadoktori kraad eesti keele alal, on uurinud suulist eesti keelt. Jakobi 2, 51014 Tartu, Estonia
tiit.hennoste@ut.ee

Riina Kasterpalu (Tartu Ülikool) on uurinud tagasisidepartikleid eesti keeles. Jakobi 2, 51014 Tartu, Estonia
riina.kasterpalu@ut.ee

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

Kirsi Laanesoo (Tartu Ülikool), magistrikraad eesti keele alal, on uurinud küsimusi suulises eesti keeles. Jakobi 2, 51014 Tartu, Estonia
kirsi.laanesoo@ut.ee

Krista Mihkels (Tartu Ülikool), filosoofiadoktori kraad üldkeeleteaduse alal, on uurinud parandusi suulises eesti keeles. Jakobi 2, 51014 Tartu, Estonia
krista.mihkels@ut.ee

Andriela Rääbis (Tartu Ülikool), filosoofiadoktori kraad üldkeeleteaduse alal, on uurinud suulist eesti keelt. Jakobi 2, 51014 Tartu, Estonia
andriela.raabis@ut.ee

SOFTWARE FOR PRAGMATIC ANALYSIS OF DIALOGUES

Sven Aller, Olga Gerassimenko, Tiit Hennoste, Riina Kasterpalu, Mare Koit, Krista Mihkels, Kirsi Laanesoo, Andriela Rääbis

University of Tartu

We are investigating written Estonian dialogues – transcripts of human-human spoken dialogues as well as human-computer dialogues – with the aim of determining which dialogue acts are used in interaction, which communicative strategies are followed in order to achieve communicative goals, and which structural parts a dialogue includes. In order to simplify the pragmatic analysis of dialogues, we have developed software that makes it possible to recognise and annotate the dialogue acts, the communicative strategies and the structure of a dialogue. In recognition of dialogue acts, a data-driven method is implemented. Determination of the communicative strategies and the dialogue structure is based on rules. The software is used by linguists in dialogue studies the further aim of which is to develop a dialogue system that interacts with a user in Estonian and follows norms of human-human communication. The paper introduces the software tool and gives examples of its implementation.

Keywords: dialogue, dialogue act, communicative strategy, structure, Estonian

SÕNALIIGID EESTI LAPSE KÕNES JA LAPSELE SUUNATUD KÕNES

Reili Argus, Helen Kõrgesaar

Ülevaade. Küsimus, kas ja miks on laste varase sõnavara hulgas ülekaalus nimisõnad, on olnud paljude kirjutiste huvi keskmes. Siinses artiklis vaadeldakse sõnaliikide jaotumist ja esinemissagedust eesti lapse kõne varasel arenguperioodil vanuses 1;3–2;3 ning lapsele suunatud kõnes. Analüüsiks on kasutatud kahe lapse spontaanse kõne lindistusi mahus kokku 16 tundi. Materjalist ilmneb, et ka eesti lastel on sarnaselt inglise keelt kõnelevate lastega varase sõnavara hulgas ülekaalus nimisõnad. Samas ei anna vanemate kõne neile sellist sõnaliikide jaotumist ette, vanemate kõnes moodustavad nimisõnad umbes veerandi kõikidest sõnedest. Vaadeldud aasta jooksul nimisõnade esinemissagedus laste keeles väheneb ning vanuses 2;3 sarnaneb laste kõne sõnaliikide jaotumine vanemate omaga.

Võtmesõnad: sõnaliigid, lastekeel, lapsele suunatud kõne, eesti keel

1. Sissejuhatuseks: keeleüksuste kasutussageduse roll keeleomandamises

Keeleüksuste kasutussagedust on täiskasvanud keelekasutajate puhul peetud oluliseks keele igal tasandil, nii toetab muutemorfoloogias näiteks sagedus ebareeglipäraste vormide elujõulisust (Bybee 1995: 428) ning on oluline keelemuutuste tekkepõhjus (Diessel 2007: 108). Diesseli (2007: 123) sõnul põhineb keeleüksuste esinemissageduse mõju sellisel psüühilisel mehhanismil, mis seisneb mingi (keelise) esituse ehk vormi tugevdamisel. Elena Lieveni väitel muudab sagedus ühed esitused n-ö tugevamaks kui teised: tugevamad esitused ehk keeleüksused tunneme kiiremini ära ning nende põhjal loome ka üldistusi ehk paigutame uusi elemente mingite reeglite alla (Lieven 2010: 2546). Kindlasti võib sama väita ka laste alles omandatava keele kohta, kus sagedusest sõltub paljuski näiteks nii käändevormide, muuttüüpide kui ka üksikute lekseemide omandamine (vt nt Vihman, Vija 2006, Argus 2008).

Kasutuspõhisest lähenemisest lähtuvalt loovad lapsed oma grammatika fonoloogilis-leksikaalsetest jadadest ehk sõna(vormi)dest, mida nad kuulevad, mitte ei analüüsi sisendkeelt mingitest abstraktsetest sünnipärastest kategooriatest lähtuvalt (Lieven 2010: 2547). Seetõttu on leitud, et sagedus on omandamise võtmetegur, sest mis tahes reeglid tuleb omandaja sisendkeele distributiivsetest omadustest või mallidest (Ellis 2002: 144). Redingtoni jt (1998) järgi pakub tegelik keekekasutus ehk lapsele suunatud keel hulgaliselt distributiivset infot, mis aitab lapsel omandada peamiste sõnaliikide esinemissageduse.

Sageli on laste varast sõnavara uurides leitud, et selles on ülekaalus nimisõnad (nt Fenson jt 1994, Nelson 1973). Põhjuseks on autorid enamasti pidanud nimisõnade suuremat esinemissagedust lapsele suunatud keeles, kuid ka verbide kognitiivset keerukust võrreldes nimisõnadega (Genter 1982: 301). Mõne uurija (Genter 1982, Genter, Boroditsky 2001) arvates on nimisõnade ülekaal varase sõnavara hulgas omane mis tahes keele omandamise varasele staadiumile ning nimisõnu on rohkem, sest need aitavad lapsel keelt kiiremini omandada.

Ometi on uuemad uuringud näidanud, et keeltes, kus lapsele suunatud kõne sisaldab rohkem tegusõnu (nt mandariini k), on ka lapse varases kõnes enam just tegusõnu (Sandhofer jt 2000). Ka nimisõnade ja tegusõnade omandamise hõlpsuse või keerukuse kohta on saadud vastakaid andmeid (Stoll jt 2012: 286). Stoll jt väitel võib tegusõnade vähesus tuleneda eri kultuuridele omastest vestlusstrateegiatest, nt oletavad nad, et chintangi keeles on laste kõnes rohkem nimisõnu kui vanematel seepärast, et lastel lastakse sageli objekte nimetada (Stoll jt 2012: 316). Kuna väga paljud uuringud (nt Bates jt 1994, Fenson jt 1993, eesti keele kohta Schults jt 2012) põhinevad lapsevanema täidetavatel küsimustikel (nt MacArthur-Bates Communicative Development Inventory), mis on tugevalt nimisõnakesksed, võib arvata, et spontaanse kõne andmed ei pruugi sellist nimisõnade ülekaalu näidata.

Sandhofer jt (2000: 580) väidavad veel, et nimisõnad on keeles n-ö ühtlasema jaotumusega, nimisõnu on küll rohkesti, kuid iga üksik sõna ei esine keekekasutuses kuigi sageli ehk ükski nimisõna ei tõuse teistest esinemissageduse poolest esile. Oluliseks peavad Sandhofer jt ka seda, et lapsele suunatud nimisõnad on kontseptuaalselt sarnased, nad tähistavad väga sageli mingeid kindlaid nähtavaid objekte. Lapsed õpivad need kergesti ära kui konkreetsete asjade nimetused. Sandhoferi jt järgi kuulevad lapsed väiksel hulgal verbilekseme, mis esinevad aga sagedamini, st et üht verbi korratakse lapsele rohkemates vormides. (Sandhofer jt 2000: 578) Seda väidet võib kinnitada ka olemasoleva eesti materjali põhjal: eesti laste varased verbiparadigmad on vormirohkemad kui nimisõnade omad (Argus 2009: 150, Argus 2010). Lisaks tuleb lastel omandada väga erineva sagedusega verbe (nt väga sagedased *tegema-* ja *olema-*verb, mida kuuleb laps tunni aja jooksul kümneid kordi, ja mõni selline verb, mis esineb väga harva). Seega peavad lapsed suhteliselt väiksema valiku verbivormide põhjal tegema rohkem üldistusi kui nimisõnade puhul.

Sagedus ei ole kindlasti ainuke tegur, mis omandamist mõjutab. Näiteks Stoll jt (2012: 289) on varase sõnavara sõnaliigilist jaotumist vaadeldes leidnud, et nimisõnade ja tegusõnade esilduvus ning vormimoodustuse keerukus eri keeltes on üks põhjusi, miks on laste varases sõnavaras keeliti erinev hulk tegu- ja nimisõnu.

Eesti keele sõnaliikide jaotumise ja esinemissageduse kohta ei leidu kuigi palju materjali. Sõnavarastatistilisi töid küll on, kuid neis on enamasti keskmes kas näiteks ilukirjandusproosa autorikõne sõnavarastatistika (Kaasik jt 1977) või

seaduste keele grammatiliste kategooriate sagedus (Viks, Hein 2001) või on esitatud lemmade loend sageduse järjekorras (nt sagedussõnastikes). Andmeid eestikeelsete igapäevavestluste sõnaliigilise jaotumise kohta on samuti vähe, ainukese pidepunktina võib mainida Kristel Uiboaja jt (2013) verbide sageduse andmeid eesti murrete dialoogides. Eesti keele sõnaliikide omandamise kohta on MacArthur-Batesi testi põhiseid uurimusi, nt Schultsi jt (2012) uurimuses on vaadeldud eesti laste sõnavara vanuses 0;8–1;4. MacArthur-Batesi testiga kogutud keelematerjal on sõnad jaotatud suhtlussõnadeks, nimisõnadeks, predikaatideks (sellesse kategooriasse kuuluvad nii verbid kui ka adjektiivid) ja abisõnadeks. Nii on ka eesti laste kohta leitud, et kui lapse sõnavara koosneb rohkem kui 20 sõnast, hakkab suurenema nimisõnade hulk, kui lapse sõnavara on suurem kui 50 sõna, ongi seal ülekaalus nimisõnad ja alles siis, kui sõnavara veelgi kasvab, suureneb ka tegusõnade hulk (Schultsi jt 2012: 678). Schultsi jt uuringu puhul tuleb muidugi arvestada, et tegemist on meetodiga, kus vanemad panevad lapse sõnu ise kirja, ja nii võib see anda tulemused, mis erinevad paljuski spontaanses suhtlussituatsioonis saadutest.

Inglise keele materjalil põhinevas Li ja Fangi uurimuses (2011) on keskpunktis küll sõnaliikide jaotumine, kuid seda vaadeldakse nii, et kogu materjal laste vanusest 1;7–2;0 on liidetud üheks andmekogumiks ja analüüsitud sõnaliikide esinemissagedust koos, sõnaliikide osakaalu muutumine laste kõnes selle uuringu tähelepanu keskmes seega ei ole.

Siinse kirjutise eesmärk ongi esmalt anda ülevaade, milline on sõnaliikide jaotumine ja esinemissagedus spontaansetes igapäevavestlustes lapse ja täiskasvanu vahel. Vaadeldakse, kuidas see jaotumine lapse keelelise arengu varasel perioodil umbes aasta jooksul muutub ning kas ja kuivõrd muutub eri sõnaliikidesse kuuluvate sõnade esinemissagedus lapsega kõneleva vanema kõnes. Lisaks proovitakse välja selgitada, kas ja kuidas lapsele suunatud kõnes toimuvad muutused peegelduvad laste enda kõnes.

Kuna seniste uuringute nimi- ja tegusõnade esinemissageduse andmed suuresti lahknevad, on siinses kirjutises lähema tähelepanu all just nimi- ja tegusõnade osakaal nii lapse kui ka vanema kõnes. Keeleomandamisel on peetud oluliseks sedagi, kui keerukas või lihtne on kas nimi- või tegusõnade vormimoodustus, seepärast võib oletada, et ka neisse sõnaliikidesse kuuluvate sõnade struktuuriline keerukus või lihtsus võib omandamist mõjutada. Näiteks võib perioodil, kui mingi sõnaliigi esinemissagedus vanema kõnes oluliselt ei muutu, ometi muutus eksisteerida ja olla pigem kvalitatiivne ehk teisisõnu: vanema kõne võib sisaldada järjest keerukama ehitusega keeleüksuseid (nt tuletisi, liitsõnu või ühend- ja väljendverbe). Seetõttu vaadeldakse siinses kirjutises lähemalt tegusõnade jaotumist nende keerukusest lähtuvalt nii lapse enda kui ka lapsele suunatud kõnes.

2. Meetod ja keeleandmestik

Analüüsi aluseks on spontaanse kõne andmed: kokku 16 tundi igapäevastes situatsioonides lindistatud dialooge. Kasutatud on kaht CHILDESi eesti andmekogu, Andrease¹ (vanusest 1;7–2;3) ja Martina² (vanuses 1;3–2;3) vestluste transkriptsioone. Kuigi laste lindistused algavad eri vanusest, on nad siiski võrreldavad, sest

¹ Andrease andmekogu on koostanud Maigi Vija ning see asub aadressil <http://childes.psy.cmu.edu/browser/index.php?url=Other/Estonian/Vija/> (3.2.2014).

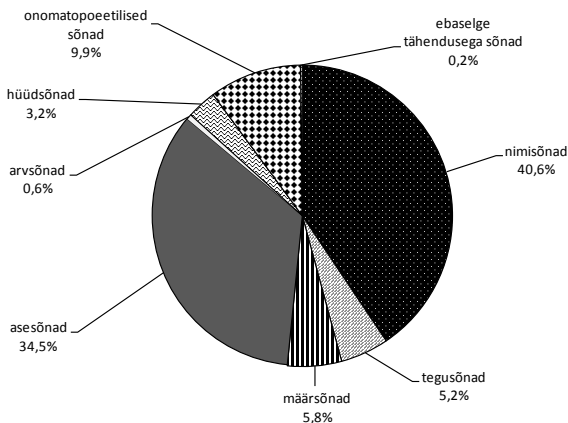
² Martina andmekogu on koostanud Airi Kapanen ning see asub aadressil <http://childes.psy.cmu.edu/browser/index.php?url=Other/Estonian/Kapanen/> (3.2.2014).

laste väljendite keskmine pikkus on mõlemal rohkem kui üks ja vähem kui kaks sõnet lausungis – seega on mõlemad keelelise arengu sellises etapis, kus nende kõnes leidub esimesi ühest sõnast pikemaid lausungeid.

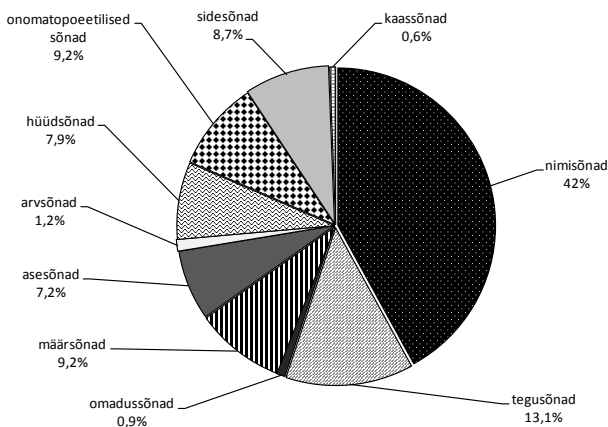
Keelematerjali töötlemisel on kasutatud programmi Clan³ võimalusi, nt sõnavormide sagedusloendit, kuid põhisos on andmed töödeldud käsitsi, eelkõige laste kõnes sel perioodil leiduvate ebakindla häälikulise koostise või ebaselge tähendusega üksuste ning lalina ja keelemängu rohkuse tõttu. Sõnaliikide jaotusena on kasutatud eesti grammatikate traditsioonilisemat liigitust nimi-, tegu-, omadus-, määr-, arv-, ase-, side-, hüüd- ja kaassõnadeks. Määrsõnade hulka on arvatud ka asemäärsõnad, nagu *siin*, *seal* jts. Eraldi on arvestatud onomatopoeetilised sõnad, mille sõnaliigiline kuuluvus on sageli ebamäärane, ebaselge tähendusega sõnad ning vaatluse alt välja jäänud üneemid. Keksõnavormide liigitamisel on lähtutud nende rollist lauses: öeldisena toimivad on loetud tegusõnadeks (nt *see on ära söödud*), täiendina toimivad omadussõnadeks (*võta soolatud võid*). Analüüsitud on nii laste kui ka laste vestluspartnerite ehk täiskasvanute kõne sõnaliikide esinemissagedust, enamasti on tähelepanu keskpunktis sõnade esinemisjuhud ehk sõned, sõnavara ulatuse näitamiseks on mõnes vanusepunktis esitatud ka eri lekseemide hulk.

3. Sõnaliikide jaotumine vaatlusperioodi alguses

Sõnaliikide jaotumisest annavad ülevaate joonised 1 ja 2. Joonistelt võib näha, et lindistamisperioodi alguses on mõlema lapse kõnes üsna sarnaselt suur hulk nimisõnu, mis moodustavad Andreasel 40% ja Martinal 41% kõikidest sõnedest. Üsna ühesugune (9–10%) on mõlema lapse kõnes ka onomatopoeetiliste sõnade sagedus. Osa neist sõnadest on lastel samad, näiteks leidis nii Andrease kui ka Martina kõnes koera tähistamiseks sõna *aua*, kuid osa imitatiive olid mõlemal lapsel erinevad, näiteks linnuhääle kohta kasutas Andreas *kakaa*, Martina aga *tiu*.



Joonis 1. Andrease kõne sõnaliikide jaotumine arvatuna sõnade esinemisjuhtudes ehk sõnedes vanuses 1;7



Joonis 2. Martina kõne sõnaliikide jaotumine arvatatuna sõnade esinemisjuhtudes ehk sõnedes vanuses 1;3

Oluline erinevus kahe lapse keelematerjalis puudutab tegusõnade osakaalu: Andreasel on tegusõnu vähe, kõigest 5%, kuid Martina materjalis leidub neid mõnevõrra rohkem, 14%. Torkab silma ka see, et kui laps on vaadeldud ajahetkeks omandanud suurema hulga tegusõnu, kasutab ta ka rohkem määrsõnu, nii on Martina kõnes tunduvalt enam määrsõnu kui Andrease kõnes. Samas on Andreasel määrsõnu isegi protsendi jagu rohkem kui verbe. Kuna Andrease lausungid on sel perioodil veel enamasti ühesõnalised, kasutab laps mingi suhte väljendamiseks pigem adverbki kui verbi:

- (1) *AND: *sii [=siin].*
 *EMA: *sinna vaatased jah? (Andreas vanuses 1;7)*

Samuti on väga erinev asesõnade hulk, mis on Andreasel lausa 35% kõikidest sõnedest. Selline asesõnade suur kasutussagedus tuleneb esimeses lindistuses sellest, et vaatlusaluse lapse kõne koosnes sel perioodil suures osas küsimustest *kesse* 'kes see on?' ja *misse* 'mis see on?'⁴. Järgmistes lindistustes väheneb selliste lausungite arv ja sel moel ka asesõnade hulk tunduvalt. Kaassõnu ja omadussõnu on Martina kõnes mõni üksik, Andreasel need esimeses lindistuses puudusid.

Mõlema lapse keelekasutuses on esialgu veel ebaselge tähendusega elemente: Martina kõnes on neid mõnevõrra rohkem, sest tema suhtlusstiili iseloomustab rohke keelemäng ja justkui hääldusorganite treenimiseks mõeldud sõnalaadsete üksuste kordamine:

- (2) *MAR: *ko ku ke kue kuk, emme.* (Martina vanuses 1;3)

Sageli on dialoogist näha, et ka vanemad ei saa lapse öeldust aru:

- (3) *MAR: *toppibäli, kappibäl.*
 *EMA: *mida?*
 *ISA: *ei saand aru.*
 *MAR: *topbäl.*

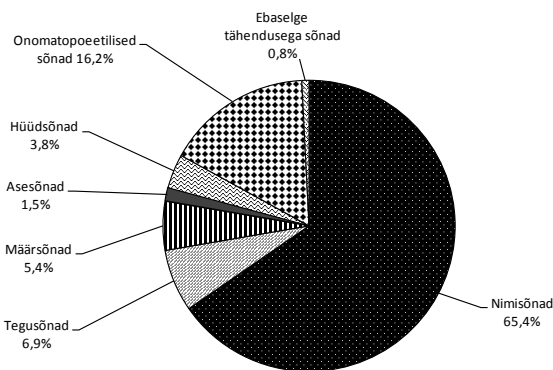
⁴ Sellised kokku hääldatud ühendid arvestati materjalis eraldi, st *kesse* läks arvesse kui pronomeenid *kes* ja *see*.

%com: ISA ütleb EMale sõna järgi, püüdes aru saada, mida MAR on öelnud.

*ISA: ütle veel, ei saand aru.

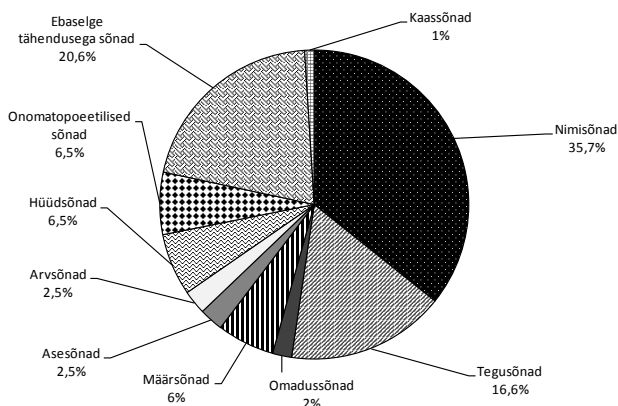
*MAR: *kap käi, idäh.* (Martina vanuses 1;5)

Kui vaadelda laste varase sõnavara **eri lekseemide arvu** sõnaliigiti, saab ülevaate sellest, kui rikas või siis ka kui piiratud on mingi kindla sõnaliigi sõnavara. Joonistel 3 ja 4 on kujutatud laste sõnavara koostist vaatlusperioodi alguses nii, et arvutatud on iga sõnaliigi eri lekseemide hulk. Nagu joonistelt näha, on sõnavara rikkus sõnaliigiti erinev. Andrease sõnavaras torkab silma, et nimisõnu on tema kõnes sel perioodil palju, 65% kogu sõnavarast. Küllaltki mitmekesine on ka onomatopoeetiline sõnavara, vähem on tegu- ja määrsõnu, vastavalt ainult 7% ja 5% (vt joonis 3).



Joonis 3. Andrease sõnavara koostis arvatuna lekseemides lindistusperioodi alguses vanuses 1;7

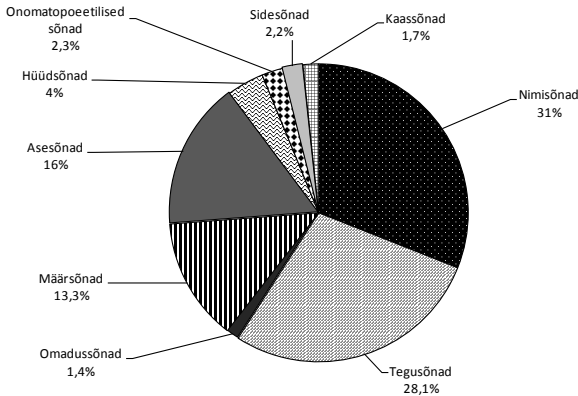
Martina sõnavara puhul torkab silma, et tema kõnes on kõige rikkalikum valik nimi- sõnu ning seejärel ebaselge tähendusega sõnu. Tegusõnu oli Martinal varasel perioodil rohkem nii esinemisjuhtudes kui ka lekseemides arvatuna (vrd joonis 2 ja 4). Need kolm sõnaliiki ongi Martina kõne lindistamise alguses kõige rikkalikuma sõnavaraga.



Joonis 4. Martina sõnavara koostis arvatuna lekseemides lindistusperioodi alguses vanuses 1;3

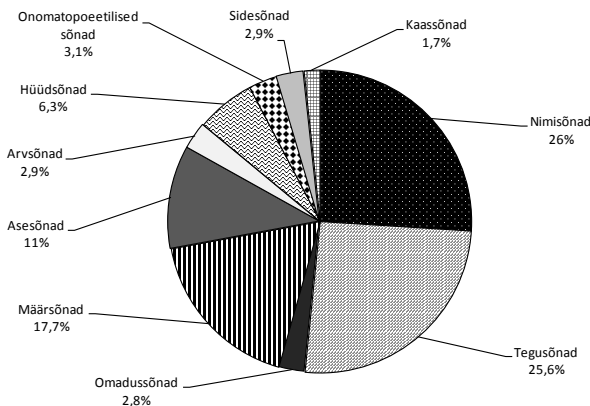
Schultsi jt (2012: 678) andmetel on eesti laste varase sõnavara hulgas vähe abisõnu (ingl *function words*) ehk määr-, kaas- ja sidesõnu. Kuna siinses analüüsis kasutatud materjal pärineb mõnevõrra vanemate laste kõnest kui Schultsi jt materjal, siis võib oletada, et näiteks määrsõnade üsna suur hulk vaatlusaluste laste kõnes võib tuleneda vanusest, aga ka andmekogumise meetodilistest erinevustest (lapse kasutatud abisõnu ei ole ilmselt vanemal kuigi kerge tähele panna ja üles kirjutada).

Kui kõrvutada eri sõnaliikide jaotumist laste kõnes **lastele suunatud kõne andmetega**, paistab kõigepealt silma, et võrreldes sama dialoogi lapse kõnega sisaldab vanema kõne tunduvalt vähem nimisõnu ja palju enam tegusõnu (vt joonis 5 ja 6 võrrelduna joonistega 1 ja 2). Suur erinevus on ka onomatopoeetiliste sõnade osakaalus: vanema kõnes moodustavad need kõigest 2–3% kõigist esinenud sõnedest.



Joonis 5. Andreasele suunatud kõne sõnaliikide jaotumine arvatuna sõnedes vaatlusperioodi alguses lapse vanuses 1;7

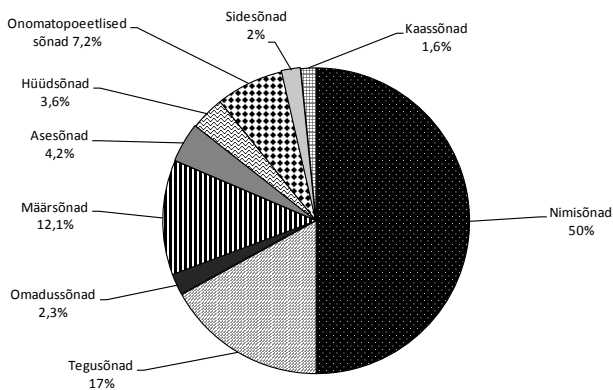
Martinale suunatud kõne on võrreldes lapse enda kõnega samuti tegusõnarikkam, tegusõnad moodustavad selle lindistuse kõikidest sõnedest veerandi. Teise veerandi moodustavad nimisõnad ning võrreldes lapse kõnega on sagedased ka määrsõnad. Samas leidub Martinale suunatud kõnes võrreldes lapse enda kõnega tunduvalt rohkem asesõnu. Nii nagu Andreasele suunatud kõnes, on ka Martinale suunatud kõnes võrreldes lapse kõnega vähe onomatopoeetilisi sõnu.



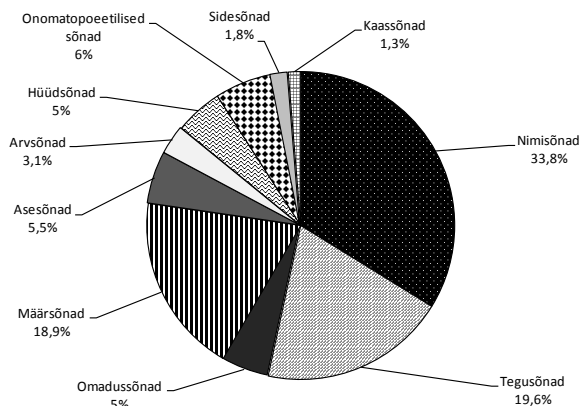
Joonis 6. Martinale suunatud kõne sõnaliikide jaotumine arvatuna sõnedes vaatlusperioodi alguses lapse vanuses 1;3

Võrreldes kahe vanema kõne sõnaliikide jaotumist, võib väita, et see on üldjoontes sarnane: nimi- ja tegusõnad moodustavad enam-vähem sama suure hulga, neid on pisut üle poole kõikidest tekstisõnadest ehk sõnedest. Teistesse sõnaliikidesse kuuluvate sõnade hulgas on ülekaalus määrsõnad ja asesõnad, väga vähe kasutatakse omadussõnu. Kui vaadata nimi- ja tegusõnade osakaalu lastele suunatud kõnes, võib väita, et lapsed ei pruugi sugugi omandada tegusõnu hiljem kui nimisõnu pelgalt seepärast, et nad kuulevad neid vähem. Ilmselt on tunduvalt olulisem tegur tegusõnade kas kognitiivne või vormiline keerukus. Kõige suurem erinevus laste endi ja lastele suunatud kõne vahel ilmnebki just tegusõnade esinemissageduses. Vanemad kasutavad tunduvalt sagedamini tegusõnu kui lapsed ning ei saa üldiseltki väita, et varasel perioodil peegeldub vanema kõne sõnaliikide jaotumine kuigi selgelt lapse kõnes: erinevusi on mitte ainult tegu- ja nimisõnade, vaid ka onomatopoeetiliste sõnade, asesõnade ja side- ning kaassõnade esinemissageduses.

Kui arvutada vanemate kõne sõnaliikide esinemissagedust lekseemides, on proportsioonid pisut erinevad. Nimisõnavara on rikkam, moodustades Andrea-sele suunatud kõne sõnavarast 50% ja Martinale suunatud kõnes 34%, tegusõnu on Andrea-sele suunatud kõnes aga ainult 17% ja Martinale suunatud kõnes 20%. Seega võib ka eesti keele kohta väita, et lastele suunatud kõnes (ja võimalik, et ilmselt üldse suulises kõnes) kasutatakse sagedamini vähemat hulka tegusõnu ja suuremat hulka nimisõnu nii, et iga nimisõna esineb vestluse jooksul ainult mõnel korral. Andmetest nähtub näiteks, et lindistusperioodi esimeses vestluses kasutas Andrea-se ema üht nimisõna keskmiselt 3,3 korral ja üht tegusõna 8,9 korral, Martina ema kõnes esines üks nimisõnaleksee keskmiselt 5,4 korda ja tegusõna 9,2 korda. Siit võib oletada, et nimi- ja tegusõnu saavad lapsed tõepoolest omandada hoopis erinevalt: nimisõnade kohta tehakse üldistusi suurema hulga eri lekseemide põhjal, tegusõnade kohta aga väiksema hulga lekseemide, kuid sagedasemate esinemisjuhtude põhjal.



Joonis 7. Andrea-sele suunatud kõne sõnavara jaotumine arvutatuna lekseemides lapse vanuses 1;7



Joonis 8. Martinale suunatud kõne sõnavara jaotumine arvatuna lekseemides lapse vanuses 1;3

4. Laste sõnaliikide esinemissageduse ja jaotumise muutumine vaatlusperioodi jooksul

Laste kõne sõnaliikide esinemissagedus muutub vanuse kasvades. Vaatlusperioodi keskel, vanuses 1;11–2;1 ilmuvad lapse kõnemes näiteks sidesõnad ja kaassõnad, mis annavad märku grammatika arengust (vt tabel 1 ja 2)⁵.

Tabel 1. Sõnaliikide esinemissagedus arvestatuna sõnedes Andrease keelematerjalis vanuses 1;7–2;3

Lapse vanus	Nimi-sõnad	Tegu-sõnad	Omadus-sõnad	Määr-sõnad	Ase-sõnad	Hüüd-sõnad	Arv-sõnad	Side-sõnad	Kaas-sõnad	Onomato-poeetilised sõnad	Ebaselge tähendusega sõnad	Kokku
1;7	216	28	0	31	184	17	3	0	0	53	1	74
1;8	323	92	30	67	128	46	3	0	0	71	10	130
1;9	138	45	19	18	56	11	0	0	0	10	1	22
1;10	347	146	38	110	89	63	18	0	12	14	0	837
1;11 ⁶	90	51	14	18	22	30	0	0	5	4	1	40
2;0	373	281	35	127	108	89	80	7	12	14	28	1154
2;1	242	250	23	93	112	51	22	4	12	11	0	820
2;3	140	202	18	148	77	41	2	9	13	1	13	664
Kokku	1869	1095	177	612	776	348	128	20	54	178	54	5311

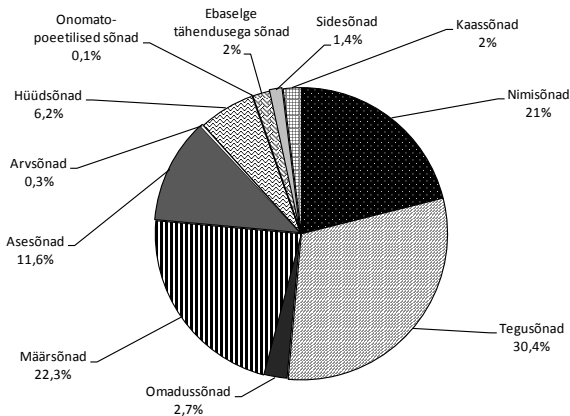
⁵ Tabelis 1 ja 2 on esitatud ahel- ja tugiverbides ning seriaalkonstruksioonides esinenud verbitüved eraldi (nt *tule pane* on loetud kaheks verbiks).

⁶ Selle dialoogi lindistamise ajal oli laps haige, sellest võib tuleneda tunduvalt väiksem hulk lausungeid ja seetõttu ka sõnesid selles vanusepunktis.

Tabel 2. Sõnaliikide esinemissagedus arvestatuna sõnedes Martina kõnes vanuses 1;3–2;3

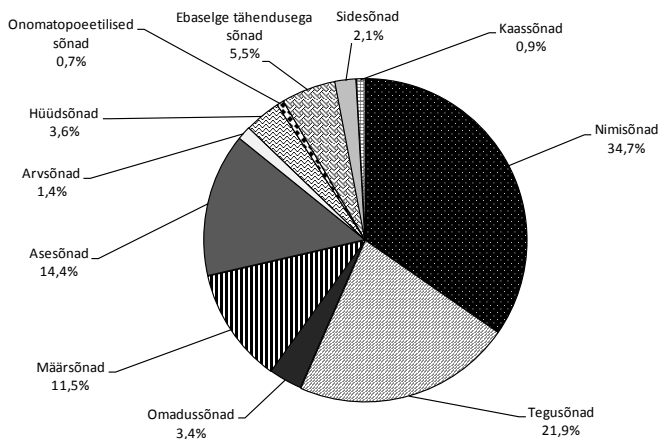
Lapse vanus	Nimi-sõnad	Tegu-sõnad	Omadus-sõnad	Määr-sõnad	Ase-sõnad	Hüüd-sõnad	Arv-sõnad	Side-sõnad	Kaas-sõnad	Onomato-poeetilised sõnad	Ebaselge tähendusega sõnad	Kokku
1;3	343	117	7	75	59	65	10	0	5	75	71	827
1;5	478	214	16	102	46	58	1	1	32	30	63	1041
1;6	287	140	18	53	63	20	4	1	13	5	25	629
1;9	411	409	19	137	223	89	20	8	23	29	35	1403
1;10	452	390	55	244	266	55	6	8	6	16	40	1538
1;11	363	314	38	128	160	38	6	10	29	2	41	1129
2;1	455	476	56	283	151	55	24	54	28	2	44	1628
2;3	547	346	53	181	227	56	22	33	15	11	86	1577
Kokku	3336	2406	262	1203	1195	436	93	115	151	170	405	9772

Tabelid 1 ja 2 annavad ülevaate laste sõnaliikide esinemissagedusest kogu vaatlusperioodi lõikes. Tabeleist on näha, kuidas onomatopoeetiliste sõnade hulk laste kõnes enne kaheaastaseks saamist (1;10) väheneb ning umbes samal ajal hakkab suurenema tegusõnade ja sidesõnade esinemisjuhtude arv. Arvsõnade hulk tundub sõltuvat pigem vestlusteemast, sest see kõigub lindistuse üsna oluliselt. Kui nimisõnade hulk on üsna suur pidevalt ja kõigub lindistuse peaaegu kogu aeg, siis tegusõnade hulk teeb järsu hüppe ja seda mõlemal lapsel üsna samas vanuses: Andreasel vanuses 1;10 ja Martinal vanuses 1;9. Samasugune järsk tõus peegeldub samas vanuses ka määrõnade arvus.



Joonis 9. Andree kõne sõnaliikide jaotumine arvatuna sõnedes vaatlusperioodi lõpus vanuses 2;3

Vaatlusperioodi lõpuks, vanuseks 2;3 on Andree kõnes nimisõnade esinemisjuhtude osakaal vähenenud 40%-lt 21%-le ning tegusõnade osakaal suurenenud 5%-lt lausa 30%-le (vrd joonis 1 ja 9). Sarnaselt tegusõnadega on suurenenud määrõnade sagedus. Lisandunud on omadus-, side- ja kaassõnad, mida esimeses lindistuses ei leidunud.

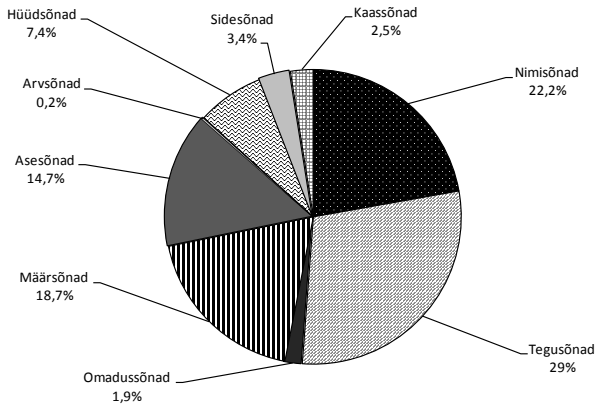


Joonis 10. Martina kõne sõnaliikide jaotumine arvatuna sõnedes vaatlusperioodi lõpus vanuses 2;3

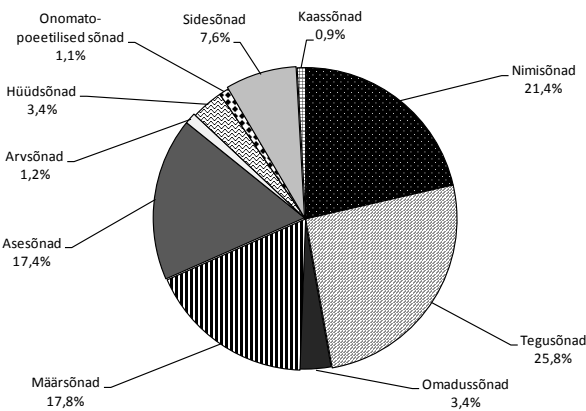
Ka Martina kõnes on selgelt näha nimisõnade esinemisjuhtude osakaalu vähenemine (41%-lt 35%-le) ning tegusõnade osakaalu tõus (14%-lt 22%-le). Samuti moodustavad onomatopoeetilised sõnad lapse kõnes tunduvalt väiksema osa kui lindistusperioodi alguses (0,7% võrreldes varasema 9%-ga). Teistesse sõnaliikidesse kuuluvate sõnade esinemissagedus on muutunud mõnevõrra vähem (vrd joonis 2 ja 10). Mõlema lapse kõne on eri sõnaliiki kuuluvate sõnede osakaalult sarnane.

Ka inglise keelt omandavad lapsed kasutavad sellises vanuses vähe omadussõnu (vanuses 2;0 moodustasid omadussõnad kõigest 2,83% kõikidest sõnedest; Li, Fang 2011: 101). Siinses uurimuses vaadeldud laste kõne andmed on üsna sarnased – vanuses 2;1 moodustavad Martinal omadussõnad 3,4%, Andreasel vanuses 2;0 aga 3,2% kõikidest sõnedest. Lisaks väidavad Li ja Fang, et lapsed kasutavad märksa sagedamini määrsõnu kui omadussõnu. See peab paika eesti andmetegi puhul, kõikides lindistustes on määrsõnu rohkem kui 10% kõikidest sõnedest. Nimisõnade sagedus paistab eesti lastel olema pisut suurem kui inglise keelt omandavatel lastel: kui inglise keelt omandavate laste kõnes moodustasid nimisõnad vanuses 2;0 umbes 20% kõikidest sõnedest, siis Andrease kõnes on samas vanuses 32,3% kõikidest sõnedest nimisõnad ja Martina kõnes on vanuses 2;1 nimisõnu pisut vähemgi, 27,6%.

Kui vaadelda aga **lapsele suunatud kõne** sõnaliikide jaotumist samas vanuses lindistatud dialoogis (vrd joonis 11 ja 12), võib näha, et ühe vanema, Andrease ema, kõnes esinenud ainult õige pisut suurem hulk tegusõnu (vt joonis 11) võib peegelduda ka lapse kõnes (joonis 9). Loomulikult võib seda tingida ka lindistussituatsioon ning praeguse väikese materjali hulga põhjal vanemate nimi- või tegusõnalembuse kohta väga selgeid järeldusi teha ei saa.



Joonis 11. Sõnaliikide jaotumine arvatuna sõnedes Andreasele suunatud kõnes vaatlusperioodi lõpus vanuses 2;3



Joonis 12. Sõnaliikide jaotumine arvatuna sõnedes Martinale suunatud kõnes vaatlusperioodi lõpus vanuses 2;3

Kuigi võis oletada, et vanemate kõne muutub lapse kasvades, ei saa seda siiski sinne materjali põhjal eri sõnaliikide jaotumise kohta väita. Ainukese olulise muutusena võib täheldada nimisõnade osakaalu väikest langust (ühel vanemal 5%, teisel 9%). Üks põhjus võib olla selles, et kuna lindistused olid tehtud igapäevastes situatsioonides, nagu söömine, riietumine, mängimine, võib olla, et kõne arengu varasel perioodil ongi sellises situatsioonis toimuv vestlus üsna ühesugune, sõltumata kuigi palju lapse vanusest. Niisugune jaotumine ei pruugi olla omane isegi mitte lapsele suunatud kõnele, vaid võib suure tõenäosusega olla iseloomulik üldiselt argivestlusele. Näiteks on leitud, et eesti murretes on vestluses esinevate tegusõnade hulk umbkaudu 20% kõikidest sõnedest (Uiboaed jt 2013). Võib oletada, et lapsele suunatud kõne on õige pisut tegevusekesksem ja seetõttu ulatub sinne materjali tegusõnade esinemisjuhtude osakaal üle 20%. Põhjalikum võrdlus teiste täiskasvanult täiskasvanule suunatud argivestlustega lubaks siinkohal teha ehk ka selgemaid järeldusi.

Inglise keele omandamise andmetega võrreldes (Li, Fang 2011: 103) torkab olulise erinevusena silma, et inglise vanemad kasutavad lastega suheldes rohkesti asesõnu, nt moodustavad nende kõnes asesõnad rohkem kui 21% sõnedest, eesti lastele suunatud kõnes ei ole see osakaal suurem 17%-st.

5. Tegusõnade jaotumine nende struktuuri keerukusest lähtuvalt

Kuigi lapsele suunatud kõnes leidus vaatlusperioodi alguses tegusõnu pea sama palju kui perioodi lõpus, oli verbide hulk laste keelekasutuses ikkagi see, mis muutus vaatlusperioodi jooksul kõige enam. Kontrollimaks oletust, et kui tegusõnade hulk lapsele suunatud kõnes ei muutu, muutub ometi nende keerukuse aste, ongi järgnevalt kõik nii lapse enda kui ka lapsele suunatud kõne tegusõna sisaldavad konstruktsioonid jagatud ülesehitusest lähtuvalt kolmeks. Esimese rühma moodustavad juurverbid ehk ainult üht verbitüve sisaldavad verbid (nt *jooma*), teise jaotuse verbitüvest ja tuletusliitest koosnevad ehk tuletatud verbid (nt *söö-t-ma*) ning kolmanda rühma kahest lekseemist ehk kas kahest verbitüvest (nt seriaalkonstruktsioonid, nagu *mine vii*) või verbist ja abimäärsõnast, nimi- või asesõnatüvest koosnevad konstruktsioonid (ühend- ja väljendverbid, nagu *silma torkama* või *katki minema*).

Vaatlusperioodi alguses esines ühe lapse, Andrease kõnes tegusõnu 28 korral (9 eri lekseemi) ning kõik need olid eranditult juurverbid. Alles kaks kuud hiljem ilmusid tema kõnesse esimesed ühendverbid (*sai otsa, lahti tee* vanuses 1;9). Teise lapse, Martina kõnes leidus ka esimestes lindistustes 115 verbi esinemisjuhu hulgas siiski üks tuletatud verb (*näitama* vanuses 1;3) ning juba teisest lindistusest on registreeritud 7 tuletatud verbi 13 esinemisjuhuga. Samuti ilmub juba vanuses 1;3 tema kõnesse esimene ühendverb (*tee lahti*). Siiski moodustavad ka tema tegusõnade hulgas ülekaaluka osa juurverbid.

Vaatlusperioodi lõpuks on laste tegusõnakasutus keerukusest lähtuvalt väga sarnane: suurema osa verbisõnedest moodustavad ikkagi juurverbid: 85% Martina kõnes ja 89% Andrease kõnes. Tuletatud verbid moodustavad laste kõnes vaatlusperioodi lõpus u 3–5% kõikidest verbide esinemisjuhtudest. Ühendverbide ja seriaalkonstruktsioonide hulk on kasvanud, kuid laste kõnes pisut erinev: Andrease kõnes on neid 6% kõikidest verbi esinemisjuhtudest, kuid Martinal poole rohkem ehk ligi 12%. Seega võib väita, et võrreldes vaatlusperioodi algusega, mil lapsed kasutasid pea eranditult juurverbe, on nende verbikasutus siiski muutunud, esineb nii tuletisi kui ka mitmeosalisi tegusõnu.

Vaadeldes **lapsele suunatud kõnes** esinevate tegusõnade keerukuse astet, on proportsioonid üsna sarnased sellega, mida laste endi keeles näha võib. Vaatlusperioodi alguses on küll vanemate kõnes rohkem tuletisi kui laste enda kõnes, kuid nende hulk ei ületa 6% piiri kõikidest verbisõnedest. Ühendverbe ja seriaalkonstruktsioone on Martinale suunatud kõnes mõnevõrra rohkem – 20% kõikide verbide esinemisjuhtudest – kui Andreasele suunatud kõnes, kus neid on 9%. Tabelites 3 ja 4 on esitatud kõik laste ja neile suunatud kõne tegusõnad keerukusest lähtuvalt. Kui lapse kõnes suureneb keerukamate tegusõnade hulk, siis täiskasvanu kõnes sõltub eri tüüpi verbide sagedus ilmselt vestlussituatsioonist, sest ühtlast muutumissuunda

materjalist kummagi vanema puhul ei ilmne. Vaatlusperioodi lõpuks on tuletiste osakaal mõlema vanema keelekasutuses ainult paari protsendi võrra suurenenud, kuid jääb siiski 10% piiresse. Mitmest üksusest koosnevate tegusõnade hulk on suurenenud ainult Andreasele suunatud kõnes ja seda 17%-ni. Martinale suunatud kõnes on seriaalkonstruksioonide ning ühend- ja väljendverbide hulk kõikidest verbisõnadest isegi pisut väiksem kui esimeses lindistuses, 18%.

Tabel 3. Martina ja talle suunatud kõne tegusõnade esinemissagedus lähtuvalt nende keerukusest arvatuna sõnedes

Tegusõnad		Vanus							
		1;3	1;5	1;6	1;9	1;10	1;11	2;1	2;3
Martina	Juurverbid	115	200	134	365	348	247	381	274
	Tuletatud verbid	1	13	4	9	12	13	56	9
	Seriaalkonstruksioonid, ühend- ja väljendverbid	1	1	2	21	21	32	27	38
	Kokku	117	214	140	395	381	292	464	321
Martina ema	Juurverbid	438	553	336	520	381	394	620	312
	Tuletatud verbid	45	30	18	22	40	23	41	24
	Seriaalkonstruksioonid, ühend- ja väljendverbid	121	133	50	107	127	124	160	68
	Kokku	604	716	404	649	548	541	821	404

Tabel 4. Andrease ja talle suunatud kõne tegusõnade esinemissagedus lähtuvalt nende keerukusest arvatuna sõnedes

Tegusõnad		Vanus							
		1;7	1;8	1;9	1;10	1;11	2;0	2;1	2;3
Andreas	Juurverbid	28	91	41	135	48	213	181	178
	Tuletatud verbid	0	1	0	5	3	34	20	10
	Seriaalkonstruksioonid, ühend- ja väljendverbid	0	0	4	6	0	22	27	12
	Kokku	28	92	45	146	51	269	228	200
Andrease ema	Juurverbid	393	242	195	230	115	257	221	94
	Tuletatud verbid	11	8	8	23	4	47	13	13
	Seriaalkonstruksioonid, ühend- ja väljendverbid	40	39	48	31	13	105	63	23
	Kokku	444	289	251	284	132	409	297	130

Seega ei saa väita, et tegusõnade keerukuse aste lapsele suunatud kõnes lapse kasvades väga palju muutub, sest juurverbide esinemissagedus on mõlemale lapsele suunatud kõnes endiselt väga suur. Väikesed erinevused lapsele suunatud kõne perifrastiliste verbide sageduses peegelduvad siiski laste kõnes – suurem hulk ühendverbide ja seriaalkonstruksioonide esinemisjuhte Martinale suunatud kõnes ilmselt mõjutab seda last selliseid konstruksioone mõnevõrra sagedamini kasutama.

6. Kokkuvõtteks

Sõnaliikide jaotumise ja esinemissageduse kohta eesti laste varases kõnes võib esmase analüüsi põhjal öelda, et eesti laste kõne on keele omandamise varases staadiumis nimisõnakeskne ning aasta jooksul see nimisõnakesksus mõnevõrra väheneb. Sarnaselt varasemate, teiste keelte kohta saadud uuringutega, paistab ka siinsest materjalist, et laste nimisõnakesksus ei tule otse sisendkeele nimisõnade suuremast osakaalust: tegusõnu on lapsele suunatud kõnes vaatlusperioodi alguses pea sama palju kui vaatlusperioodi lõpus ja tunduvalt rohkem kui lastel vaatlusperioodi alguses. Kinnitust leidis ka see, et tegusõnade osakaal lekseemiti on tunduvalt suurem kui sõneti ehk teisisõnu: ka eesti lapsed kuulevad nimisõnu, mis esinevad vähem kordi, ning enam kordi ühtesid ja samu tegusõnu.

Sõnaliikide jaotumiselt on laste sõnavara vanuses 2;3 juba üsna sarnane täiskasvanute omaga, seda nii tegu- ja nimisõnade osakaalu kui ka onomato-poeetiliste sõnade hulka silmas pidades, väga väikesed erinevused on näha ainult ebaselge tähendusega sõnade hulgas. Individuaalsed erinevused vanemate kõnes peegelduvad mingil määral ka laste kõnes. Loomulikult võib tegemist olla üksiku vestlussituatsiooni eripäraga, jutuainesega, mis tingib teatud liiki sõnade sagedama kasutuse. Selline tingitus võib aga suunata nimisõnade kasutamist ja vähem verbide oma. Ometi paistab näiteks Martinale suunatud kõne suurem perifrastiliste tegusõnade esinemissagedus mõjutavat ka lapse keelekasutust.

Võrreldes eesti andmeid inglise keelt omandavate laste ja neile suunatud kõne andmetega ei ole erinevused kuigi suured: eesti laste ja lastele suunatud kõnes on õige pisut enam tegusõnu ja veidi vähem asesõnu.

Lapsele suunatud kõne sõnaliikide jaotumine muutus vaatlusperioodi vältel väga vähe, samuti ei muutunud oluliselt verbide keerukuse aste. Seega võib väita, et vanemad ei kohanda lapse kasvades, vähemalt ühe vaadeldud aasta jooksul, oma keelekasutust kuigi palju. Selgitamaks välja, kas eesti vanemad tõepoolest lapsega kõneledes oma keelekasutust lapse vanuse kasvades oluliselt ei muuda, tuleks analüüsida pikema perioodi jooksul tehtud lindistusi ning võrrelda saadud andmeid täiskasvanult täiskasvanule suunatud kõnega, eriti argivestluste andmetega.

Viidatud kirjandus

- Argus, Reili 2008. Eesti keele muutmorfoloogia omandamine. Tallinna Ülikooli humanitaarteaduste dissertatsioonid 19. Tallinn: Tallinna Ülikooli kirjastus.
- Argus, Reili 2009. The early development of case and number in Estonian. – Maria D. Voeikova, Ursula Stephany (Eds.). Development of Nominal Inflection in First Language Acquisition. A Cross-Linguistic Perspective. Berlin: Mouton De Gryter, 111–152.
- Argus, Reili 2010. Mida teeb *tegema*-verb hoidjakeeles. – Eesti ja soome-ugri keeleteaduse ajakiri ESUKA / Journal of Estonian and Finno-Ugric Linguistics JEFUL, 2, 17–34.
- Bates, Elisabeth; Marchman, Virginia; Thal, Donna; Fenston, Larry; Dale, Philip; Reznick, Steven; Reilly, Judy; Hartung, Jeff 1994. Developmental and stylistic variation in the composition of early vocabulary. – Journal of Child Language, 21 (1), 85–123. <http://dx.doi.org/10.1017/S0305000900008680>
- Diessel, Holger 2007. Frequency effects in language acquisition, language use, and diachronic change. – New Ideas in Psychology, 25 (2), 108–127. <http://dx.doi.org/10.1016/j.newideapsych.2007.02.002>

- Ellis, Nick C. 2002. Frequency effects in language processing. A review with implications for theories of implicit and explicit language acquisition. – *Studies in Second Language Acquisition*, 24 (2), 143–188. <http://dx.doi.org/10.1017/S0272263102002024>
- Fenson, Larry; Dale, Philip; Reznick, Steven; Thal, Donna; Bates, Elisabeth; Hartung, Jeff; Pethick, Stephen; Reilly, Judy 1993. *The MacArthur-Bates Communicative Development Inventories: User's guide and technical manual*. Baltimore: Paul H. Brookes Publishing Company.
- Fenson, Larry; Dale, Philip; Reznick, Steven; Bates, Elisabeth; Thal, Donna; Pethick, Stephen 1994. *Variability in Early Communicative Development*. Monographs of the Society for Research of Child Development, Serial No. 242, Vol. 59 (5).
- Genter, Dedre 1982. Why nouns are learned before verbs. Linguistic relativity versus natural partitioning. – S. A. Kuczaj (Ed.). *Language Development*. Vol. 2: Language, Thought and Culture. Hillsdale, NJ: Lawrence Erlbaum, 301–334.
- Gentner, Dedre; Boroditsky, Lera 2001. Individuation, and early word learning. – Melissa Bowerman, Stephen C. Levinson (Eds.). *Language Acquisition and Conceptual Development*. New York: Cambridge University Press, 215–256.
- Kaasik, Ülo; Tuldava, Juhan; Viilup, Astrid; Ääremaa, Kuldav 1977. Eesti tänapäeva ilukirjandusproosa autorikõne lekseemide sagedussõnastik. – *Keelestatistika* 2. TRÜ toimetised, vihik 413. Tartu, 5–140.
- Li, Hanhong; Fang, Alex C. 2011. Word frequency of the CHILDES corpus: Another perspective of child language features. – *ICAME Journal*, 35, 95–116.
- Lieven, Elena 2010. Input and first language acquisition: Evaluating the role of frequency. – *Lingua*, 120 (11), 2546–2556. <http://dx.doi.org/10.1016/j.lingua.2010.06.005>
- Nelson, Katherine 1973. *Structure and Strategy in Learning to Talk*. Monographs of the Society for Research of Child Development, Serial No. 149, Vol. 38.
- Redington, Martin; Chater, Nick; Finch, Steven 1998. Distributional information: A powerful cue for acquiring syntactic categories. – *Cognitive Science*, 22 (4), 425–469. [http://dx.doi.org/10.1016/S0364-0213\(99\)80046-9](http://dx.doi.org/10.1016/S0364-0213(99)80046-9)
- Sandhofer, Catherine M.; Smith, Linda B.; Luo, Jun 2000. Counting nouns and verbs in the input: Differential frequencies, different kinds on learning. – *Journal of Child Language*, 27 (3), 561–585.
- Schults, Astra; Tulviste, Tiia; Konstabel, Kenn 2012. Early vocabulary and gestures in Estonian children. – *Journal of Child Language*, 39 (3), 664–686. <http://dx.doi.org/10.1017/S0305000911000225>
- Stoll, Sabine; Biskel, Benjamin; Lieven, Elena; Banjade, Goma; Bhatta, Toya N.; Gaenszle, Martin; Paudyal, Netra P.; Pettigrew, Judith; Rai, Ichchha Purna; Rai, Manoj; Rai, Novel Kishore 2012. Nouns and verbs in Chintang: Children's usage and surrounding adult speech. – *Journal of Child Language*, 39 (2), 284–321. <http://dx.doi.org/10.1017/S0305000911000080>
- Tulviste, Tiia 2007. Variation in vocabulary development among Estonian children as a function of child's gender, birth order, child-care, and parental education. – M. Eriksson (Ed.). *Proceedings from the First European Network Meeting on the Communicative Development Inventories*. Gävle, Sweden: University of Gävle, 16–21.
- Uiboaed, Kristel; Lindström, Liina; Muischnek, Kadri 2013. Frequencies of verbs and verbal constructions across Estonian dialects and standard language. – *Ettekanne, The relative frequencies of nouns, pronouns, and verbs in discourse. An international workshop in Leipzig*, 12.–13. august 2013.
- Vihman, Marilyn May; Vija, Maigi 2006. The acquisition of verbal inflection in Estonian. – Natalia Gagarina, Insa Gülzow (Eds.). *The Acquisition of Verbs and their Grammar: The Effect of Particular Languages*. *Studies in Theoretical Psycholinguistics* 33. Dordrecht: Springer, 263–295.
- Viks, Ülle; Hein, Indrek 2001. *Sõnavormide kasutus seadusetekestides*. Tallinn: Eesti Keele Instituut. http://www.eki.ee/teemad/seadused_kategooriad/kategooriad.pdf (30.9.2013).

WORD CLASSES IN THE CHILD'S SPEECH AND IN THE CHILD-DIRECTED SPEECH

Reili Argus, Helen Kõrgesaar

Tallinn University

This paper discusses the early phases of the acquisition of Estonian. The longitudinal data (recordings of spontaneous speech, 16 hours) of two Estonian typically-developing monolingual children aged 1;3–2;3 and their caregivers was analyzed and compared in relation to frequency and distribution of different word classes and complexity of verbs.

The main aim is to specify how this distribution changes when the child gets older and what kind of effect has the usage of verbs with different structure in the input to the child language.

Distributional information about the occurrence of different word classes shows that in the beginning of the observation period the nouns were the most frequently used word class in children's speech. At the same time, the speech of adults consists of approximately same amount of nouns and verbs. The frequency of usage of different word classes increases during the observation period: in the middle of their third year of life children start to use more verbs, adverbs, conjunctions, adjectives and adpositions, while the usage of verbs and adverbs increases remarkably during the observation period.

The distribution of different word classes in the input does not change significantly during the observation period. Still the usage of verbs with different complexity (root verbs, derived verbs, phrasal verbs and serial verb constructions) increases in the speech of adults and this reflects also in the speech of children. At the end of the observation period more verbs with complex structure (derived verbs and phrasal verbs) were used by both children and the bigger usage of phrasal verbs by one adult can be considered to be reflecting also in the speech of the child.

Keywords: first language acquisition, word classes, child-directed speech, child's speech, Estonian

Reili Arguse (Tallinna Ülikool) uurimisvaldkonnaks on esimese keele omandamine, huvikeskmes on siiani olnud muetomorfoloogia ja leksikaal-grammatiliste kategooriate omandamisega seonduv.
Narva mnt 25, 10120 Tallinn, Estonia
reili.argus@tlu.ee

Helen Kõrgesaar (Tallinna Ülikool) uurib eesti hoidjakeelt, selle dünaamikat ja mõju lapse kõne arengule.
Narva mnt 25, 10120 Tallinn, Estonia
helen.korgesaar@gmail.com

MORFOSÜNTAKTILISE JA LEKSIKAALSE VARIEERUMISE PIIRIDEST: ILUKIRJANDUS- JA ÕPPIJAKEELE KASUTUSMUSTRITE VÕRDLU

Pille Eslon

Ülevaade. Artiklis avaldatud klasteranalüüsi tulemused on saanud eesti ilukirjandustekstide ja õppija kirjutatud loomingulist laadi esseede keelekasutusmustrite võrdlemisel, aluseks formaalgrammatilise ja funktsionaalse lähenemise sidusus kasutuspõhises keelekäsitluses. Ühelt poolt eristatakse keelt kui hierarhilist süsteemi ja teisalt keelekasutust kui sõnade ja vormide koosinemist tekstis, kus keeleüksused on seotud kindlat liiki leksikaalsete ning morfosüntaktiliste piirangute ja varieerumise piiridega. Tekstide automaatne töötlus toob esile keelekasutusmustrid, võimaldades teha mh järeltõlki nii emakeelekõneleja kui ka õppija sõna- ja vormieelistuste ratsionaalsuse kohta.*

Võtmesõnad: kasutuspõhine keelekäsitlus, teksti automaatne töötlus, lingvistiline klasteranalüüs, morfosüntaks, õppijakeel, eesti keel

1. Uurimuse lähted

Uurimus tugineb ühelt poolt Saussure'i dihhotoomiate dialektilisele loogikale ja teisalt formaalgrammatilise ning funktsionaalse lähenemise ühtsusele kasutuspõhises keelekäsitluses (nt Barlow, Kemmer 2000: 7–28, Bybee 2007, Bybee, Eddington 2006, Bybee, Hopper 2001, Verhagen 2009: 119–152, Croft 2003, Goldberg 2006). Ka eesti keele uurimustes on lähtutud konstruktsioonigrammatikast ja kasutuspõhisest keelekäsitlusest (Sahkai 2008, Penjam 2008, Veismann 2009, Klavan 2012, Uibo 2013 jt).

Käesoleva uurimuse olulised üldmõisted on invariant, sagedus ja keele funktsionaalne potentsiaal, mis kerkivad esile vajadusest eristada või integreerida grammatikat ja keelekasutust.

Grammatika alla kuulub morfoloogiliste paradigmade ja kategooriate süsteemne esitus. Kaasaegne muutemorfologia mudel on mitmemõõtmeline: tugineb

* Tööd on toetanud riikliku programmi "Eesti keel ja kultuurimälu (2009–2013)" projekt "REKKi käsikirjaliste materjalide digiteerimine, Eesti vahekeele korpuse alamkorpuste loomine ja korpuse kasutusvõimaluste populariseerimine (2009–2013)"; ETF-i grant 8222 "Ülekantud tähenduses fraasid eesti keele korpustes (2010–2013)" ja ETF-i grant 8240 "Lähtekeele mõju ja teise keele omandamine: korpuspõhine uurimus (2010–2013)".

erinevate tunnuste ristumisele (nt Roman Jakobsoni universaalne verbikategoriate kirjeldus), binaarsetele tunnustele (nt Nikolai Trubetzkoy fonoloogiasüsteem), algoritmilisele esitusele, kus iga järgnev tunnustepaar tuleneb eelneva(te)st, võimaldades kirjeldada grammatikaseoseid ja vormide funktsionaalset potentsiaali integreeritult ning tuua esile oluliste/mitteoluliste tähenduste eristamisele tuginev semantiline võrgustik, mille alusel üldistuvad invariantid tähendused (nt Mihhail Šeljagini vene verbi integreeritud kirjeldus).

Grammatikas on vormi invariantseks tähenduseks üldine tunnus, mis seda vormi kõigist teistest vormidest eristab ja samas nendega ühendab. Bondarko väidab, et Jakobson on avanud invariantseuse mõiste just selles aspektis ehk grammatikasüsteemi keskselt, samas aga üldistuvad invariantid tähendused tekstikasutuse põhjal, vormi funktsionaalne potentsiaal avaneb. (Bondarko 2002: 159–163 jj) Grammatikas ning tekstikasutuses regulaarselt ilmnevate kategoriaalsete seoste ja semantika vahel on kindlat liiki piirangud, mis eksplitseeruvad erinevates kombinatsioonides ja erineva tekstisagedusega. See, mis on invariant ühtedes seostes, võib teistes olla variant või fakultatiivne nähtus – piir keelesüsteemi ning keelekasutuse vahel on olemuselt hajus.

Kui keele kategoriaalseid seoseid ja vormistikku saab esitada erinevate hierarhiiliste tunnuste ristumisele ning loogilisele vastandatusele rajaneva paradigmaatilise mudelina, seletada grammatikareeglitega, siis vormide funktsionaalne potentsiaal avaneb keeleüksuste lineaarse kooskasutuse tingimustes, kus grammatikareeglid üksüheselt ei rakendu. Nt ülekantud tähenduses vormikasutust ja ekspressiivsete värvingute lisandumist ei saa seletada morfoloogiareeglitega, küll aga on abi tekstikasutuse morfosüntaktiliste, semantiliste ja pragmaatiliste piirangute tundmisest. Need näitavad elementide kombineerumisvõimalusi ja regulaarsust tekstis (dominantne, sage, harv, kasutamata). On põhimõtteline erinevus, kas lingvistiline kirjeldus rajaneb semantiliste tunnuste loogilisel hierarhial ja kategoriaalsetel seostel (grammatikakirjeldus) või lingvistilistel tervikobjektidel, millel on kindlad leksikaalse ja morfosüntaktilise varieerumise piirid ning sagedus (kasutus põhine keelekirjeldus).

Sageduse puhul eristatakse sõnastiku- ja tekstisagedust, mõlemad on kesksel kohal semantika ja grammatika integreerimisel, omades seetõttu teoreetilist väärtust (Greenberg 1990: 219–220). Käesolevas uurimuses on vaadeldud keeleüksuste tekstisagedust. See on samasugune üldmõiste kui invariant grammatikas: mõlemat saab võrdväärselt rakendada nii vormide kui ka nende funktsionaalse potentsiaali avanemisel. Tekstikasutuses sagedased struktuurid on seotud kindlate tekstiliste funktsioonidega. Selge dominandi ehk invarianti puudumisel võivad selle staatusele pretendeerida ka nt kaks ja enam keelekasutusmustrit.

Haspelmath (1999, 2007) on märkinud, et keelekasutusest lähtuvad uurimused vajavad teooriat, mis seletaks, mille põhjal sõnu ja vorme valitakse ning reastatakse; mida tekstikasutuses eelistatakse kombineerida ja kuidas seda tehakse; missugused leksikaalsed ja morfosüntaktilised piirangud või varieerumise võimalused seejuures avanevad. Croft (1999: 206) on pidanud oluliseks süntagmaatilise kombineerimise printsiipi, mis tema arvates võimaldavad Haspelmathil integreerida formaalgrammatilise ja funktsionaalse keelekäsitluse. Küsimus on selles, mis tingib keelekasutaja valikuid, missugust kooskasutust peetakse loomulikuks. Ellise (2006) arvates seostub loomulikkus valikute ratsionaalsuse ja optimaalsusega, mh ka keele valdamise tasemega.

Siinses uurimuses on tuginetud eespool sõnastatud arusaamale üldmõistete osast keele elementide funktsionaalse potentsiaali avanemisel ja tekstikasutuses oluliste tervikobjektide kirjeldamisel.

2. Uurimuse eesmärk, objekt ja meetod

Avaram eesmärk on seotud keele elementide funktsionaalse potentsiaaliga eesti kirjakeeles, mis väljendub morfoloogilistes, morfosüntaktilistes ja leksikaalsetes piirangutes ning varieerumise piirides. Kitsam eesmärk on keelekasutusmuustrite automaatne tuvastamine, et välja tuua emakeelekõneleja ja õppija tekstiloomed morfoloogilised, morfosüntaktilised ja leksikaalsed eelistused ning neid omavahel võrrelda. Rakenduslik olulisus seisneb võimaluses näidata, kui võrd adekvaatne või nihkes on õppija tekstiloomed võrreldes emakeelekõneleja omaste leksikaalsete ja morfosüntaktiliste valikutega.

Vaadeldud ei ole kõiki tekstikasutusi: uurimisobjektiks on verbist vasakule jääv kontekst (nt *metsa vari ulatus, veel ei liikunud, ise pidid tegema, kord mul oli*). Ilukirjandustekstid (umbes 200 000 sõnet)¹ ja õppijakeele esseed (umbes 500 000 sõnet)² sisestati Sander Otsa programmeeritud klastrileidjasse, mis töötab andmekaeve põhimõttel, otsides valimitest sageduse alusel sarnaseid morfo- ja süntaksimärgendite lineaarseid järjendeid ehk n-gramme. Klastrileidja koostamisel on olnud eeskujuks WordSmith Tools, kfNgram ja Xaira (vt Ots 2012: 6, 16). Käesoleva uurimuse tarvis on valitud trigrammid, mis korduvad tekstis viis ja enam korda. Samalaadselt vormistatud trigrammide kogumeid on autor veidi tinglikult nimetanud lingvistilisteks klastriteks ja meetodit lingvistiliseks klasteranalüüsiks.³ Kuigi võrreldavate valimite maht erineb, ei ole vormide esinemus korrelatsioonis sõnede arvuga (vt Biber jt 1998: 33) – erineva mahuga valimid on võrreldavad vormide osakaalu põhjal (vt tabel 1).

Klasteranalüüsi tulemusena leiti ilukirjandustekstide valimist kokku 904 trigrammi, mis jagunevad 18 klastri ja 6 klassi vahel; õppijakeele valimist 13941 trigrammi, mis jagunevad 38 klastri ja 8 klassi vahel (vt tabel 1). Tegu on hierarhilise jaotusega, mille aluseks trigrammi esimese ja viimase komponendi sõnaliigiline samalaadsus/erinemine ning keskmise komponendi varieerumine. Nt substantiiviga (S) algavate ja verbiga (V) lõppevate trigrammide keskmise komponendi sõnaliigilise varieerumise põhjal eristuvad klastrid SKV ehk substantiiv + adpositsioon + verb (*maja juures on*), SDV ehk substantiiv + adverb + verb (*siida natuke kripeldas*), SSV ehk substantiiv + substantiiv + verb (*metsa vari ulatus*) jne.⁴ Need klastrid reastuvad sageduse alusel substantiivialgulisse morfoloogilisse klassi S–V. Tabelis 1 on näha, et klassis S–V on ilukirjanduskeeles viis ja õppijakeeles kuus klastrit. Trigrammid, mis ei alga substantiiviga, moodustavad keskmise komponendi sõnaliigilise varieerumise põhjal uusi klastreid, mis jagunevad omakorda uute morfoloogiliste klasside vahel. Seega on klass ülemmõiste, mis ühendab morfoloogilised klastrid nende sageduse ja osakaalu alusel, ning klaster ülemmõiste, mis ühendab trigrammid nende morfosüntaktilise sarnasuse alusel.

¹ Eesti keele süntaktilisel analüüsitud ja ühestatud ilukirjanduskorpus: http://lepo.it.da.ut.ee/~heli_u/SA/ (15.8.2013).

² Eesti vahekeele korpus: http://evkk.tlu.ee/Search/search_results.html (15.8.2013).

³ Sama programmi on eestii ilukirjanduskeele keelekasutusmuustrite uurimisel rakendanud J. Trainis ja K. Allkivi (2014, vt käesolev kogumik).

⁴ Juhised eesti keele morfo- ja süntaksimärgendite kohta vt Tartu Ülikooli arvutilingvistika töörühma kodulehel: math.ut.ee/~kaaili/Korpus/pindmine/labels.pdf (4.1.2013).

Artiklis võrreldakse ilukirjandus- (ILU) ja õppijakeele (Õ) morfoloogiliste klasside invariantsete klastrite invariantseid trigramme, kuna trigrammide tasandil avalduvad keelekasutuse leksikaalsed ja morfosüntaktilised piirangud, mis võimaldab siduda invariandi, sageduse ja osakaalu mõisted (invariant tuleb esile kõige sagedamini, selle osakaal valimis on kõige suurem, vt tabel 1).

Tabel 1. Morfoloogiliste klasside klastrivarieeruvus, sagedus ja osakaal. S – substantiiv, V – verb, D – adverb, P – pronoomen, A – adjektiiv, J – konjunktsioon, G – genitiivatribuut, K – adpositsioon

Ilukirjanduskeel				Õppijakeel			
Klassid	Klastrid	Sagedus	Osakaal	Klassid	Klastrid	Sagedus	Osakaal
D-V 412	DDV	206	50%	J-V 3590	JPV	1558	44%
	DVV	116	28%		JSV	977	27%
	DJV	69	17%		JVV	582	16%
	DSV	11	3%		JDV	473	13%
	DAV	10	2%				
J-V 230	JDV	114	50%	P-V 2803	PVV	1881	67%
	JPV	67	29%		PSV	558	20%
	JSV	49	21%		PDV	177	6%
			PKV		67	3%	
			PPV		64	2%	
S-V 150	SDV	45	30%	S-V 2430	SVV	1329	55%
	SVV	43	29%		SSV	412	17%
	SKV	29	19%		SKV	273	11%
	SSV	20	13%		SJV	204	8%
	SJV	13	9%		SPV	116	5%
P-V 53				V-V 2402	SDV	96	4%
	PDV	28	53%		VVV	1025	42%
	PVV	15	28%		VDV	576	24%
	PSV	10	19%		VJV	305	13%
			VPV		264	11%	
A-V 44				D-V 2169	VSV	211	9%
	ASV	44	100%		VAV	21	1%
					DVV	832	38%
					DPV	590	27%
					DDV	468	22%
V-V 15				A-V 347	DSV	224	10%
					DJV	42	2%
					DAV	13	1%
					ASV	308	89%
	VDV	15	100%	K-V 124	AVV	17	5%
					AJV	16	4%
					ADV	6	2%
					KSV	50	40%
					KVV	46	37%
			G-V 76		KJV	18	15%
					KDV	5	4%
					KPV	5	4%
			G-V 76		GSV	76	100%

Morfoloogiliste klasside ja klastrite võrdlus on siinses artiklis ülevaateline, tugineb klasside-klastrite osakaalu läheduse-erinemise-vastandlikkuse skaalale. Võrdlus- aluseks on ilukirjanduskeel.

3. Analüüsi tulemused

3.1. Ülevaade ilukirjandus- ja õppijakeele morfoloogilistest klassidest

Morfoloogiliste klasside võrdlus rajaneb nende osakaalule ILU ja Õ valimites (vt tabel 2). Emakeelekõneleja tekstiloomes piisab suhtlusvajadusteks kolmest morfoloogilisest klassist: tavapäraselt kasutatakse adverbialgulist (D–V); harvem konjunktsiooni- (J–V) ja substantiivialgulist (S–V) klassi. Pronoomeni- (P–V), adjektiivi- (A–V) ja verbialgulist (V–V) klasside osakaal on väike, adpositsiooni- (K–V) ja genitiivatribuudialgulist (G–V) klasside osakaal sisuliselt olematu.

Tabel 2. Morfoloogiliste klasside osakaal ilukirjandus- (ILU) ja õppijakeele (Õ) valimites

Korpus	Morfoloogilised klassid							
	D–V	J–V	S–V	P–V	A–V	V–V	K–V	G–V
ILU	45%	25%	17%	6%	5%	2%	0%	0%
Õ	16%	26%	17%	20%	2%	17%	1%	1%

Õppija tekstiloomes on morfoloogiliste klasside varieerumine avaram, erinevate klasside osakaal ühtlasem. Eelistatakse konjunktsiooni- (J–V) ja pronomeni- algulist (P–V) klassi; järgnevad substantiivi- (S–V), verbi- (V–V) ja adverbialguline (D–V). Seega on emakeelekõneleja ja õppija morfoloogilised valikud erinevad, millest annab täpsema ettekujutuse morfoloogiliste klasside klastrivarieeruvuse võrdlus.

3.2. Ülevaade ilukirjandus- ja õppijakeele klastrivarieeruvusest

Klasside morfoloogilise varieerumise piirid tulevad esile klastrite osakaalu läheduse-erinemise-vastandlikkuse skaalal (täpsem kirjeldus Eslon 2013). ILU ja Õ võrdluse tulemused näitavad nii sarnasust kui ka olulisi erinevusi. Nt paigutub substantiivi- ja pronomeni- algulist klasside S–V ning P–V klastrikasutus läheduse-erinemise skaalale; konjunktsioonialgulist klassi J–V klastrikasutus on erinev; adjektiivialgulist klassi A–V puhul vastandlik; verbialgulist klassi V–V kasutus on täiesti vastandlik; adverbialgulist klassi D–V klastrikasutus aga oluliselt vastandlik. Tuleb rõhutada, et need erinevused ei tähista mittestandardset, vaid nihestunud keelekasutust, mis ei sega suhtlust, kuid on emakeelekõneleja seisukohast ebaratsionaalne.

ILU morfoloogiliste klastrite kasutust iseloomustab rangepiirilisus. Õ klastrivarieeruvus on avaram, morfoloogiliselt mitmekesisem, samas ei saa seda tõlgendada morfoloogilise keerukusena. Pigem on see märk õppija vähesest oskusest kasutada keelt sama ratsionaalselt kui emakeelekõneleja.

3.3. Invariantsete trigrammide morfosüntaktiline ja leksikaalne varieerumine

Invariantsete trigrammide võrdlus tugineb viie invariantse klasteri vastandusele (v.a ASV): DDV vs. DVV, JDV vs. JPV, SDV vs. SVV, PDV vs. PVV, VDV vs. VVV. Nagu klasside ja klasterite ülevaates, nii on ka invariantsete trigrammide funktsionaalse potentsiaali kirjeldamisel aluseks ilukirjanduskeel (vt tabelid 3 ja 4). Neis tulevad selgelt esile emakeelekõnelejale ja õppijale omased valikud (ILU adverbialgulised või -kesksed morfoloogilised struktuurid; Õ verbi- ja pronoomenikesksed struktuurid), nt DDV vs. DVV (*väga tublisti kosunud vs. sageli ei vali*); JDV vs. JPV (*ja kindlasti oli vs. et see on*); SDV vs. SVV (*süda natuke kripeldas, vee peale valanud vs. riik ei hakka*); PDV vs. PVV (*ma lihtsalt ei vs. ma ei saa*); VDV vs. VVV (*oli seni olnud vs. ei saa elada*).

Tabel 3. Eesti ilukirjanduskeele invariantsete klasterid ja trigrammid

Klass	Invariantne klaster	Invariantne trigramm	Näited
D-V	DDV	_D_// @ADVL _D_// @ADVL _V_ main partic past ps #Intr // @-FMV	<i>kuskile ära sõitnud, väga tublisti kosunud</i>
J-V	JDV	_J_crd // **CLB @J _D_// @ADVL _V_ main indic impf ps3 sg ps af #FinV #Intr // @+FMV	<i>ja kindlasti oli, ning siis kadus, kuid äkki tuli</i>
S-V	SDV	_S_com sg nom // @SUBJ _D_// @ADVL _V_ main indic impf ps3 sg ps af #FinV #Intr // @+FMV	<i>süda natuke kripeldas, põlv ebameeldivalt tuikas</i>
		_S_com sg gen // @OBJ _D_// @ADVL _V_ main partic past ps #NGP-P // @-FMV	<i>vee peale valanud, mehe üle võtnud</i>
P-V	PDV	_P_pers ps 1 sg nom // @SUBJ _D_// @ADVL _V_ aux neg // @NEG	<i>ma lihtsalt ei, ma veel ei, ma ju ei</i>
A-V	ASV	_A_pos sg nom // @AN> _S_com sg nom // @SUBJ _V_ main indic impf ps3 sg ps af #FinV #Intr // @+FMV	<i>täielik hullumeelsus oli, vintske vanamees suri</i>
V-V	VDV	_V_ aux indic impf ps3 sg ps af #FinV #Intr // @+FCV _D_// @ADVL _V_ main partic past ps #Intr // @-FMV	<i>oli kokku tulnud, oli seni olnud</i>

Viie klasterivastanduse ja ASV invariantsete trigrammide võrdlusest saab ettekujutuse, kuidas funktsionaalne potentsiaal erinevates keelekasutusvariantides avaneb, missuguseid morfosüntaktilisi ja kombinatoorseid piiranguid tekstikasutuses esineb ning sõnavara ja semantiliste kitsendustega seostub.

Tabel 4. Eesti õppijakeele invariantseid klastrid ja trigrammid

Klass	Invariantne klaster	Invariantne trigramm	Näited
J-V	JPV	_J_sub // **CLB @J _P_dem sg nom // @SUBJ _V_main indic pres ps3 sg ps af #FinV #Intr // @+FMV	<i>et see on, sest see toob</i>
P-V	PVV	_P_pers ps1 sg nom // @SUBJ _V_aux neg // @NEG _V_main indic pres ps neg #FinV #NGP-P // @+FMV	<i>ma ei saa, ma ei taha</i>
S-V	SVV	_S_com sg nom // @SUBJ _V_aux neg // @NEG _V_main indic pres ps neg #FinV // @+FMV	<i>tunne ei ole, riik ei hakka</i>
V-V	VVV	_V_aux neg // @NEG _V_main indic pres ps neg #FinV #NGP-P // @+FMV _V_main inf // @-FMV	<i>ei saa elada (peita, unustada, teha, töötada)</i>
D-V	DVV	_D_ // @ADVL _V_aux neg // @NEG _V_main indic pres ps neg #FinV #NGP-P // @+FMV	<i>sageli ei vali, kohe ei saa</i>
A-V	ASV	_A_pos sg nom // @AN> _S_com sg nom // @SUBJ _V_main indic pres ps3 sg ps af #FinV #Intr // @+FMV	<i>täpne koht on, tänane põlvkond on</i>

Invariantsete trigrammide võrdlus toob esile emakeelekõneleja ja õppija keelekasutusele omaseid valikuid. Kirjelduse keskmes on morfosüntaktilise ja leksikaalse varieerumise piirid.

3.3.1. Adverbiaalgulised klastrid DDV vs. DVV

Adverbiaalgulise klastri kasutus on vastandlik: adverbikesksele muustrile (ILU) vastandub verbikeskne (Õ), keerulisele morfosüntaktilisele ja rikkalikule leksikaalsele varieerumisele (ILU) piiratud vormi- ja sõnakasutus (Õ).

- 1) Emakeelekõneleja eelistab jaatavat kõnet: kaks leksikaalselt varieeruvat adverbi + mineviku liitaja vormi kuuluv *nud*-partitsiip (*seal varemgi olnud, väga tublisti kosunud*). Õppija seevastu eelistab eitavat kõnet: adverb + eituspartikkel + verbi finiitvorm (*sageli ei vali, kohe ei saa*);
- 2) ILU adverbide leksikaalne varieerumine on avaram, kui õppijakeeles. Samas on mõningad adverbid, mida eelistatakse korrata: ILU – trigrammi alguses *enam, nüüd, siis, mitte* ja keskel *põhjalikult, täpselt, õigeaegselt* (enamasti modaalpartikli funktsioonis); Õ – trigrammi alguses (*üldse, üldiselt, ka, kunagi, kindlasti, veel, isegi, lihtsalt*). Esile tuleb ka õppija tüüpviga **juba* <*enam*>, kui adverbide *juba* ja *enam* valikul pole mõistetud vaatepunktide erinevust;
- 3) ILU verbi leksikaalne varieerumine on rikkalik, eriti analüütiliste verbide puhul. Enamus neist on ainukordsed ühendverbid (nt *ette tulema (teatama)*, *edasi minema (tormama)*, *järele kuulama (hõikama)*, *kinni mähkima*). Õppija on aga eelistanud piiratud hulka sageli korduvaid sünteetilisi verbe (*saama, ütleva, teadma, unustama, mõtlema, meeldima, ilmuma, esinema*);

- 4) ILU verbivormid on keeruka morfosüntaktilise varieerumisega: kõige enam kasutatakse *nud-* ja *tud-*partitsiipe (valdavalt perfekti koosseisus); *da-*infinitiiv on tavaline kas infiniitse öeldise või objekti funktsioonis; finiitverbi vormivalik on piiratud ainsuse 3. pöördega imperfektis ja preesensis; *ma-*infinitiivi on kasutatud infiniitse öeldisena.

Õ verbi ajavormide valik on leksikaalselt piiratud: eksistentsiaalverb *olema* ja modaalverb *saama* eitava kõne indikatiivi preesensis (*ei ole, ei saa, ei ütle, ei tea*), harvem imperfektis (*ei olnud*). Tegu on tüüpilise trigrammiga (*üldiselt, ka, kunagi, kindlasti, veel*) *ei ole / ei saa ~ ei olnud*.

3.3.2. Konjunktsioonialgulisel klastrid JDV vs. JPV

Konjunktsioonialgulistel klastrite invariantsetel trigrammide võrdlus toob esile ühelt poolt vastanduse adverbikesksus (ILU) – pronoomenikesksus (Õ) ja teisalt trigrammide leksikaalse samalaadsuse (*olema*-verbi eelistamine).

- 1) ILU klaster JDV invariantsetes trigrammis on koordineerivat sidesõna kasutatud võrdselt nii lause kui ka osalause alguses, järgnevad adverb ja finiitverb öeldise funktsioonis (*ja kindlasti oli, ning siis kadus, Kuid kahtlemata oli, Aga ilmselt jäi*). Õ klaster JPV invariantne trigramm on osalause alguses: subordineeriv sidesõna + demonstratiivpronoomen *see* subjekti funktsioonis + finiitverbi öeldisena (*et see on, et see viib, sest see annab*). Seega eelistab emakeelekõneleja koordinatsiooniseost lause ja osalause alguses, õppija subordinatsiooniseost osalause alguses;
- 2) ILU invariantse trigrammi leksikaalne varieerumine on piiratud. Sagedaimad on koordineerivad sidesõnad *ja, ning, aga, ent, või, kuid* ja adverb *siis*, ent samas leidus rohkesti ka ainukordseid adverbe (nt *ilmselt, kahtlemata, kindlasti*). Õ invariantsetes trigrammis on kasutatud subordinatsiooni (valdavalt *et*). Verbidest on mõlemas keelekasutusvariandis sagedaim *olema*, ülejäänud väike hulk verbe on ainukordsed (nt ILU: *tuli, meenus, astus, ilmus*; Õ: *aitab, võib, jääb, annab, võimaldab, teeb*);
- 3) ILU sagedaim verbivorm on imperfekti ainsuse 3. pööre, järgnevad preesensi ainsuse 3. pööre ja imperfekti mitmuse 3. pööre (*ja siis oli/on*). Vähesel määral esines ka eitava kõne näiteid (*aga järele ei, ent tookord ei*). Õ sagedaim verbivorm on preesensi ainsuse 3. pööre, harva imperfekti mitmuse 3. pööre (*et see on/oli*). Seega emakeelekõneleja eelistab *olema*-verbi imperfektis ja õppija preesensis;
- 4) ILU invariantne trigramm on morfosüntaktiline tervikobjekt, mida struktuurianalooogia põhjal on lihtne produtseerida, eriti eitavas kõnes, kus trigrammi kõik kolm komponenti on muutumatud sõnad (*ent tookord ei*). Tegu on pealtnäha lihtsa keelestruktuuriga, millel on oma osa tekstilise terviku loomisel. Ent kui arvestada, et eituspartikkel kuulub siin aktiivi/passiivi mineviku liitaja koosseisu, siis pole see morfosüntaktiline struktuur sugugi lihtne (vrd *ent tookord ei tulnud* mida teha?; *ent tookord ei tulnud* + adverb). Sellele vastupidiselt on Õ invariantse trigrammi näol tegu stereotüübiga *et see on (oli)* ja *et (ja) see ei*, millele järgnevas kontekstis on öeldistäide (tavaliselt lekseem *hea*).

3.3.3. Substantiivialgulised klastrid SDV vs. SVV

Substantiivialgulist klastrit on kasutatud erinevalt: ILU jätkuvalt adverbikeskne, Õ verbikeskne ja eitavat kõnet eelistav.

- 1) ILU klastri SDV invariantse trigrammi staatust jagavad kaks morfosüntaktistliku struktuuri: subjektialguline (*süda natuke kripeldas, põlv ebameeldivalt tuikas*) ja objektialguline (*šokolaadi ära söönud, mehe üle võtnud*). Õ klastri SVV invariantne trigramm on eitavas kõnes (*tunne ei ole, riik ei hakka*);
- 2) ILU objektiga algava trigrammi leksikaalne varieerumine on avar: nii substantiive kui ka analüütilisi verbe on kasutatud ainukordsetena, v.a üksikjuhtumid (nt *suu(dki) lahti teha, ukse lahti teha, tervis tagasi tuli*). Seevastu Õ klastri SVV invariantse trigrammis on tüüpsõnavara: substantiivid *inimene, tulemus, baar, mobiiltelefon, probleem, eestlane, keel, töö, raha* ja verbid *olema, saama, tohtima, sobima, unustama, jooma*. Ainukordsete ja kaks korda kasutatud lekseemide osakaal on kesine;
- 3) ILU invariantse trigrammis varieerub substantiivi käänne ja süntaktiline funktsioon: totaalobjekt ainsuse genitiivis (*inimese ära tapnud, hobuse siit leidnud*), harva partsiaalobjekt ainsuse partitiivis (*meest ärkvele raputada, teekonnaajuttu peale suruda*); subjekt ainsuse nominatiivis (*mõte ainult vilksatas, talumees muudkui muheles*). Totaalobjektile järgneva analüütilise verbi vorm varieerub morfosüntaktiliselt: mineviku liitaja koosseisu kuuluv *nud*-partitsiip, *da*-infinitiiv. Subjektiga algavates konstruktsioonides on finiiitverb imperfekti ainsuse 3. pöördes, variandidid on *ma*-infinitiiv öeldise funktsioonis (<*tundus*> *aeg möödab olema*, <*hakkas*> *sisemus isemoodi sügelema*) ja eitav kõne (*jutt enam ei <sobinud>, ema kergelt ei <andesta>*). Substantiiv ainsuse komitatiivis on käändsõnaline viisi- või kaasnevusmäärus (*kiiruga alla jooksnud, vestlusega kaasa läinud*).

ILU klastri SDV invariantsete trigrammide keeruline morfosüntaktiline varieerumine avab seoseid, mida eesti keele funktsionaalne potentsiaal võimaldab: substantiiv objekti, subjekti ja käändelise määruse funktsioonis, millest sõltub öeldisverbi valik. Ilmneb seaduspärasus: totaalobjekt + analüütilise verbi vorm; subjekt + sünteetilise verbi vorm. Keerukust lisab sõnajärg, kuna adverbi kasutatakse subjekti/objekti ja predikaadi vahel. Verbi ja substantiivi avar leksikaalne varieerumine ning idiomaatilist laadi analüütiliste verbide kasutamine viitab sõnavara rikkusele. Struktuur SDV on eesti keeles morfosüntaktiliselt keerukam ja leksikaalselt varieeruvam kui DDV.

Õ klastri SVV invariantse trigrammi kasutuses ilmnevad piirangud substantiivi funktsionaalse potentsiaali kasutamisel: subjekt on tavaliselt ainsuse (*tunne ei ole, riik ei hakka*), harvem mitmuse nominatiivis (*inimesed ei saa, noored ei tea*), semantiliselt käänded väljendavad määruslikke tähendusi (ainsuse inessiiv – kohamäärus, ainsuse adessiiv – valdajamäärus; ajamäärus esines harvem). Verbi kasutatakse eitava kõne preesensis (*sõber ei joonud, kirik ei olnud*), mitmusliku subjektiga ka imperfektis. Trigrammi vähene morfosüntaktiline varieerumine ja tüüpsõnavara toovad esile stereotüübi *inimene ei ole/ei saa*. Seega ILU trigrammide morfosüntaktilisele keerukusele ja leksikaalsele rikkusele vastandub Õ stereotüüpne keelekasutusmuster.

3.3.4. Pronoomenialgulised klastrid PDV vs. PVV

Pronoomenialguline klaster näitab ILU adverbi- ja Õ verbikesksust. Ühisjooned on eitava kõne eelistamine, personaalpronoomen *ma* subjekti funktsioonis ja kitsa leksikaalse varieeruvusega morfosüntaktiline struktuur.

- 1) ILU klatri PDV invariantne trigramm on ainsuse 1. isiku personaalpronoomeni lühivorm subjekti funktsioonis + adverb + eituspartikkel (*ma lihtsalt ei, ma veel ei, ma ju ei*); Õ klatri PVV invariantne trigramm on ainsuse 1. isiku personaalpronoomeni lühivorm subjekti funktsioonis + eituspartikkel + liitpredikaati kuuluv modaalne abiverb (*ma ei saa, ma ei taha*);
- 2) pealtnäha on ILU invariantne trigramm lihtne, kuid eituspartiklile järgnev kontekst näitab, et tavaliselt on öeldiseks eitava kõne imperfekt või preesens (*ma lihtsalt ei <saanud, saa> + infinitiiv või ma veel ei <teadnud, tea>*). Trigrammi keskmise komponendi leksikaalne varieerumine on piiratud, lekseemide kasutussagedus suur. Järelikult peab trigramm esile tooma adverbide funktsionaalse potentsiaali. Neid kasutataksegi kas rõhu-, modaalpartikli, viisi-, aja- või hulgamäärusena (nt ajamäärus *parajasti*, modaalpartikkel *tõepoolest*, rõhupartikkel *ju*). Trigrammi esimese ja viimase komponendi muutumatu kasutus + piiratud hulk sageli kasutatud adverbide moodustavad morfosüntaktilise tervikobjekti, mida on lihtne struktuurianalooogia põhjal produtseerida ja kasutada tagasiviite vahendina.

Õ klatri PVV invariantse trigrammi alguses on samuti personaalpronoomen *ma*, kuid oluline erinevus ILU trigrammist lisaks verbikesksusele seisneb selleski, et subjekti funktsioonis tuleb laialdaselt esile erinevaid pronoomeni liike, eriti sageli demonstratiivpronoomen *see*. Järgneb eitava kõne preesens + infinitiiv. Verbide leksikaalne varieerumine on kitsas (*saama, tahtma, olema, harvem pidama, nägema*).

- 3) ILU invariantse trigrammi morfosüntaktiline varieerumine on seotud pronoomeni ja eitava/jaatava kõneliigi valikuga, jaatavas kõnes ka verbivormi varieerumisega. See on hea näide pealtnäha lihtsa konstruktsiooni järkjärguliselt keerukamaks muutumisest. Nii võib eitavas kõnes subjektina esineda lisaks ainsuse 1. isikule ka ainsuse 3. isik (*tema küll ei, ta siis ei*); jaatavas kõnes kasutatakse subjektina enamasti ainsuse 3. isikut ja finiiitverbi imperfekti ainsuse 3. pöördes (*ta äkitselt kadus*). Veelgi keerulisem varieerumise juhtum on indefiniitne pronoomen *miski* partsiaalobjektina koos analüütilise verbi *da*-infinitiiviga, kui trigrammile eelnevas kontekstis seisab modaalne abiverb (<*tuleb*> *midagi ette võtta*, <*vaja*> *midagi välja pressida*).

Õ klatri PVV invariantse trigrammi morfosüntaktiline varieerumine on dikteeritud kõneliigist: a) eitava kõne vormivalik toob esile seose modaalverbide tähenduse ja finiiitverbi kõneviisi-ajavormi valiku vahel: *saama, teadma* + indikatiivi preesens (*ma ei saa, ma ei tea*); *saama* + konditsionaali preesens (*ma ei saaks*) ja indikatiivi imperfekt (*ma ei saanud*); b) jaatavas kõnes domineerib konstruktsioon *tahtma* + *da*-infinitiiv totaalobjekti funktsioonis (*ma tahan õppida*), harva esines *ma*-infinitiiviga liitpredikaat (*ma pean ostma*). Esile on tulnud ka üks jaatava kõne veamuster, mis tingitud finiiitverbi semantikast tuleneva valikuga *da*- ja *ma*-infinitiivi

vahel (nt *ma pean *parandada* (**õppida*, **minna*, **lugeda*, **vaadata*), verbide häälduspärasest kirjaviisist (**suitsida*, **soitsida* <*suitsetada*>, **õpida* <*õppida*>) või isikupärasest sõnakasutusest (**praktida* <*praktiseerida*>).

3.3.5. Adjektiivialguline klaster ASV

Adjektiivialgulise klatri invariantseid trigramme on kasutatud vastandlikult, kuid nii ILU kui ka Õ invariantne trigramm on substantiivikeskne.

- 1) ILU klaster ASV koosneb vaid ühest trigrammist, mida on kasutatud nii lause sees kui ka alguses (*vintske vanamees suri*, *Liikumatu muie püsis*); Õ klaster sisaldab rohkem trigramme, kuid invariantne struktuur on sama (*täpne koht on*, *suur rühm on*, *suur palk sõltub*). Erinevused ilmnevad trigrammi komponentide leksikaalses ja morfosüntaktilises varieerumises;
- 2) kõik ILU adjektiivid ja substantiivid on ainukordsed, kuid verbikasutuses tuleb esile selge dominant, kuna pool näidetest sisaldab *olema*-verbi. Sama tendents ilmneb ka õppijakeeles, kus *olema*-verb esineb kahes kolmandikus näidetest. Seega ILU tüüpilisele leksikaalsele piirangule vastab Õ leksikaalne dominant;
- 3) ILU trigrammis järgneb adjektiivsele eestäiendile tavaliselt substantiiv ainsuse või mitmuse nominatiivis (subjekt) ja *olema*-verb imperfekti ainsuse või mitmuse 3. pöördes (*täielik hullumeelsus oli*, *Kuldpruunid palkseivad olid*). Variant: substantiiv ainsuse partitiivis (partiaalobjekt) + *da*-infinitiiv öeldisena (*verbaalset vastet leida*, *õiget naist saada*). Õ trigrammis on substantiiv enamasti ainsuse nominatiivis ja finiitverb preesensi ainsuse 3. pöördes (*täpne koht on*, *suur palk sõltub*). Esineb ka mitmeid teisi arvu- ja ajavormide kombinatsioone, kuid need on harvad. Kuigi emakeelekõneleja ja õppija kasutavad sama morfoloogilist struktuuri ja eelistavad *olema*-verbi, erinevad trigrammide leksikaalse varieerumise piirid ja morfosüntaktiline lihtsus/keerukus. ILU tekstikasutuses on tüüpiline *olema* imperfekti ainsuse 3. pöördes, Õ tekstikasutuse stereotüüp *olema* preesensi ainsuse 3. pöördes. Eakeelekõneleja jaoks on tavaline ka substantiiv partiaalobjektina + infiniitne öeldis, mis õppija keelekasutuses puudub.

3.3.6. Verbialgulised klastrid VDV vs. VVV

Verbialguliste klastrite kasutamine on täiesti vastandlik: ILU adverbi- ja Õ verbi-keskne, emakeelekõneleja eelistab jaatavat ja õppija eitavat kõnet.

- 1) ILU klatri VDV trigramm koosneb verbi pluskvamperfekti liitaja vormist, mille "lõhub" adverb (*oli kokku tulnud*, *oli seni olnud*). Tegu on eesti keelele omase sõnajärjega, mis võib tekitada probleeme nii keeleõppijale kui ka emakeelekõnelejale. Õ klatri VVV invariantne trigramm langeb kokku eitava kõne liitpredikaadiga, kus modaalsele abiverbile preesensis (harvem imperfektis, konditsionaali preesensis, väga harva impersonaalis) järgneb *da*-infinitiiv infiniitse öeldise funktsioonis, harvem totaalobjektina (*ei saa elada*, *ei tohi suitsetada*);

- 2) ILU trigrammis on kõiki adverbe ja verbe kasutatud ainukordselt. Õ trigrammis vastandub sellele tüüpsõnavara: preesensis on korratud seitset verbi (*saama, tohtima, harvem tahtma, oskama, paaril korral viitsima, suutma, armastama*), viimane komponent on *da*-infinitiiv, mis on küll avarama leksikaalse varieerumisega, kuid ka siin tuleb esile selge erinevus sageli (*teha, elada, öelda/ütelda, suitsetada*) ja vaid paar korda esinenud verbide vahel (*loobuda, istuda, olla*). Niisiis on invariantsete trigrammi-de leksikaalne varieerumine ILU ja Õ tekstides täiesti erinev: leksikaalne rikkus *vs.* tüüpsõnavara.

4. Arutelu

Eesti ilukirjandus- ja õppijakeele võrdleva analüüsi põhjal selgus, et emakeelekõneleja eelistab piiratud hulka sagedasi morfoloogilisi klastreid ning morfosüntaktilisi trigramme. Nende seas tulevad esile adverbialgulised ja -kesksed mustrid, kus adverbil on erinevaid tekstilisi funktsioone, mis viitab adverbi avarale funktsionaalsele potentsiaalile. Määruse funktsiooni taandumine tekstiliste funktsioonide ees on keeleomane nähtus, mida on seotud diakroonilise arengu üldisemate seaduspärasustega (vt Valdmets, Habicht 2013, Veismann 2009). Seetõttu on ka siinse arutelu keskmes kõik adverbiga seonduv.

Adverbi sisaldavad mustrid on üsna keerulised. Nii eesti kirjakeele⁵ kui ka siinses artiklis vaadeldud verbist vasakule jääva konteksti klasteranalüüs toob esile adverbiliste struktuuride tekstilised funktsioonid ja nende seose sõnajärgjega: lause alguses tagasiviite ja eelneva lause sisu konkretiseerimise või hinnangu vahend; lause sees modaal- või rõhupartikkel. Sõltuvalt funktsioonist võib adverb oma kohta lauses vabalt varieerida, paigutudes subjekti/objekti ja liitpredikaadi või verbi liitjavormi vahele. Seetõttu püüab õppija neid struktuure vältida, kasutades adverbi enamasti *aja-* ja *kohamäärusena*.

Konjunktsioonialguline adverbikeskne trigramm esineb emakeelekõneleja tekstides regulaarselt lause, harvem osalause alguses, õppijal aga osalause alguses. ILU trigrammis kasutatakse kuut koordineerivat sidesõna (sagedaim *ja*); adverbide valik on leksikaalselt piiratud (sagedaim *siis*), verbidest on sagedaim *olema* imperfekti ainsuse 3. pöördes (*ja siis oli*). Seda struktuuri on lihtne produtsseerida, verbi saab leksikaalselt vabalt varieerida, ent emakeelekõneleja pole seda võimalust eriti kasutanud. Ka õppijal on konjunktsioonialguline struktuur sage, kuid koordineerimisest välditakse, adverbi asendab verb, *olema* on preesensi ainsuse 3. pöördes (*et see on*). Erinevused on tingitud sellest, mis otstarbel kirjeldataud trigramme kasutatakse. Emakeelekõnelejal on see tekstilise terviku loomise vahend, kus järgnev lause täiendab informatiivselt eelneva sisu, õppijal aga vahend kõrvallause sidumiseks pealausega (põhjus-tagajärg seos). See seletab kummagi keelekasutusvariandi eelistusi (koordineatsioon *vs.* subordinatsioon ja tagasiviitevahendina sünonüümne asendus demonstratiivpronoomeniga *see*). Siin peitubki nihke olemus emakeelekõnelejaga võrreldes. Ning kui me tahame, et õppija oskaks kasutada adverbistrukture sama oskuslikult nagu emakeelekõneleja ning neid mitte vältida, siis tuleb õpetada, kuidas kõige ratsionaalsemal viisil sidusaid tekste

produtseerida, subjektiivseid hinnanguid ja muid informatsiooni aktualiseerimise vahendeid kasutada.

Siinkohal paar üldisemat märkust. Esiteks on adverbilised mustrid hea näide grammatika, semantika ja pragmaatika sidususest tekstis. Selle alusel kujunevad keeleüksuste lineaarse kooskasutuse piirangud ehk reeglid, samas kui eesti keele õpe tugineb jätkuvalt grammatika reeglitele. Sagedaste lingvistiliste tervikobjektide asemel õpetatakse vormidest ja sõnadest vaba valiku alusel moodustatud grammatilisi konstruktsioone. Uurimused on näidanud, et olenemata keelest moodustavad kahest-kolmest elemendist koosnevad sagedased terviküksused, nt vormelid (ingl *formulaic sequences*) keskmiselt pool teksti mahust, mistõttu piiratud sõna- ja vormivalikuga tekstilistel struktuuridel on kindel koht keele omandamisprotsessis. Need jäävad hästi meelde, neid loetakse ja produtseeritakse kiiremini. (Conklin, Schmitt 2008: 72–77) Kuid eesti keele õpikutes leidub niisuguseid struktuure vähe (Puhtvend 2009). Ei osata arvestada, et idiomaatilise printsiipi domineerib keeleüksuste vaba varieerumise üle, et idiomaatilist tähendust võivad omada “ka produktiivselt moodustatud väljendid” (Sahkai 2008: 171–172). Keeles on “süntaktilisi püsimalle, [---] [mida] on võimalik täita kümnete, sadade või tuhandete erinevate figuratiivsete vm leksikaalsete elementidega” (Krikmann 2004: 116). See, mis on invariant grammatikas ning mille alusel reegleid tuletatakse, ei ole tekstikasutuse invariant.

Teine üldisemat laadi märkus tuleneb eelnevast ja seostub keelekasutuse ratsionaalsusega. Siinse klasteranalüüsi tulemused näitavad, et emakeelekõneleja eelistab lihtsa struktuuriga morfosüntaktilisi terviküksusi, milles saab lekseme vabalt varieerida ning mida produtseeritakse struktuurianalooogia alusel. Just niisugune on adverbialguline invariantne klaster DDV, mille alusel saab genereerida ühendverbe. Järelikult on struktuuri DDV kasutusele mõttekas rajada ka eesti ühendverbide produtseerimis- ja kasutamissoetus. Nähtus on universaalsemat laadi, nt itaalia keeles käsitletakse iseseisvat tähendust omavate mitmesõnaliste predikatiivsete üksuste arengut semantika-süntaksi koostoimena (Masini 2005). Sama kehtib saksa keele partikkelverbide, vene ja inglise perifrastiliste ja fraasiverbide kohta. Analooiselt on võimalik kasutada ka substantiivseid trigramme, et genereerida väljendverbe. Mõistetavalt polnud verbist vasakule jäävas kontekstis nende kasutus sage, kuid verbist paremale jäävas kontekstis, prototüüpelt Ø-markeeritusega objektifraasis, ilmnevad need trigrammid regulaarselt (Esilon 2012: 23–26). Tuumverbid *tegema*, *andma*, *saama* ja *võtma* moodustavad koos elutu partitiivse objektiga nii idiomaatilisi terviküksuseid (*nalja tegema*, *nõu andma*, *asja saama*, *sõna võtma*) kui ka vabu ühendeid (*leiba tegema*, *eratunde andma*, *tööd saama*, *ravimeid võtma*), mida õppija on genereerinud normikohaselt. Sageli kasutatuna võivad ka vaba valiku alusel moodustatud ühenditest aja jooksul kujuneda idiomaatilised terviküksused, mis täiendavad eesti väljendverbide loendit. Ka see nähtus pole eesti keele keskne. Nt hispaania keeles genereeritakse *saama*-tähendusega verbe kooskasutuses adjektiiviga ja elusa subjektiga. Seda struktuuri rakendatakse võrdväärselt nii muutust tähistavate idiomaatiliste kui ka mitteidiomaatiliste terviküksuste produtseerimiseks (Bybee, Eddington 2006). Kirjeldatud morfosüntaktiliste struktuuride funktsionaalne potentsiaal avaneb tänu struktuurianalooogiale, mis on kõige loomulikum ja ratsionaalsem viis analüütiliste verbi(konstruktsiooni)de produtseerimiseks ning omandamiseks. Keele valdamiseks oluline teave ei pärine grammatikast, vaid piirangutest keeleüksuste lineaarses kooskasutuses.

Kuna teadmised eesti keele kasutusmustrite kohta ei ole piisavad, pole ka õpetamis- ja õppimisprotsess kuigi optimaalne, vaid pigem intuiitivne. See tingibki leksikaalsete ja morfosüntaktiliste stereotüüpide teket ning mittevajalikku variatiivsust. Nt õppijakeeles on ebaratsionaalselt avar pronoomenite varieerumine. Konjunktsioonialgulise klatri JPV invariantne trigramm sisaldab stereotüüpselt demonstratiivpronoomenit *see*, lisaks veel 43 sagedat trigrammi personaalpronoomeniga. Valdab *mina*-kesksus, õppija tekstikasutuses vastanduvad *ma-nad*, *ta-nad*. Eesti ilukirjanduskeeles ei ole *see* esile tulnud. Pronoomenialguline morfoloogiline klass on küll olemas, kuid selle osakaal jääb adverbiaalgulisele klassile tunduvalt alla (vt tabel 2). Samas eristub *mina*-kesksus nt vene- ja ingliskeelses tekstiloomes. Kuna eesti vahekeele korpuse loominguist laadi esseede autorid on valdavalt vene ja vähesel määral inglise emakeelega eesti keele õppijad, siis saab pronoomenikesksust seletada keelemeelisusega. Struktuuris JPV piirab pronoomen ka verbi leksikaalset ning morfosüntaktilist varieerumist: eitava kõne korral kasutatakse modaalverbide *saama* ja *tahtma* ning jaatavas kõnes *tahtma* finiiivormi. Koos ainsuse 1. isiku personaalpronoomeniga moodustub stereotüüp *ma ei saa*. Seega on õppija keelekasutuses pronoomenikesksusel otsene seos modaalsete hinnangute leksikaalse väljendamisega. Isegi ajakirjanduskeeles, kus hinnangulisus on informeerimise järel teine oluline pragmaatiline funktsioon, ei tule modaalkonstruksioonid niisuguse regulaarsusega esile kui õppijakeeles (Eslon 2009: 46). Samas on emakeelekõneleja hinnangute edastamisel stereotüüpselt trigrammile eelistanud leksikaalset ja morfosüntaktilist varieeruvust. Õppija lihtsalt ei valda seda keelelist mitmekesisust, olgugi et mõningate variatsioonide ja eripäraga on see olemas ka tema emakeeles.

Eesti keeles kõige keerulisema morfosüntaktilise varieerumisega klaster on SDV ja selle kaks invariantset trigrammi. Substantiivi on kasutatud nii objekti, subjekti kui ka käändelise määрусena, substantiivi funktsioonist oleneb verbi valik: totaalobjekt + analüütilise verbi vorm; subjekt + sünteetilise verbi vorm. Tegu on ratsionaalse kasutusreegliga, mida ei teata – osaliselt seetõttu kasutab õppija stereotüüpset eitava kõne konstruksiooni *inimene ei ole/ei saa*.

Eespool kirjeldatu on otseselt seotud vormikasutuse keerukuse/ratsionaalsuse dialektikaga. Morfoloogiliselt ja morfosüntaktiliselt varieerub ilukirjanduskeel vähem kui õppijakeel, eristudes viimasest vormikeerukuse ja sõnajärjeraskuse poolest. Emakeelekõneleja sõnastus on kahtlemata varieeruvam ja keerulisem, kuid tehtud valikute poolest ratsionaalsem. Dominantsete trigrammide kasutus klattrites ASV ja VDV toob esile korrelatsiooni leksikaalse ja morfosüntaktilise varieerumise vahel: maksimaalse leksikaalse varieerumisega kaasneb vähene morfosüntaktiline varieeruvus. Õppijakeeles pole korrelatsioon esile tulnud, vormide varieerumine võib küll olla avaram, kuid sõnavara ikkagi piiratud. Nihkeid ja nende tekkepõhjusti on erinevaid, kuid tekstide produtseerimisoskuse kujundamisel on oluline tugineda tekstikasutuses sagedate mustrite leksikaalgrammatilistele piirangutele, mitte grammatikale ja sõnastikule.

Viidatud kirjandus

- Barlow, Michael; Kemmer, Suzanne (Eds.) 2000. Usage-Based Models of Language. Stanford: CSLI.
- Biber, Douglas; Conrad, Susan; Reppen, Randi 1998. Corpus Linguistics: Investigating Language Structure and Use. Cambridge University Press. <http://dx.doi.org/10.1017/CBO9780511804489>
- Bondarko, Aleksandr 2002. Teorija znachenija v sisteme funkcional'noj grammatiki: Na materiale russkogo jazyka. [The Meaning Theory in System of the Functional Grammar: On the Russian Language.] Moskva: Jazyki slav'anskoj kul'tury, 159–356.
- Bybee, Joan; Hopper, Paul (Eds.) 2001. Frequency and the Emergence of Linguistic Structure. Amsterdam: John Benjamins. <http://dx.doi.org/10.1075/tsl.45>
- Bybee, Joan; Eddington, David 2006. A usage-based approach to Spanish verbs of 'becoming'. – Language, 82 (2), 323–355. <http://dx.doi.org/10.1353/lan.2006.0081>
- Bybee, Joan 2007. Frequency of Use and the Organization of Language. Oxford: Oxford University Press. <http://dx.doi.org/10.1093/acprof:oso/9780195301571.001.0001>
- Conklin, Kathy; Schmitt, Norbert 2008. Formulaic sequences: Are they processed more quickly than nonformulaic language by native and nonnative speakers? – Applied Linguistics, 29 (1), 72–89. <http://dx.doi.org/10.1093/applin/amm022>
- Croft, William 1999. Adaptation, optimality and diachrony. – Zeitschrift für Sprachwissenschaft, 18 (2), 206–208. <http://dx.doi.org/10.1515/zfs.1999.18.2.206>
- Croft, William 2003. Typology and Universals. Second edition. Cambridge: Cambridge University Press.
- Ellis, Nick C. 2006. Language acquisition as rational contingency learning. – Applied Linguistics, 27 (1), 1–24. <http://dx.doi.org/10.1093/applin/ami038>
- Eslon, Pille 2009. Eestikeelses tekstiloomes eelistatud konstruktsioonid ja käändevormid. [Contextual preferences the use of case forms in text production in Estonian.] – Korpusuuringute metodoloogia ja märgendamise probleemid. Tallinna Ülikooli eesti keele ja kultuuri instituudi toimetised 11. Tallinn: TLÜ Kirjastus, 30–53.
- Eslon, Pille 2012. Objekti ja tegevuse leksikaalne markeeritus eesti õppijakeeles. [The lexical markedness of object and action in Estonian learner language.] – Lähivõrdlusi. Lähivertailuja, 22, 15–42. <http://dx.doi.org/10.5128/LV22.01>
- Eslon, Pille 2013. Kahe keekekasutusvariandi võrdlus: morfoloogilised klassid ja klastrid. [The comparative study of language use: Morphological classes and clusters.] – Lähivõrdlusi. Lähivertailuja, 23, 13–38. <http://dx.doi.org/10.5128/LV23.01>
- Goldberg, Adele E. 2006. Constructions at Work. The Nature of Generalization in Language. Oxford: Oxford University Press.
- Greenberg, Joseph 1990. The relation of frequency to semantic feature in a case language (Russian). – Keith Denning, Suzanne Kemmer (Eds.). On Language. Selected writings of Joseph H. Greenberg. Stanford: Stanford University Press, 207–226.
- Haspelmath, Martin 1999. Optimality and diachronic adaptation. – Zeitschrift für Sprachwissenschaft, 18 (2), 180–205. <http://dx.doi.org/10.1515/zfs.1999.18.2.180>
- Haspelmath, Martin 2007. Creating economical morphosyntactic patterns in language change. – Jeff Good (Ed.). Language Universals and Language Change. Oxford: Oxford University Press. <http://email.eva.mpg.de/~haspelmt/CreatingEconomy.pdf> (3.2.2013).
- Klavan, Jane 2012. Evidence in Linguistics: Corpus-Linguistic and Experimental Methods for Studying Grammatical Synonymy. Dissertationes Linguisticae Universitatis Tartuensis 15. Tartu: University of Tartu.
- Krikmann, Arvo 2004. “Sai hea obaduse vastu obadust”: löömist ja peksmist märkivad väljendid eesti keeles. [Multi-word expressions denoting hitting in Estonian]. Reetor 3. Tartu: Eesti Kirjandusmuuseum Folkloristika osakond, Eesti Kultuuriloo ja Folkloristika Keskus.

- Masini, Francesca 2005. Multi-word expressions between syntax and the lexicon: the case of Italian verb-particle constructions. – *SKY Journal of Linguistics*, 18, 145–173.
- Ots, Sander 2012. Statistikapõhise tarkvara loomine morfoloogiliste kollokatsioonide eraldamiseks eesti keele tekstidest. [Software for morphosyntactic cluster extraction from Estonian texts.] Bakalaureusetöö. Tallinn: TLÜ informaatikainstituut.
- Penjam, Pille 2008. Eesti kirjakeele *da-* ja *ma-*infinitiiviga konstruktsioonid. [The constructions of *da-* and *ma-*infinitives in written Estonian.] *Dissertationes Philologiae Estonicae Universitatis Tartuensis* 23. Tartu: Tartu Ülikooli Kirjastus.
- Puhtvend, Kädi 2009. I–IV klassi laste kujundlik keelekasutus. Õpetamine ja oskused. [The use of figurative language in elementary school (1st–4th class): Teaching and skills.] Bakalaureusetöö. Käsikiri Tallinna Ülikooli eesti keele ja kultuuri instituudis.
- Sahkai, Heete 2008. Konstruktsioonipõhine keelemudel ja sõnaraamatumudel. [Some lexicographic implications of a construction-based model of language.] – *Eesti Rakenduslingvistika Ühingu aastaraamat*, 4, 177–186. <http://dx.doi.org/10.5128/ERYa4.11>
- Trainis, Jekaterina; Allkivi, Kais 2014. Ilukirjanduskeelest uue pilguga. [On belletristic language from a new perspective.] – *Eesti Rakenduslingvistika Ühingu aastaraamat*, 10, 283–306. <http://dx.doi.org/10.5128/ERYa10.18>
- Uiboed, Kristel 2013. Kollostruktsioonilised meetodid ja konstruktsioonilise varieerumise tuvastamine. [Collostructional methods and verb constructions in Estonian dialects.] – *Eesti ja soomeugri keeleteaduse ajakiri / The Journal of Estonian and Finno-Ugric Linguistics (ESUKA–JEFUL)*, 4 (1), 185–204.
- Valdmets, Annika 2011. Kahe eesti kirjakeele modaalpartikli arengust viimase sadakonna aasta jooksul. [On the development of two modal particles in written Estonian over the last century.] – *Keel ja Kirjandus*, 10, 764–776.
- Valdmets, Annika; Habicht, Külli 2013. Episteemilistest modaalpartiklitest eesti kirjakeeles. [About epistemic modal particles in Estonian.] – *Eesti ja soomeugri keeleteaduse ajakiri / The Journal of Estonian and Finno-Ugric Linguistics (ESUKA–JEFUL)*, 4 (1), 205–222.
- Veismann, Ann 200. Eesti keele kaas- ja määrsõnade semantika võimalusi. [Semantics of Estonian adpositions and adverbs.] *Dissertationes Linguisticae Universitatis Tartuensis* 11. Tartu: Tartu Ülikooli Kirjastus.
- Verhagen, Arie 2009. The conception of constructions as complex signs. Emergence of structure and reduction to usage. – *Constructions and Frames*, 1 (1), 119–152. <http://dx.doi.org/10.1075/cf.1.1.06ver>

Pille Esloni (Tallinna Ülikool) peamine uurimisvaldkond on morfosüntaks, eesti kirjakeele ja õppijakeele morfosüntaktiliste kasutusmustrite võrdlus.
Narva mnt 25, 10120 Tallinn, Estonia
peslon@tlu.ee

CONSTRAINTS ON MORPHOSYNTACTIC AND LEXICAL VARIABILITY

Pille Eslon

Tallinn University

This study uses cluster analysis in order to describe the functional mobility of a language system. It compares Estonian fiction and learner language texts with regard to the use of language preceding the verb (e.g. *ema lähemale tuli* ‘mother nearer came’, *kõhn keha vabises* ‘scrawny body trembled’, *elu ei ole* ‘life not is’, *ja kindlasti oli* ‘and certainly was’, *sest see aitab* ‘because it (that, this) helps’). The aim of this study is to show the constraints on morphosyntactic and lexical variability of the invariant n-grams. It is presumed that the functional potential of a language system is used more rationally by native speakers than by learners. Results of the analysis indicate that native speakers make optimal choices, which entail morphosyntactic simplicity but at the same time complexity and richness of vocabulary. Learners’ language usage shows on the one hand a broader range of structural variability than in the language of fiction, but on the other hand, it reveals oversimplification based on stereotypical language structures, overuse and underuse of morphological classes and clusters and poor vocabulary.

Keywords: usage-based language description, parsing, cluster analysis, syntax-morphology interaction, constraints and variability on language use, learner’s language, Estonian

KÄÄNDEVORMIDE KASUTAMISE OSKUS EAKOHASE ARENGUGA VENE-EESTI KAKS- KEELSETEL NING SPETSIIFILISE KÕNEARENGU PUUDEGA ÜKSKEELSETEL LASTEL

Merit Hallap, Marika Padrik, Signe Raudik

Ülevaade. Artikkel otsib vastust küsimusele, kas ja mille poolest sarnaneb 5–6-aastaste eakohase arenguga vene-eesti kakskeelsete ja spetsiifilise kõnearengu puudega eesti ükskeelsete laste oskus kasutada käändevorme. Andmestik pärineb 101 lapse katsetulemustest eesti keele ainsuse käändevormide (v.a nominatiiv) ja mitmuse partitiivi moodustamisel. Selgitame, kas kakskeelsete ja kõnepuudega laste areng järgib eakohase arenguga ükskeelsete laste arengut. Vaatleme, milliseid vigu lapsed vormide moodustamisel teevad ning milliseid asendusvorme eeldatud käändevormide asemel kasutavad.

Võtmesõnad: lastekeel, keele omandamine, morfoloogia, kakskeel-
sus, keele ja kõne arengu patoloogia, eesti keel

1. Sissejuhatus

Eesti keele käändeparadigma on morfoloogilise süsteemi mahukas osa. Käändevormide kasutamise oskus annab väärtuslikku infot nii ükskeelse lapse kõnearengu kui ka kakskeelse lapse eesti keele oskuse taseme kohta. Samas on andmeid eesti keele käändevormide kasutamise kohta koolieelses eas vähe. Lastekeele uurijad on nüüdseks kirjeldanud peamiselt üksikjuhtumite näitel eesti keele grammatiliste vormide kujunemist spontaanses kõnes ükskeelsetel (vt nt Argus 2008, Lipp 1977, Salo 1994) ja sünnist saati kahte keelt omandavatel, nn simultaanselt kakskeelsetel lastel (eesti-inglise kakskeelsus vt Vihman 1982, soome-eesti vt Hassinen 2002). Nime-
tatud uurimustest ilmnevad mõned ühised tendentsid vormimoodustusoskuste kujunemisel: esimeste muutevormidena eelistatakse lõputa partitiivi ja lühikese illatiivi vormi, sisekohakäänete vorme hakatakse kasutama enne kui väliskohakäändeid, kohakäändevormidega väljendatakse enne siht- kui lähtekohta, varakult ilmuvad mitmuse nominatiivi ja partitiivi vormid, vormimoodustuse kujunemise

algetappidel esinevad tüvevahelduse vead, morfeemid liituvad esmalt lühikestele, fonoloogiliselt lihtsatele tüvedele. Reili Argus (2008) kirjutab, et noomenivormide ilmumise järjekorda mõjutab muuhulgas muuttüübi produktiivsus ning lapse vajadus väljendada mingit semantilist rolli või suhet. Kõige olulisemaks peab Argus (2008: 42) aga sisendkeele mõju, millest oletatavasti (osaliselt) tulenevad ka juba mainitud uurimustes osalenud laste vormimoodustuse arengu erinevused.

Eesti keele käändevormide omandamise kohta on seega esmased andmed olemas ükskeelsete ja nn simultaanselt kakskeelsete laste osas. Koolieelses eas lastest moodustavad aga arvestatava osa ka nn astmelise kakskeelsusega lapsed (KK) ja spetsiifilise kõnearengu puudega lapsed (SKAP). Nende laste keelelise arengu uurimine ja hindamine on olulise väärtusega eesti keele õpetajate, lasteaiaõpetajate ning logopeedide jaoks. Samas Eestis praktiliselt puuduvad uuringud SKAP-lastest ja kakskeelsete laste kõnearengu ja -profiili kohta. Spetsiifilise kõnearengu puudega laste oskust kasutada käändevorme on uuritud vaid üliõpilastööde raames (nt Heina 2011). Marika Padrik (2010) analüüsis doktoritöös SKAP-lastest sõnamoodustusoskust. Eesti keele kui teise keele omandamist on käsitletud peamiselt täiskasvanud keeleõppijate näitel (nt Pool 2007). Ka keelekümbeluse programmi raames¹ pole eesti keele kui teise keele omandamist koolieelses eas süsteemselt uuritud.

Et eakohase arenguga astmelise kakskeelsusega (KK-) ja ükskeelsete SKAP-lastest keelelist arengut on Eestis vähe uuritud, on siinse uurimuse eesmärk selgitada, milline on eesti keele käändevormide kasutamise oskus vene-eesti kakskeelsetel ja spetsiifilise kõnearengu puudega eesti ükskeelsetel lastel. Alameesmärkideks on selgitada: 1) kas KK- ja SKAP-lastest käändevormide kasutamise oskus sarnaneb eakohase arenguga ükskeelsete laste oskustele; 2) milliseid vigu teevad KK- ja SKAP-lapsed käändevormide kasutamisel; 3) milliseid asendusvorme kasutavad KK- ja SKAP-rühmad käändevormide moodustamisel. Veatüüpide ja asenduste analüüs aitab otsustada, kas rühmade tulemuste vahel on kvalitatiivseid või kvantitatiivseid erinevusi.

2. Spetsiifiline kõnearengu puue ja astmeline kakskeelsus

Spetsiifiline kõnearengu puue (ingl *SLI = specific language impairment*)² avaldub eripärase kõnearenguna. SKAP puhul ei esine muude arengualdkondade mahajäämist, st nendel lastel ei ole kuulmislangust, neuroloogilist ega kognitiivset arengu esmast kahjustust (Leonard 1998, 2009). Juba varases eas ilmneb märgatav kõnearengu mahajäämus: hilineb esimeste sõnade ilmumine ning mitmesõnaliste ütluste loome. SKAP-ile iseloomulikud vead kõnes muutuvad lapse arengu käigus: koolieelses eas avalduvad need eelkõige häälduses ja grammatikas, hiljem teksti tasandil ja kirjalikus kõnes (Kohnert 2008, Padrik 2006). SKAP diagnoositakse soovitatavalt alles peale lapse nelja-aastaseks saamist, sest siis on juba võimalik hinnata, kas kõne areng on teiste laste omast erinev või mitte (Tomblin 2009: 424).

Ehkki arengu käigus on mõjutatud kõik kõneloome tasandid, peetakse SKAP-ile iseloomulikuks just grammatiliste oskuste puudulikkust arengut (Kohnert jt 2009, Paradis 2010, Leonard 2009). Grammatiliste vigade põhjuseks on uurijate arvates

¹ <http://kke.meis.ee/keelekumblusest/teadusuuringud> (25.9.13).

² Termin *spetsiifiline kõnearengu puue* puhul on autorid lähtunud põhimõttest, et keele kui vahendi omandamisraskused avalduvad kõnes.

kõnelise info töötamise probleemid, mistõttu on SKAP-lastel raskusi sisendkeele põhjal keeleliste üldistuste loomisega (Padrik 2010). Laste kõnes ilmnevad grammatilise tunnuse ärajätu, asendamise või lisamise vead. Et vead on keelespetsiifilised, võib eesti lastelt oodata eelnimetatutele lisaks n-ö morfofonoloogilisi vigu, mis avalduvad allomorfi vales valikus või moonutamises (Padrik 2006, Padrik 2010: 49).

Suuremates indoeuroopa keeltes on läbi viidud palju uurimusi selgitamaks, millised tunnused eristavad SKAP-lastete kõnet eakohase arenguga laste kõnest. Inglise keelt kõnelevatele SKAP-lastetele valmistab enim raskusi verbimorfoloogia: ajakategooria ja tegu- ning nimisõna ühildumine arvus (Kohnert 2008, Leonard 2009, Paradis 2005). Inglise keelest rikkama morfoloogiaga saksa keeles on SKAP-lastetele rasked nii verbi- kui noomenimorfoloogia (Dannenbauer 2001). Ungari (Lukács jt 2010) ja türgi (Rothweiler jt 2010) keeles teevad SKAP-lapsed rohkem vigu käändemorfoloogia kui verbivormide kasutamisel. Veel on andmeid, et nt hollandi keeles eksivad SKAP-lapsed enim sookategooria (Orgassa, Weerman 2008), hispaania keeles artiklite ja derivatsioonimorfeemide (Morgan jt 2013) kasutamisel. Võib teha esialgse järelduse, et rikka morfoloogiaga keeltes on SKAP-lastete noomenimorfoloogia verbimorfoloogiast keerukam.

Astmelise kakskeelsuse (ingl *sequential bilingualism*) puhul hakkab laps teist keelt (K2) omandama siis, kui esimeses keeles on saavutatud juba mingi taseme oskus. Astmelise ja simultaanse kakskeelsuse piir ei ole kindlalt fikseeritud. Johanne Paradisi (2010: 228) arvates on astmelise kakskeelsusega tegemist juhul, kui K2 hakatakse omandama peale 3-aastaseks saamist. Erineval arvamusel ollakse selles, kas kakskeelse lapse areng järgib ükskeelse arengut või mitte. Monika Rothweileri jt (2010: 542) järgi sarnaneb enne 3–4 aasta vanuseks saamist teist keelt omandama hakanud laste keeleline areng ükskeelsete arengule ja peale kuue aasta vanust täiskasvanud K2 omandajatele. Marilyn May Vihman ja Barry McLaughlin (1982: 54) märgivad, et kakskeelsete laste keelise arengu mustrid võivad olla, aga ei pea olema sarnased esimese keele omandajate omadega.

Sarnaselt ükskeelsete SKAP-lastega on ka astmelise kakskeelsusega (edaspidi: kakskeelsete) laste keelelist arengut uurides täheldatud rohkete vigade esinemist keelise arengu teatud etappidel. Uurijad (Paradis 2005, Orgassa, Weerman 2008, Kohnert 2008, Morgan jt 2013) on jõudnud järeldusele, et eakohase arenguga kakskeelsete laste vead morfosüntaktilisel tasandil sarnanevad ükskeelsete SKAP-lastete vigadele. Paradis (2005: 179, 2010: 245) ja Kathryn Kohnert (2008: 95) märgivad, et kahe rühma sarnasus esineb morfoloogia varase arengu etapil, s.o peamiselt kahe esimese aasta jooksul kakskeelsetel lastel teise keele omandamise algusest ja SKAP-lastel grammatiliste morfeemide ilmumisest kõnesse. Kolmandal aastal võib juba täheldada erinevusi nende kahe rühma vahel.

Laurence B. Leonardi (2009) eri keeli kõnelevate SKAP-lastete vigade analüüsis selgus, et keeliti võivad esineda eri tüüpi vead. Inglise ja itaalia keeles on täheldatud peamiselt morfeemide ärajätu kohustuslikust kontekstist, aga nt hispaania SKAP-lapsed nii jätavad morfeeme ära kui ka asendavad neid. Paradis (2005) kirjeldas kakskeelsetel lastel sarnaseid veamustreid ükskeelsete SKAP-lastete omadega, sh esines neil lastel rohkem morfeemi ärajätu- kui asendusvigu. Antje Orgassa ja Fred Weerman (2008) uurisid hollandi keele morfoloogiliste kategooriate kasutamist 7-aastastel ükskeelsetel, kakskeelsetel eakohase arenguga ja SKAP-lastel. Nende tulemused näitasid, et ükskeelsed SKAP- ja eakohase arenguga kakskeelsed lapsed

tegid samu vigu mida nooremad eakohase arenguga ükskeelsed lapsed. Rühmade tulemused olid siiski kvantitatiivselt erinevad, sh sooritasid ükskeelsed SKAP-lapsed katse paremini kui kakskeelsed lapsed. Gareth P. Morgani jt (2013) uurimuse kohaselt jäi eakohase arenguga hispaania-inglise kakskeelsete laste tulemus hispaania ükskeelsete eakohase arenguga ja SKAP-lastete tulemuste vahele: katsed paremini sooritanud kakskeelsed sarnanesid grammatilistelt oskustelt ükskeelsetele eakohase arenguga lastele ja halvemini sooritanud ükskeelsetele SKAP-lastele. Vigade analüüs näitas rühmades sarnaseid mustreid: ärajätuvigu oli rohkem kui asendusvigu.

Uurimuste tulemused on seega keeltest sõltuvalt erinevad. Teise keele omandajate ja SKAP-lastete keelelise arengu sarnasus (teatud ajahetkel) raskendab nende markerite leidmist, mis eristaksid kõnepuudega kakskeelseid lapsi eakohase arenguga lastest. Just varase K2 omandamise etapi puhul on raske otsustada, kas tegemist on eakohase arenguga või kõnepuudega (Paradis 2005: 185). Enne, kui saab hinnata, kas kakskeelse lapse keeleline areng toimub oodatud viisil, on vaja andmeid selle kohta, mis on eripärane eakohase arenguga kakskeelse lapse arengus võrreldes ükskeelse lapse arenguga ning mis on iseloomulik kõnepuudega ükskeelse lapse keelelisele arengule. Et grammatilised oskused, sh vead on keelespetsiifilised, ei saa teiste keelte uurimuste järeldusi eesti keelele üle kanda.

3. Katse ja informandid

Uuringus kasutasime kohandatult Maarja Heina (2011) metoodikat käändevormide kasutamise oskuse uurimiseks. Katses tuli lastel moodustada 13 käändevormi (kõik käänded, v.a nominatiiv) ainsuses ja mitmuse partitiivi vorm. Viimane on erandlikult haaratud katsesse, sest on nii praktika kui ka Heina (2011: 32) uuringu alusel osutunud lastele üheks keerulisemaks vormiks oletatavasti käändetunnuse morfeemivariantide rohkuse tõttu. Et osa käändevormide kasutust uuriti erinevates funktsioonides (vt osa 4), oli laste ülesandeks moodustada kokku 21 vormi.

Keelematerjali valikul arvestati eesti keele spetsiifikat: tüve- ja lõpuvariantide varieeruvust ning käändevormide erinevaid funktsioone. Stiimulsõnad olid 1) lastele tuttavad, 2) nii suurema kui väiksema kasutussagedusega, 3) nii astmevahelduseta kui ka laadi- ja vältevahelduslikud, 4) piltide või reaalsete esemete abil näitlikustatavad. Näiteks genitiiv ja partitiiv tuli moodustada sõnadest *põder*, *kringel*, *õis*, *aken*, *uks*, *kirss*, *lind*, *pliiats*, *jalg*, *saabas*, *padi* ja *täht*. Nende vormide moodustamise ülesandes välditi sõnu, millel esineb vormihomonüümiat.

Meetodina kasutati vormikasutuse suunatud esilekutsumist (ingl *structured elicitation method*). Iga vormi puhul esitati lastele kaks näidist, et kindlustada ülesandest arusaamine. Lapsel tuli vastata uurija esitatud küsimusele ja/või lõpetada lause eeldatud sõnavormiga, toetudes esitatud piltidele või uurija poolt loodud mängulisele situatsioonile (tegutsemisele mänguasjadega).

Näiteks genitiivi (objekti funktsioonis) ning ainsuse ja mitmuse partitiivi vormide (hulgasõna laiendi funktsioonis) moodustamisel kommenteeris uurija mängukaru tegevust ja suunas last lauseid lõpetama: *Vaata, mõmmil on lind* (uurija osutab pildile). *Mõmmi võtab veel ühe* (osutab pildile) (linnu). *Nüüd on mõmmil kaks* (osutab kahele pildile) (lindu). *Aga siin* (osutab pildile, kus on palju linde) *on palju* (linde).

Kohakäänete kasutamisel kohta, adressaadi ja lähteallika funktsioonis ning komitatiivi ja terminatiivi vormide kasutamisel pidi laps lõpetama uurija lause, toetudes küsimusele. Nt *Kuhu ma palli panen?* (uurija asetab palli korvi). *Panen palli* (korvi sisse/korvi); nt *Milleni kiisu jookseb?* (uurija imiteerib mänguasja liikumist objektini). *Kiisu jookseb* (aiani).

Ülejäänud vormide kasutuse uurimisel pidi laps lause lõpetama eeldatud vormiga pildi, lauselise konteksti (ja küsimuse) alusel. Nt *Vaata pilti! Laual on puudu jalg* (uurija osutab pildile). *Laud on ilma* (jalata); nt *See on kass ja see on koer* (uurija osutab mänguasjadele). *Ütle kiisule, kellest ma räägin! Ta valvab maja. Ta elab kuudis. Kellest ma rääkisin?* (uurija osutab koerale). *Rääkisin* (koerast).

Katse viidi läbi iga lapsega eraldi kahes jaos 2013. a kevadel. Testüksuste järjekord oli kõigil lastel sama. Järjekorra määramisel pidasime silmas, et 1) ühe või teise vormi kasutus ei mõjuta järgmise (sarnase funktsiooniga või sama vormi teise funktsiooniga) vormi kasutust (nt ei esitatud järjest allatiivi, adessiivi, ablatiivi), 2) suurema kasutussagedusega vormide (nt partitiiv) moodustamine vaheldub väiksema sagedusega (nt essiiv) vormide moodustamisega.

Valimi moodustasid 101 last. Lapsed jagunesid nelja rühma (vt tabel 1): 5–6-aastased eakohase arenguga vene-eesti kakskeelsed lapsed (KK), 4–6-aastased ükskeelsed SKAP-lapsed (SKAP) ning võrdlusrühmadena eakohase arenguga ükskeelsed 4–6-aastased (ÜK5-6) ning kolmeaastased lapsed (ÜK3). Kakskeelsete laste valikul arvestati järgmiste kriteeriumitega: 1) lapsed räägivad kodus ainult vene keelt, 2) lapsed käivad kolmandat õppeaastat eestikeelses lasteaias, 3) laste vaimne ja kõne areng on lasteaiaõpetajate ja logopeedide hinnangul eakohane. Tegemist oli seega lastega, kes hakkasid eesti keelt omandama kolmeaastaselt loomulikus suhtluskeskkonnas. SKAP-lapsed olid logopeedide poolt identifitseeritud kui kõnepuudega lapsed, kel ei esine muid arenguprobleeme.

Tabel 1. Katses osalenud lasterühmad

Lasterühm	SKAP	KK	ÜK5-6	ÜK3
Laste arv	21	29	30	21
Vanusevahemik kuudes	56–77	61–80	57–76	32–39
Keskmine vanus kuudes	67	69	69	37
Sooline jaotuvus	N = 5 M = 16	N = 16 M = 13	N = 20 M = 10	N = 11 M = 10

KK- ja SKAP-lastepuhul püüti arvestada sellega, et nende laste grammatiliste oskuste areng oleks kestnud ajaliselt sama kaua. Lähtudes teadmised, et SKAP-lastepuhul aktiivne morfoloogia omandamine algab tavaliselt kolme aasta vanuses (Dannenbauer 2001, Leonard 1998), valiti uuringusse KK-lapsed, kes olid eesti keelt hakanud omandama samuti kolmeaastaselt. Uurimuses kasutati võrdlusrühmadena 1) SKAP- ja KK-lastepuhul ühevanuseid eakohase arenguga ükskeelseid lapsi (ÜK5-6), st uuritavatega potentsiaalselt sama vaimse arengu taseme ja suhtlemisoskustega lapsi; 2) nooremaid ükskeelseid lapsi (ÜK3), st uuritavatega potentsiaalselt sarnase keelilise arengu tasemega lapsi.

4. Tulemused

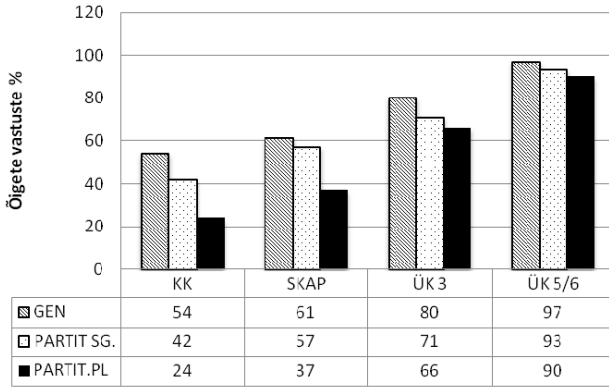
Andmeid analüüsisime nii kvantitatiivselt kui ka kvalitatiivselt, kasutades andmetöötlusprogrammi STATISTICA 7.0. Kokku moodustasid lapsed 12 928 sõnavormi, mida rühmitasime mitmel alusel: 1) õigete ja valede vastuste andmekogum; 2) vigadega moodustatud vormide rühmad. Õigeks lugesime vastuse, milles laps moodustas eeldatud vormi, kasutades seejuures ka õiget tüveallomorfi. Hääldusvigadest tingitud sõnade moonutusi tõlgendasime kahel viisil: 1) lugesime valesti hääldatud sõnavormi õigeks (nt hääliku ärajätt kaashäälikuühendist: *kingit pro kringlit*), kui oli võimalik määrata, mis vormi laps kasutas, või 2) märkisime moonutatud sõnana, kui vormi polnud võimalik ära tunda (nt *ussus pro ust*). Valed vastused jaotasime järgmisteks tüüpideks: 1) vale tüveallomorfi kasutus õige käänevormiga; 2) vale käänevormi kasutus õige tüveallomorfiga; 3) vale käänevormi kasutus vale tüveallomorfiga; 4) algvormi, s.o ainsuse nominatiivi kasutus eeldatud vormi asemel; 5) tõlgendamatu vorm sõnast või vastamata jätmine; 6) käändetunnuse vale morfeemivariandi kasutus (esines ainult mitmuse partitiivi puhul).

Esmalt pakkus meile huvi, kas KK- ja SKAP-lastel käänevormide kasutamise oskuse areng järgib eakohase arenguga ükskeelsete laste arengut. Et võrrelda kõigi uuritud vormide kasutuse edukuse seost neljas lasterühmas, kasutasime Spearmani korrelatsioonanalüüsi. Õigete vastuste protsentuaalse osakaalu alusel vastuste koguarvust võrdlesime omavahel nii SKAP- ja KK-rühmade kui ka kummagi nimetatud rühma tulemusi ÜK3 ja ÜK5-6 tulemustega. Kõikide rühmade tulemuste vahel ilmnes tugev statistiliselt oluline korrelatsioon ($\rho = 0,61-0,89, p < 0,001$), mis oli kõige tugevam KK ja SKAP ning nõrgem ÜK5-6 ja KK tulemuste vahel. Tulemustest järeldub, et kõikide rühmade puhul avaldus käänevormide kasutuse edukuses sarnane tendents: edukamad oldi samade vormide moodustamisel (nt allatiiv ja adessiiv koha tähistajana) ning raskusi valmistasid samuti samad vormid (nt essiiv oleku väljendajana, mitmuse partitiiv hulgasõna laiendina). Järgnevalt analüüsisime täpsemalt vormide kasutust käänderühmade kaupa, püüdes leida nii kvantitatiivseid kui kvalitatiivseid erinevusi lasterühmade tulemuste vahel. Kvantitatiivses analüüsis kasutasime oluliste erinevuste leidmiseks korduvmõõtmistega dispersioonanalüüsi (ingl *repeated measures ANOVA*), kvalitatiivselt analüüsisime veatüüpe ja vormide asendusi. Dispersioonanalüüs näitas, et kõigi käänevormide puhul erinesid rühmad oluliselt vaid õigete vastuste alusel, veatüüpides ei olnud erinevused ühegi vormi puhul olulised.

4.1. Grammatiliste käänete vormide kasutus

Genitiivi vormide kasutamine osutus jõukohaseks ÜK3- ja ÜK5-6-lastele, kuid raskeks KK- ja SKAP-lastele. Genitiivi kasutamisel avaldusid rühmade tulemuste vahel olulised erinevused ($F(6; 95) = 9,278; p < 0,001$). *Post hoc* analüüs Scheffé testiga kinnitas, et oluliselt erinesid õigete vastuste alusel KK- ja SKAP-lapsed ÜK3- ja ÜK5-6-lastest. SKAP- ja KK-lastel tulemused, samuti ÜK3- ja ÜK5-6-lastel tulemused olid õigete vastuste alusel sarnased. Vigade analüüsil ilmnes SKAP- ja KK-lastel sage algvormi kasutus genitiivi asemel. Kakskeelsed lapsed moodustasid genitiivi asemel paljudel juhtudel veel ainsuse partitiivi (nt *võtab veel ühe ust pro*

ukse), samas kui SKAP-lapsed kasutasid asendustes valdavalt mitmuse nominatiivi (nt võtab veel ühe ukse pro ukse).



Joonis 1. Grammatiliste käänete vormide kasutuse tulemused

Ainsuse partitiivi kasutamine osutus kõigile lastele raskemaks kui genitiivi kasutamine. Dispersioonanalüüs näitas rühma kui faktori olulist mõju ($F(6; 95) = 11,22; p < 0,001$). *Post hoc* analüüsil selgus, et KK-lastel edukus on oluliselt madalam kõigist teistest rühmadest. SKAP-lastel tulemused olid sarnased ÜK3 tulemustega. Vaadeldes veatüüpide vahelisi proportsioone, selgus, et SKAP-lapsed kasutasid eeldatud vormi asemel palju algvormi, samas kui KK-lapsed moodustasid rohkem valesid vorme õige tüveallomorfiga (nt *Mõmmil on nüüd kaks ukse pro ust*) ja valesid vorme vale tüveallomorfiga (nt *Mõmmil on nüüd kaks õised pro õit*). Kahe rühma asendusvormid erinesid: SKAP-lapsed kasutasid ainsuse partitiivi asemel mitmuse partitiivi, KK-lapsed mitmuse nominatiivi ja ainsuse genitiivi vormi.

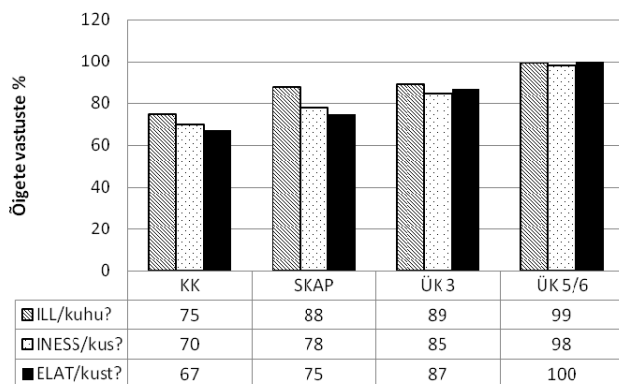
Mitmuse partitiiv osutus grammatilistest käänetest raskeimaks kõigile lastele. Dispersioonanalüüs näitas jällegi rühmade tulemuste vahel olulist erinevust ($F(6; 95) = 11,38; p < 0,001$), *post hoc* analüüsi järgi ei olnud KK ja SKAP tulemused erinevad, küll aga erinesid mõlema rühma tulemused ÜK3 omadest. Vigade analüüsil ilmnes juba mainitud tendents: KK-lapsed moodustasid mitmuse partitiivi asemel teise vormi: peamiselt mitmuse nominatiivi (*siin on palju linnud pro linde/lindusid*) ja ainsuse partitiivi (*siin on palju lindu pro linde*). SKAP-lapsed eelistasid mitmuse partitiivi asemel ainsuse partitiivi. SKAP-lastel esines veatüüpidest kõige enam partitiivi tunnuse vale morfeemivariandi kasutust (nt *siin on palju saapaid pro saapaid*).

ÜK3-lapsed eksisid genitiivi kasutamisel tüveallomorfi ja partitiivi käändetunnuse morfeemivariandi valikul. Mõne teise vormi kasutust grammatiliste käänete asemel ÜK3-lastel ei täheldatud.

4.2. Sisepääkäänete vormide kasutus

Lapsed kasutasid illatiivi ja inessiivi vorme ainult koha funktsioonis, elatiivi lisaks materjali (*on tehtud millest?*) ja objekti (*räägib kellest?*) tähistajana. Vormide moodustamine koha ja materjali tähistamiseks osutus küllaltki jõukohaseks kõikidele

rühmadele (vt joonised 2 ja 3). Elatiivi kasutamisel objekti märkijana eksisid SKAP ja KK kõrval küllaltki palju ka ÜK3-lapsed (vt joonis 3). Dispersioonanalüüs näitas sisekohakäänete vormide kasutamisel rühmade vahel olulisi erinevusi ($F(4; 97) = 6,3841; p < 0,001$). *Post hoc* analüüsi kohaselt ei erinenud KK- ja SKAP-laste tulemused oluliselt vormide moodustamisel koha ja materjali tähistajana, küll aga olid KK tulemused oluliselt halvemad elatiivi kasutamisel objekti funktsioonis. SKAP rühm erines oluliselt ÜK3-st elatiivi moodustamisel materjali tähistajana, muudes funktsioonides vormide kasutamisel olid SKAP-laste tulemused sarnased ÜK3-le. KK tulemused olid sisekohakäänete kõigi funktsioonide puhul ÜK3-st oluliselt nõrgemad.



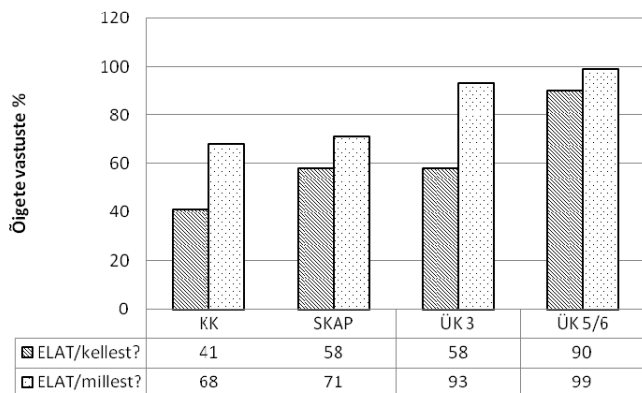
Joonis 2. Sisekohakäänete vormide kasutuse tulemused (kohta tähistav funktsioon)

Veatuüpide analüüsist selgus, et illatiivi moodustamisel kasutasid SKAP-lapsed küllaltki palju algvormi. KK-lapsed eksisid tüveallomorfi valikul ning moodustasid ka vale vormi (peamiselt allatiivi ja genitiivi). Inessiivi kasutamisel olid SKAP ja KK vead ühesugused: mõlemad kasutasid peamiselt valet tüveallomorfi. Elatiivi moodustamisel koha funktsioonis eksisid nii SKAP kui KK peamiselt tüveallomorfi valikul. Lisaks moodustasid mõlemad rühmad eeldatust erinevaid käändevorme, sh asendasid elatiivi peamiselt inessiivi vormiga (nt *võtan palli taskus* pro *taskust*). ÜK3-laste vead olid eranditult seotud vale tüveallomorfi kasutusega.

Elatiivi kasutamisel objekti tähistajana (vt joonis 3) said kõik rühmad sisekohakäänete vormide nõrgimad tulemused. Nii SKAP kui KK puhul oli eelkõige tegemist vale tüveallomorfi kasutusega. SKAP-lastel esines küllaltki palju ka algvormi kasutust. KK- ja SKAP-lapsed moodustasid ka teisi käändevorme, kusjuures kahe rühma vormiasendused olid erinevad: SKAP-lapsed kasutasid komitatiivi (*räägib koeraga* pro *koerast*) ja allatiivi (*räägib koerale* pro *koerast*), KK-lapsed aga ainsuse partitiivi ja ablatiivi (*räägib koeralt* pro *koerast*). ÜK3 vähene edukus tulenes taas vale tüveallomorfi kasutusest.

Elatiivi kasutus materjali tähistajana (vt joonis 3) valmistas SKAP-lastele pisut rohkem raskusi kui koha funktsioonis, KK-rühma tulemus oli nimetatud funktsioonide puhul sarnane. Tulemusi analüüsides ilmnes teistsugune tendents kui eelmiste vormidega: SKAP-lastel esines rohkem vale vormi moodustamist kui KK-lastel. Vormide asendused olid kahes rühmas erinevad: SKAP kasutas elatiivi asemel komitatiivi (*on tehtud kummiga* pro *kummist*), translatiivi (*on tehtud paberiks*

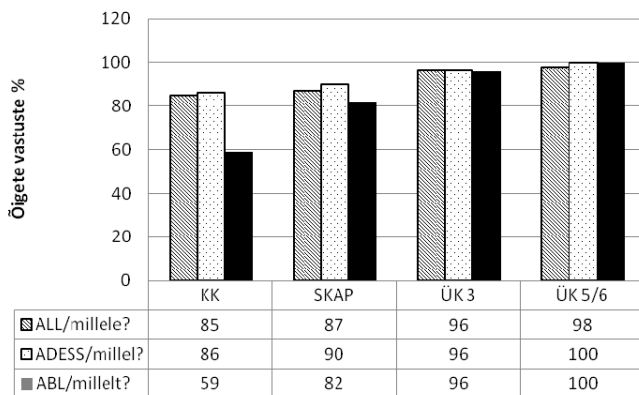
pro *paberist*) ja inessiivi (*on tehtud klaasis pro klaasist*) vorme, KK inessiivi ja genitiivi. ÜK3-l esinesid üksikud tüveallomorfi vead.



Joonis 3. Sisekohakäänete vormide kasutuse tulemused (objekti ja materjali tähistav funktsioon)

4.3. Väliskohakäänete vormide kasutus

Lapsed pidid allatiivi kasutama koha ja adreassaadi (*kingib kellele?*), adessiivi koha, omaja (*kellel?*) ja aja ning ablatiivi koha ja lähteallika (*sai kellelt?*) tähistamiseks. ÜK5-6-lastele ei valmistanud raskusi ühegi vormi moodustamine (vt joonised 4 ja 5). Teiste laste tulemused sõltusid käändevormi funktsioonist.

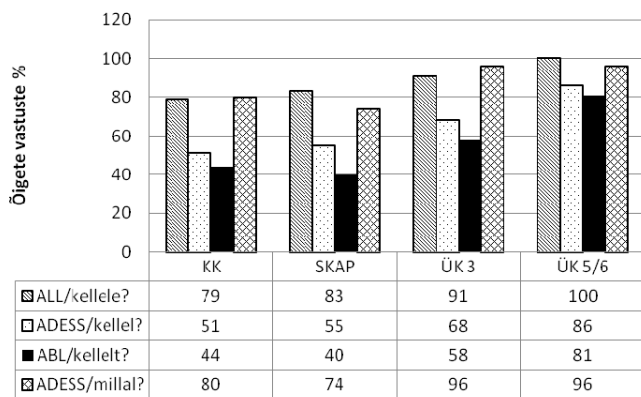


Joonis 4. Väliskohakäänete vormide kasutuse tulemused (kohta tähistav funktsioon)

Vormide kasutus kohafunktsioonis oli kerge kõigile, v.a KK-lapsed, kes olid raskustes ablatiivi moodustamisega. Oluliselt raskem oli KK-, SKAP- ja ÜK3-lastele adessiivi kasutamine omaja funktsioonis (*kellel?*) ja ablatiivi kasutamine lähteallika funktsioonis (*sai kellelt?*). Dispersioonanalüüs näitas rühmade tulemuste vahel statistiliselt olulist erinevust ($F(6; 96) = 5,647; p < 0,001$). *Post hoc* analüüsil selgus, et KK tulemused olid SKAP omadest oluliselt nõrgemad ablatiivi moodustamisel

kohafunktsioonis, muude vormide osas ei olnud erinevused olulised. KK-laste tulemused olid oluliselt nõrgemad ka ÜK3 tulemustest adessiivi ja ablatiivi kasutamisel kohafunktsioonis. SKAP-laste tulemused ei erinenud oluliselt ÜK3 omadest, v.a adessiivi kasutamisel aega tähistavas funktsioonis.

Vigade analüüsisist selgus, et nii allatiivi kui adessiivi moodustamisel kohafunktsioonis oli KK- ja SKAP-lastel peamiseks veatüübiks vale tüveallomorfi kasutus. Ablatiivi puhul aga kasutasid mõlemad rühmad eeldatud vormi asemel muud käändevormi. Nii SKAP kui KK kasutasid kõige enam adessiivi (*võtab vaibal* pro *vaibalt*), sageduselt teisena esinenud asendusvorm oli SKAP-lastel genitiiv, KK-l elatiiv. ÜK3 nende vormide moodustamisel praktiliselt vigu ei teinud. Ka adessiivivormide kasutamisel omaja tähistajana eksisid KK- ja SKAP-rühmad ülekaalukalt tüveallomorfi valikul. SKAP-lapsed kasutasid küllaltki palju ka algvormi. ÜK3-lastel esinesid mõned tüveallomorfi vead. Aja väljendamine adessiivivormiga ei olnud raske ühelegi rühmale. Tegemist oli ilmselt tervikvormina omandatud sõnadega (nt *talvel*, *hommikul*).



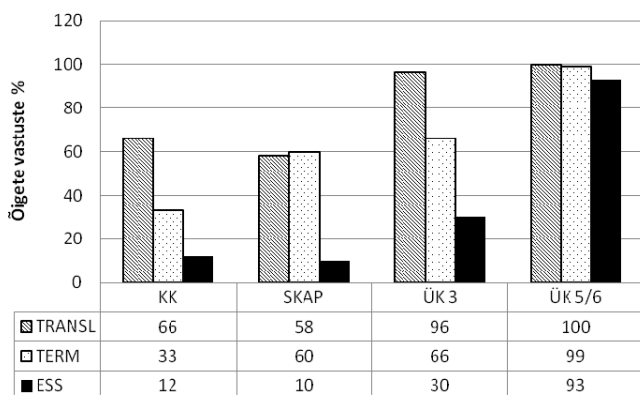
Joonis 5. Väliskohakäänete vormide kasutuse tulemused (adressaati, omajat, lähteallikat ja aega tähistav funktsioon)

Allatiivi moodustamisel adressaadi tähistamiseks eksisid KK-lapsed peamiselt tüveallomorfi valikul, SKAP-lastel esines nimetatu kõrval ka vale vormi kasutust – peamiselt genitiivi. ÜK3-lapsed eksisid mõne üksiku tüveallomorfi valikul.

Väliskohakäänetest valmistas kõigile rühmadele enim raskusi ablatiivi kasutus lähteallika (*sai kellelt?*) märkimisel. Nii SKAP kui KK eksisid peamiselt tüveallomorfi valikul, aga ka valet vormi nii õige kui vale tüveallomorfiga esines küllaltki palju. Vormiasendused olid rühmadel sarnased: peamiselt kasutati elatiivi (*sai kondi kassist* pro *kassilt*), aga ka adessiivi (*sai kondi kassil* pro *kassilt*). ÜK3-lapsed eksisid ülekaalukalt ikka tüveallomorfi kasutamises, kuid esinesid ka üksikud valed vormid (elatiiv pro ablatiiv).

4.4. Muude käändeformide kasutus

Translatiivi, terminatiivi ja essiivi kasutamise tulemused olid edukalt toime ÜK5-6-lapsed, teistele valmistati suuri raskusi essiivi moodustamine (vt joonis 6). Rühmadevahelised erinevused olid statistiliselt olulised ($F(2; 99) = 22,244; p < 0,001$). *Post hoc* analüüs näitas, et KK-laste tulemused olid oluliselt madalamad SKAP-omadest vaid terminatiivi moodustamisel. ÜK3-lastest oluliselt vähem kasutasid SKAP-lapsed õigesti translatiivi ning KK terminatiivi.



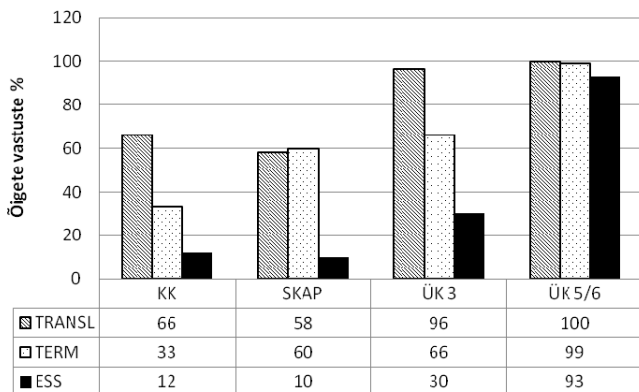
Joonis 6. Translatiivi, terminatiivi ja essiivi vormide kasutuse tulemused

Vigade analüüsil ilmnus, et translatiivi kasutamisel oli nii SKAP- kui KK-laste peamiseks veatüübiks vale vormi moodustamine. SKAP-asendusvormid olid väga erinevad, st ei ilmnunud ühte kindlat asendusvormi, samas kui KK-lapsed kasutasid translatiivi asemel ülekaalukalt genitiivi (*siil tahab ka saada lenduri pro lenduriks*).

Terminatiivi puhul kasutas KK-rühm nii valet tüveallomorfi kui ka valet vormi – õige ja ka vale tüveallomorfiga. Ka SKAP-lastel esines veatüüpidest enim vale vormi moodustamist, kuid SKAP ja KK-asendused olid erinevad: SKAP-lapsed kasutasid peamiselt genitiivi ja KK-allatiivivormi (nt *jookseb munale pro munani*).

Essiivi tulemused olid SKAP-, KK- ja ÜK3-lastel kõikidest vormidest nõrgimad. Kõigil oli peamiseks veatüübiks vale vormi kasutus õige tüveallomorfiga. Küllaltki palju esines ka algvormi kasutust, st sisuliselt kordasid lapsed uurija poolt öeldud lausest ametit märkivat sõna (nt *See on eesel. Ta on lendur. Kellena eesel töötab? Eesel töötab ... lendur pro lendurina*). Ühine kolmele rühmale oli allatiivi (*töötab juuksurile pro juuksurina*) ja komitatiivi (*töötab juuksuriga pro juuksurina*) moodustamine essiivi asemel. ÜK3-lastel olidki need ainsad asendusvormid. SKAP-lapsed kasutasid aga erinevaid asendusi – peale eelnimetatute esines veel inessiivi, adessiivi ja elatiivi kasutust, KK-l enam translatiivi ja genitiivi.

Abessiivi ja komitatiivi kasutuse tulemustes ilmnus dispersioonanalüüsi järgi rühmade vaheline statistiliselt oluline erinevus ($F(2; 99) = 11,74; p < 0,001$). *Post hoc* analüüs näitas ÜK5-6-tulemuste olulist erinevust teiste rühmade tulemustest; SKAP, KK ja ÜK3-tulemused olid sarnased.



Joonis 7. Abessiivi ja komitatiivi vormide kasutuse tulemused

Abessiivi kasutuse edukus oli SKAP-, KK- ja ÜK3-rühmades küllaltki madal (vt joonis 7). Vigade analüüsisist selgus, et nii SKAP- kui KK-lastel esines peamiselt vale vormi moodustamist õige või vale tüveallomorfiga. SKAP kasutas abessiivi asemel põhiliselt komitatiivi (*maja on ilma aknaga pro aknata*), KK ainsuse partitiivi ja genitiivi. ÜK3-lastel esines peamiselt vale tüveallomorfi kasutust, kuid täheldasime ka üksikuid vormiasendusi, sh kasutati eranditult komitatiivi.

Komitatiivi moodustamise suhteliselt madalad tulemused seostusid kõigil rühmadel vale tüveallomorfi kasutusega. Muid veatüüpe esines väga vähe.

5. Arutelu ja kokkuvõte

Püüdsime selgitada, kas ja kuidas sarnaneb KK- ja SKAP-lastel käändevormide kasutamise oskuse areng ükskeelsete laste keelelise arenguga. Eeldatult kasutasid ÜK5-6-lapsed kõiki uuritud käändevorme peamiselt õigesti (õigete vastuste osakaal 78–100%). ÜK3-lapsed erinesid mitmete käänete puhul edukuselt ÜK5-6-lastest, eksides valdavalt tüveallomorfi valikul, millele kui iseloomulikule veale vormimoodustuse kujunemise algetappidel viitab ka Argus (2008: 23). Võrreldes KK ja SKAP tulemusi kvantitatiivselt selgus, et enamiku käändevormide (v.a essiiv, translatiiv, ablatiiv lähteallika ning adessiiv aja funktsioonis) kasutamisel said madalamad tulemused KK-lapsed (õigete vastuste osakaal 12–86%). SKAP-lapsed (õigete vastuste osakaal 10–90%) paigutasid oma tulemustelt KK ja ÜK3 vahele, olles viimastest oluliselt nõrgemad genitiivi, mitmuse partitiivi, materjali tähistava elatiivi, aega tähistava adessiivi ning translatiivi osas. Kakskeelsete nõrgemaid tulemusi SKAP-lastega võrreldes täheldasid oma uurimuses ka Orgassa ja Weerman (2008). KK-lastel olid SKAP-omadest kvantitatiivselt oluliselt madalamad järgmistes vormides: ainsuse partitiiv, terminatiiv, elatiiv objekti ja ablatiiv kohta tähistavas funktsioonis. Samas oli mõnede vormide kasutamine nii KK- kui SKAP-lastel ka jõukohane (õigete vastuste osakaal üle 70%; nt sise- ja väliskohakäänded siht- ja asukohta tähistavas funktsioonis, allatiiv adressaadi funktsioonis, samuti adessiiv aega tähistavas funktsioonis – viimase puhul on sõnad ilmselt omandatud valmisvormidena). Nimetatud käänete puhul mängib olulist rolli vormide

kasutamissagedus, seega ka sisendkeele mõju, mille olulisust morfoloogiliste vormide ilmumise järjekorrale lapse kõnes on rõhutanud Argus (2008).

Oletust, et KK- ja SKAP-lastest morfoloogiline areng järgib ÜK laste oma, kinnitab ka korrelatsioonanalüüs, mis näitas tugevat seost kõigi lasterühmade tulemuste vahel käänete järjestuses nende moodustamise edukuse alusel. Kvantitatiivse analüüsi alusel leidis seega kinnitust Paradisi (2005, 2010) ja Kohnerti (2008) seisukoht, et nende kahe lasterühma morfoloogia varane areng on sarnane. Samuti leiab toetust oletus, et KK-lastest areng järgib ÜK-lastest arengut, milles uurijad (Rothweiler 2010, Vihman, McLaughlin 1982) ei ole üksmeelel. Siiski ei saa teha ühest järeldust, analüüsimate laste viga.

Nagu eespool mainitud, eksisid ÜK₃-lapsed valdavalt tüve- või lõpuallomorfe kasutuses, st ilmnis lapse varasele grammatikale iseloomulik nn üleüldistamine. Võib oletada, et KK- ja SKAP-lapsed teevad samuti selliseid vigu (Lukács jt 2010, Leonard 2009). Olgu mainitud, et statistiliselt olulisi erinevusi veatüüpide sageduses SKAP- ja KK-lastel ei ilmnenu. SKAP ja KK vigade analüüsi alusel on võimalik välja tuua järgmist.

- Nii KK- kui SKAP-lastel esines sama tüüpi viga: peamiselt vale tüveallomorfi (nt *hobune kingib kaalika lammasele pro lambale*) ja vale vormi (nt *lammas sai kaalika hobusest pro hobuselt*) kasutust, kusjuures nimetatud vead võisid samas sõnavormis ka kombineeruda (nt *Aga siin on palju lindud pro linde/lindusid*). Tulemused erinevad seega inglise keele baasil tehtutest, kus morfeemide ärajätuvigu esineb nii KK- kui ka SKAP-lastel rohkem kui asendusi (vt Paradis 2005: 182-183, Leonard 2009: 309).
- Käändevormide puhul, mille kasutuse edukus oli kõrgem ja mis ilmuvad lapse kõnesse varem (nt sisekohakäänded kohta tähistavas funktsioonis, komitatiiv), domineerisid mõlemal rühmal tüveallomorfi kasutusvead. N-ö raskemate käändevormide (translatiiv, terminatiiv, essiiv, abessiiv) puhul olid mõlemal rühmal ülekaalus vormide asendused.
- Enamasti asendati vorme sama käänderühma piires, näiteks grammatiliste käänete puhul asendasid SKAP-lapsed ainsuse partitiivi mitmusega, KK-lapsed kasutasid nii ainsuse kui mitmuse partitiivi asemel mitmuse nominatiivi (nt *Aga siin on palju saapad pro saapaid*). Kohakäänete puhul asendati vorme teise kohakäände vormiga (nt ablatiivi adessiiviga, elatiivi inessiiviga). Terminatiivi puhul esines mõlemal rühmal asendusi allatiiviga (nt *Kiisu jookseb aiale pro aiani*), st säilis käändevormi ruumi tähistav funktsioon. Ágnes Lukácsi jt (2010) andmetel on sellised lähedased asendused (ingl *near miss errors*) iseloomulikud SKAP-lastele rikka morfoloogiaga keeltes, st lapsed eksivad vormi moodustamisel vaid ühes tunnuses (nt arv, lähte koht vs. asukoht). Niisiis võib väita, et SKAP- ja KK-lastel avalduvad käändemorfoloogia omandamisel sarnased vead, mis tulenevad omandatava keele eripärast.
- Asenduste analüüs võimaldab välja tuua ka erisusi SKAP- ja KK-lastest vormikasutuses, mis avaldusid selgemini raskemate käänetega. Näiteks abessiivi asemel kasutasid SKAP-lapsed komitatiivi (*laud on ilma jalaga pro jalata*), mida võib selgitada asendusega tähenduse alusel (tunnuse olemasolu vs. puudumine), KK-lapsed aga genitiivi või partitiivi, mida võib selgitada vene keele mõjuga (*без чего? → ilma jala; ilma jalga pro jalata*).

Elatiivi asemel (materjali funktsioonis) kasutasid SKAP-lapsed erinevalt KK-lastest komitatiivi (*on tehtud kummiga pro kummist*) ja translatiivi (*on tehtud rauaks pro rauast*), mis võib seostuda tegusõna piiratud tähendusega (valentsusega). KK-lapsed kasutasid elatiivi asemel sageli genitiivi, mis võib tuleneda n-ö tõlkimisest (*tehtud uz uezto? → kummi pro kummist*). Sama tendents avaldus elatiivi kasutamisel objekti tähistavas funktsioonis: SKAP-lapsed kasutasid komitatiivi või allatiivi (*räägib koeraga/koerale pro koerast*), KK-lapsed aga ablatiivi (*räägib koeralt pro koerast*, s.o tähenduselt lähedast vormi). Translatiivi puhul kasutasid SKAP-lapsed erinevaid asendusi, milles ei olnud võimalik tuvastada reeglipärasust. KK-lapsed kasutasid translatiivi asemel valdavalt genitiivi, mida võib põhjendada tegusõna tähenduse mittemõistmisega (*tahab saada kozo/kelle? → arsti pro kelleks? → arstiks*). Sama käändevormiga esines SKAP-lastel ka moonutatud vormide kasutamist, mida ei esinenud kordagi KK-lastel.

- SKAP-lastele on KK-ga võrreldes iseloomulikum algvormi kasutamine, mis osutus peamiseks veatüübiks genitiivi, ainsuse partitiivi ja illatiivi puhul, kuid esines sageli ka elatiivi (*räägib kellest?*), adessiivi (*kelle!*) ja essiivi puhul. KK-l esines algvormi vaid genitiivi, translatiivi ja essiivi puhul.

Niisiis võib välja tuua erinevused SKAP- ja KK-lastel asenduste tüüpides. SKAP-lapsed kasutasid KK-lastega võrreldes sagedamini algvormi, nende asendused olid juhuslikud n-ö raskete käänete korral (nt translatiiv, essiiv), aga võisid tuleneda ka tegusõna piiratud tähendusest. Nimetatud eripära SKAP-lastel vigade muustris on võimalik selgitada teistsuguse keele omandamise mehhanismiga (Padrik, 2010, Leonard 2009). SKAP-ile on omased süntagmaatiliste seoste piiratus, eriti tegusõna puhul (Dannenbauer 2001), raskused keele fonoloogilise süsteemi omandamisega (Padrik 2006, Dannenbauer 2001), mis põhjustab vormide häälduslikku moonutamist ja osaliselt ka algvormi või genitiivi kui häälduslikult lihtsamate vormide kasutamist. Viimast, samuti vormide ebapüsivat kasutust, on kirjeldanud ka Padrik (2010), Leonard (1998), Dannenbauer (2001). KK-lastel esines aga hulgaliselt asendusi, mida võib selgitada vene keele mõjuga. Võib oletada, et laps kasutab sobiva vormi leidmisel n-ö kodeerimist/tõlkimist vene keelest. Siiski ei ole alati võimalik vea põhjust üheselt selgitada: üheaegselt võivad mõju avaldada nii keeleline kui mitteverbaalne kontekst, teine keel või vormi häälduslik raskus.

Eeltoodud järelduste üldistamisel tuleb arvestada mitmete piirangutega. Et informantide rühmad olid väikesed, keskendusime kvalitatiivsele analüüsile, püüdes välja tuua tendentse, mida on vaja kontrollida suurema valimi baasil. Samuti võis tulemusi mõjutada keelematerjali valik: nt translatiivi ja essiivi puhul kasutatud elukutsete nimetused on häälduslikult rasked sõnad (*õmbleja, juuksur, arst, lendur, ehitaja*), sõnade arv ühe käändevormi kasutuse uurimisel oli paljude vormide puhul väike (4–5 sõna). Grammatiliste käändevormide kasutamisel esines nii SKAP- kui KK-lastel asendusi, mida võis selgitada katsematerjali mõjuga. Võib oletada, et eeldatud vormide kasutamise asemel kirjeldasid lapsed pildil kujutatut (nt eeldatud: *võtab veel ühe põdra → laps: see on põder*; eeldatud: *palju põtru → laps: need on põdrad*). On loomulik, et teist keelt omandav või alakõnega laps ei toetu ülesannet täites alati verbaalsele kontekstile, vaid vastab tajutavast situatsioonist lähtuvalt. Seetõttu oleks otstarbekas edaspidi nende vormide verbaalset konteksti muuta.

Arvestama peab ka seda, et tulemused on saadud katsetingimustes, mistõttu vormikasutus võib erineda spontaanses kõnes esinevast. Seepärast tuleks meie uuringut käsitleda kui pilootuuringut esmase ülevaate saamiseks probleemist.

Uuringu põhjal selgus seega, et nii vene-eesti KK-lapsed, kes on eesti keelega kokku puutunud vähemalt kaks aastat, kui SKAP-lapsed järgivad käändemorfoloogia omandamisel ükskeelsete laste arengut. Olles käändevormide kasutamise edukuselt SKAP-lastest nõrgemad, sarnanesid KK-lapsed nendega vormikasutusvigade osas. Mõlemal rühmal esinesid valdavalt tüveallomorfivead ja vormide asendused; esines ka tunnuse ärajätte, mis oli siiski SKAP-lastele iseloomulikum veatüüp. Vigade analüüs võimaldas selgitada mõningaid käändevormide kasutuse kvalitatiivseid iseärasusi, mis vajavad kontrollimist edasiste uuringute käigus.

Lühendid

KK	astmelise kakskeelsusega laps
M	poiss
N	tüdruk
SKAP	spetsiifiline kõnearengu puue, spetsiifilise kõnearengu puudega laps
ÜK	eakohase arenguga ükskeelne laps
ÜK3	eakohase arenguga ükskeelne kolmeaastane laps
ÜK5-6	eakohase arenguga ükskeelne 5-6-aastane laps

Viidatud kirjandus

- Argus, Reili 2008. Eesti keele muutemorfoloogia omandamine. [Acquisition of Morphology in Estonian.] Tallinna Ülikooli humanitaarteaduste dissertatsioonid 19. Tallinn: Tallinna Ülikooli Kirjastus.
- Dannenbauer, Friedrich Michael 2001. Spezifische Sprachentwicklungsstörung. – Manfred Grohnfeldt (Hrsg.). Lehrbuch der Sprachheilpädagogik. Band 2. Stuttgart: Kohlhammer, 48–74.
- Hassinen, Sirje 2002. Simultaaninen kaksikielisyys. Läheiset sukukielet viro ja suomi rinnakkain. Oulu: Oulu University Press. <http://herkules.oulu.fi/isbn951426648X/isbn951426648X.pdf> (17.8.2013).
- Heina, Maarja 2011. Käändevormide kasutamine ja mõistmine spetsiifilise kõnearengu puudega lastel. [Usage and Comprehension of Case Forms by Children with Specific Language Impairment.] Magistritöö. Tartu Ülikool. http://dspace.utlib.ee/dspace/bitstream/handle/10062/23025/Heina_Maarja.pdf?sequence=1 (13.1.2013).
- Keelekümluskeskus. <http://kke.meis.ee/keelekumblusest/teadusuuringud> (25.9.2013).
- Kohnert, Kathryn 2008. Language Disorders in Bilingual Children and Adults. Second printing. San Diego, Oxford: Plural Publishing.
- Kohnert, Kathryn; Windsor, Jennifer; Ebert, Kerry Danahy 2009. Primary or “specific” language impairment and children learning a second language. – Brain and Language, 109 (2–3), 101–111. <http://dx.doi.org/10.1016/j.bandl.2008.01.009>
- Leonard, Laurence B. 2009. Cross-linguistic studies of child language disorders. – Richard G. Schwartz (Ed.). Handbook of Child Language Disorders. Hove and New York: Psychology Press, 308–324.
- Leonard, Laurence B. 1998. Children with Specific Language. Cambridge: A Bradford Book.
- Lipp, Ellen 1977. The acquisition of Estonian inflections. – Journal of Child Language, 4 (2), 313–319. <http://dx.doi.org/10.1017/S0305000900001689>

- Lukács, Ágnes; Leonard, Laurence B.; Kas, Bence 2010. Use of noun morphology by children with language impairment: the case of Hungarian. – *International Journal of Language and Communication Disorders*, 45 (2), 145–161. <http://dx.doi.org/10.3109/13682820902781060>
- Morgan, Gareth P.; Restrepo, M. Adelaida; Auza, Alejandra 2013. Comparison of Spanish morphology in monolingual and Spanish-English bilingual children with and without language impairment. – *Bilingualism: Language and Cognition*, 16 (3), 578–596. <http://dx.doi.org/10.1017/S1366728912000697>
- Orgassa, Antje; Weerman, Fred 2008. Dutch gender in specific language impairment and second language acquisition. – *Second Language Research*, 24 (3), 333–364. <http://dx.doi.org/10.1177/0267658308090184>
- Padrik, Marika 2010. Word-Formation Skills in Estonian Children with Specific Language Impairment. *Dissertationes pedagogicae Universitatis Tartuensis* 8. Tartu: Tartu Ülikooli Kirjastus.
- Padrik, Marika 2006. Milles seisneb kõnearengu puude spetsiifilisus? [What is specific in language impairment?] – *Eripedagoogika. Logopeedia ja emakeel*, 26, 13–15.
- Paradis, Johanne 2005. Grammatical morphology in children learning English as a second language: Implications of similarities with specific language impairment. – *Language, Speech, and Hearing Services in Schools*, 36 (July), 172–187. [http://dx.doi.org/10.1044/0161-1461\(2005/019\)](http://dx.doi.org/10.1044/0161-1461(2005/019))
- Paradis, Johanne 2010. The interface between bilingual development and specific language impairment. – *Applied Psycholinguistics*, 31 (2), 227–252. <http://dx.doi.org/10.1017/S0142716409990373>
- Pool, Raili 2007. Eesti keele teise keelena omandamise seaduspärasusi täis- ja osasihitise näitel. [The Acquisition of Total and Partial Objects by Learners of Estonian as a Second Language.] *Dissertationes philologiae estonicae Universitatis Tartuensis* 19. Tartu: Tartu Ülikooli Kirjastus.
- Rothweiler, Monika; Chilla, Solveig; Babur, Ezel 2010. Specific language impairment in Turkish: Evidence from case morphology in Turkish-German successive bilinguals. – *Clinical Linguistics and Phonetics*, 24 (7), 540–555. <http://dx.doi.org/10.3109/02699200903545328>
- Salo, Age 1994. Muutelõppude ilmumine ühe Eesti lapse keelde vanuses 1.5–2.5. [The emergence of inflectional endings in one Estonian child's speech, age 1;5–2;5.] – Jari Toivainen, Kirsti Toivainen (Toim.). *Ensikielenä suomalaiskieli. Turun yliopiston suomalaisen ja yleisen kielitieteen laitoksen julkaisuja* 46. Turku: Turun yliopisto, 22–28.
- Tomblin, J. Bruce 2009. Children with specific language impairment. – Edith L. Bavin (Ed.). *The Cambridge Handbook of Child Language*. Cambridge University Press, 417–431.
- Vihman, Marilyn May 1982. The acquisition of morphology by a bilingual child: A whole-word approach. – *Applied Psycholinguistics*, 3 (2), 141–160. <http://dx.doi.org/10.1017/S0142716400006676>
- Vihman, Marilyn May; Barry, McLaughlin 1982. Bilingualism and second language acquisition in preschool children. – Charles J. Brainerd, Michael Pressley (Eds.). *Progress in Cognitive Development Research: Verbal Processes in Children*. Berlin: Springer Verlag, 35–58.

Merit Hallapi (Tartu Ülikool) uurimistöö valdkonnad on üks- ja kakskeelsete laste kõne areng ja kõnepuuded.
Näituse 2, 50409 Tartu, Estonia
merit.hallap@ut.ee

Marika Padriku (Tartu Ülikool) uurimistöö valdkonnad on kõnearengupuudega laste kõne areng ja arendamine.
Näituse 2, 50409 Tartu, Estonia
marika.padrik@ut.ee

Signe Raudiku (Tartu Ülikool) uurimistöö valdkonnad on kõnepuudega laste kõne areng ja arendamine.
Näituse 2, 50409 Tartu, Estonia
signe.raudik@ut.ee

ESTONIAN CASE MORPHOLOGY IN SECOND LANGUAGE ACQUISITION AND SPECIFIC LANGUAGE IMPAIRMENT

Merit Hallap, Marika Padrik, Signe Raudik

University of Tartu

This paper presents findings from a study examining Estonian case morphology usage by typically developing Russian-Estonian sequential bilingual and monolingual Estonian speaking children with specific language impairment (SLI) aged 5–6. The control group included 5–6-year-old and 3-year-old monolingual typically developing Estonian children. Performance across groups was compared using the structured elicitation method. The findings suggest that children with SLI and typically developing bilingual children follow similar developmental paths in the early stages of grammatical acquisition. Moreover, the results reveal that the developmental patterns of both groups approximate those of monolingual Estonian language learners. Overall, the accuracy rates of case morphology production by bilingual children were lower than those of children with SLI. Both groups experienced more difficulties with the usage of the genitive, partitive, essive, terminative, abessive, and translative cases. The usage of locative grammatical cases (e.g. ablative, elative) in the secondary functions was limited as well. Both child groups basically made the same type of errors. However, qualitative analysis revealed that errors of omission of grammatical morphemes were more frequent in SLI children, while commission errors were more typical of bilingual children.

Keywords: child's speech, language acquisition, morphology, bilingualism, language and speech pathology, Estonian

SWITCHING BETWEEN ESTONIAN SIGN LANGUAGE AND SPOKEN ESTONIAN IN A SCHOOL ENVIRONMENT

Kadri Hein

Abstract. The study focuses on the variables and functions that affect switching (code-switching and code-blending) between Estonian Sign Language and spoken Estonian among bimodal bilinguals in a school environment. It was found that deaf students at Tallinn Helen's School predominantly code-blended (simultaneously spoke and signed), whereas their peers from Tartu Hiie School tended to code-switch (stopped speaking and started signing or vice versa). A deaf student's level of switching was influenced most by the interlocutor's level of switching, which highlights the importance of input provided by teachers. Although the most often used function of switching was to emphasise a constituent in a clause, the students also switched to express themselves in an original way (humour, language play) and to attract attention, while the teachers sometimes used switching for pedagogical reasons or to side-comment. In cases where it appeared problematic to apply a single function to a switch, a sequential turn-by-turn analysis (Auer 1995) was used.*

Keywords: sociolinguistics, unimodal bilingualism, bimodal bilingualism, code-switching, code-blending, Estonian Sign Language

1. Introduction

This paper investigates switching, i.e. code-switching and code-blending between Estonian and Estonian Sign Language in two deaf schools in Estonia: Tallinn Helen's School and Tartu Hiie School, and shares some aspects with research into the contact between two spoken/written languages, such as Estonian-Russian (Verschik 2008, Zabrodskaja 2005), and on the contact between a spoken and a sign language, such as English and American Sign Language (Emmorey et al. 2008,

* The present paper is based on a PhD thesis written at Birkbeck, University of London (2011) and it has been conducted with the help of the following funding bodies: Kristjan Jaak scholarship and state doctorate studies scholarship from Estonia, College Research Studentship from Birkbeck College, scholarship from Estonian Students' Fund in USA, Margot and Herbert Linna scholarship from the Estonian World Council, a scholarship from Alfred Kordelin Foundation, and a bursary from the Estonian Educational Trust in the UK.

Bishop, Hicks 2005). As the majority of the studies on bimodal bilingualism have focused on switching by hearing bilinguals, this paper takes a different perspective, including deaf and hearing teachers, as well as deaf students.

The most noteworthy aspect of the contact between a spoken and a sign language is that elements of the two languages from different modalities (a visual-spatial and an auditory-vocal language) can be combined both sequentially and simultaneously. Following researchers such as Emmorey et al. (2008), it is expected that the deaf students simultaneously produce signs and speech, i.e. they code-blend, and that the Matrix Language of code-blends is mainly Estonian. Also, it is hypothesised that the majority of the word classes in code-blends and code-switches are nouns (Muysken 2004: 153, Van den Bogaerde, Baker 2008).

The analysis of variables and functions that affect switching between spoken Estonian and Estonian Sign Language (ESL) in a school environment constitutes the heart of this paper. The relationship between variables such as hearing status, level of hearing loss, parental hearing status, hearing status of the interlocutor, Estonian language proficiency, interlocutor's switching and the number of switches was investigated. In addition, the functions of switching and speech acts were analysed. In cases where it appeared problematic to determine a single function of a switch, the interactions were investigated turn by turn.

2. Code-switching in unimodal and bimodal bilingualism

In unimodal bilingualism¹, several researchers have focused either on the grammatical (Myers-Scotton 1997) or the sociolinguistic view (Auer 1984, 1995, Li Wei 2002) of code-switching. This paper includes features from each framework with the aim of giving a more objective perspective of the topic studied. According to Myers-Scotton's (1997) Matrix Language Frame (MLF) model for code-switching, one language dominates (Matrix Language, or ML) over the non-dominant language (Embedded Language, or EL) by supplying more words and morphemes to the sentence. The grammatical aspects of switching in bimodal bilingualism, i.e. in the contact between a spoken/written language and a sign language, have primarily been studied by relying on Coda² communication. Emmorey et al. (2008) found that Codas prefer code-blending to code-switching when conversing with other bimodal bilinguals with English being the Matrix Language. For bimodal bilinguals, the most often used word classes in code-blends and code-switches were verbs, as opposed to unimodal bilinguals, who show more code-blending in nouns (Muysken 2004: 153). The results are mixed though, as according to Van den Bogaerde and Baker (2008) nouns dominated over verbs in code-blends produced by bimodal bilinguals. Besides word classes, semantics and frequency of use have to be taken into account (Backus, Verschik 2012). Bishop and Hicks (2005) studied the written language of Codas and found that Coda-talk is characterised by systematic code-blending with some features of American Sign Language.

The sociolinguistic approaches have been concerned with different issues, including the functions of code-switching. Zabrodska (2005) compiled the

¹ In the present article, unimodal bilingualism refers to the contact between two spoken/written languages. In principle, the contact between two sign languages is also unimodal (Adam 2013: 21).

² The term Coda is used to refer to a hearing child of deaf adults.

following taxonomy for switches: reported speech; language play; expressive function; side-comment; reiteration; relationship between the participants; and insufficient language skills. Other reasons for code-switching include addressee specification, emphasising a constituent in a sentence, attracting and retaining attention (e.g. McClure 1981), presenting an attitude or opinion (Praakli 2009: 105), participant make-up, activity type, topic shift, and topic-comment structure (Auer 1995: 120).

Although researchers have studied particular functions for switching in bimodal bilingualism (e.g. Gerner de Garcia 1995), they have been more involved in discussing the effect of different variables on code-switching, such as the hearing status (Van den Bogaerde 2000), hearing status of the parents (Hoffmeister, Moores 1987) and the interlocutor (Lucas, Valli 1995, Grosjean 1996), and language proficiency (Kuntze 2000). This study focuses on the relationship between hearing status, level of hearing loss, parental hearing status, hearing status of the interlocutor, Estonian language proficiency, interlocutor’s switching and the amount of switching.

The proponents of the Conversation Analysis (CA) approach have emphasised the need to analyze how meaning emerges in interaction through code-switching (Li Wei 2002: 167), as a general taxonomy of the functions cannot possibly account for all cases of switching. Auer (1995) distinguishes four patterns of code choice in conversation (Table 1): type I switches are discourse-related, type II switching is seen as preference-related, type III switching may be both discourse and participant-related, and type IV is termed *transfer*.

Table 1. The patterns of code-switching (Auer 1995: 125–126)

Name	Pattern
Ia	A1 ³ A2 A1 A2//B1 B2 B1 B2
Ib	A1 A2 A1 A2 A1//B1 B2 B1 B2
IIa	A1 B2 A1 B2 A1 B2 A1 B2
IIb	A1 B2 A1 B2 A1//A2 A1 A2 A1
IIIa	AB1 AB2 AB1 AB2
IIIb	AB1//A2 A1 A2
IV	A1[B1]A1

In bimodal bilingualism, researchers have also stressed the need to consider the context in which the conversation takes place, e.g. Kuntze (2000: 293) filmed classes in a classroom with both a bilingual teacher and bilingual students, which may have triggered more code-switching than in other contexts.

3. Estonian Sign Language

Estonian Sign Language⁴ with the SIL code ESO (Sutrop 1998) is a language used by the members of the Estonian Deaf community (1500–2000 signers). Users of ESL are concentrated around bigger towns in Estonia, such as Tallinn, Tartu and Pärnu. ESL was officially recognised on 1 March 2007 (Language Act 2007).

³ The letters refer to the languages A and B; the numbers indicate the speakers 1 and 2.

⁴ Paabo (2010) gives a thorough overview of ESL.

Estonian Sign Language is a topic-comment language with a relatively flexible sign order (Laipea 1992: 2101). There is a tendency to place interrogative signs at the end of the sentence in ESL. The three main types of questions in ESL are yes/no questions, alternative questions and *wh*-questions. It is also possible to form rhetorical questions, which are technically questions, but do not request an answer. In ESL, there are three different ways to modify nouns (Miljan 2001: 176): the combination of a modifier and a noun; bracketing, in which case the modifier precedes and follows the head noun or the base sign; pre-nominal and post-nominal modifiers.

Systematic teaching of ESL began in 1990 when Regina Toom started teaching it to the students of the Special Education Department at the University of Tartu. Courses have also been taught by the Estonian Association of Parents with Hearing Impaired Children, the Estonian Association of the Deaf, and Tallinn Pedagogical College. ESL interpreter training has been offered at the University of Tartu since 2006.

4. Database

The material was videotaped in two deaf schools in Estonia: Tartu Hiie School and Tallinn Helen's School⁵. Tartu Hiie School mainly uses the oral method in educating deaf students, relying on speaking, lip-reading, fingerspelling and writing. The Estonian language is taught as the mother tongue and English as a foreign language. Tallinn Helen's School provides bilingual education for the deaf students, which means that ESL is taught as the first language and Estonian as the second language. Both hearing and deaf teachers work at the school, instructing mainly in ESL and (written) Estonian⁶. In lessons where the hearing teacher uses Estonian, sign language interpreting service is offered.

The filmed lessons lasted approximately 299 minutes and consisted of 5322 utterances⁷. Seven videotaped lessons were filmed in Tallinn Helen's School (195 minutes), and three in Tartu Hiie School (104 minutes). The lessons filmed in Tallinn Helen's School consisted of interactions between students and teachers in seven lessons: Estonian as a second language (three lessons), mathematics, history, Estonian Sign Language, and biology. Four different classes of 27 deaf students aged 12–21 were videotaped. The lessons were taught by six different teachers, four deaf and two hearing. In Tartu Hiie School, the filmed material consisted of interactions between the students and the teachers in three lessons: Estonian, social studies, and English, which were taught by two hearing teachers. Eight students aged 14–16 participated in all three lessons.

5. Transcription and analysis

Three aspects were analysed in depth by using the ELAN software (Language Archiving Technology 2008): ESL, Estonian, and word classes⁸. Pauses and latching were separately marked (Table 2).

⁵ Loit (2013) provides more information about teaching deaf children in Estonia.

⁶ The situation is changing, as more hearing impaired students enter Tallinn Helen's School – their speaking skills need to be developed.

⁷ An utterance was usually surrounded by longer pauses.

⁸ The distinction of word classes in the current study was mainly made considering semantics and pragmatics.

Table 2. Transcription conventions

[start of simultaneous talk
]	end of simultaneous talk
=	latching (no interval between turns)
(.)	micro-pause (max. 0.2 seconds)
(2.0)	length of silence in seconds
h h h	laughter
{--}	incomprehensible or inaudible utterance

The names of students were replaced by codes, including S for student, T for teacher, a number, and D or H, referring to the hearing status of the informant's parents. The students' names in the excerpts were replaced with common names in Estonia. The analysis of each guided discussion started quantitatively followed by the functional analysis of excerpts. ML and EL in code-blends were determined by considering features of ESL and Estonian, as EL and ML in code-blends of Estonian and Estonian Sign Language cannot be determined by counting the number of morphemes, because the modalities and morphologies differ significantly. ESL was considered to be the ML if role-shift⁹ and classifiers¹⁰ were used. Estonian was the ML if a word was fingerspelled¹¹ instead of signing. Some aspects were analysed in more detail: word order,¹² expressing time, question formation, the category of number, adjectival modification, and agent nouns, due to differences in the two languages.

Pearson's product-moment correlation coefficient or Pearson's *r* was employed to investigate the relationship between the number of switches and the following variables: level of hearing loss, parental hearing status, hearing status of the deaf person's interlocutor, scores on an Estonian language proficiency test, and the interlocutor's switches.

The functions of code-switching were determined by combining the taxonomies of Zabrodskaja (2005) and McClure (1981), to which the category of pedagogical reasons was added. Speech acts were determined according to Searle's (1976) typology (statement, question, correction, request, doubt, order, refusal, explanation, recommendation, and thanking). Sometimes Austin's (1962) classification of speech acts into locutionary, illocutionary, and perlocutionary acts was also used.

The language background of the students was investigated by using a questionnaire, and the Estonian Language level was measured with a test designed on the basis of an exit exam for year 9¹³ students in Tallinn Helen's School¹⁴. The sign language levels of the deaf and hearing participants were determined by a deaf expert (Maret Õun) according to the Estonian Sign Language Proficiency Scale (compiled by Regina Toom). The deaf students from Tartu Hiie School with slight hearing loss received the best results in the Estonian language proficiency test ($M = 57$, $SD = 22.7$), followed by the deaf students from Tartu Hiie School with profound hearing loss ($M = 44.8$, $SD = 25.0$), and deaf students from Tallinn Helen's School

⁹ Role-shift is the possibility of shifting between the characters in a narration (Laipea 2007: 59).

¹⁰ Classifiers are morphemes used for indicating spatial relations, movement, and the shape and size of the objects (Sandler, Lillo-Martin 2006).

¹¹ Fingerspelled words were not considered as code-switches in this study. Fingerspelling is an outcome of language contact between a sign language and a spoken language and should be treated separately.

¹² The word order of an affirmative sentence in Estonian is usually SVO. ESL is a topic-comment language (Laipea 1992: 2101).

¹³ Year 9 corresponds to year 8 in regular schools.

¹⁴ The questionnaire and the language test can be found in Hein (2012: 329–350).

($M = 44.5$, $SD = 24.8$). The deaf students from Tallinn Helen’s School had better results (B1-B2) in ESL than the ones from Tartu Hiie School (A1-A2) and the hearing students from Tallinn Helen’s School (A1-A2).

6. Quantitative aspects of switching

The percentages of code-switches, code-blends and switches in Table 3 show that the students from Tallinn Helen’s School mainly code-blended (23.0%), whereas the students from Tartu Hiie School produced more code-switches (43.2%). Altogether, the percentage of switches was 28.5 in Tallinn Helen’s School and 60.2 in Tartu Hiie School.

Table 3. Students’ mean percentages of utterances containing code-blends, code-switches and switches

Students	Code-blends	Code-switches	Switches
Tallinn Helen’s School lessons	23.0	8.1	28.5
Tartu Hiie School lessons	27.2	43.2	60.2
Mean	25.1	25.7	44.4

As seen from the table, interlocutor’s switches influenced the interaction most: in Tallinn Helen’s School, Pearson’s r shows a significant positive correlation between the number of switches produced by the teachers and students, $r = .93$, $p \leq .01$ (one-tailed), and in Tartu Hiie School, there is a strong positive correlation between the number of teachers’ switches and the mean number of the students’ switches, $r = .90$, $p > .05$ (one-tailed).

When analysing the lessons separately, it appeared that most of the students’ switched utterances in Tallinn Helen’s School occurred in the lessons taught by hearing teachers ($M = 47.7\%$), whereas the main form of communication with deaf teachers was Estonian Sign Language, as only 9.3% of the students’ utterances contained some type of switching. In Tartu Hiie School, the percentage of switches was at the same level in all three lessons: Estonian (60.6%), social studies (63%), and English (58.7%).

ESL proficiency did not significantly influence the results in either school; the students with lower levels of ESL (B1 and less) switched slightly less than the students with higher level of ESL (B2 and above) in Tallinn Helen’s School. The former produced 22% switches, while the latter produced 28% switches on average. In Tartu Hiie School, the students with a lower level of ESL proficiency (A1) switched slightly more often than those with higher levels of ESL proficiency (A2 and above). The mean percentage of switches was 61.5 in the former and 56.9 in the latter group.

Table 4. Correlation between variables and the level of switching

Variables	Tallinn Helen's School lessons	Tartu Hiie School lessons
Level of hearing loss	$r = .10$	$r = -.29$
Parental hearing status	$r = -.32^*$	–
Estonian language proficiency	$r = -.06$	$r = .04$
Interlocutor's switches	$r = .93^*$	$r = .90$

As indicated in Table 5, the students from Tallinn Helen's School had slightly more verbs in code-blends (30.1%), whereas more nouns occurred in the code-blends of the students from Tartu Hiie School (44.2%). No major difference occurred between the nouns and verbs in code-switches.

Table 5. Percentages of nouns and verbs in code-switches and code-blends

Variables	Tallinn Helen's School lessons		Tartu Hiie School lessons	
	Verbs (2185)	Nouns (2000)	Verbs (1996)	Nouns (2391)
Code-blends	30.1	25.0	26.8	44.2
Code-switches	28.4	29.0	28.1	27.9

The Matrix Language (ML) of the utterances produced by the students proved to be Estonian (R¹⁵ 61.0%, E1¹⁶ 48.3%, E2¹⁷ 54.4%), and the Accompanying Language in the case of code-blends was Estonian Sign Language (R 14.0%, E1 18.0%, E2 14.3%), which may partly be caused by the nature of the data, as some of the videotaped data came from Estonian language lessons. In Tartu Hiie School, most of the code-blended utterances also appeared to have Estonian as the ML (R 20.4%, E1 29.5%, E2 28.9%) rather than Estonian Sign Language (R 6%, E1 9.1%, E2 8.6%).

The ML could not always be determined, due to too short utterances and semantically non-equivalent code-blends¹⁸, some of which consisted of lexical items produced with differing word order in each language, making the utterance syntactically correct both in Estonian and Estonian Sign Language.

- (1) T1H: shake – VASTAMA MITTE
 ei sa ei vasta
 not you: 2SgNom NEG/answer answer/no
 'You don't answer.'

In the first line there are signs in ESL, the second line contains spoken Estonian, the English glosses¹⁹ are in the third line, and the fourth line translates the utterance (example 1) The negative particle *ei* 'not' precedes the verb *vastama* 'to answer' in Estonian, while in ESL negation is indicated post-verbally. This means of code-blending enables access to grammatically correct input for both a hearing and deaf audience. As both the characteristics of ESL and Estonian are present, determining the ML proves to be problematic, as also found for some spoken language pairs, e.g. in Russian-Estonian contact (Zabrodskaia 2009: 60).

¹⁵ Researcher's opinion.

¹⁶ First expert.

¹⁷ Second expert.

¹⁸ See Emmorey et al. (2005, 2008) for the categories of semantically non-equivalent code-blends in American Sign Language-English hearing bimodal bilinguals.

¹⁹ See Paabo (2010) for ESL glossing conventions.

7. Functions of switching

7.1. Emphasising a constituent

The students frequently switched in order to emphasise a constituent in a clause, which often resulted in fingerspelling in ESL and frequently occurred with proper names, such as names of people and places, as well as words that the subjects were not familiar with.

A new term is emphasised in (2) which originates from Tartu Hiie School. Here, S31H utters the recommendation in Estonian only, which remains unclear to the hearing teacher, so he/she fingerspells the new term to emphasise it in a question after a 0.5-second pause. S31H and S29H then explain the term to the teacher by code-blending.

(2)

- S31H: metroo või-ks ka olla
metro: can-Cond also be: Inf
NomSg
'We could also have a metro.'
(0.5)
- T8D: – – – – m-e-t-r-o-o
metroo (.) kohvik saa-b teha metroo
metro: café: can-3Sg make: metro:
NomSg NomSg Inf NomSg
'The café can make a metro?'
- S31H: – (.) SAI LÕIKAMA PANEMA
jah (.) sai lõika-n poole-ks pane-n pea-le
yes baguette: cut-Pr1Sg half-Tr1Sg put-Pr1Sg on-All
NomSg
'Yes. I cut a baguette in half and put stuff on it.'
- S29H: TEADMA m-e-t-r-o-o
tea-d metroo
know-Pr2Sg metro: NomSg
'You know metro?'

7.2. Expressive function

Deaf students sometimes switched for expressive purposes: to make fun, to be original, or to play with language. In (3), the hearing teacher and students from Tallinn Helen's School are discussing if it is possible to guess the character of a person by reading his/her self-description. The first utterance by T2H functions as a question.

(3)

T2H: või kui ta ütle-b, et (.) [mulle
or if he: 3SgNom say-Pr3Sg that I: AllSg
meeldi-b kan-da kleit-e ja seelikui-d ja
like-Pr3Sg wear-Inf dress-PartPl and skirt-PartPl and
ma ei armasta eriti pükse]
I: NomSg NEG love particularly trousers:
PartPl

‘Or if she says that I like to wear dresses and skirts and I do not particularly like trousers.’

I1D: [ÜTLEMA KLEIT MEELDIMA PÜKSID MITTE-MEELDIMA]
say dress like trousers not like

‘If she says that I like to wear dresses and I do not like trousers.’

S12H: BEIB

beib

babe: NomSg

‘Babe.’

I1D: beib

babe: NomSg

‘Babe.’

T2H: [h h h]

S12H: [h h h]

In answering the teacher’s question, the student emphasises a specific type of image-obsessed girl (BEIB ‘babe’) by signing and speaking at the same time, which serves an expressive and humorous function that would not be revealed if only one language were employed. The interaction, which in terms of speech acts consists of statements only, ends with the simultaneous laughter of T2H and S12H.

7.3. Attracting attention

Another interaction where switching is used for expressive purposes occurs at the end of the Estonian class in Tallinn Helen’s School taught by a hearing teacher fluent in ESL.

(4)

S1H: kas kodust ülesanne-t ei ole vä
if home: PartSg task-PartSg not be or
‘There’s no homework?’

(1.5)

S1H: MARI

Mari

‘Mari!’

(0.5)

S1H:	KODU^ÜLESANNE		POLE
	homework		NEG + be
	‘There’s no homework?’		
T1H:	–	KODU^ÜLESANNE	POLE
	ee (.)	koduülesanne-t	ei ole
	ee	homework-PartSg	NEG + be
	‘There’s no homework.’		
S1H:	JESS	(.) –	
	–	(.) bingo	
	yes	bingo	
	‘Yes. Bingo!’		

First, S1H asks T1H in colloquial Estonian if there is homework, and as there is no answer in 1.5 seconds he/she calls the teacher’s name to attract attention. After another pause (0.5 seconds) the student reiterates the utterance in ESL to receive attention. The teacher first fills the pause and then produces the answer by code-blending, which is warmly received by the students, as seen in the utterance by S1H. The fact that both Estonian and ESL are represented in this utterance indicates that in language negotiation, the student is moving towards the teacher’s code choice – code-blending. As for speech acts, this excerpt mainly follows the question-statement pattern.

7.4. Side-comment

The teachers at Tartu Hiie School also produced side-comments, which were mainly uttered in Estonian.

(5)				
T8D:	–	–	–	SALAT
	valik	on	kas	sinu-le salat
	choice: NomSg	be: Pr3	if	you-AllSg salad: NomSg
	MEELDIMA		MITTE_MEELDIMA	–
	meeldi-b	või	ei	meeldi kirjuta
	like: Pr3Sg	or	NEG	like write: Imp2Sg
	‘The choice is: do you like salad or not; write.’			
	(1.0) (schoolbell is ringing)			
S28H:	salat	ei	taha	
	salad: NomSg	NEG	want	
	‘I don’t want salad.’			
T8D:	no		siis	kirjuta
	well		then	write: Imp2Sg
	‘Write then.’			
	(5.5) (students are writing)			

T8D: pea-me selle ikka vist valmis tege-ma
 have to-Pr1Pl it: GenSg still probably ready make-Inf
 ega me muidu ei pääse
 or we: NomPl otherwise NEG get away
 ‘We probably have to finish this; we don’t get away without it.’

There are two functions of the code choice in example (5): the make-up of the participants involved and side-commenting. The teacher starts with code-blending in giving the order, but as S28H answers in Estonian, he/she also switches to Estonian. This choice may be influenced by the fact that the student sits in the first row and can clearly see the teacher’s mouth patterns. After that there is a 5.5-second silence, as the students are filling in their questionnaires. The bell rings, and the teacher states that they cannot end the lesson at this stage, as the questionnaire has not been completed yet. This side-comment is produced in Estonian only, and after that the teacher continues talking in the code-blended mode.

7.5. Pedagogical reasons

A teacher from Tartu Hiie School occasionally switched to make information explicit or asked a student fluent in ESL to explain a term to other signers. In (6) the teacher asks a student in the code-blended mode if he/she knows what an X-ray is with an illocutionary force to receive an answer about an X-ray. As a perlocutionary effect, he/she receives an affirmative answer even before the end of the utterance. The teacher then asks S30H to explain the term to others, which can be seen as an order, as the word *please* is not used. S30H’s interruption of T8D in the middle of the utterance allows the teacher to continue talking, as it is signed. S30H then explains to the other students in ESL what an X-ray is in a question format.

(6)

T8D: – – – – – r-ö-n-t-[g-e-n]
 kas sina tea-d mis on rönt[gen]
 if you: 2SgNomknow-Pr2Sg what: NomSg be: Pr3 X-ray: NomSg
 ‘Do you know what an X-ray is?’

S30H: [PILT]
 [pilt]
 picture: NomSg
 ‘Picture.’

T8D: – – SELETAMA –
 voh(.) [näe-d] seleta nei-le
 well see-Pr2Sg explain-Imp2Sg they-AllPl
 ‘See? Tell them.’

S30H: [TEADMA]
 know
 ‘I know.’

8. Conclusions

The results show that the students in Tallinn Helen's School predominantly code-blend rather than code-switch, as also found in previous studies on bimodal bilingualism (e.g. Emmorey et al. 2008), whereas the students from Tartu Hiie School code-switch more often than code-blend, probably due to their confidence in using spoken Estonian in an oral deaf school.

Most of the switched utterances in Tallinn Helen's School occurred in the lessons taught by hearing teachers, while the main form of communication with deaf teachers was ESL. Pearson's *r* showed a strong positive correlation between the number of switches produced by teachers and students in the two schools, which should be borne in mind when teaching deaf students – the available input influences the language production to a great extent. Thus, if it is aimed that the students use both ESL and Estonian, then these two languages should also be employed in the classroom.

In both Tallinn Helen's School and Tartu Hiie School there was no major difference between the percentage of verbs and nouns in code-switches. In code-blends, verbs were more often used than nouns in Tallinn Helen's School, and in Tartu Hiie School, the results were in the opposite direction. Previous research shows mixed results: Emmorey et al. (2008) found that verbs tend to occur significantly more often than nouns in the code-blends and code-switches of hearing bimodal bilinguals, and Van den Bogaerde and Baker (2008) found the opposite: more nouns than verbs were code-blended. Besides word classes, other aspects have to be considered, such as frequency and meaning, which are linked to different types of switching (Backus, Verschik 2012).

Although it was sometimes difficult to determine the ML and EL, as also noted in some spoken language pairs (Zabrodskaia 2009), the analysis shows that the Estonian syntactic structure was utilised more often than that of Estonian Sign Language. The results are in accordance with previous findings in bimodal bilingualism, as Emmorey et al. (2008) state that the spoken/written language was the ML in most of the utterances.

The most often used function of switching in the two schools was emphasising a constituent in a clause. The students in Tallinn Helen's School also switched to attract the attention of the teacher and to express themselves in an original way (humour, language play). The teachers in the two schools occasionally switched to side-comment in Estonian, perhaps because deaf people do not often sign alone, but it is normal for the hearing people to side-comment in Estonian. The use of this function has not been thoroughly researched in the previous studies on bimodal bilingualism and should be further investigated.

The main difference in the functions of switching between the two schools lies in the fact that the students at Tartu Hiie School used switching less for attracting the attention of the teacher, as the teachers called on the students. In Tallinn Helen's School, some lessons consisted of independent work by the students, during which they had to attract the teacher's attention in order to ask for advice.

In conclusion, the analysis of the lessons shows that the deaf students and teachers at Tallinn Helen's School and Tartu Hiie School switched for various reasons. Besides emphasising the constituent, the informants also code-switched to express

themselves in an original way, for the purpose of humour, and for pedagogical reasons. Thus, the teachers should be aware that the lexical gap does not influence switching much; other functions play a greater role. As in some utterances the code choice was caused purely by the code choice of the previous turn, they were analysed by using the sequential model of code-switching compiled by Auer (1995), which has not been done in studying sign language – spoken language switching before. The results show that code choice may change in time, as in both deaf schools, accommodation was observed at the end of the lesson: the students spoke more often and the teacher signed more.

This is the first study on switching between Estonian Sign Language and spoken Estonian, which hopefully sheds some light on the nature of bimodal bilingual communication and may assist in teaching deaf children. Future research could focus on the functions of switching between Estonian Sign Language and spoken Estonian in conversations between hearing bimodal bilinguals.

References

- Adam, Robert 2013. Abstract book of the 9th International Symposium on Bilingualism (ISB9). Singapore: ISB9 Local Organising Committee.
- Auer, Peter 1984. *Bilingual Conversation*. Amsterdam: John Benjamins. <http://dx.doi.org/10.1075/pb.v.8>
- Auer, Peter 1995. The pragmatics of code-switching: A sequential approach. – Lesley Milroy, Pieter Muysken (Eds.). *One Speaker, Two Languages: Cross-disciplinary Perspectives on Code-Switching*. Cambridge: Cambridge University Press, 115–135.
- Austin, John Langshaw 1962. *How to Do Things with Words*. Oxford: Oxford University Press.
- Backus, Albert Marie; Verschik, Anna 2012. Copiability of (bound) morphology. – Lars Johanson, Martine Robbeets (Eds.). *Copies versus Cognates in Bound Morphology*. Leiden-Boston: Brill, 123–149.
- Bishop, Michele; Hicks, Sherry 2005. Orange eyes: Bimodal bilingualism in hearing adults from deaf families. – *Sign Language Studies*, 5 (2), 188–230. <http://dx.doi.org/10.1353/sls.2005.0001>
- ELAN = Language Archiving Technology. Max-Planck-Institute for Psycholinguistics. <http://www.lat-mpi.eu/tools/elan> (1.1.2012).
- Emmorey, Karen; Borinstein, Helsa B.; Thompson, Robin; Gollan, Tamar H. 2008. Bimodal bilingualism. – *Bilingualism: Language and Cognition*, 11 (1), 43–61. <http://dx.doi.org/10.1017/S1366728907003203>
- Gerner de Garcia, Barbara 1995. Communication and language use in Spanish-speaking families with deaf children. – Ceil Lucas (Ed.). *Sociolinguistics in Deaf Communities*. Washington, D.C.: Gallaudet University Press, 221–252.
- Grosjean, François 1996. Living with two languages and two cultures. – Ila Parasnis (Ed.). *Cultural and Language Diversity and the Deaf Experience*. Cambridge: Cambridge University Press, 20–37.
- Hein, Kadri 2012. *Switching Between Estonian Sign Language and Spoken Estonian: Functions and Factors Affecting Language Use among Deaf and Hearing Bimodal Bilinguals*. Saarbrücken: LAP Lambert Academic Publishing.
- Hoffmeister, Robert J.; Moores, Donald F. 1987. Code switching in deaf adults. – *American Annals of the Deaf*, 132 (1), 31–34. <http://dx.doi.org/10.1353/aad.2012.0770>
- Kuntze, Marlon 2000. Codeswitching in ASL and written English language contact. – Karen Emmorey, Harlan Lane (Eds.). *The Signs of Language Revisited*. Mahwah, NJ: Lawrence Erlbaum Associates Publishers, 287–302.
- Laiapea, Vahur 1992. Mis on viipekeel. [What is Sign Language.] – *Akadeemia*, 4, 2098–2136.

- Laiapea, Vahur 2007. Keel on lahti. Tähendusi viipekeelest. [Notes on Sign Language.] Tallinn: Eesti Keele Sihtasutus.
- Language Act. <https://www.riigiteataja.ee/ert/act.jsp?id=12796913> (20.11.2007).
- Li Wei 2002. 'What do you want me to say?' On the Conversation Analysis approach to bilingual interaction. – *Language in Society*, 31 (2), 159–180. <http://dx.doi.org/10.1017/S0047404501020140>
- Loit, Raili 2013. Viipekeelne kurt laps koolis. Juhendmaterjal õpetajale. [Signing Deaf Child at School. A Guide for Teachers.] Tallinn: SA Innove. <http://www.hev.edu.ee/?id=246> (1.2.2014).
- Lucas, Ceil; Valli, Clayton 1995. *Linguistics of American Sign Language. An Introduction*. Second edition. Washington, D.C.: Gallaudet University Press.
- McClure, Erica 1981. Formal and functional aspects of the codeswitched discourse of bilingual children. – Richard P. Duran (Ed.). *Latino Language and Communicative Behavior*. Norwood, NJ: ABLEX, 69–94.
- Miljan, Merilin 2001. Adjectival modification in Estonian and Estonian Sign Language. – *Estonian Typological Studies V*. Publications of the Department of Estonian of the University of Tartu 18. Tartu: University of Tartu, 169–188.
- Muysken, Pieter 2004. Two linguistic systems in contact. – Tei K. Bhatia, William C. Ritchie (Eds.). *The Handbook of Bilingualism*. Oxford: Blackwell, 147–168.
- Myers-Scotton, Carol 1997. *Duelling Languages: Grammatical Structure in Codeswitching*. Second edition. Oxford: Clarendon Press.
- Paabo, Regina 2010. Viibelda on mõnus. Käsiraamat eesti viipekeele õppimiseks. [Signing is Pleasant. A Handbook for Learning Estonian Sign Language.] Tallinn: Eesti Keele Sihtasutus.
- Praakli, Kristiina 2009. Esimese põlvkonna Soome eestlaste kakskeelne keelekasutus ja koodikopeerimine. [Bilingual Language Use by First Generation Estonians in Finland and Code-Copying.] Tartu: Tartu Ülikooli Kirjastus.
- Sandler, Wendy; Lillo-Martin, Diane 2006. *Sign Language and Linguistic Universals*. Cambridge: Cambridge University Press. <http://dx.doi.org/10.1017/CBO9781139163910>
- Searle, John Rogers 1976. A Classification of Illocutionary Acts. – *Language in Society*, 5 (1), 1–23. <http://dx.doi.org/10.1017/S0047404500006837>
- Sutrop, Urmas 1998. *Estonian Sign Language*. – Raymond Gordon Jr. (Ed.). *Ethnologue: Languages of the World*. Dallas, Texas: SIL International. <https://www.ethnologue.com/language/eso> (2.2.2014).
- Van den Bogaerde, Beppie 2000. Input and interaction in deaf families. – *Sign Language and Linguistics*, 3 (1), 143–151. <http://dx.doi.org/10.1075/sll.3.1.12bog>
- Van den Bogaerde, Beppie; Baker, Anne 2008. Code-mixing in signs and words in input to and output from children. – Carolina Plaza-Pust, Esperanza Morales-López (Eds.). *Sign Bilingualism. Language Development, Interaction, and Maintenance in Sign Language Contact Situations*. Amsterdam/Philadelphia: John Benjamins Publishing Company, 1–27.
- Verschik, Anna 2008. *Emerging Bilingual Speech: From Monolingualism to Code-Copying*. London: Continuum Publishers.
- Zabrodskaja, Anastassia 2005. Vene-eesti koodivahetus Kohtla-Järve vene emakeelega algkoolilastel. [Russian-Estonian Code-Switching among Russian-Speaking Primary School Children in Kohtla-Järve.] Tallinn: Tallinna Ülikooli Kirjastus.
- Zabrodskaja, Anastassia 2009. *Russian-Estonian Language Contacts: Grammatical Aspects of Language Use and Change*. PhD thesis. Tallinn: Tallinna Ülikooli Kirjastus.

EESTI VIIPEKEELE JA SUULISE EESTI KEELE VAHELINE KOODIMUUTUS KOOLIKESKKONNAS

Kadri Hein

Tallinna Heleni Kool, Tallinna Ülikool

Artiklis uuritakse, mis näitajad ja funktsioonid mõjutavad koodimuutust (koodivahetust ja koodisegunemist) eesti viipekeele ja eesti keele vahel bimodaalsete kakskeelsete hulgas. Tundidest kogutud andmed näitasid, et Tallinna Heleni Kooli õpilased eelistavad koodisegunemist, st nad viipevad ja räägivad samaaegselt, ning Tartu Hiie Kooli õpilased kasutavad sagedamini koodivahetust, st lõpetavad viipemise ja alustavad rääkimist või vastupidi. Kuigi koodisegunemise maatrikskeel oli enamasti eesti keel, jäi see mõnikord määramata, kuna bimodaalsetel kakskeelsetel on võimalus kasutada eesti viipekeelt ja eesti keelt samaaegselt erineva sõnajärgiga.

Kurdi õpilase koodimuutuse tase sõltus kõige enam sellest, mil määral õpetaja koodi muutis. Koodimuutuse peamine funktsioon õpilaste seas oli lauseosa rõhutamise, kuid koodi muudeti ka huumori ja keelemängu eesmärgil. Lausungites, kus koodimuutuse funktsiooni määramine osutus keeruliseks, analüüsiti vestlust voor vooru järel (Auer 1995).

Võtmesõnad: sotsiolingvistika, bimodaalne kakskeelsus, koodivahetus, koodisegunemine, eesti viipekeel

Kadri Hein's (Tallinn Helen's School, Tallinn University) main research interests are sociolinguistics, bilingualism and Estonian Sign Language.

Ehte 7, 10318 Tallinn, Estonia

kadri.hein@helen.edu.ee

TÕHUSUSE TAHUD SÕJANDUSTERMINOLOOGIA NÄITEL: ÜLEDIFERENTSIMINE VÕI ERINEV TUNNETUSVÄÄRTUS?

Reet Hendrikson

Ülevaade. Artikkel käsitleb terminoloogilist eristamisvajadust: sõjanduse tekstikatkete abil otsin vastust küsimusele, kust läheb piir ülediferentsimise ja erialase mõttetäpsuse huve teeniva eristamisvajaduse vahel. Klassikalise terminiteooria põhiseisukohti on, et erialakeele toimivuse ja ühemõttelisuse tagab fikseeritud kood. Seetõttu on iga-sugust varieerumist kas taunitud või aktsepteeritud mõõndustega ning eeldatud, et mõistete vahel on selged piirid.

Tulemustest selgub, et spetsialistide arusaamad tunnetusväärtusest ja diferentsimisvajadusest erinevad vägagi, mistõttu hägustuvad oskuskeelenähtuste piirid (nt piir sünonüümia, hüponüümia ja polüseemia vahel), tekitades terminipuntraid. Keerukate mõistesuhete puhul võib just vaste laenuandvas keeles olla spetsialistile võti mõiste tuvastamisel, kuigi praktikas on laenuandva keele eeskujul siiani sageli taunitud. Diferentsimise ja ülediferentsimise piiri leidmist komplitseerivad termini kujunemislugu ja kaasnenud mõistem muutused, samuti sotsiokultuurilised tegurid (nt võimusuhted). Ühene pole ka terminikorrastuse mõju: korrastamise nimel langetatud otsused võivad otsuse leviku algfaasis segadust juurdegi tekitada. Analüüsist järeldub, et diferentsimise ja ülediferentsimise piiri otsingul on üldkehtivate reeglite paikapanek ebaotstarbekas. Pigem tuleks küsida, milliseid aspekte iga konkreetse otsuse puhul arvestada.

Võtmesõnad: sõjanduskeel, terminoloogia, oskuskeelekorraldus, terminivariantide eristamine, terminivariantide tunnetusväärtus, eesti keel

1. Teoreetiline taust ja probleem

Esmalt kolmest siinse artikli kesksest mõistest: tõhusus, ülediferentsimine ja terminivariantide tunnetusväärtus. Erialakeele tõhusust on võimalik määratleda erineval moel, kuid siinses lähtun eeldusest, et **tõhusus** on hea erialakeele põhilisi

defineerivad tunnuseid (Cabrè 1999: 47, Mereste 2000: 224, Erelt 2007: 18, Tavast 2008: 134). Kui eeldada, et erialakeele põhifunktsioone on võimalikult täpne infoedastus, siis on tõhus selline erialasuhtlus, mille puhul kõneleja (sõjanduskontekstis nt käsuandja) saavutab oma kõneteoga suhtluseesmärgi.

Tõhususe saavutamisel on oluline ühelt poolt lahus hoida mõisteid, mida pole üldkeeles tarvidust eristada. Teisalt on vähemalt sama tähtis mitte luua põhjendamatu veelahet seal, kus see pole (keeleliselt) võimalik või (erialaselt) vajalik. Niisiis seostub tõhususega järgmine märksõna – **(üle)diferentsimine** ehk (üle)eristamine. Ülediferentsimist on nii termini(korrastus)töodes kui teistes valdkonnakäsitlustes tavaks defineerida Mereste (2000 [1969]) järgi kui nähtust, mille puhul “püütakse teha terminoloogilist vahet selliste samatähenduslike sõnade vahel, samuti selliste sõnakujude vahel, mis keele üldseaduspärasuste põhjal tähendusisusi ei muuda” (samas, 79). Sellest definitsioonist näib olevat üks oluline osa siiski puudu. Lähtudes eeldusest, et erialakeel on eeskätt erialainimese tööriist, on selle olulisimaid nõudeid funktsionaalsus just erialainimeste seisukohalt – see, et erialane suhtlустаhe saaks võimalikult täpselt-tõhusalt edastatud. Kõnealust definitsiooni võiks laiendada näiteks nii: “[---] ja mille eristamise järele pole eriala(keele)s tege likku vajadust.” Mereste edasisest käsitlusest (vt samas, 79–82) võib küll oletada, et erialast, mõistelist aspekti on ehk peetud nii iseenesest mõistetavaks, et seda pole loetud tarvilikuks definitsioonis eraldi mainida. Erialases kontekstis on ometi ilmne, et *vaid* keeleliste näitajate põhjal määratletuna jääb kajastamata üks – ja ehk tähtsaimgi – mõistetunus. Lähtudes defineerimise põhitõdedest peaks mõiste olulisimad tunnused siiski definitsioonis kajastuma. Iseküsimus on, kes otsustab selle vajaduse üle, aga sellest juba edaspidi.

Erialakeele kui erialainimese töövahendi viimistlemisel tuleks niisiis teada ka peamise kasutajaskonna – praegusel juhul kaitseväelaste – vaated ja mõistetusviisi. Ohvitseride oskuskeelekäsitust uurides selgub, et nende vaated kalduvad olema äärmuslikult konservatiivsed (nt arvamus, et variatiivsusele pole sõjanduskeeles kohta, mistõttu tuleks välja kuulutada terministandard, mida kõik peavad kohustuslikuks korras järgima; vt Hendrikson 2012). Seisukohavõttudest selgub, et täiseitav seisukoht igasuguse varieerumise suhtes tuleneb põhjenduse kõrval, et variandid koormavad mälu (vrd Tauli 1968: 94), sageli just ülediferentsimishirmust (vrd Erelt 2007: 117–118): kartusest, et terminivariante võidakse hakata erinevalt tõlgendama, millest johtuvad segadus ja väär mõistmine.

Ohvitserid on variatiivsusest kõneldes seega sama meelt klassikalise terminiteooriaga, mis eeldab, et iga sõna nõuab keelekasutajalt lisajõudu, millest omakorda tuleneb variatiivsuse taunimine (vt nt ISO 1999: 22, Tauli 1968: 94). Uuemad terminiteooriad, nii kommunikatiivne, sotsiokognitiivne, freimipõhine kui sotsioteooria, rõhutavad terminikasutuse sotsiaalset ja pragmaatilist aspekti: termin on keelesüsteemi dünaamiline, kontekstisõltuv osa (Cabrè 1999: 112–114, Temmerman 2000: 15, Faber 2009: 114–115, 2011: 15, 25, Rey 1995: 99, 103, Gaudin 1993: 216). Temmerman (2011b: 108–109, 117) läheb veelgi kaugemale, leides, et varieerumisel on erialainimeste jaoks sageli tunnetuslik põhjus ja harva on kaalutlused puhtstilistilised. Nii võib liigne ühtlustamispüüe pädida ülestandardimisega (Bowker, Hawkins 2006: 180): sama termin võib erinevates kontekstides viidata veidi erinevatele denotaatidele (Temmerman 2000: 56). Teisalt ei välista Coxi ja Temmermani (2011: 6) hinnangul ka standard väär mõistmisi, kuna olulist rolli

mängivad kultuurikontekst ja keelespetsiifilised faktorid (vrd Tavast 2012: 227). **Terminivariandi tunnetusväärtuse** mõiste lähtubki arusaamast, et varieerumine peegeldab tunnetuse ja kategoriseerimise mitmetahulisust: variantidel on veidi erinev tunnetuslik taust ja valik sõltub parasjagu fookuses olevast mõistenüansist (vrd Fernández-Silva jt 2011: 51–52).

Eesti oskuskeelekorralduses on traditsiooniliselt jäädud nende kahe äärmuse vahepeale. Näiteks sünonüümiast kõneldes räägib Kull (2000: 101) küll sallitavast ja suisa soovitatavast sünonüümiast, ent seegi peab olema allutatud kindlale korrapärare (vt ka Kull, Saari 1970, Saari 1980: 662, 665). Mereste (2000: 83) järgi on homonüümia ja sünonüümia teatud ulatuses vältimatud ja isegi positiivsed (kuigi teisel viitab sünonüümia halbusele, vt samas, 87). Polüseemia seevastu on Mereste (samas, 83) hinnangul läbini taunitav (vt ka Saari 1980: 662). Erelt (2007: 120) kõneleb terminivariantide erinevast stilistilisest väärtusest. Variantide juurde-mõtlemisest tuleks siiski hoiduda seal, kus nende järele “ei ole tegelikku tarvidust” (samas, 123): kui mõiste tarvis on juba olemas sobiv termin või kui on tegu juhutuju või maitseotsusega. Viimane mõttekäik näib ilmselt nende jaoks, kes terminitöö praktilise poolega mingilgi moel kokku puutunud, pikemalt mõtlemata eelduspärane. Meenutades eelmainitud diferentsimisdefinitsiooni ja selle võimalikku täpsustust, tekib ometi küsimus: mis peaks olema “tegeliku tarviduse” mõõdupuu? Eeldades, et erialasuhtluses tuleb esiplaanile seada võimalikult täpne teabeedastus, siis võiksid terminivariandid justkui soositavad-soovitatavad olla – kuidas muidu oleks võimalik väljendada mõistenüansse ja oma suhtlustahet täpselt edastada?

Artikli eesmärk ongi otsida vastust küsimusele, kust läheb piir ülediferentsimise ja valdkondliku eristamisvajaduse (*resp.* erialainimeste tajutud tunnetusvajaduse) vahel ning milliseid aspekte tuleb otsuse tegemisel arvestada. Selleks et leida sobivad otsustuskriteeriumid, tuleb esmalt saada kõnealuselt nähtusest ülevaade, viimane aga eeldab diferentsimisega külgnevate oskuskeelenähtuste käsitlemist ning uuritava materjali ja meetodi lühitutvustust.

2. Materjal ja meetod

Temmerman (2011b: 109–110) rõhutab, et terminivarieerumise ja selle tagamaade uurimisel on üks paremaid allikaid erialainimeste metakeelised arutelud konkreetsete terminite sobivuse kohta. Sarnane on ka siinsete tekstikatkete valik: tegu on ohvitseridest sõjandusspetsialistide arvamusalaldustega, mis on pärit väljaõppe-materjalidest, ohvitseridest magistrikursuse õppurite esseedest, magistritöödest jm. Vajadusel olen võrdluseks kasutanud ka (eriala)sõnaraamatutes esitatud andmeid. Kuna valim on väike, ei saa teha suuri üldistusi, kuid ometi on toodavad näited minu hinnangul tüüpilised, võimaldades kombata käsitletavate mõistete piirialasid. Terminiuurija jaoks paljastavad sellised metakeelised seisukohavõtud erialakeele selliseid tahke, mida on üsna keerukas muul moel dokumenteerida.

Originaalkujul tekstinäidetes on olulisemad ja/või kõnealused märksõnad alla joonitud. Tsitaatidele järgneb allikaviide (magistritöödel teema, esseedel kursus).

3. Diferentsimine ja külgnevad oskuskeelenähtused

Et avada paremini terminoloogilise diferentsimise olemust, on kõigepealt paslik näidete varal põgusalt peatuda sellega seotud ehk külgnevatel oskuskeelenähtustel. Kõikidest neist tuleb 4. peatükis ka pikemalt juttu. Külgnevatena käsitletakse olukordi, mille puhul aktualiseerub küsimus mõistete eristamisvajadusest. Võimalusi on *teoreetiliselt* neli (vt tabel 1).

Tabel 1. Diferentsimisega külgnevad oskuskeelenähtused

	ÜKS MÕISTE	KAKS VÕI ENAM MÕISTET
ÜKS TERMIN	IDEAAL	POLÜSEEMIA
KAKS VÕI ENAM TERMINIT	SÜNUNÜUMIA	IDEAAL VÕI MÕISTESEGADIK

Mõiste-termini üksühene vastavus on ikka olnud terminikorrastuse (*resp.* klassikalise terminoloogiateooria) ideaal, ent sama seisukoht peegeldub ohvitseride arvamusalaldustes. Näiteks jalaväetaktika termineid korrastavas magistritöös (2007: 26) rõhutatakse, et sõjandusterminoloogias peab valitsema ühtsus, kuna siin on “ühtne mõistmine ja konkreetne iseäranis oluline”. Logistikatermineid korrastavas uurimuses (2009: 4) tõdetakse, et terminivariantide eristamine on sageli tarbetu ja johtub sellest, et mõistetest saadakse erinevalt aru. Ja vastupidi: kui ühe mõiste tarvis on kasutusel kaks terminit, hakatakse neid nagunii põhjendamatuult eristama (Pioneeriterminid 2011: 68). Leitakse, et ühetähenduslikkuse nimel tuleb kasutada vaid üht terminit ja “see peab olema rangelt teistest eristatud” (KV MLA alg: 76). Kerge (2012: 67, 72) järgi on rangete normingute pooldamine ja keeletundlikkus eestlasele loomumood: sageli ei mõisteta, et aeg nõuab tolerantust rööbikute suhtes – juba ainuüksi seetõttu, et terminivariantide puhul ei pruugi maitseotsused kokku langeda.

Oskuskeelenähtuse **sünonüümia** all vaatlengi olukordi, mis taanduvad küsimusele, kas tegu on ikkagi samatähenduslike terminitega. Nii mõnigi kord võivad mõisted näida esmapilgul (ka definitsiooni süvenedes) kattuvatena, kuid erialainimese jaoks võib neid eristada oluline mõistetunus või -nüanss. Mereste järgi avaldub ülediferentsimine sageli juba laenuandvas keeles ning ilmneb eeskätt suhteliselt kõrvaliste stiili-, murde- vms varjundite järeleaimamises ja teises keeles eri viisidel eluõiguse võitnud täissünonüümide eristamises (Mereste 2000: 81). Praktikast seostub mõlema aspektiga hulk *aga*’sid, nt millised on – erialakeele täpsustatult arvestades – *kõrvalised* stilistilised varjundid ning mil moel on täissünonüümid (?) teises keeles siiski eluõiguse võitnud? Milliste tunnuste järgi on määratletav täissünonüümia? Vaadeldagem neid küsimusi esmalt näite (1) varal. Tsitaadi autor rõhutab, et taktikaterminid *tiibamine* ja *haaramine* ei tähista sama mõistet, kuigi neid on sageli samastatud.

- (1) “Näiteks terminipaar *tiibamine* (ingl k *envelopment*) ja *haaramine* (ingl k *turning movement*). Esmapilgul tunduvad need sünonüümsed, kuid esimese puhul on eesmärk rünnata vastase positsioone küljelt ning teise puhul mööduda vastase positsioonidest ning liikuda sügavusse (tagalasse).” (Jalaväetaktika terminid 2007: 13)

Kui kõrvutada eelnenud seisukohta taktikaõpikus esitatuga,¹ siis selgub, et seal nähakse mõistete vahel hoopis hüponüümiat, nii et *tiibamine* (taktikaõpikus ingl *flanking*) on koos *süvahaaramisega* üks *haaramise* erivorme. Näide (1) ilmestab praktikas nii mõnigi kord ettetulevat olukorda: näilised sünonüümid võivad tähistada hoopis lähimõisteid (vrd Cruse 1986: 151) ja viidata – sageli spetsialistide erimeelsuste tõttu tekkivale – mõistesegadikule.

Niisiis on viimasel juhul tegu **mitme mõiste ja mitme termini suhestumisega**. Sõjanduses on levinud seisukoht, et kui tegu on erinevate mõistetega, peavad neid tähistama *selgelt eristuvad* keelevariandid – võimalusel erinevad tüved (vt tabeli 1 lahtreid “ideaal” ja “ideaal või mõistesegadik”). Sama seisukohta näitlikustab tekstikatke (2).

- (2) “[---] mission, mis meie kontekstis omab väga spetsiifilist tähendust ning selle tõlkimine “ülesandeks” on küll lihtne ja eestimeelne, aga loob lahinguväljal sõdurite peas palju segadust, sest samuti kasutatakse sõna “task”- ülesanne-, ning lisaks on meil juba niigi olemas ‘määratud ülesanded’, ‘kaasnevad ülesanded’, ‘põhiülesanne’ ja veel ‘võtmeülesanded’. On võimatu garanteerida, et sõjategevuse äärmuslikes tingimustes ei hakata neid omavahel sassi ajama.” (KV MLA alg: 21)

Polüseemiana vaatlen olukordi, kus erialainimeste jaoks on tegu selgelt erinevate mõistetega, aga tekib tõrge, kuna tähistajaid on üks. Mereste (2000: 87) on polüseemia suhtes kompromissitu: “Positiivset mitmetähenduslikkust pole olemas”. Kõnekas on, et mõiste-termini suhteid käsitledes ei maini nt ISO 1999 (22–23) polüseemiat üldse, mõtestades seda ilmselt homonüümia alaliigina. Erelt (2007: 104) märgib ühetähenduslikkuse nõudega seoses, et iial pole raske püstitada mingeid nõudeid – raske on neid nõudeid optimaalsetes piirides rakendada. Cabré (1999: 108) leiab, et ehkki terminoloogias on ideaal üks-mõiste-üks-termin-põhimõte, esineb tegelikkuses sellist olukorda vähe. Näide (3) on esitatud selle kohta, mis võib ohvitseride arvates juhtuda, kui mitut mõistet tähistatakse sama terminiga (siinses ülemmõiste ja üks kaasalluvatest mõistetest).

- (3) “Kui näiteks rahvusvahelise väekontingendi ülem annab Eesti allüksuse ülemale informatsiooni, et naaberüksus julgestab (*screen*) neid, siis [---] võivad alluvad ülemad seda oluliselt erinevalt tõlgendada, kuna terminit *guard* võib samuti tõlkida kui julgestamistegevust. Tähenduse erinevus on siin selles, et esimesel juhul annab naaberüksus informatsiooni vastase kohta ja võitleb ainult enesekaitseks, teisel juhul saame olla kindlad, et nad võitlevad igal juhul, et anda meile aega ettevalmistamiseks. Selline situatsioon on äärmiselt ohtlik ning võib ebasoodsate asjaolude kokkulangemisel maksta meile palju inimelusid.” (2. KAKi essee)

Ohvitserid ei väsi rõhutamast, kui katastroofiliseks võib olukord osutuda, kui ei järgita termini-mõiste üksühese vastavuse põhimõtet. Samas: hoolimata sellistest, enamasti stiilis mis-siis-kui-näidetest tundub erialakeel tegelikult suhtluses ometi üldjuhul funktsioneerivat. Tavasti (2012: 227–228) järgi ei ole teada, et oleks

¹ *Haarama* (ingl *envelop*) – pealetungimanööver, mille käigus põhijõud liiguvad ümber või üle vastase kaitse põhi positsioonide, et vallutada ja julgestada ründeobjekti tema tagalas (KV MLA 2010: 232).

Süvahaaramine (ingl *turning movement*) – haaramise erivorm, milles mõeldutakse või liigutakse üle vastase kaitsepositsioonidest, et vallutada ja julgestada määratud objektid vastase tagala sügavuses, ähvardades sellega nende äralõikamist, mis sunnib vastast oma positsioone maha jätma (KV MLA 2010: 247).

Tiibamine (ingl *flanking*) – haaramise erivorm, mille korral rünnatakse vastase positsioone küljelt, liikumata tema tagalasse (KV MLA 2010: 249).

dokumenteeritud *“ehtsaid* [esiletõst lisatud] väärnimõistmise juhtumeid pädevate ja koostööle orienteeritud suhtlejate vahel”.

Järgmise tekstikatke (4) autor soovib *mission-task*-probleemi (vt näide 2) lahendamiseks võtta sõjanduses informaatika eeskujul kasutusele sõna *tegum* – niisiis kasutada **siselaenamist** (täendusülekanne), mida klassikaline terminiteooria käsitleb polüseemia erijuhuna (vt ISO 2002: 12). Katkendi autor viitab keelilise kuju ja sellest johtuvate seoste sobivusele (sh analoogia inglise terminikujuga, vt siinses artiklis ptk 4.3) ning mõistestruktuuride analoogiale (vt viimast lauset).

(4) Esialgu tundus sõna [tegum] imelik, kuid järele mõeldes tundus täitsa hea ja kohane. [---] Tegum seostub kohe tegutsemisega ning mingil määral välise sarnasuse tõttu ka inglisekeelse sõnaga task. Tegumi saab lihtsalt siduda definitsiooniga, seega [on sõna] täpne ning vaieldamatult ka lühike. Selle süsteemne kasutamine ei tekita segadust (põhi) ülesande alajaotistena: kästud, tuletatud, oluline ja võtmetegum. [---] Keeleliselt on juba olemas ka sarnane sõna – kogum ja seega alluks tegum samadel alustel ka keelereeglitele. [---] Nimelt kasutatakse Tallinna Tehnikaülikoolis juba mõnda aega õpetamisel terminit tegum ning sellele on olemas ka sarnaste seostega definitsioon.” (6. KAKi essee)

Temmermani (2000: 44) järgi on polüseemia juured mõistete prototüübistruktuuris ja tähendusmuutustes – niisiis on polüseemia erialakeele funktsionaalne osa (vrd Temmerman 2002: 211, Black 1962: 37). Tavast (2002: 45–46) nendib, et kuigi ametlikult on siselaenamine tunnustatud terminisaamisviise, on praktikas seda sageli taunitud,² põhjuseks ilmselt võõristusefekt, mis tekib sõna kasutamisel nihkunud tähenduses (vt näite (4) esimene lause). Samas on selgi küsimusel mitu tahku: ohvitserid leiavad, et juurdunud terminite kõrvale või asemele uue loomisel võib ajendiks olla muudki kui vaid püüdlemine suurema suhtlustõhususe poole (nt võimusuhted jm sotsiokultuurilised tegurid, vt näiteid (9) ja (10)).

Ilmneb, et diferentsimisega seostub rida teisi oskuskeelenähtusi, mis omakorda komplitseerib probleemi käsitlemist. Jättes kõrvale sõjandusspetsialistide seisukohad, mis ühest küljest taunivad variatiivsust, ent teisalt rõhutavad suhtlustäpsuse ja -tõhususe olulisust, jääb ikkagi alles küsimus: kuidas ja mille põhjal otsustada, kas konkreetsel juhul on tegu ülediferentsimisega või põhjendatud eristamisega, mis teenib erialase mõttetäpsuse huve?

4. Kriteeriumiotsingud ehk kas diferentsida või mitte?

Eelnenust ja erialastest arvamusedest on tuvastatavad võimalikud kriteeriumid, mis võiks esmapilgul aidata eelmise peatüki lõpus esitatud küsimusse selgust tuua. Siinkohal käsitlen järgmisi:

- õigekeelsussõnaraamat ja kirjakeelenorming,
- erialaselt tunnetatud vajadus ning
- laenuandva keele eeskjuu.

² Selle seisukoha paikapidavust näib kinnitavat tõik, et inglise *task* vastena on sõjanduses senini kasutusel *ülesanne*. Konservatiivset joont järgis oma otsuses ka sõjandusterminoloogia komisjon (vt Militerm; nimetatud andmebaasi ja komisjoni kohta vt ka 7. joonealust märkust).

Kindlasti on teisi mõjureid (nt traditsioon ja isiklik harjumus), kuid järgnevad näited hõlmavad vähemalt mõningaid neist. Samas, nagu järgnevast nähtub, on kõigi nende kolme probleemide.

4.1. Õigekeelsussõnaraamat ja kirjakeelenorming

Terminiaruteludes on ikka kõlanud seisukohti, et mõistete eristamisel peab olema esmakriteeriumiks õigekeelsussõnaraamat (ÕS). Kui ÕS-i järgi on kaks sõna samatähenduslikud, siis ei saavat anda neile ka oskuskeelis erinevaid tähendusi – hoolimata ÕS-i saatesõnast, milles on rõhutatud, et sõnastik hõlmab eelkõige “kirjakeele kui üldrahvalikult kasutatava ja ühtseima keelekuju sõnavara”. Sellised arvamused näivad tuginevat ÕS-i funktsiooni ja laiemalt kirjakeelenormingu väärtõlgendusele.³ Kõigi erialakeelte ammendavale kirjeldamisele ÕS ju ei pretendeeri: terminoloogias tekib sageli vajadus eristada üldkeeles sünonüümsete sõnade tähendusi. Teoorias seda vajadust üldjuhul tunnetatakse ja tunnustatakse, kuid praktikas on küllalt näiteid vastupidisest.⁴

Sagedased on vaidlused sellegi üle, kas erialakeeles ikka võib anda sõnadele ÕS-ist erineva tähenduse.⁵ Teoreetilistes seisukohavõttudes küll aktsepteeritakse tähendusülekannet ja siselaenamist, kuid praktikas kaldutakse põhjendama terminivariandi mitte-meeldimist väitega, et ÕS-i järgi tähendab sõna hoopis muud. Taoliste põhjenduste üldsõnalisust illustreerib hästi NATO terminitöö juhend (NATO Terminology Guidance 2007: 13). Viimase järgi on sobimatu (ingl *deprecated*) selline terminivariant, mis on eksitav või ebatäielik (“[---] unsuitable, for example it is misleading or incomplete”). Samas ei selgitata kuskil täpsemalt, kuidas neid omadusi tuvastada. Ebasobiva terminivariandi kirjeldamisel on selline üldsõnalisus tüüpiline: ka terminivaidlustes kasutatakse viitamist kirja-keelenormingule sageli pigem nn katteargumendina, mille taga on individuaalne maitseotsus ja tekkiv võõristusefekt. Termin, mis näib ühele erialainimesele (nt väärseoste tõttu) eksitav, ei pruugi seda olla teisele (vt ka Hendrikson 2012, Hendrikson 2013).

Ei ole läbini mõistetud, et kirjakeele sõnastik ei näe ette ega saagi ette näha kõiki kasutusjuhte ega kõigi erialakeelte vajadusi (ka Tavast 2013: 154–155). On ilmne, et oskuskeelt vaesestab kalduvus järgida mis tahes sõnaraamatut või normingut pimesi, vaid vormi tasandil, arvestamata keele siseomast ülekandevõimet. Lähtekoht võiks pigem olla erialane vajadus ja suhtlusfunktsioon. Aga nagu edasist nähtub, ei ole needki alati ühesed.

³ Vt Eesti kirjakeele normi rakendamise kord (§ 2 lg 1): “Kirjakeele norm on määratud Eesti Keele Instituudi uusima õigekeelsussõnaraamatuga”, ent samas paragrahvis (lg 3) seisab: “Oskussõnavara osas täiendavad ja täpsustavad kirjakeele normi *oskussõnastikud* [esiletõst lisatud]”.

⁴ Näiteks esitab ÕS 2013 sõna *tempo* sünonüümvaste/tähendussetuse/tähendusvihje väljal sõna *kiirus* ja sellest lähtudes olen kuulnud väidet, et mainitud sõnu ei tohiks ka oskuskeeles eristada. Sõjandusterminina tähistab *tempo* järjestikuste tegevuste rütmi ja tihedust (KV MLA 2010: 249), sõnal *kiirus* on sama tähendus mis üldkeeleski. On ilmne, et tähenduste eristamise on tinginud erialane vajadus.

⁵ Näiteks esitab ÕS 2013 sõnale *kanaliseerima* tähendussetusena ‘kanalisatsiooniga varustama’, mistõttu ei saavat seda kasutada tähenduses ‘sundima vastast liikuma kitsale maa-alale’. Samas tõestavad küsitlustulemused (Jalaväetaktika terminid 2007: 60), et termin on nimetatud mõiste tähistajana sõjanduses ilmselgelt juurdunud. Kujundlik terminikasutus võimaldab anda edasi nüansse, mis oleks muul moel raskendatud, ja aitab eristada lähimõisteid (nt sinisel juhul ei saa termin olla *suunama*, kuna viimane on kasutusel juba teise mõiste tähistajana; pikemalt Hendrikson 2013).

4.2. Erialaselt tunnetatud vajadus

Erialaselt tunnetatud vajadus haakub otseselt kahe küsimusega, mis terminitöös nii mõnigi kord esile kerkivad. Esiteks: kas tegu on ikkagi terminiga? Juhul kui keelevariandid on selgelt terminid, võib osutada küsitavaks, kas need viitavad ikkagi samale mõistele, ning kui mitte, siis milline on nende suhe. Vaadeldgem mõlemat varianti näidete varal.

4.2.1. Termin või mitte?

Teoorias tundub termini ja üldkeele sõna vahetegur olevat klaar (seda enam et ka erinevad terminiteooriad näivad olevat ses suhtes üsna ühte meelt – erinevalt nt mõiste määratlusest). Termin on mõiste/mõistmisüksuse keeleline kuju; sõna(d), millele antakse erialakeeles täpsem, piiritletud tähendus (vt nt Saari 1983: 251, ISO 1999: 22, Cabré 1999: 39, Mereste 2000: 54, Erelt 2007: 62, NATO Terminology Guidance 2007: 5–6).

Kummatigi võib praktikas vahetegemine olla päris keerukas. Sarnasele tõdemusele jõuab Kerge (2011: 111), märkides, et enamik väljendeid mõjub oskuskeelendina vaid oma stiilitunnuste tõttu (ka Erelt 2007: 69, 120, vrd Leete 2012). Cabré (1999: 80–81) leiab, et termini ja sõna erinevus saab ilmsiks, kui vaadelda termineid suhtlusüksustena. Näib niisiis, et ka keelendi n-ö terminilisuse üle otsustamisel võiks otsustusaluseks olla *teoreetiliselt* erialainimeste vajadus. Terminiarutelusid vaadeldes pole seegi mitte alati ühene ja selge.

Kaitseväetekstides esineb termin *ülema kavatsus*⁶ (ingl *commander's intent*), millega paralleelselt on kasutusel *ülema tahe*. Tekstikatke (5) autor väidab, et õigupoolest on tegu kahe erineva mõistega.

- (5) “Ülema tahe. Saagu kohe öeldud, et siinjuures rõhutan ma sõna ‘tahe’, mitte kavatsus, sest ülema tahe on operatsiooni täideviimise n.ö viimane juhis, kui miski muu enam ei toimi. Aga kui ma plaanin minna päikest võtma või äkki õlut rüüpama või, mine tea, jalgpalli vaatama, siis see on kavatsus. Kavatsus on plaan, mis sisaldab endas ebakindlust ja muutlikkust, ülema tahe ent ei ole tingitav – see kuulub vastuvaidlematule järgimisele.” (KV MLA alg: 64)

Kirjutaja leiab, et keelendite tunnetusväärtus on erinev, ja soovib neid täpsust taotledes eristada. See seisukoht oli ilmselt liiga uudne ka tsiteeritud õppematerjali jaoks – miks muidu on õppematerjali lõppversioonist see lõik ära jäetud ja kasutatakse terminit *ülema kavatsus*? Samal kujul esineb see ka Militermis,⁷ *ülema tahet* aga terminiks ei peeta. Ometi esineb *ülema tahe* siiani näiteks seadusetekstis – küll vaid mõiste *käsk* definitsioonis (vt KVKS § 29) ja mõnes väljaõppematerjalis paralleelselt nii *kavatsus* kui *tahe* (nt Jao instruktori käsiraamat 2008: 33).

Sõjanduses esineb keelendite kõrval, mille terminiväärtuses ollakse ühte meelt, küllalt selliseid, mille terminilisus tekitab erimeelsusi. Erelt (2007: 123) leiab, et

⁶ *Ülema kavatsus* – sõjakäigu või operatsiooni soovitud eesmärkide ja lõpptulemuse saavutamise lühikirjeldus (AJP-3.2 tõlge; Militerm). Olgu lisatud, et informaaalses vestluses on mitu ohvitseri olnud veendunud, et *ülema kavatsus* ja *ülema tahe* on erinevad mõisted, kuid ühelgi juhul ei osatud leida neile kahte erinevat ingliskeelset vastet (nn taandatavuskriteeriumi kohta vt ptk 4.3).

⁷ Sõjandusterminoloogia andmebaas Militerm on Kaitseministeeriumi sõjanduse ning julgeoleku- ja kaitsepoliitika terminoloogiakomisjoni (edaspidi *sõjandusterminoloogia komisjon*) töö väljund. Andmebaas sisaldab termineid alates julgeolekupoliitikast kuni sõjanduse taktikatasandini.

terminivariantide juurdemõtlemisest tuleks hoiduda siis, kui tegu on juhutuju või maitseotsusega. Eelnenud näide illustreerib seda, et juhutuju ja maitseotsust on sageli keeruline eristada tegelikust erialasest vajadusest – juba seetõttu, et ka viimase suhtes ollakse eri seisukohtadel. Niisiis näib, et keelendi terminilisus on pigem skalaarne, mitte bipolaarne suurus ja kaksikjaotuse *termin-mittetermin* asemel on täpsem kõnelda pidevustikust. Põhjuseks võib olla spetsialisti(de) või terminikomisjoni käsitusviiside erinevus ja nende muutumine ajas. Neist kõigist tuleb juttu ka järgnevas alapeatükis.

4.2.2. Kas samad mõisted või mitte?

Kui eelmises näites taandus küsimus ühe variandi n-ö terminilisusele, siis järgnevad on komplitseeritumad. Nagu näites (3), nii tekitab ka järgmistes keerukust tõlgenduste paljusus (segadus mõistetasandil), mille lähteks on võimalik terminite hüponüümia, kasutuskontekst, kasustraditsioon ja mõistete kujunemislugu.

Sõjanduskeeles käis kaua vaidlus, kas *engagement area* (EA) ja *kill(ing) zone* (KZ) on samaviitelised või mitte. Arvamuste paljususe illustreerimiseks olgu nimeetatud seisukohti, mida olen kohanud:

- 1) EA on mõistemahult laiem kui KZ (viimane on seotud vaid tuletegevusega, seal ei toimu sõbralike üksuste manöövrit vastupidiselt EA-le);
- 2) tegu on ühe ja sama mõistega, erinevus on sõjatasandis (EA on operatsioonitasandi termin, KZ taktikatasandi oma);
- 3) KZ on EA vananenud variant;
- 4) EA on KZ-i eufemistlik variant (vrd *hävitusala/tapmisala* ja *löögiala*);
- 5) vaatepunkti erinevus (üks vastase, teine sõbralike üksuste vaatepunktist, vt näidet 7);
- 6) kasutuskonteksti erinevus (eristuse aluseks lahinguliik, vt näidet 6).

Olgu järgnevalt esitatud mõni dokumenteeritud arvamusalvaidus. Näites (6) leitakse, et kasutuskonteksti erinevusest hoolimata on tegu sama mõistega, mistõttu soovitatakse tarvitada sama terminit ja loobuda sünonüümist.

- (6) “Lisaks terminile *engagement area* on kasutusel ingliskeelne termin *kill zone* (FM 3-90 2001: 5-31), mis mõiste järgi samuti tähistab ala, kus toimub vastase hävitamine. FM 7-8 (1992: 2-24-b) järgi on mõiste üks, kuid varitsuse (ingl k *ambush*) puhul nimetatakse seda *kill zone*. Käesoleva töö autor arvab, et segaduste vältimiseks tuleks eesti keeles nendele kahele ingliskeelsele terminile leida ühine eestikeelne vaste.” (Jalaväetaktika terminid 2007: 51)

Kontekstuaalne faktor – küll teises võtmes – on argumendiks ka näites (7), kuid siin jõutakse terminite eristamisvajaduses vastupidisele seisukohale.

- (7) “Arendatav sõjanduskeel ei tohi kaotada nüansse, mille väljendamine ühe sõnaga võib vägagi raskendatud olla. Pisidetallide kadumisele võib viidata näiteks arutelu sõnade *kill(ing) zone* ja *engagement area* erinevuste üle. Tapetakse mõlemas alas, aga nüansiks on see, et ühes kavatseme seda meie teha ja teises plaanitakse meid tappa.” (2. KAKi essee)

Olgu öeldud, et samale arusaamale nagu näites (6) jõuti spetsialistide kaasabil lõpuks ka sõjandusterminoloogia komisjonis (vt Militermi kirjet *hävitusala*).

Teine, komplitseeritum näide (8) on sõjandusvaldkonnast, mida nimetatakse pioneerinduseks. Terminid, millest juttu tuleb, on *maskeerimine* ja *moondamine*.

- (8) “Praegu käib pioneerispetsialistide seas jätkuvalt vaidlus, kumba [*maskeerimine* või *moondamine*] eelistada. [---] Kuigi nii “Pioneeritegevuse käsiraamat” kui ka V. Pöder oma magistritöös “Pettetegevus ja selle rakendamine Eesti Kaitseväes” toovad välja, et tegemist on kahe erineva mõistega (*maskeerimine* – värvuse muutmine; *moondamine* – kuju ja suuruse muutmine), siis inglisekeelne terminoloogia (camouflage) neil vahet ei tee. Sama kehtib ka kasutajaskonna kohta, kes kasutavad mõlemat terminit läbisegi (*maskeerimis-*, *moondamisvõrk*; *maskeerimis-*, *moondamisvahendid*). [---] teeb autor ettepaneku kasutada kõnealuse mõiste tähistamiseks vaid terminit *maskeerimine* ja kaotada Militermis sünonüüm.” (Pioneeriterminid 2011: 68)

Kuigi spetsialistid on *moondamise* ja *maskeerimise* eristamisvajaduse suhtes eri meelt (vt näites (8) esitatud kõrvutavaid viiteid), arvavad mõlemad, et katustermin võiks olla *pettetegevus* (Pioneerindus 2010: 13, Pioneeriterminid 2011: 66–68). Viimane on nüüdseks kinnitatud üldterminina ka sõjandusterminoloogia komisjonis, *maskeerimine* ja *moondamine* selle samatähenduslike hüponüümidena.⁸ *Moondamise* ja *maskeerimise* samastab ka Ernits (1998), defineerides neid üldterminitena (vrd Ernits 1997, kus mõistete vahel on alluvussuhe: *maskeerimine* on *moondamise* hüperonüüm). *Pettetegevus* mõlemast puudub. “Pioneeritegevuse käsiraamatus” (2010: 8) on *maskeerimine* defineeritud taas üldterminina. Sellega kõrvuti esineb samas tähenduses, kuigi erineva definitsiooniga *pettetegevus*, puudub *moondamine*.

Kõnealusel probleemil on seega mitu tahku. Temmermani (2011b: 109) järgi on terminite varieerumisel sageli tunnetuslik tagapõhi, mistõttu võib pealtnäha samaviiteliste, näiliselt põhjusetu varieeruvate terminite taga olla erinev mõiste ja erialainimeste täpsustaotlus. Ometi nähtub, et asjatundja arvamus võib olla vaid ühe inimese isiklik, mitte valdkonnas üldaktsepteeritav seisukoht (vt ka Valdre 2012: 269). Lisaks võib terminiasenduste põhjus ohvitseride endi hinnangul olla sotsiokultuuriline, nt võimusuhted või erudeerituse näitamine (vt näiteid (9) ja (10); lähemalt Hendrikson 2012).

- (9) “Pahatihti on meil liigselt võimust võtnud lapsik “põidlajälgede vajutamise sündroom”. Nii on tekkinud olukord, kus samu mõisteid nimetatakse erinevalt, mille tõttu muutub ka nende tähendus hägusaks ning mõnikord isegi vastupidiseks.” (4. KAKi essee)
- (10) “Väga moodne on kasutada suhtluses inglise keelseid väljendeid, et näidata ennast targemana.” (3. KAKi essee)

Nagu eespool öeldud, on mõistenüansside eristamine funktsionaalne mõistagi vaid siis, kui adressaat neist nüanssidest aru saab ja (nt käsuandmissituatsioonis) vastavalt käitub. Ometi on ilmne, et klassikalisele terminoloogiateooriale omane käsitus (tähenduste täielik kattuvus *versus* selgepiiriline eristuvus; vt nt ISO 1999: 22) kirjeldab pigem erandolukordi kui reeglit.

⁸ *Pettetegevus* – looduslike või tehismaterjalide kasutamine isikkoosseisu, objektide või lahingupositsioonide maskeerimiseks, et vaenlast segadusse ajada, eksiteele viia või vältida (Militerm). Olgu öeldud, et 1914. aastal ilmunud “Sõjamehe sõnastikus” on mõlemat mõistet tähistatud ilusa eestikeelse terminiga *sarnastama* (vt Pitka 1914).

4.3. Taandatavus laenuandvale keelele

Laenuandva keele moodustusmalli järgimine on tegur, mida on traditsioonilises keelekorralduses tavaks taunida ja võrdsustada teise keele kopeerimisega. Näiteks kõneleb Mereste (2000: 85, 173) kolmandasest terminituletusest (*sic!*), seostades seda vaid lähtumisega võõrkeelest, süvenemata mõiste olemusse. Valdre (2012: 258) leiab tehnikakeele valdkonnas, et sageli on eesmärgiks seniseid terminineid parandada, lähendades neid inglise omadele, kuid olukorda see paremaks muuta ei pruugi, kuna uued terminid lihtsalt kopeerivad teise keele väljenduslaadi ja tekitavad paralleelvariante (vt ka Erelt 2007: 213–219). Erelt tõdeb siiski, et tõlke-laenude hange ei tohi olla mehaaniline ja peab johtuma väljendusvajadusest (vt Erelt 2007: 218).

Rangelt võttes ei saa muidugi väita, nagu tähendaks või tähistaks üks sõna teist (teise keele) sõna (vrd Tavast, Taukar 2013: 62). Ometi on sageli just vaste laenuandvas keeles erialainimesele mõiste olulisim indikaator (vt ka Tavast 2002: 75, Maanso, Unt 2005). Kui tekstist avastatakse uus termin või kuulatakse seda vestluses, on vähemalt sõjanduses esimesi küsimusi sageli: kuidas see inglise keeles kõlab? On see keele allakäigu ja valdkonnakaotuse märk? Või pelgalt üks keele järjekordsetest arengujärkudest, millega tuleb tugeva liiderkeele mõjualas teatava piirini leppida – näiteks seni, kuni oma terminid on juurdunud?

Järgnev näide (11) käsitleb ühte mõistepundart, mille juures on kirjutajal olnud abi just nimelt sellest võttest. Termin on ingl *main effort* (ME) 'jõudude või vahendite koondamine piirkonda, kus ülem soovib saavutada lahenduse' (Militerm), mille eestikeelsete vastetena levinud paralleelvariandid on erialainimesi eksitanud: *raskuspunkt*, *põhirõhk*, *põhipingutus*, *põhijõupingutus*, *peamine jõupingutus*, *põhipingutusüksus*. Tekstidest selgub, et aja jooksul on neid hakatud tähenduslikult eristama. Olukorra teeb põnevaks asjaolu, et on ka *centre of gravity* (COG) 'omadused, võimed või paigad, millest tulenevad riigi, liidu, sõjajõudude või muu rühmituse tegutsemisvabadus, füüsiline jõud või võitlustahe' (Militerm). COG-d on aga vastendatud nii terminiga *raskuspunkt* kui *raskuskese*. Niisiis on tegu näitega selle kohta, kuidas kaks mõistet on hakanud väidetavalt segi minema, kuna tähistajad on hakanud kasutuses kokku langema. Taktikaõpiku algversioonis (11) leitakse, et *põhirõhk* ja *peamine jõupingutus* tähistavad sama mõistet, ning kirjutaja argumendiks on just vastav inglise terminikuju.

- (11) "Mis suhe on siis põhirõhul peamise jõupingutusega? On üsna kindel, et nad on ühetähenduslikud – tehke tagasikäik inglisekeele sõnapaarile *main effort* – kas on kahtlusi? Et nende paralleelsest kasutamisest tingitud segadust vältida, peame ühest neist loobuma, sest sõjalise eneseväljenduse ilu ei peitu mitte tema mitmekülgsuses, vaid ühetimõistetavuses. Niisiis teeme valiku ja käesolevast hetkest alates eelistame põhirõhku." (KV MLA alg: 76)

Jalaväetaktika termineid korrastavas magistritöös pakutakse ME vasteks aga *raskuspunkt*.

- (12) "Lähtuvalt definitsioonist on vaste sõnastamine [inglise terminile *main effort*] üsna problemaatiline – see peab mahutama korraga (abstraktse) tegevustähenduse kui (konkreetse) üksuse või maa-ala tähenduse [---],

sõna *raskuspunkt* [tähistab] ülekantud tähenduses tähtsaimat osa (ÕS 2006: 266, 766). Terminivariant *raskuspunkt* annab selgemini tähenduse edasi [kui *peamine* või *põhijõupingutus*] ning on ka ökonomsem ja keeleliselt mugavam.” (Jalaväetaktika terminid 2007: 41)

Taktikaõpiku algversioonis (13) kõneldakse terminist *raskuspunkt* enesestmõistetavalt hoopis inglise COG tähenduses. Seejuures tuginetakse taas vastele laenuandnud keeles.

- (13) “Nüüd aga kolmas mõiste – *raskuspunkt*. Tuleb tunnistada, et ka *see mõiste on saanud alguse samast n.ö allikast*; ilmselt on see *otsetõlge saksakeelsest sõnast *schwerpunkt** ja nagu me eelnevalt selgeks tegime, *ei peaks me enda sõjalist keelekasutust sünonüümidega risustama*. Ometi ei tarvitse siinjuures prügikasti kaane avamisega kiirustada. Kui me ise suudame anda sellele *mõistele uue sisulise tähenduse, mille on praktiline vajadus, ja see juurutada, siis oleme me ratsionaalselt lahendamas põhjendatud vajadust*. [---] Teisisõnu – üritame jätta alles alternatiivi, millega väljendada lahingutegevuse n.ö rõhuasetust eelkõige lahingutoetusüksustele, samas *eristades seda üksuse põhirõhust*.” (KV MLA alg: 76)

Termini dünaamikat iseloomustab see, et nii taktikaõpiku lõppversioon (KV MLA 2010: 244) kui Militerm esitab ME kahe erineva mõistena (*põhipingutus* ja *põhipingutusüksus*), COG vasteks on aga *raskuspunkt*, sünonüümina *raskuskese*.

Niisiis on ka taandatavusel laenuandvale keelele mitmeid *aga*'sid. Lisaks sellele, et klammerdumine liiderkeelde võib kaasa tuua põhjendamatu teise keele matkimise ja asendada mõistekeskuse sõnakeskusega, ei pruugi ka siin olla olukord ühene erialainimeste endigi jaoks.

5. Kokkuvõte ja järeldused

Artikli eesmärk oli otsida vastust küsimusele, kust läheb piir terminoloogilise ülediferentsimise ja erialasest tunnetusvajadusest johtuva eristamise vahel. Sõjanduskeel on nii tõhususe kui eristamisvajaduse käsitlemisel tänuväärne uurimisobjekt: selle üldeesmärk on nagu teistelgi erialakeelteil täpne erialane teabevahetus (tegu on ühega oskuskeele allkeeltest). Samas on sõjanduskeeles iseäranis oluline see, et adressant saavutaks kõneteoga oma suhtluseesmärgi (alluv peab üksikasjaliselt aru saama, mida oodatakse, ja vastavalt tegutsema).

Ilmnes, et sissejuhatuses täiendatud diferentsimise määratlus ühest küljest küll raskendab nähtuse käsitlemist, ent teisalt aitab täpsemalt tabada selle olemust. Tõhususe poole pürgimisel on oluline arvestada võimalusega, et näilistel terminivariantidel võib olla erialainimeste jaoks erinev tunnetusvääratus, st veidi erinev tähendus(varjund). Seetõttu võib pealtnäha samaviiteliste, näiliselt põhjusteta varieeruvate terminite taga olla erinev mõiste ja erialainimeste täpsustaotlus. Samas näitavad uurimistulemused, et asjatundja arvamus võib olla vaid ühe inimese isiklik, mitte valdkonnas üldaktsepteeritav seisukoht, teiseks on arusaamad ka isikuti muutumises. Keerukust lisavad termini kujunemislugu ja sellega kaasnenud mõistemuuatus(ed), samuti sotsiokultuurilised tegurid (nt võimusuhted). Pealegi on mõistenüansside eristamine tõhus ikkagi siis, kui adressaatki neid mõistab.

Ehkki ohvitseride seas valitseb täiseitav seisukoht igasuguse variatiivsuse suhtes, tõdevad nemadki, et praktikas on olukord teistsugune. See, kuivõrd varieerumine erialasuhtluses tegelikke mõistmiskaskusi tekitab, vajab veel edasist uurimist. Samuti pole ühene termini ja sõna vahekord: keelendite kõrval, mis on selgelt terminid, leidub küllalt selliseid, mille terminiväärtuses ei olda ühte meelt. Seega on keelendi terminilisus pigem skalaarne, mitte bipolaarne suurus. Triviaalsusena võib kajada tõde, et otsus keelendi terminilisuse ega mõistete eristamisvajaduse üle ei saa lähtuda *vaid* sõnaraamatust ega keeletasandist, kuigi samas võib laenuandnud keele terminivaste olla spetsialistile võti mõiste tuvastamisel.

Terminivariantide rohkest on peetud kujunemisejärgus eriala jooneks. See kehtib ka sõjanduse kui kiirelt ja tugeva liiderkeele mõjualas areneva valdkonna kohta. Näidetest selgus, et terminikorrastusel võib siinjuures olla ambivalentne roll: hoolimata ühtlustamis- ja segaduste vältimise taotlusest võidakse – vähemalt otsuse leviku algfaasis – segadust juurdegi tekitada. Tegu on ehk terminikorrastuse paratamatu kõrvalmõjuga, kuid siiski on oluline seda võimalust juba otsustamisel arvestada.

Ehkki eelnenud näited aitasid heita valgust diferentsimise piirialadele, näib püüde paika panna universaalseid otsustuskriteeriume ebaotstarbekas. Pigem tuleks küsida, milliseid aspekte iga konkreetse terminiotsuse puhul arvestada, ja näidete varal neid aspekte artiklis vaadeldigi. Temmerman (2011a: 56) leiab, et reaalses tekstides tuleb divergentsuse kui paratamatusega lihtsalt arvestada. Ühtlasi viitab ta paradoksile, et just n-ö ühtsed mõisted ongi sageli segaduste allikaks, kuna arvatakse, et nende suhtes ollakse ühte meelt.

Viidatud kirjandus

- Black, Max 1962. *Models and Metaphors*. Ithaca: Cornell University Press.
- Bowker, Lynne; Hawkins, Shane 2006. Variation in the organization of medical terms. Exploring some motivations for term choice. – *Terminology*, 12 (1), 79–110. <http://dx.doi.org/10.1075/term.12.1.05bow>
- Cabrè, Marie Teresa 1999. *Terminology Theory, Methods and Applications*. Amsterdam: John Benjamins. <http://dx.doi.org/10.1075/tlrp.1>
- Cox, Antoon; Temmerman, Rita 2011. Can EU harmonisation efforts of administrative forms co-exist with cultural and linguistic diversity? – *Proceedings of Conference “Languages Across Cultures”* 31. Venice, 1–10.
- Cruse, Alan D. 1986. *Lexical Semantics*. Cambridge: Cambridge University Press.
- Eesti kirjakeele normi rakendamise kord. [Standard Estonian Act Implementation Act.] – *Riigi Teataja I* 2011, 3.
- Erelt, Tiiu 2007. *Terminiõpetus*. [Terminology.] Tartu: Tartu Ülikooli Kirjastus.
- Faber, Pamela Benitez 2009. The cognitive shift in terminology and specialized translation. – África Vidal, Javier Franco (Eds.). *A (Self-)Critical Perspective of Translation Theories*. *Monografías de Traducción e Interpretación*, 1. Alicante: University of Alicante, 107–134. http://rua.ua.es/dspace/bitstream/10045/13039/1/MonTI_01_10.pdf (10.9.2013).
- Faber, Pamela 2011. The dynamics of specialized knowledge representation. Simulational reconstruction or the perception-action interface. – *Terminology*, 17 (1), 9–29. <http://dx.doi.org/10.1075/term.17.1.02fab>
- Fernández-Silva, Sabela; Freixa, Judit; Cabrè, M. Teresa 2011. A proposed method for analysing the dynamics of cognition through term variation. – *Terminology*, 17 (1), 49–73. <http://dx.doi.org/10.1075/term.17.1.04fer>

- Gaudin, F. 1993. Pour une socioterminologie: Des problèmes pratiques aux pratiques institutionnelles. Rouen: Publications de l'Université de Rouen.
- Hendrikson, Reet 2012. Keelesõjas sõjakeel? Ohvitseride oskuskeelevaadetest. [Military Language at War? The Views of Officers about Military Terminology.] – Kaitseväe Ühendatud Õppeasutuste toimetised, 16, 15–39.
- Hendrikson, Reet 2013. Sõjanduskeele ja kujundkeele ristteel ehk Kas kujund (oskus)keelt pigistab? [At the Crossroads of Military and Figurative Language.] – Eesti Rakenduslingvistika Ühingu aastaraamat, 9, 61–73. <http://dx.doi.org/10.5128/ERYa9.04>.
- ISO 1999 = Terminology work – Principles and methods. 1999. ISO/TC 37/SC 1 N. 185. Sweden: The International Organization for Standardization.
- ISO 2002 = Terminoloogiatoõ. Sõnastik. Osa 1. Teooria ja rakendus. [Terminology work. Vocabulary. Part 1. Theory and application.] Eesti standard EVS-ISO 1087–1:2002.
- Kerge, Krista 2011. Tõlkevõhik keele- ja lugemisoskuse valdkonna terminivarast. [Translation ignorant mediating CEFR and PISA special terms.] – Peep Nemvalts (Toim.). Eesti teaduskeel ja terminikorrastus. Tallinna Ülikooli eesti keele ja kultuuri instituudi toimetised 13. Tallinn: Tallinna Ülikool, 91–114.
- Kerge, Krista 2012. Keelkorraldus – kas ka täna? [Language Planning – Is It Still Actual?] – Maire Raadik, Tiina Leemets (Koost. ja Toim.). Sõnaga mõeldud mõte. Pühendusteos Tiitu Ereltile 20. aprillil 2012. Tallinn: Eesti Keele Sihtasutus, 67–75.
- Kull, Rein; Saari, Henn 1970. Die Entwicklung des estnischen terminologischen Gedankens und die Förderung der Fachsprachen in den letzten Jahren. – Congressus tertius internationalis Fenno-ugristarum Tallinnae habitus 17.–23. VIII 1970. Pars 1. Acta linguistica, 245–249.
- Kull, Rein 2000. Kirjakeel, oskuskeel, üldkeel. [Standardised Language, Specialized Language, General Language.] Tallinn: Eesti Keele Sihtasutus.
- KVKKS = Kaitseväe korralduse seadus. [Estonian Defence Forces Organisation Act.] 20. märts 2013. – Riigi Teataja I, 18.
- Leete, Epp 2012. Sõjandusteksti tõlgendamise NATO maaväe lahingutegevuse aluste “Allied Land Tactics” näitel. [Interpretation of a Military Text Based on the NATO Allied Land Tactic Foundations. Conference theses.] – KVÜÕA konverentsi “Sõjandustekstide tõlkimine ja tõlgendamine” ettekandeteesid. Tartu, 3.–4. mai 2012.
- Maanso, Viivi; Unt, Inge (koost) 2005. Valik ülddidaktika termineid (2). Kontroll ja hindamine. [A Selection of General Didactics Terms (2). Checking and Assessment.] Tallinn: Tallinna Ülikooli Kirjastus.
- Mereste, Uno 2000 [1969]. Oskuskeel ja seaduste keeleline rüü. Artikleid ja lühiaurumusi. [Specialized Language and Language of Legal Acts. Articles and short studies.] Tallinn: Eesti Keele Sihtasutus.
- NATO Terminology Guidance 2007 = Guidance for the Development and Publication of NATO Terminology. North Atlantic Council. C-M (2007) 0023.
- Rey, Alain 1995. Essays on Terminology. Translated by Juan C. Sager. Amsterdam: John Benjamins. <http://dx.doi.org/10.1075/btl.9>
- Saari, Henn 1980. Omasõna ja võõrsõna paarid eesti oskussõnavaras (1). Eelduste loomine. [Pairs of genuine and loan words in Estonian specialised language (1). Providing conditions.] – Keel ja Kirjandus, 11, 654–666; 12, 737–743.
- Saari, Henn 1983. Mida nimetada terminiks? Terminoloogilisi mistselle (II). [What is a Term? Terminological remarks.] – Keel ja Kirjandus, 5, 249–252.
- Tauli, Valter 1968. Keelekorralduse alused. [A Theory of Language Planning.] Stockholm: Vaba Eesti.
- Tavast, Arvi 2002. Onomasioloogia ja semasioloogia vahekorraldus oskuskeeles. [The relationship between onomasiology and semasiology in LSP.] Magistritöö. Juhendaja Mati Erelt. Tartu: Tartu Ülikooli eesti keele õppetool.

- Tavast, Arvi 2008. The Translator is Human Too: A Case for Instrumentalism in Multilingual Specialised Communication. *Dissertationes philologiae estonicae Universitas Tartuensis* 21. Tartu: Tartu University Press.
- Tavast, Arvi 2012. Keelehoolde teoreetiliste aluste otsingul: naabervaldkonnad appi. [In search of a theoretical justification for language planning: contributions from neighbouring subject fields.] – Maire Raadik, Tiina Leemets (koost ja toim). *Sõnaga mõeldud mõte*. Pühendusteos Tiiu Eretille 20. aprillil 2012. Tallinn: Eesti Keele Sihtasutus, 224–234.
- Tavast, Arvi; Taugar, Marju 2013. Mitmekeelne oskussuhtlus. [Multilingual specialised communication.] <http://tavast.ee/public/opik/opik.pdf> (1.7.2013).
- Temmerman, Rita 2000. Towards New Ways of Terminology Description: the Sociocognitive Approach. Amsterdam/Philadelphia: John Benjamins. <http://dx.doi.org/10.1075/tlrp.3>
- Temmerman, Rita 2002. Metaphorical models and the translation of scientific texts. – *Linguistica Antverpiensia*, 1, 211–226.
- Temmerman, Rita 2011a. Stars, problem children, dogs and cash cows: Evocative terminology in multilingual business communication. – *Synaps: A Journal of Professional Communication*, 25 (2), 48–61.
- Temmerman, Rita 2011b. Ways of managing the dynamics of terminology in multilingual communication. – *Scolia*, 25, 105–122.
- Valdre, Andres 2012. Tehnikakeel kui keel, mida päriselt ei osata. Eesti tehnikakeele olukord. [Technical Estonian as a poorly spoken foreign language: The situation of technical Estonian.] – Maire Raadik, Tiina Leemets (koost ja toim). *Sõnaga mõeldud mõte*. Pühendusteos Tiiu Eretille 20. aprillil 2012. Tallinn: Eesti Keele Sihtasutus, 236–273.

Allikad

- Ernits 1997 = Valik sõjandustermineid 1997. [Selection of Military Terms.] Koostaja Heino Ernits. Tallinn: Eesti Entsüklopeediakirjastus.
- Ernits 1998 = Valimik sõjandustermineid 1998. [Selection of Military Terms.] Koostaja Heino Ernits. Tallinn: Eesti Entsüklopeediakirjastus.
- Jalaväetaktika terminid 2007 = Paimla, Raigo 2007. Jalaväetaktika terminite ühtlustamine. [Infantry Tactics Terminology Harmonisation in Estonian.] Magistritöö. Juhendajad Reet Hendrikson-Pungar ja mjr Martin Herem. Tartu: Kaitseväe Ühendatud Õppeasutused.
- Jao instruktoriga käsiraamat 2008. [Infantry Section Leader Handbook.] Võru: Kaitseväe Võru Lahingukool.
- KV MLA alg = Eesti Kaitseväe maavägede lahingutegevuse alused. Maaväe ohvitseride ja staabitöö väljaõppeks. 2009. [Army Manual of the Estonian Defence Forces (draft).] Tartu: Kaitseväe Ühendatud Õppeasutused.
- KV MLA 2010 = Eesti Kaitseväe maaväe lahingutegevuse alused. Maaväe ohvitseride ja staabitöö väljaõppejuhend 2010. [Army Manual of the Estonian Defence Forces.] Kaitseväe Ühendatud Õppeasutuse taktika õppetool.
- Logistikaterminid 2009 = Tameri, Tarmo 2009. Eesti Kaitseväe logistikaterminoloogia korraldamine. [Logistics Terminology Harmonisation in Estonian.] Magistritöö. Juhendajad Reet Hendrikson ja kpt Erlend Zirk. Tartu: Kaitseväe Ühendatud Õppeasutused.
- Militerm = Sõjandustermnoloogia andmebaas Militerm. [The Database of Military Terminology.] <http://termin.eki.ee/militerm> (1.10.2013).
- Pioneerindus 2010 = Pöder, Vesse 2010. Pettetegevus ja selle rakendamine Eesti Kaitseväes. [Deception and its Application in the Estonian Defence Forces.] Magistritöö. Juhendajad Tõnu Tomberg ja Reet Hendrikson. Tartu: Kaitseväe Ühendatud Õppeasutused.

- Pioneeritegevuse käsiraamat 2010. [Engineers Handbook.] Tallinn: Maaväe Staap. <http://172.18.48.32/frame.html> (1.10.2013).
- Pioneeriterminid 2011 = Afanasjev, Ainar 2011. Kaitseväe pioneeriterminoloogia korrastamine. [Combat Engineering Terminology Harmonisation in Estonian.] Magistritöö. Juhendajad Reet Hendrikson ja mjr Raul Kütt. Tartu: Kaitseväe Ühendatud Õppeasutused.
- Pitka 1914 = Sõjamehe sõnastik 1914. [Warrior's Vocabulary.] Irkutski polgu kapten Pitka kokkuseatud. Tallinn: G. Pihlaka raamatukaupluse kirjastus.
- ÕS 2013 = Eesti õigekeelsussõnaraamat ÕS 2013. [Normative Dictionary of Estonian Language.] Toim Maire Raadik. Tallinn: Eesti Keele Sihtasutus. <http://www.eki.ee/dict/qs/> (5.2.2014).

Reet Hendrikson (Kaitseväe Ühendatud Õppeasutused, Tartu Ülikooli doktorant), peamised uurimisvaldkonnad on ilukirjanduslik ja erialane kujundkeel, sõjandusterminoloogia ning erialakeele tõhusus.

Ropka 19–13, 50111, Tartu, Eesti
reet.hendrikson@mil.ee

ASPECTS OF SPECIALISED LANGUAGE EFFICIENCY: OVER-DIFFERENTIATION OR COGNITIVE MOTIVATION?

Reet Hendrikson

Estonian National Defence College

This article addresses terminological (over)differentiation based on military terminology. The aim of the article is to discuss the boundary between over-differentiation and cognitively motivated term variants whose function is to ensure accurate specialised communication. The traditional view states that the functionality and unambiguity of SL (specialised language) is guaranteed by invariability of terms (established standards) and thus “unnecessary” term variants are rejected.

There are a number of other LSP (language for special purposes) topics directly associated with differentiation (e.g. synonymy, polysemy, internal loans, conceptual hotchpotch). My study has shown that officers’ opinions about the necessity of differentiation vary considerably, blurring the boundaries between differentiation and over-differentiation. The changes in conceptual relations over time and the impact of sociocultural factors (e.g. power relations) both add complexity to the issue. The role of a donor language is ambivalent too: although interference is largely deprecated, in the case of complex relationships between terms, reduction to the source language often serves as a key for specialists for identifying concepts. This study suggests that attempts towards terminology harmonisation may cause even more confusion at first stage.

As current results indicate it is not functional to define generally established rules when aiming to find boundaries between differentiation and over-differentiation. Instead, we need to ask which aspects of each specific case to consider. Further research is needed to investigate the real impact of term variations in the field.

Keywords: military terminology, LSP planning, (over)differentiation of term variants, cognitive motivation, Estonian

THE TRANSLATION OF IDIOMS IN CHILDREN'S CARTOONS: A COMPARATIVE ANALYSIS OF ENGLISH DIALOGUES AND LITHUANIAN SUBTITLES

Ligita Judickaitė-Pašvenskienė

Abstract. The article is focused on the translation of English idioms in the Lithuanian subtitles of children's cartoons. The aims of the article are to find out whether the element of meaning or the composition of meaning and form gets preference in the Lithuanian translation; to present the perception and use of the term *idiom* in English and Lithuanian; and to show the processes which take place during the translation of English idioms containing proper nouns. The article is a case study and refers to the analysis of five children's cartoons.

Keywords: idioms, phraseology, subtitling, cartoons, English, Lithuanian

1. Introduction

The rendering of idiomatic expressions is considered one of the most problematic aspects of translation which has been widely discussed by various scholars all over the world, such as Charles Hockett (1958), Jürg Strässler (1982), Mona Baker (1992), Rosamund Moon (1998), to mention but a few. For instance, Baker (1992), among other difficulties, points out the difficulty of noticing idiomatic expressions in a source language (SL) text. As regards Lithuanian idioms, they may contain words of 'non Lithuanian origin, barbarisms, and rarely used vernacularisms' (Paulauskas 2003: 7), all of which may result in miscomprehension of such idioms. Apart from the widely discussed issue of the absence of appropriate equivalents in a target language (TL), Ghazala (2003: 213) notes that different cultures may have taboos related to religion, social and cultural practices; such taboos may lead to the elimination of SL idioms in a TL text.

The rendering of idioms in audiovisual translation (AVT), which is a comparatively new branch of translation studies, has also received considerable attention

internationally. However, idiomatic expressions in the subtitles of children's cartoons have to date received little attention from researchers internationally and it is a virtually unexplored topic in Lithuania.

It has to be noted that dubbing is the preferred mode of cartoon translation for child audiences in Lithuania, especially in cinemas where all children's cartoons are dubbed. Yet, the situation is different in the DVD sphere, where subtitling is still present as the form of translation. Moreover, Judickaitė-Pašvenskienė (2013) notes that some 'children tend to choose the subtitled versions of cartoons, since they are interested in hearing the original soundtrack with the attempt to try to understand the dialogue themselves'; in addition, there are cases 'when parents buy subtitled cartoons for their children with the purpose of exposing the children to a foreign [---] language in this way accelerating foreign language acquisition' (Judickaitė-Pašvenskienė 2013: 165). Thus, subtitling for children is an important sphere for investigation, as it may serve as a mean of foreign language learning.

2. Aim, material and method

The main aim of the article is to find out which element, the meaning or the composition of meaning and form, is given preference in the translation of English idioms into Lithuanian in five children's cartoons. In other words, the article investigates whether the SL idioms are translated in the subtitles with an attempt to retain their meaning, without regard for their idiomatic form, or whether an attempt is made to maintain both idiomatic form and meaning. The study also aims to indicate the differences in the use of the term 'idiom' in English and Lithuanian, as well as to show how Lithuanian subtitlers handle the challenge of the translation of culture-specific SL idioms which contain proper nouns.

This article presents a comparative analysis of 192 (including repetitions) English language idioms and their Lithuanian translations in the subtitles of five children's cartoons on DVD which are on release in the UK or the USA: *Valiant* (2005), *The Wild* (2006), *Over the Hedge* (2006), *Horton Hears a Who* (2008) and *Rio* (2011). It should be noted that phrasal verbs, common discursal expressions such as *how do you do?* and foreign phrases are not included in the corpus. English phrasal verbs usually have as their counterpart a single word in Lithuanian, which cannot normally be ascribed to idioms, whereas discursal expressions rarely present any idiomatic problems in translation. In order to indicate the levels of idiomaticity and to highlight their possible influence on the TL translations, SL idioms are subdivided into three categories. The first two categories, *semantically unmotivated* (also called pure or opaque by different authors) and *semantically motivated* (also called literal or transparent), make use of Evalda Jakaitienė's (1980: 100) classification and are discussed in more detail in section 3.1. The last category of idioms in this study is culture specific idioms which contain proper nouns; they are discussed in section 3.2. Despite the fact that researchers of idioms tend to outline one more group, semi-idioms or semi-opaque idioms, this has been avoided in the present study since in many cases the distinction between semi-idioms and semantically motivated (or literal) idioms is arguable. Therefore, the category of motivated idioms in this study contains both semantically motivated and

semi-idioms. It should be mentioned that the analysis also includes idioms which appear in transformed versions in the cartoons. For instance: *read my beak* is the transformation of *read my lips*; the use of such transformations corresponds with the situations in the cartoons, where the characters are mainly animals.

There were two criteria which governed the choice of the cartoons: they needed to have been played in cinemas in Lithuania (presumably, such cartoons have attracted more attention from the viewers than those released only in DVD format and not played in cinemas, as the former are much more widely advertised) and to have been released during the last ten years. The cartoons selected for the study are the subtitled DVD versions of the dubbed ones that had previously been played in cinemas. All the cartoons are treated as separate individual translations since the absence of the names of their subtitlers, except in one case, prevented the present author from tracing the practices employed by the same translator or searching for similarities and differences between the translations done by different translators.

This article is a descriptive case study and makes use of the product-oriented approach by James S. Holmes in that it is aimed at analyzing a product – children’s cartoons – and presenting a description of translation processes in a single ST (English dialogue) – TT (Lithuanian translation) pair (Holmes 1988 [1972]: 177). Also, the method of coupled pairs by Gideon Toury is used, as only specific language units, i.e. idioms, are analyzed and compared instead of concentrating on the SL and TL texts as a whole (Toury 1995: 81).

3. Theoretical background

3.1. The notion of idiom

An *idiom* can be traditionally defined as ‘a group of words whose meaning is different from the meanings of the individual words’, as stated in Oxford Advanced Learner’s Dictionary (2000: 643), or ‘a type of collocation’ or ‘multiword lexeme’, as indicated by Jackson and Amvela (2001: 65). However, a deeper analysis of the term shows that its definition varies. Moon (1998) looks at the use of the term *idiom* by various researchers and states:

In broader uses, idiom is a general term for many kinds of multiword item, whether semantically opaque or not... Makkai uses idiom to cover non-compositional polymorphemic words such as *blackbird* as well as collocations and constructions that are not freely formed (1972). Hockett’s view is still broader, embracing even single morphemes, since their meanings cannot be deducible (1958: 171 ff.) [---] Sadock (1972) draws attention to the ambiguity of utterances which have more than one pragmatic function: ambiguity leads to status as idioms. (Moon 1998: 4)

Thus, the perception of *idiom* differs with different authors and in many cases a deviation from the traditional ‘dictionary’ definition can be noticed, yet some generalizations are still possible. For example, Ghazala (2003: 204) summarizes various dictionary definitions of idioms and states that several common points can be outlined: idioms are metaphorical and not literal; they have fixed syntactic form; their meanings are invariable; idioms are usually culture-specific and informal.

Similarly to the definition and perception, the classification of idioms is also a complex and variable matter. For instance, Baker (1992) separates idioms from such units as collocations and fixed expressions. Such a division is based on two criteria, namely the flexibility of patterning and the transparency of meaning. Baker sees idioms and fixed expressions as more or less frozen units, while collocations, on the other hand, are seen as being flexible in form. In contrast to idioms, fixed expressions and collocations are considered fairly transparent in meaning (Baker 1992: 63-64). Moon (1998: 2) notes that some authors use the term *fixed expression* as an umbrella term for such categories as idioms, frozen collocations, grammatically ill-formed collocations, proverbs, routine formulae, sayings and similes. It is also important to stress that despite the traditional understanding of an idiom as a metaphorical unit, such items as jargon, common phrases and terms, which do not present any metaphorical problems, are sometimes ascribed to the category of idioms by some authors (Ghazala 2003: 206).

In comparison to the English language, the term *idiom* is not so widely used in Lithuanian. Instead, an umbrella term *phraseologism* or *phraseological unit* is more common. According to one of the main Lithuanian lexicologists, Jakaitienė (1980), a *phraseologism* is understood as an opposition to a free word unit and defined as a phrase ‘whose meaning is totally or partially different from the meanings of the constituent words [---] for example *to sit in somebody’s pocket* means “to be indebted” (Jakaitienė 1980: 95).¹ In this respect the Lithuanian terms *frazologizmas* ‘phraseologism’ or *frazologinis junginys* ‘phraseological unit’ is similar to the English *idiom*.

From the semantic point of view, Lithuanian phraseologisms are divided into two main groups, namely, *semantically unmotivated* and *semantically motivated* ones. The former group includes *idioms*, which are understood as ‘totally fixed, self-contained phraseologism whose meaning is very distant from the meaning of its constituent words’, for instance, the Lithuanian idiom *a wind’s whip* refers to a frivolous person (Jakaitienė 1980: 100). Thus, the denotative (direct) meanings of the constituent words do not motivate the overall meaning of the idiom as none of the words are used literally; ‘the image presented in the idiom is not real [---] and its relation with the denotation is not tangible’ (Jakaitienė 1980: 101). In opposition, semantically motivated phraseologisms are word combinations ‘in which the connection between the meaning of the individual words and that of the whole expression is tangible, although the connection is more or less obscure’ (Jakaitienė 1980: 102). The group of semantically motivated phraseologisms includes tropes, which are understood as metaphoric or metonymic phraseologisms, similes and phraseological combinations, which are considered to be the loosest in form and are in the periphery of Lithuanian phraseology (ibid.). For example, the expression *to convert something into ashes* means ‘to burn’; here the connection between the image created by the expression and the idiomatic meaning is clear as ashes are the result of burning (Jakaitienė 1980: 105). Although different traditions in the use of terminology can be noticed in English and Lithuanian, for the sake of clarity the term *idiom* is used in this article to refer to figurative expressions which otherwise would be called *phraseologisms*, *similes*, *proverbs* or *sayings* in various reference books.

3.2. Idiom as a multifunctional item in translation

One of the reasons why idioms are difficult to translate is their multifunctionality. Indeed, it can be claimed that idioms are comprised of several constituent elements which serve different functions. This article is focused on three main elements: meaning, syntactic form and cultural specificity.

Baker (1992: 66) states that '[a] large number of idioms in English, and probably all languages, have both a literal and an idiomatic meaning [---]'. In cases where idioms are grammatically ill-formed they prevent themselves from possible literal meanings, as the grammatical relationship between the words do not allow for literal interpretations. However, when idioms are formed according to the rules of grammar, they may create a false impression of being unidiomatic (1).

- (1) No. I'm not falling for any more of your smooth talk!
I don't know what you're up to, but my entire shell is tingling.
Know what? I'm listening to it this time,
– and **I'm putting my foot down.** (OH)

In this scene the turtle character, in his quarrel with the raccoon character, uses the idiom *to put one's foot down* to show his objection. Considering the grammar, the idiom can be interpreted both literally and idiomatically. However, according to the situation in the scene, the idiomatic meaning is the one which is important in this case. Thus, one of the functions of idioms is to carry a message or to convey a meaning which is important to be understood by the recipients.

There also appear cases when both the literal and the idiomatic meanings of the idiom are present at the same time (2).

- (2) I took a few clippings out of my quills to do a little comparison.
Look at this, **the grass seems to be greener** over here. (OH)

In this episode the viewers are shown a group of animal characters who cross the hedge which separates their forest from the newly built suburb and enter the to-them-undiscovered world of human civilization. As they stand on the neatly cut lawn in one of the yards of the suburb, the porcupine character uses a modified version of the proverb *the grass is always greener on the other side of the fence*. In this case both the literal and the idiomatic meanings are present since the character is pointing at the grass and, at the same time, expressing the idea that the life on this side of the hedge is better than their life in the forest. Such polysemy serves a comic function.

One more element of the idiom is specific syntax. As stated in Paulauskas (2001: VII), the meaning of the phraseologism can 'often be expressed by a single, totally different word', thus, by its specific syntactic form the idiom 'often expresses the speaker's attitude towards a certain issue [---] in this way serving a stylistic function in a language'.

- (3) Hang on a sec!
I've got miracles **comin' out my ears!** (W)

The meaning of the idiom *coming out of one's ears* (3) can also be expressed by such stylistically neutral phrases as *a huge amount of something, more than one*

needs, etc. Yet, these phrases lack the vividness and the dynamism of the idiom. Thus, the form is important as it is used for the specific purpose of creating a particular style.

The last element of idioms which is considered in the present study is culture specificity. Baker (1992) equates culture specificity with various patterns of social behavior, fixed linguistic formulae or references to specific places or objects (e.g. *to carry coals to Newcastle*), etc. (Baker 1992: 68–69). Often, such idioms have no equivalents in a TL since social behavior may differ in different cultures and the referents of SL idioms are totally unknown or have different connotations in a TL. This article presents the analysis of only a small fraction of culture specific idioms, namely those which contain proper nouns and refer to places, objects or people. They are close to Pedersen's notion of *extralinguistic culture-bound references* (ECR) since they pertain 'to realia, or cultural items, which are not part of a language system' (Pedersen 2005: 2). Although Pedersen excludes idioms from the notion of ECR, it has to be noted that references to places, objects or people in idioms call more for knowledge of an SL culture rather than (its) language.

3.3. Idioms in subtitling

Subtitling is a mode of audiovisual translation in which spoken language is changed into writing and is presented at the bottom of the screen in the form of one or two lines of text. Given the fact that the translation needs to fit into a maximum of two lines which consist of roughly up to 40 characters each and have to be synchronic with and match the visual information on the screen, the otherwise problematic task of idiom translation in literary texts becomes even more challenging in subtitling. The multimodality of the film comes into play as apart from the verbal and the audio codes, the visual code becomes very prominent. In order to preserve the entirety of the film, it is important that the three codes be presented in harmony with each other. Example (2) in the previous section shows that the image of grass is present not only in the idiom which the porcupine character utters, but is also shown on the screen; thus, such interconnectedness of the two codes makes the image of grass important to be retained in subtitles.

Considering cartoon subtitling for children, several problematic aspects specific to this field should be emphasized. Professional subtitler Tortoriello (2006: 55) notes that, unlike adults, children are a specific audience 'with a rather limited knowledge of the world' who want to 'get the message immediately'. In cases when SL idioms are culture specific and have no equivalents in a TL the translation may be in danger of being misunderstood. Tortoriello (2006: 56) also emphasizes that reading speed is different for adult and children audiences (160 and 120 words per minute respectively), and thus dialogues need to be presented in an extremely condensed form. Consequently, the opportunities for explicitation, while limited in subtitling in general, become even more limited in subtitling for children. In addition, idioms function as a feature of spoken language and are important to be retained in subtitles according to the requirement presented in *Code of Good Subtitling Practice* (Carroll, Ivansson 1998), which states that '[t]he language register should be appropriate and correspond with the spoken word'.

Despite the fact that subtitling and audiovisual translation in general is a comparatively new branch of translation studies which emerged only in the 20th century, researchers of subtitling have shown a great interest in idioms; Henrik Gottlieb's book *Subtitles, Translation & Idioms* (1997) is probably the most prominent work in the field so far. A number of smaller case-studies also show a great distribution among different languages. For instance, Abd-el-Kareem (2010) investigates idioms in Arabic–English subtitled films; Theodoropoulou (2006) presents the analysis of English idioms in Greek subtitles; Chen (2009) devotes a part of his article to the translation of idioms from English into Chinese in the film *Spellbound*, etc. However, idioms in subtitling, like AVT on the whole, are still an unexplored sphere in Lithuania. During the period 2005–2009 several authors published their articles on both national and international levels. Baravykaitė (2005, 2006, 2007) addresses such topics as the general position of AVT in Lithuania, subtitling as a means of language learning and general translation problems in subtitling. Baranauskienė and Blaževičienė (2008) take sentences as units for their analysis with the aim of finding the most frequently used translation procedures in the voiced over Lithuanian version of the film *Bridget Jones's Diary*. Judickaitė-Pašvenskienė (2009a, 2009b, 2013) concentrates on cartoon subtitling: she has analyzed the translation of culture specific items and presented the analysis of spoken English discourse and devoted part of the analysis to the translation of idioms; she has also considered subtitled cartoons in the context of language teaching and learning. Thus, even though initial investigations have been made, the translation of idioms in AVT, especially in children's cartoons, still lacks scholarly attention in present-day Lithuania.

4. Translation of idioms

As mentioned in section 2, the principal aim of the present study is to find out which element, the meaning or the composition of meaning and form is given preference in the translation of idiomatic SL expressions. Given that subtitling is a constrained form of translation and idioms are one of the most challenging items in translation, it should be noted that the classification of the Lithuanian renderings do not follow the traditional patterns. To make it clearer, if the Lithuanian translations of the SL idioms preserve the meaning of the original and contain one or more words which are used in their figurative sense, such renderings are treated as containing both the idiomatic form and the meaning. The renderings which present the meaning of the SL idioms and contain the words which are used in their denotative meaning are treated as having no qualities of idiomatic form.

4.1. Unmotivated idioms

The analysis shows that the dialogues of the cartoons contain 34 unmotivated idioms, which comprise 18% of the total number of idioms which appear in the cartoons. Table 1 illustrates which elements of the SL unmotivated idioms are given preference in the Lithuanian subtitles.

Table 1. The elements of the SL idioms which are transferred in the TL subtitles

Elements transferred	Findings
Composition of idiomatic meaning and form	6
Meaning	25
Changes in idiomatic meaning	3
Omission	0
Total number of unmotivated idioms	34

According to the data presented in Table 1, the translators' major attempts are to retain the meaning of the English idioms in the Lithuanian subtitles; in this process the idiomatic form is not retained.

- (4) English dialogue: I am beat. Time **to hit the hay**, I guess. (HHW)
Lithuanian subtitle: Visai nusivariau nuo kojų. Ko gero, laikas **eiti miegoti**. (HHW)
Back-translation: I'm completely worn out. I guess it's time **to go to sleep**

As example (4) shows, the SL idiom *to hit the hay* is reduced to its meaning in the Lithuanian subtitle and the idiomatic form is not retained. It should be noted that in this case the translator might have been constrained by the number of characters, as the subtitle already consists of 29 characters. The analysis shows that in half of the cases in which the SL unmotivated idioms are reduced to meaning the subtitlers are constrained by a limited number of characters or a fast rate of speech.

In six cases both the meaning and the idiomatic form are retained in the Lithuanian subtitles.

- (5) English dialogue: I've got miracles **comin' out my ears!** (OH)
Lithuanian subtitle: Stebuklų turiu **per akis!** (OH)
Back-translation: I've got miracles **over my eyes!**

In the example (5) the English idiom, which has a meaning 'a huge amount of something', is translated using the Lithuanian idiom of the same meaning.² It is noticeable that the main image of the SL idiom is changed in the translation (*ears-eyes*). However, both the figurative meaning and the idiomatic form are retained, and such a change does not contradict the images on the screen.

It should also be noted that there appear three cases in which changes in idiomatic meaning of the SL idioms can be noticed, i.e. the TL idiom has a different meaning from that of the original; the image created by the TL rendering is obscure to the TL audience, evokes different connotations in comparison to the SL idiom or does not correspond with the plot line of the episode, e.g. (6).

- (6) English dialogue: If we don't find some more birds fast, **our goose is going to be cooked.** (V)
Lithuanian subtitle: Jei mes greitai nesurasime daugiau paukščių... **mūsų žąsis bus iškepta.** (V)
Back-translation: If we don't find some more birds fast, **our goose is going to be cooked.**

² Lietuvių kalbos žodynas [Dictionary of the Lithuanian Language], <http://www.lkz.lt/startas.htm> (27.12.2013).

The cartoon *Valiant* is set in 1944, when Britain fights against Germany in World War II. In the scene presented in example (6) Wing Commander Gutsy receives a message from a seagull near Dover that all of their top flyers were killed and did not manage to come back to Britain. Wing Commander realizes that in order to be able to continue fighting they need more flyers and indicates that by using the idiom *our goose is going to be cooked*, which has a meaning of ‘one is finished’ or ‘our plans will be spoilt’. The word-for-word translation in the subtitle reflects only the literal meaning of the idiom, which is different from the idiomatic one. The overall image created by the Lithuanian rendering does not carry the negative connotations which are present in the original. The *Corpus of the Contemporary Lithuanian Language*³ shows that the phrase *to cook the goose* (‘(iš)kepti žąsį’) is used in culinary contexts, e.g. ‘We [---] start cooking the goose at the temperature of 180 degrees Celsius’. Although some traces of ‘being finished’ may be noted in the word *iškepta* (‘cooked’), such an interpretation may lead to an assumption that Gutsy is referring to a goose character who is going to be killed (‘cooked’), which does not correspond with the plot line of the cartoon, as none of the characters are geese, and may mislead the audience. It should be noted that in this episode the war map with various pins is shown in the background. While uttering the idiom, Gutsy takes one pin, which is in the form of a bird, off the map and throws it on the ground. Such action can be interpreted as having negative connotations (e.g. destruction) and therefore may serve as the reinforcement of the idiom. Considering the fact that all the characters in the cartoon are birds, the idiom also functions as a pun and it might be one of the reasons why the translator attempted to preserve the image of a bird in the Lithuanian subtitle.

The analysis of the three cases in which a change in idiomatic meaning of the SL idioms is observed shows that in two cases the meaning of the SL idioms is not clearly reinforced by the plot of the cartoons or the action in particular episodes and this might be one of the possible reasons for the changes which occur in the subtitles.

4.2. Motivated idioms

The SL dialogues of the cartoons contain 156 motivated idioms, which make up 81% of the total number of all SL idioms that appear in the cartoons. Considering the elements which get preference in the process of translation, the tendencies are similar to the aforementioned unmotivated idioms.

Table 2. The elements of the SL idioms which are transferred in the TL subtitles

Elements transferred	Findings
Composition of idiomatic meaning and form	68
Meaning	81
Changes in idiomatic meaning	5
Omission	2
Total number of motivated idioms	156

³ <http://tekstynas.vdu.lt/tekstynas/> (27.12.2013).

As Table 2 illustrates, the element of meaning gets preference in the translation of motivated SL idioms in the Lithuanian subtitles.

- (7) English dialogue: Nothing you say **makes any sense**. (RI)
Lithuanian subtitle: Kalbi **visiškas nesąmones**. (RI)
Back-translation: You're **talking total nonsense**.

In the example (7) the meaning of the SL idiom is transferred in the subtitle; however, considering the form, the Lithuanian translation is not an idiom and none of the words are used in their figurative sense. The comparison of unmotivated and motivated idioms shows that motivated idioms are reduced to meaning disregarding the idiomatic form 1.3 times less often in the Lithuanian subtitles. Therefore, it can be stated that motivated idioms are easier to subtitle such that both the meaning and the form are retained.

The second element which gets preference in the translation is the composition of meaning and form, as illustrated in example (8):

- (8) English dialogue: You done **runnin' your mouth** yet, Carmine? (W)
Lithuanian subtitle: Ar dar nebaigei **malti žabtais**, Karmainai? (W)
Back-translation: Haven't you finished **beating your jaws**, Carmine?

In this scene one character gets angry with the other and uses the idiom *run your mouth*, whose meaning is 'to talk a lot'. In the subtitle the idiom is replaced with the Lithuanian metaphoric phrase *malti žabtais*, in which the verb *malti* has a figurative meaning of 'to speak fast⁴' and the noun *žabtai* is a pejorative form for 'mouth' or 'lips' (ibid.). Although the *Corpus of the Contemporary Lithuanian Language* shows that the phrase *malti liežuvii*⁵ ('to beat one's tongue') is used in the contexts which refer to talking a lot, the word *žabtai* serves as an insult and thus matches the mood of the character. Thus, it can be claimed that both the idiomatic meaning and form are retained in the subtitle. The comparison of unmotivated and motivated idioms with regard to the composition of meaning and form shows that in the case of motivated idioms the meaning together with idiomatic form is retained 2.5 times more often in the subtitles.

It should be noted that in five cases changes in idiomatic meaning of the SL motivated idioms can be noticed in translation, e.g. (9).

- (9) English dialogue: Rafi has some special guests from **out of town** and let's show them some love. (RI)
Lithuanian subtitle: Rafis atsivedė ypatingus svečius **iš miesto**, būkim jiems svetingi. (RI)
Back-translation: Rafi has brought some special guests **from town**, let's be hospitable to them.

In this episode the bird character Nico introduces two macaws, one of which is a foreigner while the other is from a different town, to his friends at a discotheque and uses the idiom *out of town* to indicate that the two guests are not locals. However, in the TL subtitle one reads that the guests are *from town*, as if the opposition between the town and the suburb or a village is intended to be made, which is not present in the scene. The analysis of all cases of motivated idiom mistranslation

⁴ Lietuvių kalbos žodynas [Dictionary of the Lithuanian Language], <http://www.lkz.lt/startas.htm> (25.2.2012).

⁵ <http://tekstynas.vdu.lt/tekstynas/> (27.12.2013).

shows that one of the possible reasons for changes in idiomatic meaning could be the challenges in noticing the SL idioms.

Finally, two cases of omission appear in the subtitles, yet no information is lost as in both cases the subtitler omits the repetitions of idioms. On the whole, considering the constrained nature of subtitling, the number of omissions is very small. However, it is reasonable as most idioms are semantically important units which, in addition to their stylistic functions, carry relevant information.

4.3. Culture-specific idioms

Only two culture-specific idioms which contain proper nouns are present in the dialogues of the cartoons (10), (11). None of the proper nouns are transferred in the subtitles or replaced by Lithuanian proper nouns.

- (10) English dialogue: Just follow us and **Bob's your uncle**. (W)
Lithuanian subtitle: Sekite paskui mus ir **rūpesčių nematysit**. (W)
Back-translation: Follow us and you **won't see any troubles**.
- (11) English dialogue: I'm from New York City, **for Pete's sake**. (W)
Lithuanian subtitle: Juk aš iš Niujorko, **po šimts pypkių**. (W)
Back-translation: After all, I'm from New York, **for a hundred pipes**.

The unmotivated SL idiom *Bob's your uncle* is usually used as a concluding phrase for a simple instruction to mean, 'and there you have it,' or 'you're all set'⁶. As the idiom is culture-specific, no equivalent with the same proper noun exists in the Lithuanian language; thus the Lithuanian rendering is a motivated phraseologism in which the verb *nematysit* ('you will not see') is used in its figurative sense 'you will not experience'. The unmotivated idiom in the next subtitle, *for Pete's sake*, is used to express annoyance; similarly to the previous idiom, it has no equivalent in Lithuanian which would contain the same proper noun. Therefore, the translator uses the Lithuanian unmotivated idiom *po šimts pypkių*, which is also used as a mild curse to express annoyance.

Since only two culture-specific idioms which contain proper names appear in the selected cartoons, no conclusions can be drawn on this part of the study, nor can any tendencies be traced. A study on a larger scale is needed in order to get more revealing results.

5. Conclusion

According to the analysis, the dialogues of the cartoons contain 34 unmotivated, 156 motivated and 2 culture-specific idioms containing proper nouns. In the translation of unmotivated idioms the element of meaning gets preference, as 25 idioms are reduced to meaning without transferring the idiomatic form. One of the reasons for that is the very nature of subtitling, i.e. a limited number of characters and a fast rate of speech. Six idioms retain both the form and the meaning in the subtitles,

⁶ The Free Dictionary, <http://encyclopedia.thefreedictionary.com/Bob%27s+your+uncle> (12.1.2012).

while in three cases changes in idiomatic meaning are noticed. One of the possible causes for such change might be the fact that the meaning of idioms is not always reinforced by the images on the screen or the plot of the cartoons.

With regard to motivated idioms, in the majority of cases, 81, the element of meaning is retained in the subtitles; the composition of meaning and form is retained in 68 cases, which is 2.5 times more often than in the translation of unmotivated idioms. This shows that the subtitling of motivated idioms is less challenging. In five cases changes in idiomatic meaning are noted. Also, there appear two cases of omission, although it is partial, as only the repetitions of idioms are omitted in the Lithuanian subtitles, which causes no loss of information.

Considering the group of culture specific idioms containing proper nouns, no conclusions can be drawn as only two such idioms appear in the corpus. In order to trace any tendencies in their translation, a much larger study is needed.

Abbreviations

- AVT audiovisual translation
ECR extralinguistic culture-bound reference
SL source language
TL target language

Corpus bibliography

- HHW = Horton Hears a Who. B. Anderson, B. Gordon (Producers), J. Hayward, S. Martino (Directors) 2008. [Animated movie.] USA: Twentieth Century Fox Animation.
OH = Over the Hedge. B. Arnold (Producer), T. Johnson, K. Kirkpatrick (Directors) 2006. [Animated movie.] USA: DreamWorks Animation.
RI = Rio. B. Anderson, J. C. Donkin, C. Wedge (Producers), C. Saldanha (Director) 2011. [Animated movie.] USA: 20th Century Fox Home Entertainment.
W = The Wild. C. Goldman (Producer), S. Williams (Director) 2006. [Animated movie.] USA: Walt Disney Pictures.
V = Valiant. J. H. Williams (Producer), G. Chapman (Director) 2005. [Animated movie.] UK: Buena Vista Home Entertainment.

References

- Abd-el-Kareem, Amani Mahmoud Ahmad 2010. Towards strategies for translating idioms in Arabic English subtitling. – Canadian Journal of Communication. Thesis Abstracts. <http://www.cjc-online.ca/index.php/journal/thesis/view/125> (11.4.2012).
Baker, Mona 1992. In Other Words. London, New York: Routledge. <http://dx.doi.org/10.4324/9780203327579>
Baranauskienė, Reda; Blaževičienė, Rasa 2008. Audiovisual translation of feature films from English into Lithuanian. – Jaunųjų mokslininkų darbai, 4, 14–21.
Baravykaitė, Alina 2005. Filmų vertimo problematika. – Kalbotyra, 55 (3), 7–14.
Baravykaitė, Alina 2006. Audiovizualaus vertimo tendencijos Lietuvoje. Darbai ir dienos 45. Kaunas: VDU leidykla, 63–77.
Baravykaitė, Alina 2007. Kino filmų titrai kaip kalbų mokymosi priemonė. – Kalbotyra, 57 (3), 21–29. http://www.sub2learn.ie/downloads/audiovisual_translation_engllithuanian.pdf (19.7.2012).

- Carroll, Mary; Ivarrson, Jan 1998. Code of Good Subtitling Practice. TransEdit. <http://www.transedit.se/code.htm> (28.8.2011).
- Chen, Chapman 2009. A critical evaluation of a Chinese subtitled version of Hitchcock's *Spellbound*. – Gilbert Fong, Kenneth Au (Eds.). 2009. Dubbing and Subtitling in a World Context. Hong Kong: Chinese University Press, 107–136.
- Ghazala, Hasan 2003. Idiomaticity between evasion and invasion in translation: Stylistic, aesthetic and connotative considerations. – Babel, 49 (3), 203–227. <http://dx.doi.org/10.1075/babel.49.3.03gha>
- Gottlieb, Henrik 1997. Subtitles, Translation & Idioms. Copenhagen: University of Copenhagen.
- Hockett, Charles 1985. A Course in Modern Linguistics. New York: Macmillan.
- Holmes, James S. 1988 [1972]. The name and nature of translation studies. – James S. Holmes. Translated! Papers on Literary Translation and Translation Studies. Amsterdam: Rodopi, 67–80.
- Jackson, Howard; Ze Amvela, Etienne 2001. An Introduction to Modern English Lexicology. London, New York: Continuum.
- Jakaitienė, Evalda 1980. Lietuvių kalbos leksikologija. Vilnius: Mokslas.
- Judickaitė, Ligita 2009a. The notions of foreignization and domestication applied to film translation: Analysis of subtitles in cartoon *Ratatouille*. – Jaunųjų mokslininkų darbai, (2) 23, 36 – 43.
- Judickaitė, Ligita 2009b. Analysis of Spoken English Discourse Translation in the Lithuanian Subtitles of the Cartoon *Over the Hedge*. – Valoda – 2009. Valoda dažadu kultūru kontekstā. Daugavpils: Daugavpils University, 420–427.
- Judickaitė-Pašvenskienė, Ligita 2013. Subtitled cartoons in foreign language teaching and learning context: possible dangers. – Darnioji daugiakalbystė / Sustainable Multilingualism, 2, 161–171.
- Moon, Rosamund 1998. Fixed Expressions and Idioms in English: a Corpus Based Approach. Oxford: Clarendon Press.
- Oxford Advanced Learner's Dictionary. 6th ed. Oxford: Oxford University Press, 2000.
- Paulauskas, Jonas (Ed.) 2001. Frazeologijos žodynas. Vilnius: Lietuvių kalbos institutas.
- Paulauskas, Jonas (Ed.) 2003. Lietuvių kalbos frazeologijos žodynas. Vilnius: Šviesa.
- Pedersen, Jan 2005. How is culture rendered in subtitles? http://www.euroconferences.info/proceedings/2005_Proceedings/2005_Pedersen_Jan.pdf (21.1.2011).
- Strässler, Jürg 1982. Idioms in English: A Pragmatic Analysis. Tübingen: Narr.
- Theodoropoulou, Irene 2006. Translating English idioms into Greek subtitles: The functional variation model proposal. Unpublished Master's thesis. Cambridge, Research Centre for English and Applied Linguistics, University of Cambridge.
- Tortoriello, Adriana 2006. Funny and educational across cultures: Subtitling *Winnie the Pooh* into Italian. – The Journal of Specialized Translation, 6. http://www.jostrans.org/issue06/art_tortoriello.php (15.9.2011).
- Toury, Gideon 1995. Descriptive Translation Studies and Beyond. Amsterdam: John Benjamins.

Online references

- Lietuvių kalbos žodynas. [Dictionary of the Lithuanian language.] <http://www.lkz.lt/startas.htm> (25.2.2012).
- The Free Dictionary. <http://encyclopedia.thefreedictionary.com> (12.1.2012).
- The Corpus of the Contemporary Lithuanian Language. <http://tekstynas.vdu.lt/tekstynas/> (27.12.2013).

IDIOOMIDE TÕLKIMINE LASTE MULTIFILMIDES: INGLISKEELSETE DIALOGIDE JA LEEDUKEELSETE SUBTIITRITE VÕRDLEV ANALÜÜS

Ligita Judickaitė-Pašvenskienė

Vytautas Magnuse Ülikool

Artikkel on deskriptiivne juhtumiuuring, milles analüüsitakse kuue multifilmi subtiitreid, keskendudes inglise keele idioomide tõlkimisele leedu keelde. Uurimuse põhieesmärk oli vaadelda, millist inglise idioomide aspekti eelistatakse leedu subtiitrites: kas tähendust või tähendust ja vormi üheskoos. Lisaks annab artikkel ülevaate, mida mõistetakse idioomi all inglise ning leedu keeles ja kuidas saavad Leedu subtiitrite tõlkijad hakkama idioomide tõlkimisega. Analüüsiks kasutatud korpus sisaldas 192 ingliskeelset fraseologismi (kordused kaasa arvatud), mis omakorda olid jaotatud motiveerimata, motiveeritud ja pärisnimesid sisaldavateks kultuurispetsiifilisteks fraseologismideks. Analüüsi tulemusena selgus, et multifilmides esinenud ingliskeelsetest fraseologismidest 18% olid motiveerimata, kusjuures motiveeritud ja kultuurispetsiifilised idioomid moodustavad vastavalt 81% ja 1%. Kahe esimese rühma puhul domineerib tähenduse edasiandmine, mis on osaliselt seotud subtiitritele esitatavate piirangutega ja osaliselt sellega, et tähendus ei ole otseselt seotud sündmustikuga. Kultuurispetsiifiliste idioomide kohta ei saa järeldust teha, kuna lastefilmide dialoogides esines vaid kaks niisugust idioomi ning seega seaduspärasuste leidmiseks oleks vaja laiapõhjalisemat uurimust.

Võtmesõnad: idioomid, fraseoloogia, subtiitrid, multifilmid, inglise keel, leedu keel

Ligita Judickaitė-Pašvenskienė is the head of the Translation Centre at Vytautas Magnus University, the Institute of Foreign Languages, Kaunas, Lithuania, as well as a lecturer of English and a vice president of the Language Teachers' Association in Lithuania. Currently she is a doctoral student at Tallinn University, Estonia.

ligita.judickaite@gmail.com

SISSEVAADE VARASESSE EESTI KEELE SÜNTAKSI OMANDAMISSE

Airi Kapanen

Ülevaade. Artiklis vaadeldakse kahe lapse spontaanse kõne andmes-
tiku põhjal eesti keele süntaksi varast omandamist. Laste süntaksi
varase arengu saab jaotada kolme perioodi strateegia järgi, mida laps
esimeste mitmesõnaliste lausungite moodustamiseks kasutab. And-
metest nähtub, et eesti lapsed kasutavad mitmesõnaliste lausungite
loomisel esialgu palju grammatikavälist vahendit, kordust. Lapse
vanusega korduste hulk järk-järgult väheneb ja kõnesse ilmuvad kahest
sõnast koosnevad, kuid süntaktiliselt sidumata lausungid. Viimasel
vaatlusvahemikul leidub laste kõnes süntaktilistel baassuhetel, nagu
subjekt-verb, verb-objekt põhinevaid lausungeid.

Võtmesõnad: lastekeel, kordus, tugiverbikonstruktsioonid, partiklid,
eesti keel

1. Sissejuhatus

Kuigi on uuritud, kuidas laps lausemoodustuse omandab, pole praeguse ni
ainuõigeks peetavat lähenemist. Üksmeel olla, et täiskasvanu ja lapse gram-
matika on erinevad ning et suhtlemisel on keeleomandamises tähtis roll. Eesti
lapse süntaksiomandamist on ühe eesti keelt kõneleva lapse näitel oma diplomitöös
uurinud Reili Argus (1994), mõnda keeleomandamise aspekti on vaadeldud ühes
magistritöös (Argus 2008) ja paaris bakalaureusetöös (Lindsalu 1995, Parmas
2000). Kui lausete mõistmise oskust on võimalik uurida eksperimentaalsete mee-
toditega, siis lausete moodustamise oskuse uurimismaterjaliks sobivad ennekõike
spontaanse kõne lindistused. Siinses artiklis soovitaksegi leida spontaanse kõne
materjali põhjal vastuseid järgmistele küsimustele: 1) missugust strateegiat kasutab
eesti keelt omandav laps ühest sõnast pikemate üksuste loomiseks, 2) kas varaste
lausete puhul võib tähele panna mingeid korrapäraseid struktuure. Ühtlasi vaa-
deldakse kahe teoreetilise lähenemise sobivust eesti varase süntaksi omandamise
kirjeldamiseks.

2. Kaks vaatenurka varase süntaksi omandamisele: Michael Tomasello ja Anat Ninio

Praegustes lastekeeleuuringutes esilduvad kaks peamist süntaksi omandamise käsitlust. Ühe keskmes on konstruktsioonigrammatika, täpsemalt verb ja selle ümber koonduvad üksused (Tomasello 2003). Teine peab konstruktsioonide õppimise asemel lapse ülesandeks õppida iseseisvaid leksikaalseid üksusi ja seda, kuidas need grammatiliselt käituvad (Ninio 2011: 136). Anat Ninio järgi õpivad lapsed verbisõnavara koos verbide süntaktilise ja semantilise valentsiga; tema arvates on iga lapse varane süntaks erinev ning sõltub eri teguritest.

Michael Tomasello (2003) loodud nn verbisaare teoorias (ingl *verb island theory*) on lapse varane keel korrapäraselt koondunud üksikute verbide ümber, mille vahel seoseid ei ole: iga verb on omaette organiseeritud keeleüksus, justkui saareke muidu organiseerimata keelemustris. Keele omandamisel ja lausete moodustamisel tugineb laps kõnest kogutud verbispetsiifilistele konstruktsioonidele, milles on nn avatud koht eri nimisõnadele (ingl *open nominal slot*). Sisendkeele kaudu omandatud verbide põhjal loob laps konstruktsioone koos verbist sõltuvate süntaktiliste rollidega; iga verbi õpib laps kasutama teatud kindlat tüüpi konstruktsioonis (Tomasello 2003: 118). Nii võib laps täpselt ühe ja sama arenguperioodi vältel kasutada osa tegusõnu vaid üht tüüpi ja üsna lihtsas konstruktsioonis ning teisi mitmes ja keerukas. Näiteks kasutas Tomasello uuritud laps tegusõna “lõikama” (ingl *cut*) vaid ühes (*cut + mõni nimisõna*), kuid tegusõna “joonistama” (ingl *draw*) neljas konstruktsioonis (nt *draw ...*; *draw ... on ...*; *draw ... for ...*; *... draw on ...*) (Tomasello 1992). Keelt omandav laps liidab verbi juurde sageli just nimisõna, kuna peale nimisõnade ta muid keelelisi abstraktsioone ei valda. Kui laps kasutab sama tegusõna uuesti, siis kordab selle kasutus eelmist, kuid veidi modifitseeritud kujul. Tomasello (2003: 121) järgi saab laste süntaktilist kompetentsust kõige paremini iseloomustada kui hulka üksteisest sõltumatuid verbisaarekesi isegi kuni vanuseni 2;6, paljudel lastel ka kuni vanuseni 3;0. Holger Diessel (2013: 10) on seepeale küsinud: “Kui lapse varased verbid on seotud kindlate konstruktsioonidega, siis kuidas õpivad lapsed verbe kasutama üle nende konstruktsioonide?” See on tähtis küsimus, sest kui laps oskab verbe kasutada väljaspool sõltumatuid konstruktsioone, on see teetähis, mis näitab veel üldisema verbi-argumendi struktuuri teket, mis väga suurel määral suurendab lapse suutlikkust kasutada süntaktilisi malle produktiivselt.

Ninio (2003: 19, 171) järgi pole ükski verb saareke, vaid lapse keelesüsteem on süsteem selle esimesest päevast peale. Tema sõnul ei tähenda lapse grammatika arenemine mitte verbisaarekeste ehk kindlate konstruktsioonide kogumist, vaid pigem arendab iga tegusõna oma minisüntaksi sõltumatuna teistest verbidest. See, kuidas lapse grammatika areneb, meenutab Ninio meelest linkimist veebi, kus üksused ühenduvad sarnasuse eri tahkude järgi. Tegusõnu õpivad lapsed nende sõnade semantilise ja süntaktilise valentsi järgi (Ninio 2011: 7), seega kombineeruvad lapse varased verbid kohe eri tüüpi konstruktsioonides ega moodusta nn saarekesi.

Ninio (2003) võttis Tomasello andmed ja leidis, et laps polnud 40% verbidest, mis ilmnesisid subjekti-verbi-objekti konstruktsioonis, enne esimest subjekti-verbi-objekti lauset üldsegi kasutanud. Peaaegu kolmandikul või neljandikul verbidel, mis moodustasid subjekti-verbi-objekti konstruktsioone, polnud olemas ei subjekti-verbi, verbi-objekti või ühe sõna kordamise (ingl *single-word antecedent*) kasutusi.

Selle näitas Ninio, et Tomasello kogutud andmed tema tütre Tralise näitel ei toeta verbisaare hüpoteesi. Tema arvates loovad lapsed varased laused nii, et omandavad kohe koos verbiga selle valentsi. Laste varased laused põhinevad süntaksi baassuhetel, nagu subjekt-verb, verb-objekt, subjekt-verb-objekt, nagu eesti keeles *emme loe, poiss kala, emme loe seda*.

Kummaski vaadeldud käsitluses ei ole konstruktsioonide ja argumentstruktuuri baassuhete kõrval vaadeldud teisi tegureid, mis lapse varast süntaksit mõjutada võivad. Samas on näiteks lapse varase süntaksi omandamisel vaja vaatluse alla võtta ka grammatikavälised nähtused nagu kordus, tagasituleetus, katkestus, põhja analoogia ja sulam (vt Dressler 2000, Dressler jt 2003: 393). Grammatikavälised nähtused keeles viitavad mehhanismidele, mida arvatakse olevat väljaspool grammatikat, nagu näiteks pragmaatika ja kognitiivsed võimed (Cannizaro 2012: 3).

3. Meetod ja materjal

Analüüsimaterjalina olen kasutanud kahe eesti keelt esimese keelena omandava lapse spontaanse kõne lindistusi. Martina kõnet on lindistatud 4 tundi 36 minutit, Andrease lindistusi on vaatluse all 5 tundi. Andrease spontaanne kõne on pärit CHILDES-i lastekeele andmepanga eesti lastekeele korpusest, korpuse on koostanud Andrease ema Maigi Vija. Martina kõne on lindistanud tema ema ja helifailid on litereerinud programmiga CLAN (ingl *Computerized Language Analysis*)¹ Airi Kapanen. Nii Andreas kui ka Martina kasvavad peres, kus on ema, isa ja vend, ning kõik nende pereliikmed on eestlased.

Tabelis 1 on esitatud ülevaade salvestiste kestuse ning lausungite (sh kahe- ja enamasõnaliste) hulga kohta. Martina kõnes on kahe- ja enamasõnalisi lausungeid rohkem kui Andrease kõnes (vastavalt 1362 ja 472). Kahesõnaliste lausungite osakaal Martina kõnes suureneb vaatlusperioodi jooksul 62,7%-lt 82%-ni ja Andreasel 16,1%-lt perioodi alguses 40%-ni kõikidest lausungitest perioodi lõpus.

Tabel 1. Analüüsitud keelematerjal

Andreas						Martina					
vanus	lausungeid	VKP	kahe- ja enamasõnaliste lausungite		salvestise kestus	vanus	lausungeid	VKP	kahe- ja enamasõnaliste lausungite		salvestise kestus
			hulk	osakaal					hulk	osakaal	
1;7,24	460	1,166	74	16,1%	1.00:00	1;3,15	423	2,308	265	62,6%	56:35
1;8,10	371	1,266	114	30,7%	1.00:00	1;5,18	445	1,956	325	73,0%	1.01:02
1;8,25	308	1,195	61	19,8%	1.00:00	1;6,26	238	2,245	170	71,4%	35:10
1;9,11	263	1,530	113	43,0%	1.00:00	1;9,19	337	1,933	290	86,1%	1.00:25
1;10,03	271	1,574	110	40,6%	1.00:00	1;10,23	378	3,388	312	82,5%	1.03:28
	1673		472	28,2%	5.00:00		1821		1362	74,8%	4.36:00

¹ <http://childes.psy.cmu.edu/clan/> (5.1.2014).

4. Lausemoodustuse omandamise perioodid Andrease kõnes

Andrease varase süntaksi võib eri vanuse kõnes domineerivate moodustusstrateegiate ja lausemallide järgi tinglikult jagada kolme perioodi. Esimest perioodi (vanus 1;7,24) iseloomustab korduste rohkus. Kordusena käsitan sõna kordamist lausungis, näiteks *tudu tudu tudu* [= magab], *puppu puppu puppu* [= nuppu], samuti kuulub korduse hulka ka sõnakuju mõningane varieerumine. Teisel perioodil (vanus 1;8,10 ja 1;8,25), loob Andreas grammatiliselt sidumata konstruktsioone, nagu *hopp hopp ova* [= orav], *leli* [= lõvi] *siia*. Võrrelduna esimese perioodiga moodustab ta lauseid, milles oskab ühele sõnale lisada teise. Kolmandal perioodil (vanus 1;9,11) kasutab laps sagedasti partiklit *ka*, näiteks *kaane ka*, *teise ka*, *või ka*, kusjuures mõnes lausungis kordab laps ka partiklit ennast: *saia ka ka*, *siil ka ka*. Sel perioodil lisanduvad kordustele ja *ka*-konstruktsioonidele omadussõnad: *ilus ka*, *ilus torn*, *pähh juust* ja mäarsõnad: *jälle uppi*, *jälle kopsti*.

Kui morfoloogia omandamine on jagatud pre- ja protomorfoloogiliseks perioodiks (vt Argus 2008), siis paistab siin vaadeldud materjalist, et ka varases lausemoodustuses võib tinglikult eristada presüntaktilist ja protosüntaktilist perioodi, mida Andrease puhul esindaksid vastavalt eeltoodud esimene ja teine periood. **Presüntaktilisel** perioodil süntaksi omandamist veel ei toimu, kuid laps kasutab grammatikavälist lausemoodustusprintsipi (ingl *extragrammatical operation*), nagu kordus (vt Dressler 2000). **Protosüntaktilist** perioodi iseloomustab selliste kombinatsioonide loomine, mis ei ole omavahel grammatiliselt seostatud, näiteks *tädi kuri*. Kui presüntaktilisel perioodil kordas laps üht ja sama sõna, siis protosüntaktilisel perioodil lisandub ühele sõnale teine, nimisõnale näiteks verb, kuid sõnad on süntaktiliselt veel sidumata, st verb ei ühildu ja nimisõna pole käänatud. Täiskasvanupärasest süntaksist veel rääkida ei saa, aga süntaktiliste protsesside eelkäijaid – leksikaalsete üksuste kombineerimist – võib ometi juba märgata.

4.1. Andreas vanuses 1;7,24 – presüntaktiline periood

Andrease kõne esimeses lindistuses (1;7,24) on 460 tema enda öeldud lausungit, neist kahe- ja enamasõnalisi on 74 (16,1%). Andrease väljendite keskmine pikkus (VKP)² selles vanuses on 1,166. Kuigi grammatika ei ole veel omandatud, oskab Andreas ühest sõnast pikemate üksuste moodustamiseks kasutada lihtsat grammatikaelset võtet, **kordust**. Kõrvuti kordustega on Andrease varases kõnes palju **sulameid**, nagu *eissaa* [= ei saa], *ettaa* [= ei taha], *kesse* [= kes see on]. Ühegi korduses öeldud sõna puhul ei ole Andreas kasutanud muutelõppe, st ühest sõnast pikemate lausungite sõnad on ainsuse nimetavas ja verbide puhul kasutab Andreas vaid sõna tüve (nt *uju*).

Vanuses 1;7,24 on 74 mitmesõnalise lausungi hulgas 55 kordust (77%), seega esinevad kordused tervelt kolmveerandis lausungitest. Andreas kordab nimisõnu (1), onomatopoeetilisi sõnu (2), verbe (3), arvsõnu (4), interjektsiooni (5).

(1) *kala kala, kuts kuts* [= koer koer]

(2) *kiss kiss* [= kass], *njau njau njau*

² Süntaktilise arengu mõõtmiseks kasutatakse lastekeele uurimisel väljendite keskmist pikkust (lühendatult VKP, ingl MLU ehk *mean length of utterance*), mida mõõdetakse sõnedes (Argus 2005: 9).

- (3) opa+opa [= hüppab], uju uju uju, kiika kiika [= kiikuda], kenne kenne
[= kõndida], paia+paia [= paitama]
(4) üks üks üks
(5) aitäh aitäh

4.2. Andreas vanuses 1;8,10 ja 1;8,25 – protosüntaktiline periood

Vanuses 1;8,10 on Andrease 371 lausungi hulgas 114 kahe- ja enamasõnalist (30,7%), mille hulgas on 88 kordust (77,2%). Seega on ka selle perioodi lausungitest suurem osa selliseid, kus laps kasutab kordust. Andrease VKP on sel perioodil pisut suurem kui eelmisel, 1,266.

Vanuses 1;8,10 on Andrease kõnes 13 mitmesõnalist lausungit: *hopp hopp ova* [= orav], *siia siia maha maha maha*, *aitäh uku uku* [= nuku], *leli* [= lõvi] *siia*, *emme aska!* [= raske], *Atsu Atsu Atsu Atsu aska* [= raske], *issi äh äh issika*, *tuhh* [= tsuhh] *tuhh* [= tsuhh] *emme*, *Antsu käru*, *kiisu üks kaks*, *emme käsi*, *torni Atsu* [= Andreas] *Atsu* [= Andreas], *hallo hallo hallo ole* [= ei ole]. Võrreldes eelmise perioodi lausungitega on suur erinevus selles, millest kordustega lausungid koosnevad – nimelt ei korda laps nüüd mitte lihtsalt üht tüve, vaid kombineerib ka eri lekseeme. Kõigis neis lausungites on olemas kas tegija või objekt, üks elus või elutu olevus või ese, nt *emme*, *issi*, *Atsu*, *hallo* ('telefon'), ja teine sõna, millega selle kohta midagi öeldakse, miski, mis ilmselt on verbi asemel ja mis meenutab predikaati, nt *aska* ('raske') või *siia*.

Melissa Bowermani väitel moodustab laps paljud oma varased lausungid kahest nimisõnast (Bowerman 1973: 201). Samasugust strateegiat kasutab paljudel juhtudel ka Andreas. Nii sel viisil moodustatud lausungitele kui ka üksikule nimisõnale oskab Andreas aga liita omadussõna (6a), (6b), samuti moodustab Andreas lausungi, milles üksikule nimisõnale liidab kohasõna (7a)–(7c).

- (6a) auh auh **pai**
(6b) tädi **kuri**
(7a) emme **siia**
(7b) otsu [= notsu, merisiga] **sinna**
(7c) leli [= lõvi] **siia**

Enamik Andrease lausungeid pole grammatiliselt vormistatud ka järgmises lindsütuses ehk kuu aega hiljem. Kordusi on endiselt üsna palju, nende hulgas on kõige rohkem kahe- ja kolmesõnalisi (8), kuid on ka kolmesõnalisi kordusi (9).

- (8) sisse sisse, peitu peitu, kommi kommi, uni uni
(9) silm silm silm, kutsu kutsu kutsu, hallo hallo hallo, kata kata kata [= karda]

Kõrvuti kordustega on materjalis nimisõna ja tegusõna kombinatsioone (10a), (10b), mis annavad tunnistust protosüntaktilise perioodi algusest: kombinatsioon moodustatakse kahest grammatiliselt sidumata sõnast; vormimoodustusoskust, mis on vajalik grammatikavahendite sidumiseks, veel ei ole.

- (10a) mina mina iska [= viska]
(10b) otsu [= notsu, merisiga] paista [= paistab]

Vanuses 1;8,25 ilmub Andrease kõnesse esimene verbi ja noomeni konstruktsioon *hallo hallo hallo ole* [= ei ole] ning esimene verbi ja noomeni konstruktsioon, kus verbil on muutetunnus (*autu tuli*), samuti on Andrease kõnes üks verbi laiendiga konstruktsioon *issi tuli kop-kop*, milles ta on väljendanud ootust, et isa koju tuleks. Onomatopoeetilisele sõnale oskab Andreas lisada tegija (11a), (11b). Need lausungid annavad märku, et laps hakkab omandama esimest nn baassuhet ehk kasutab subjekti ja verbi (SV) kombinatsiooni (11b) või siis objekti ja subjekti (OS) kombinatsiooni (11a).

(11a) LAPS: tuhh tuhh [= tšuhh tšuhh] **emme**

EMA: emme peab laulma või ?

(11b) hopp hopp **ova** [= orav]

Ühel korral on ta korranud lausungi teist poolt (12), tegu võib olla soovi intensiivistamisega.

(12) LAPS: aitäh **uku** [= nuku] **uku** [= nuku]

EMA: nuku. [imitatiiv]

Vanuses 1;8,25 on 308 lausungi hulgas kahe- ja enamasõnalisi 61 (19,8%), mille hulgas on 35 kordust (59,3%). Võrrelduna perioodi algusega on kordusi 17,9 protsendi võrra vähem (77,2%-lt perioodi alguses 59,3%-le perioodi lõpus). VKP on 1,195. Sulamite (*kessee, ettaha*) hulk on palju väiksem kui vanuses 1;7,24 – neid esineb vaid paar korda.

Vaadeldud vanuses (1;8,10 ja 1;8,25) iseloomustab Andrease kõnet korduste üisna suur hulk, kuid nüüd oskab Andreas ühele sõnale lisada nimisõna (13), verbi (14) või kohasõna (15).

(13) emme ova [= orav]

(14) iska [= viska], paista

(15) siia, sinna.

Kahe- ja enamasõnaliste lausungite moodustamisel on ülekaalus kordused, kuid Andrease kõnes on tavapärase ühe ja sama sõna korduste hulgas selliseidki, mis koosnevad kahest eri sõnast, kusjuures Andreas on korranud kas mõlemat sõna (16) või neist üht (17).

(16) siia siia maha maha maha

(17) mina mina iska [= viska]

4.3. Andreas vanuses 1;9,11 – nn *ka*-periood

Vanuses 1;9,11 on Andrease öeldud 263 lausungi hulgas 113 kahe- ja enamasõnalist (43%). Korduste hulk on vähenenud: 113 kahe- ja enamasõnalise lausungi hulgas leidub vaid 15 kordust (12,3%). Uus joon Andrease kõnes on partikli *ka* rohke kasutamine (18): kahe- ja enamasõnaliste lausungite hulgas on 33 *ka*-partikliga lausungit (29,5%): *teise ka, emme ka, kaane ka, kala ka, või ka, keemi ka, seda ka*.

Partiklit *ka* võib Andrease puhul pidada verbi eelkäijaks.³ Andreas moodustab tugiverbiühendi, milles sõna *ka* asendab väljendamata jäänud tegusõna. Sõnad, mida Andreas *ka*-partikliga koos kasutab, on käändsõnad nimetavas, omastavas või osastavas käändes. Näiteks ei ütle Andreas *paneme teise*, vaid *teise ka* ning *emme paneb kreemi* asemel ütleb *keemi ka*, nimetades tugiverbiühendist vaid objekti, millele tegevus on suunatud. Osal juhtudel kordab Andreas partiklit *ka* ennast (18).

(18) teise **ka ka**, siil **ka ka**, saia **ka ka**

Vanuses 1;9,11 ilmub Andrease kõnesse lausung, milles on kolm eri sõna:

(19) Atsu [= Andreas] tuli [= diktofoni lambike] kätte

4.4. Andreas vanuses 1;10,03 – korduste hulga vähenemine, süntaksi baassuhete kujunemine

Lausungeid on selles vanuses lindistatud dialoogis kokku 271, neist mitmesõnalisi on 110. Lausungeid, milles Andreas kasutab *ka*-partiklit, on märgatavalt vähem kui eelmises vanuses – vaid 6,4% (eelmises vaatlusvanuses 29,5%).

Kahe- ja enamasõnaliste lausungite hulgas on 19 kordust (16,5%), seega on korduste hulk võrrelduna vanusega 1;8,10 vähenenud 60,7% võrra (vanuses 1;8,10 on kordusi Andrease kõnes 77,2%, vt osa 4.2). Kõnesse on ilmunud pikemad lausungid, nt (20)–(24).

(20) emme loe seda

(21) issi loe seda

(22) ettaha sõrmeküüsi

(23) ettaha [= ei taha] sasti [= sahtlisse] ära

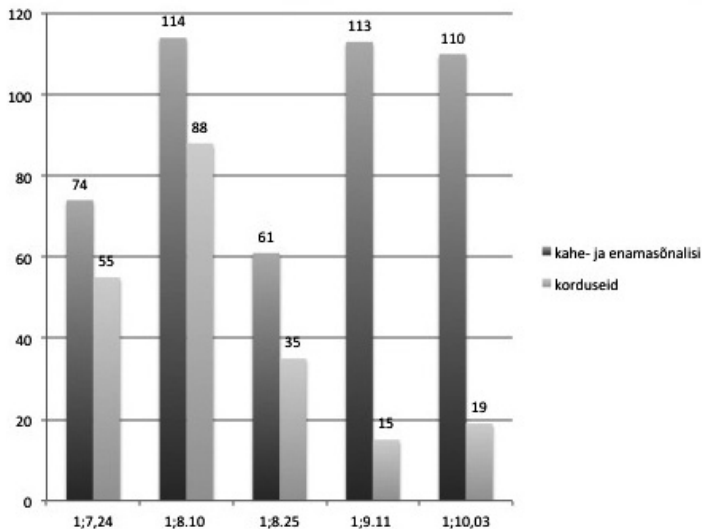
(24) emme tudu Atsu tudu

Vanuse 1;10,03 lõpus on esimesed selged märgid sellest, et laps moodustab süntaktilistel suhetel põhinevaid VO- (näited 20–23) ja SV-konstruksioone (24).

4.5. Kokkuvõte Andrease varastest lausungitest

Vanuses 1;7,24 kasutab Andreas grammatikavälist moodustuspõhimõtet kordust, et luua pikemaid üksusi (skemaatiliselt võiks neid tähistada näiteks AA, BBB). Enamasti moodustab Andreas lausungeid, milles on kaks või kolm samakujulist sõna. Lisaks kordustele on Andrease kõnes palju sulandunud sõnu. Võrdluseks: vanuses 1;8,10 on kordusi Andrease kõnes 88 (77,2%), vanuses 1;8,25 on 35 kordust (59,3%), vanuses 1;9,11 oli 113 lausungist 15 (= 12%) ja vanuses 1;10,03 oli 110 lausungist 19 (17%). Osakaal on arvatatud kahe- ja enamasõnaliste lausungite hulgast. Korduste osakaal lapse kõnes väheneb, kui laps saab vanemaks, seejuures näiteks paari kuuga üsna kiiresti: 77,2%-lt 17%-le.

³ Lapse kõnes võivad verbi funktsioonis esineda sõnad, mis kuuluvad mõnda teise sõnaklassi, eriti juhul, kui lapse sõnavara koosneb noomenitest (Argus 2004: 41).



Joonis 1. Korduste arv kahe- ja enamasõnaliste lausungite hulgas Andrease kõnes eri vanusevahemikes

Vanuses 1;8,10 korduste hulk suureneb (vt joonis 1), kuid nüüd oskab Andreas liita esimes(t)ele sõna(de)le teise lekseemi, näiteks *auh auh pai*, *tuhh tuhh emme* (AAB); *aitäh uku uku* (ABB). Perioodi lõpupoole, vanuses 1;8,25, hakkab korduste hulk vähenema – 77%-lt langeb korduste osakaal 59%-ni – ning Andrease kõnesse tekib esimene verbi-nimisõna konstruktsioon ja nimisõna-verbi-omadussõna konstruktsioon.

Nn *ka*-perioodil (1;9,11) kasutab Andreas palju partiklit *ka*, õppides sel viisil looma sihilist konstruktsiooni, näiteks *keemi ka* tähendab ‘paneme kreemi ka’. Vanuses 1;10,03 on korduste hulk võrrelduna varasemate vaatlusvahemikega märkimisväärselt vähenenud ja on hakanud kujunema süntaktilised baassuhted: verb-objekt ja subjekt-verb.

Lause moodustamist harjutab Andreas justkui tükkide kaupa: ta kasutab kordusi ja vanuse suurenedes oskab kordusele lisada uue elemendi. Prototüüpseks peetav lause, milles on alus, öeldis ja sihitis (nt *Poiss püüab kala*), on vaatlusaluses materjalisesialgu realiseerunud kahe eraldi lausungina, nt *kala püia* [= püüa] ja *poiss kala*.

Andrease kõnes esinevad verbid enamasti SV-konstruktsioonis: *issi tuli*, *notsu kadu*, *tuhti paistab*, *otsi kulli*, *kala püia*, *palli osti*, *tädi aata* [= vaata], *tooli istu*, *tita rooli*. On ka üksikuid VO-konstruktsioone: *otsa sai vann* [= vann], *osta keemi*, *ei taha juustu*.) Esimesed verbi valentsi põhised konstruktsioonid tekivad Andrease kõnesse seega vanuses 1;8,25, rohkem on neid alates vanusest 1;9,11. Andrease varastest lausungitest on näha, et ta kasutab eri verbe eri tüüpi kombinatsioonides, kindlaid ühe verbi ümber koonduvaid konstruktsioone märgata ei ole.

5. Lausemoodustuse omandamise perioodid Martina kõnes

Esimesele vaadeldud perioodile on iseloomulikud kordused. Kõrvuti tavapärase kordusega, kus Martina kordab sõna selle vormi muutmata, esineb tema kõnes sellinegi kordus, milles ta kordab üht ja sama sõna väikeste erinevustega. Sõna varieerub tavaliselt kolm korda, nt *märgi väigi väägi* [= värvi värvi värvi]. Osas lausungeist kasutab Martina peegelmalli (vt 5.1).

Teisel perioodil jätkub korduse kasutamine ja lisandub rohkelt määrsõnu: *kinni, lahti, veel, koos, siit, seal, juurde, bäl* [= veel]. Kolmas on partiklite periood, kus kõnesse ilmub suurel hulgal partikleid, nagu *ka, jälle, küll, ära*. Neljandat vaadeldud ajavahemikku Martina kõnes võib nimetada pikkade lausungite perioodiks, samasse vahemikku jääb ka sihilise konstruktsiooni moodustamine.

5.1. Martina vanuses 1;3,15 – presüntaktiline periood

Martina 423 lausungi hulgas on 265 kahe- ja enamasõnalist lausungit (62,6%), mille hulgas peaaegu pooled on kordused (127 lausungit; 48,5%). VKP vanuses 1;3,15 on 2,308.

Martina on korranud tegusõnu (25a)–(25e) ja nimisõnu (26a)–(26e).

- (25a) eita eita ei taha
- (25b) ängi ängi [= mängi mängi]
- (25c) kaia kaia [= käima käima]
- (25d) lau lau [= laulma laulma]
- (25e) aata daata [= vaata vaata]

- (26a) mai mai [= mari mari]
- (26b) võthi võthi [= mütsi mütsi]
- (26c) toma toma [= tomat tomat]
- (26d) kiidu kiisu [= kiisu kiisu]
- (26e) lill lill lill

Näidetes (27a)–(27e) ütleb Martina ühes lausungis ühe sõna kolm kuhu – tema lausungid on justkui sõna õige fonoloogilise kuhu otsimised.

- (27a) emme **märgi väigi väägi** [= emme värvi värvi värvi]
- (27b) **indi ninnis inni** [= lindistab lindistab lindistab]
- (27c) **kuutsi mõmmi paatsi uutsi** [= pluusi mõmmi pluusi pluusi]
- (27d) **ou uu õu** [= õun õun õun]
- (27e) **koos koss** [= koos koos]

Sama võtet kasutab Martina ka pikemate üksuste puhul, korrates kahesõnalist fraasi:

- (28) emme au emme auk emme auki [= emme arvuti emme arvuti emme arvuti]
- (29) kaks onu kat onu kak [= kaks onu kaks onu kaks]
- (30) ise kati ise kati ise [= ise katki ise katki ise]
- (31) te tee iia te tee te tee [= see see siia see see see see]

Martina kõnes on märgata rütmilisust (32)–(34). Osas lausungitest, nt (32)–(34), on Martina kasutanud omalaadset malli, mida võib nimetada **peegelmalliks**: ta ütleb lausungi esimese poole, lisab uue sõna ja kordab pärast seda lausungi esimest osa.

(32) bipa papa **õue õue** bipa papa [= tipa tapa õue õue tipa tapa]

(33) bipa bipa **tita** bipa papa [= tipa tapa tita tipa tapa]

(34) te te **katki katki** tee kati [= see see katki katki see katki]

Samal perioodil on Marina kõnes kahesõnalisi lausungeid, nagu *emme luue* [= loe], *emme buue* [= loe], *emme atta* [= aita], nimisõna ja tegusõnaga lausungeid *mõmmi teep* [= teeb], *eeku aata* [= seelikut ei taha], *oo pall* [= oo pall] *lendab* ning määr sõnaga lausungeid, nagu *ööbi* [= nõöbi] *kinni*, *kuku* [= luku] *lahti*.

Martina on korranud lausungites tegusõna (35) ja nimisõna (36a), (36b).

(35) diku **vaata vaata** [= diktofon vaatab vaatab]

(36a) emme **mah mah** [= emme mahl mahl]

(36b) **poiss poiss** käes [= poiss poiss käes]

Martinal on esimeses lindistuses, vanuses 1;3,15, SV-konstruktsioon *mõmmi teeb* ja juba järgmises (vanuses 1;5,18) moodustab ta SVO-konstruktsioonid *issi annab tomatit* ja *emme sööb kohvi*. Lausungite hulgas on ka VO-konstruktsioone, nagu *laul kikuu* [= laulda diktofonile], *äita* [= näita] *mängu* ja SO-konstruktsioone, näiteks *diku* [= diktofoni] *aada* [= vaatan]. Martina selle vanuse lausungites on nimisõna enamasti eespool, järgneb verb.

5.2. Martina vanuses 1;5,18 – protosüntaktiline periood

Vanuses 1;5,18 on Martina kõnes kokku 445 lausungit, mille hulgas kahe- ja enamasõnalisi on 325 (71%). Viimaste seas on 29 kordust (9,1%). Korduste hulk on võrrelduna eelmise vaatlusvanusega (neid oli 48,5%) suurel määral – 39,4 % võrra – vähenenud. Ka-partikliga lausungeid on Martina kõnes kahe- ja enamasõnaliste lausungite hulgast 14 (3,15%). Tema VKP vanuses 1;5,18 on 1,956 (vt tabel 1).

Andreasel algas protosüntaktiline periood (vt osa 4.2) vanuses 1;8,10, mil tema VKP oli 1,266. Korduste osakaal Martina ja Andree kõnes protosüntaktilisel perioodil erineb: Andree kõnes on sel perioodil kordusi 77,2% ja Martinal 9,1% kahe- ja enamasõnaliste lausungite hulgast, mille osakaal kõigist lausungitest on vastavalt 30,7% ja 71%. Kui Andreas kordab vaid üksiksõnu, siis Martina kõnes esineb lisaks seda tüüpi kordusele ka pikemate fraaside kordamist. Lisaks selgub lindistustest, et Martina kõne protosüntaktilise perioodi kõne erineb Andree protosüntaktilise perioodi kõnest selle poolest, et Martina kasutab verbe märkimisväärselt rohkem.

Martina jätkab nii üksiksõnade (37a)–(37c) kui ka fraaside (41a)–(41c) kordamist.

(37a) teda teda

(37b) põõma põõmah [= pall]

(37c) tasa tasa

(38a) emme istud emme istud

(38b) tomati kaa tomati kaa

(38c) etaha linnu [= Intsu] tomat eetahha innu [= Intsu] tommat iuu eetaha
Innu [= Intsu] tomat

Vanuse 1;15,18 kõnes on võimalik täheldada määrsonade suhtelist rohkust: *kinni, lahti, veel, koos, siit, seal, juurde, bäl* [= veel]. Peale korduste on kõnes lausungeid, nagu *emme koos õue, issi koos õue, issi anna veel tomatit, Makkus külla ilähe* [= ei lähe].

Süntaktilistest baassuhetest esineb Martina selle vanuse kõnes VO-konstruktioone *etaha* [= ei taha] *Intsu tomat*, SO-konstruktiooni *äku pokandit* [= jänku (tahab) porgandit], SV-konstruktioone *dipuki* [= diktofon] *käib*, kuid on ka SVO-konstruktioone, nagu *mina taha uid* [= võid], *issi anna veel tomatit*.

5.3. Martina vanuses 1;6,26 – pikkade lausungite kordamise periood

Vanuses 1;6,26 on Martina 238 lausungi hulgas 170 kahe- ja enamasõnalist lausungit (71,4%), Martina VKP selles vanuses on 2,245 (vt tabel 1).

Kõrvuti sõnakordusega, nagu *ukkimannaa ukkimanna* [= rukkimanna rukkimanna], *on on, nämm nämm nämm nämm, ei saa ei saa*, on Martina kõnes vanuses 1;6,26 üsna palju selliseid lausungeid, milles ta on korranud pikemat üksust:

- (39a) emme minu minu vitamiinid on üleval emme minu vitamiinid on üleval
- (39b) suurt võid ei ole siin a suurt võid ei ole siin
- (39c) ele vandidoo elevant joo elevand joo
- (39d) ja mapanin nii palju võid panin nii palju vuiid
- (39e) ei taha emaa pese selle nugaa ua pesee ära pese nugaa
- (39f) nüdon puhas emme nüdongi puhas emme nüd on emme nüdon kõik

5.4. Martina vanuses 1;9,19 – partiklite periood

Vanuses 1;9,19 on 337 lausungi hulgas 290 kahe- ja enamasõnalist lausungit (86,1%), VKP selles vaatlusvanuses on 1,933 (vt tabel 1). Martina lausungitesse ilmub suurel hulgal partikleid, ta kasutab partikleid *ka, jälle, küll, ära*, samas kui Andreasel oli ühe partikli, *ka* periood.

Lausungite üldhulgast on Martina kasutanud partiklit *ka* 23 lausungis (8,1%), partiklit *ära* 11 lausungis (3,9%) ja partiklit *küll* 45 lausungis (15,8%). Partikkel *ära* väljendab nii perfektiivsust (*lõhkus pesa ära, lahendas ülesande ära*) kui ka suunda (*sõitis ära*) (EKG II: 21). Martina kõnes väljendab partikkel *ära* perfektiivsust (*tudusin ära, emme pesid ära juba nüüd*) ja keeldu (*äla pane ise või*), mõnel juhul on Martina kasutanud sõna *ära* pöörduva eitussõnana (40d).

- (40a) ma tudusin **äraa** jubaa
- (40b) ää **äla** pane seda patsi kõvasti
- (40c) **äla** pane ise või
- (40d) emme **ära** anna see emme ära anna seda isee
- (40e) emme pühki **ära**
- (40f) tee [= see] ka **ära**

(40g) emme va mängi seda mõmmi nuusit [= nõusid] emme **ära** pese seda nõud

(40h) emme pesid **ära** uba nüüd

Partikli *ka* kasutuse puhul on märgata, et osa juhtudel kasutab Martina *ka*-partikliga samas fraasis veel verbi. Siin ilmneb erinevus Andrease ja Martina kõne vahel, Andreas lisas *ka*-partikli nimisõnale, nt *emme ka*, *keemi* [= kreemi] *ka*, *kaane ka*, Martina aga verbile:

(41a) emmee **laseb ka** endale vett

(41b) kätele **tehakse ka** massa massa

(41c) nii kas sa **tahad ka** seda kohvii?

(41d) emme **laseb ka** endale vett

Vaid üksikutes lausungites on Martina seostanud *ka*-partikli nimisõnaga:

(42) mis ma tegin teise või **ka**

(43) mul **ka** vesi

(44) kohvi **ka** kohvitass

(45) elevantid olid oma puu elevantid olid **ka** puuris olid küll

Partikkel *küll* ei saa esineda omaette vooruna, ta on rõhutav, vastust kinnitav sõna (EKSS). Sellises funktsioonis esineb partikkel *küll* Martinagi kõnes:

(46a) EMA: aa sinu võti on seal jaa

LAPS: on **küll**

(46b) EMA: vaata kui siin tu punane tuli põleb, siis ta käib, emme pani käima

LAPS: emme pani käima

EMA: mhmh

LAPS: pani **küll**

(46c) EMA: juba läks vitamiin suhuu

LAPS: läks **küll**

(46d) LAPS: kus on Saalema [= Saaremaa] või?

EMA: kapi peal

LAPS: emme või on seal seal on Saame [= Saaremaa]

LAPS: emme seon võinuga on **küll**

(46e) EMA: kas sa palun on sul mahti nüüd ööö see Martiina turvatool ümber tõsta minu autosse?

ISA: mhmh

LAPS: issi emme mme autosse tõsta sedaa [= seda] saab **küll**

Partikli *jälle* tähendus toob esile sündmuse kordumise, st olemasolevale hulgale lisandub midagi juurde. *jälle* tähendus kordumisele viitavas funktsioonis on seotud ajalise tähendusega, kuid väljendab ka lisamise aspekti (Parm 2007: 29). Martina kõnes esineb partikkel *jälle* eelviidatud funktsioonides:

(47a) tuleb **jälle**

(47b) seon [= see on] kiisu pluus **jälle**

(47c) emme teeb **jälle** endale kohvi

Süntaktilistest baassuhetest esinevad sel perioodil Martina kõnes peamiselt SV- ja VO-konstruktsioon (*idaa [= ei taha] ukkimannaputtu, võinoa ka võtta kätte, tahan ise võileiba panna*), ent on ka SVO-konstruktsioone (*mina oian koa [= hoian ka] seda, emme mina jõin täna niipalju vett, mina söön seda*).

5.5. Martina 1;10,23 – pikkade lausungite moodustamise periood

Lausungeid on Martina kõnes vanuses 1;10,23 kokku 378, millest enamiku moodustavad kahe- ja enamasõnalised (312, mis on 82,5% lausungite koguhulgast). Võrdluseks meenutame, et vanuses 1;3,15 oli kahe- ja enamasõnalisi lausungeid 62,6%). VKP on Martinal selles vanuses 3,388 (vanuses 1;3,15 oli VKP 2,308).

Kõrvuti sellega, et säilinud on kõik varasema vanuse kõnele iseloomulikud jooned – sõnade (48) ja pikemate fraaside kordamine (49a), (49b) ning partiklite *ka, ära, küll, jälle* kasutamine –, oskab Martina moodustada üsna pikki lausungeid, milles ei ole pause (50)–(52). Need on kuuldavale toodud ilma hingetõmbepausita, mistõttu ei ole võimalik lausungipiire tuvastada.

- (48) pese pese pese
- (49a) ma tulen siia emme kuvvale tulen siia emme kõrvale tasakaalu
- (49b) laa laa laa ma võtan väikse pilli võtan väikse pilli võtan väikse pilli
- (50) veel kalkan Martiina ära nüüd ma panen karbi kinni emme nüüd ma panen Hello Kiti kaabi karbi kinni teist karp kinni siis siis iise ja kuasti jäi läinud vaata see läks juba kinni kas kiisu pusle läks kinni?
- (51) jälle uhupalli kiisu jälle õhupall see käib siia sinna paikkah emme kus see käi las see olla siin las see iiah kuall nüüd paigal teise asja veel siia see sai valmis
- (52) teeb kalp kalp kalp kalp nüüd on Lote sööööök see on Lote söök ma söön Lote söögi ära

Ka vanuses 1;10,23 moodustab Martina sihilisi konstruktsioone:

- (53a) pani pudu [= pudru] potti **vett**
- (53b) ma kisu puslet **hoida**
- (53c) ma võtan hoopis **tudulapi** eest

Selle perioodi lindistuses on olemas SV-baassuhe (*issi paneb*), VO-suhe ja sageli ka SVO-suhe (*Martina võtab selle*).

5.6. Kokkuvõte Martina varastest lausungitest

Martina kasutab ühest sõnast pikemate lausungite moodustamiseks kordust ja hakkab üsna ruttu moodustama pikemaid üksuseid, mida samuti kordab. Näib, et Martina omandab süntaktilise rütmi, mille täidab igas vanuses parajasti tema käsutuses olevate leksikaalsete vahenditega. Osas lausungites kordab Martina küll üht ja sama sõna, ent varieerib sageli selle kuju, näiteks *märgi väigi väägi [= värvi värvi värvi]*, *emme au [= arvuti] emme auk [= arvuti] emme auki [= arvuti]*. Samuti

kasutab Martina kordamisel n-ö peegelmalli, kus ta ütleb lausungi esimese poole, lisab uue sõna ja kordab pärast seda lausungi esimest osa.

Martina kõnes paistab silma partiklite nagu *ka, küll, jälle, ära* rohke kasutamine, neist eriti sagedasti kasutab Martina partiklit *küll*, mis rõhutab, kinnitab vastust. Viieast vaadeldud perioodist neljanda lõpul oskab Martina luua sihelist konstruktsiooni. Tema pikimas lausungis on 38 sõna.

Nagu Andree, nii ka Martina kõnes ei leidunud ühe kindla verbi ümber koonduvad konstruktsioonid, kõik verbid kombineerusid eri hulgal muude sõnadega. Vaadeldud ajavahemiku lõpuks olid olemas VO-, SV- ja VO-konstruktsioon ning ka SVO-konstruktsioon.

6. Kokkuvõte

Eesti keelt omandavate laste varast süntaksit iseloomustab siinse uurimuse materjali põhjal kõigepealt kordus, mis on esialgu lausete moodustamise põhimoodus: laps kasutab neid leksikaalseid vahendeid, mis tal kindlas vanuses olemas on, ja kordab üht sõna mitu korda. Kui laps saab vanemaks, siis korduste hulk tema kõnes väheneb, seega on kordusel kindel roll just süntaksi omandamise varases järgus. Selgub, et rütm on peale sõnade omandamise tähtis ka varaste mitmesõnaliste lausungite puhul – laps kordab fraase, mis koosnevad kahest ja enamast sõnast, sageli justkui kindla rütmi järgi, nt kordub esimesena öeldud sõna või sõnapaar pärast uut elementi uuesti. Varasele keeleomandamisele on järgmisel etapil iseloomulik partiklite *ka, küll, ära, jälle* rohke kasutamine. Partikkel *ka* näib asendavat mõnel juhul tugiverbikonstruktsiooni verbi (vt Muischnek 2004), mida keelt alles omandav laps ei ole õppinud veel ütlema. Kui Andreas kasutas põhiliselt partiklit *ka*, siis Martina kõnes on neid vaadeldud perioodil suurem valik. Erinevus kahe lapse varase lausemoodustuse arengus puudutab veel ka omajasuhete arengut. Nimelt leidub Martina kõnes omastavakujulise täiendiga nimisõna fraase juba vanusest 1;3,15. VKP väärtuse kohta on võimalik tähele panna, et Andreasel algas süntaksi baassuhete kujunemine, kui VKP oli 1,54, Martinal, kui VKP oli 2,3.

Andree kõnes (vanuses 1;8,25) võib täheldada üksikuid lausungeid, kus üht ja sama verbi on kasutatud ühesuguses konstruktsioonis (*buss tuli, auto tuli*), kuid sedalaadi n-ö saarekesed on ainsad. Mõlema lapse kõnes esinevad üksikud tegusõnad, millega paaril korral on moodustatud ühesugune konstruktsioon, ei toeta Tomasello väidet, et varane keel on k o r r a p ä r a s e l t koondunud üksikute verbide ümber.

Süntaksi baassuhete omandamine ehk selliste lausungite moodustamine, milles on subjekt-verb, verb-objekt, algab Andreasel vanuses 1;10,03. Martinal on sellised lausungid olemas kõigis vaatlusvanustes alates salvestiste algusest (*mõmmi teeb*, 1;3,15).

Seega sobib Andree ja Martina keelematerjali põhjal eesti keele süntaksi omandamise kirjeldamiseks verbisaare teooria asemel paremini Ninio vaatenurk, mille kohaselt põhinevad lapse varased laused pigem süntaksi baassuhetel, nagu subjekt-verb, verb-objekt, subjekt-verb-objekt. Eriti hästi on seda võimalik vaadelda Martina kõnematerjali põhjal, milles salvestiste algusest peale on võimalik näha selliste lausungite moodustamist, milles on subjekt ja verb (*mõmmi teeb*), subjekt, verb

ja objekt (*issi annab tomatit, emme sööb kohvi*) või subjekt ja objekt (*issi tomatit*), kuid ka Andrease kõnes on võimalik jälgida samasugust moodustusprintsipi (*poiss püüab, poiss kala*), mitte kindlaid tervikuna omandatud korduvaid konstruktsioone.

Kokkuvõtlikult näitab eesti laste varaste lausungite esmane analüüs, et varane lausemoodustus algab grammatikaväliste elementide, ennekõike korduste kasutamisest ja jätkub süntaksi baassuhete tekkimisega. Kui kordustega on saadud selgeks pikema lausungi fonoloogiline kuju või mahupiir, saab laps hakata seda täitma verbi argumentstruktuuril põhinevate tähendus- ja vormisuhetega.

Lühendid

O	objekt
S	subjekt
V	verb
VKP	väljendi keskmine pikkus

Viidatud kirjandus

- Argus, Reili 1994. Lause areng ühe lapse keeles vanuseni 1,6–4,0. Lõputöö. Käsikiri nr 1225 Tartu Ülikooli eesti keele osakonnas. Tartu: Tartu Ülikool.
- Argus, Reili 2004. Verbi eelkäijatest esimeste kolmeliikmeliste miniparadigmadeni eesti ja soome keele võrdlusel. – Helle Metslang (koost.), Maria-Maren Sepper, Jane Lepasaar (toim.). Toimiv keel II. Töid rakenduslingvistika alalt. Tallinna Pedagoogikaülikooli eesti filoloogia osakonna toimetised 3. Tallinna Pedagoogikaülikool, Eesti Keele Instituut. Tallinn: TPÜ kirjastus, 37–52.
- Argus, Reili 2008. Eesti keele muutemorfoloogia omandamine. [Acquisition of Morphology in Estonian.] Tallinna Ülikooli humanitaarteaduste dissertatsioonid 19. Tallinn: TLÜ Kirjastus.
- Cannizzaro, Courtney Leigh 2012. Early word order and animacy. <http://www.let.rug.nl/~hendriks/papers/cannizzaro.pdf> (9.1.2014).
- Diessel, Holger 2013. Construction grammar and first language acquisition. – Graeme Trousdale, Thomas Hoffmann (Eds.). The Oxford Handbook of Construction Grammar. Oxford: Oxford University Press, 347–364.
- Dressler, Wolfgang U. 2000. Extragrammatical vs. marginal morphology. – Ursula Doleschal, Anna M. Thornton (Eds.). Extragrammatical and Marginal Morphology. Munich: Lincom Europa, 1–10.
- Dressler, Wolfgang U.; Kilani-Schoch, Marianne; Klampfer, Sabine 2003. How does a child detect morphology. – R. Harald Baayen, Robert Schreuder (Eds.). Morphological Structure in Language Processing. Trends in Linguistics, studies and monographs 151. Mouton de Gruyter, 391–427.
- EKSS = Eesti kirjakeele seletussõnaraamat I–VII. 1988–2007. [The Explanatory Dictionary of Estonian.] Eesti Keele Instituut. Tallinn: Eesti Keele Sihtasutus.
- Lindsalu, Elo 1995. Liitlaused ja nende sidumine eesti lastekeeles. Magistritöö. Tallinna Pedagoogikaülikooli (al. 2005 Tallinna Ülikool) eesti filoloogia osakond.
- Metslang, Helle 2001. On the development of the Estonian aspect: the verbal particle *ära*. – Ö. Dahl, M. Koptjevskaja-Tamm (Eds.). The Circum-Baltic Languages. Typology and Contact. Vol. 2. Grammar and Typology. Amsterdam–Philadelphia: John Benjamins, 443–480.
- Muischnek, Kadri 2004. Eesti keele tugiverbid ja nende mõju lause ehitusele. [Estonian support verbs and their impact on the syntactic structure.] – Emakeele Seltsi aasta-raamat, 50, 11–38.

- Ninio, Anat 2003. No verb is an island: Negative evidence on the verb island hypothesis. – Psychology of Language and Communication, 7 (1), 1–21. http://www.plc.psychologia.pl/plc/plc/contents/fulltext/07-1_1.pdf (23.2.2014).
- Ninio, Anat 2011. Syntactic Development, its Input and Output. Oxford: Oxford University Press. <http://dx.doi.org/10.1093/acprof:oso/9780199565962.001.0001>
- Parm, Sirli 2007. Partiklite *veel, juba, alles, jälle* tähendused eesti kirjakeeles. Magistritöö. Tartu: Tartu Ülikool. <http://dspace.utlib.ee/dspace/handle/10062/2932> (5.3.2014).
- Parmas, Viktoria 2000. Ühe eesti lapse küsilause areng (Ingel-Brit 1;6–2;2). Magistritöö. Tallinn: Tallinna Ülikool.
- Tomasello, Michael 1992. First Verbs. Case Study of Early Grammatical Development. Cambridge University Press. <http://dx.doi.org/10.1017/CBO9780511527678>
- Tomasello, Michael 2003. The New Psychology of Language. Cognitive and Functional Approaches to Language Structure, Vol. 2. Mahwah, NJ/London: Lawrence Erlbaum.

Võrgumaterjalid

- CHILDES – Child Language Data Exchange System. <http://childes.psy.cmu.edu/> (24.2.2014).
Using CLAN. <http://childes.psy.cmu.edu/clan/> (24.2.2014).

Airi Kapanen (Tallinna ülikool) uurimisvaldkond on esimese keele süntaksi omandamine eesti keelt emakeelena kõnelevates peredes.
Õismäe tee 140-64, 13511 Tallinn, Estonia
airikapanen@gmail.com

INSIGHT INTO THE EARLY ACQUISITION OF SYNTAX IN ESTONIAN

Airi Kapanen

Tallinn University

This article examines syntax acquisition in Estonian on the basis of the spontaneous speech data of two children acquiring Estonian as their native language. The early utterances and the acquisition of syntax by children can be divided into three stages according to the strategy that a child uses to produce multi-word utterances: the use of antecedents, the particle period and the formation of the basic verbal argument structure constructions.

The data indicates that Estonian children often use an extragrammatical operation, antecedents, while they iterate two- and multiword utterances. As children get older, the number of antecedents is gradually reduced and a substantial amount of particles and syntactically not bound two-word utterances appears. Simultaneously, child speech features some basic verbal argument structure constructions, subject-verb (SV), verb-object (VO) and subject-verb-object (SVO), and the number of such constructions increases gradually.

Keywords: child language, repetition, support verb constructions, particles, Estonian

TEISMEEA LOOVKIRJUTISTE SÕNAVARA JA SELLE HINDAMINE

Krista Kerge, Anne Uusen, Halliki Põlda

Ülevaade. Kõrvutatuna haritud täiskasvanud eestlastega, kes ei ole filoloogid, uuriti sõnavaramuutujaid eesti õppekeele kooli 5., 7., 9. ja 11. klassi õpilaste loovkirjutistes, sh 10% kakskeelsetelt. Kirjeldava statistika toel näidatakse kirjutiste sõnavara arengu suundumusi ea ja soo järgi, teisalt aga seda, kuidas õpetajad hindamisel sõnavara arvestavad. Tulemused panevad aluse kirjutiste ja nende hindamise edasise kvalitatiivse uurimise suundadele ja aitavad leida selleks sobivate parameetritega näidiskirjutisi. Ühtlasi vaetakse võimalusi parandada kirjutiste individuaalset hindamist üldhariduskoolis, eriti noorte elukäiku mõjutavatel eksamitel.*

Võtmesõnad: teismeiga, kognitiivne areng, sõnavara varieeruvus ja ulatus, leksikaalne tihedus, enamuskeel õppekeelena

1. Uurimisaine, eesmärgid ja materjal

Keeleareng ei lõpe lapse kooli minnes, vaid jätkub kooli ainetundides ning oluliselt ka täiseas (Nippold 2004: 1–3). Koolis pannakse alus kirjaoskusele sõna mitmekülgses tähenduses (vt Kern 2000: 16–17, Puksand, Kerge 2012), kuigi emakeeleõpetuse koht selles protsessis vajab uurimist ja läbirääkimist (Krogh 2012). Kirjutamine peegeldab õpilase üldist keelelist ja kognitiivset arengut (Brügelmann 1990: 21–25), mille seoseid on käsitleanud Edgar Krull (2000: 106–127) ja mille 29-aastase pikiuurimuse tulemusi Ingrid Schoon jt (2010). Kognitiivse arengu märgina nähakse varajase sõnavara suurenemist (Berk 2003: 357–361). Muude tegurite kõrval võib ka sõnavara varasest arengust sõltuda inimese hilisem sotsiaalne toimetulek (Schoon jt, samas). Ainult väga varane areng on sõltuvuses hoidekeelest (vt Goodman jt 2008).

Lapse sõnavara on suhteliselt ahas (vt Leiwo 1993: 98, Karlep 1998: 276) ja kasvab koos kogemusega. Kool toetab seda arengut kõigi ainete ja kogu tegevusega (vt artikleid Eesti keel 2011), kuid võrdtähtis on igasugune muu tekstikogemus (Corson 1995). Keeleareng on lahutamatu koos haridusega kasvavast keeletunnetusest (ingl *higher language cognition*, vt Hulstijn 2011).

* Uurimust on rahastanud Eesti Teadusfondi grant 8605 "Kirjutamise loomulikkuse mudel ja hindamine".

Oluliselt areneb sõnavara teismeeas; arenev kirjalik süntaks toob omakorda kaasa uusi funktsioonisõnu (Nippold 2004: 3–4). Klassist klassi peaks muutuma õpilasteksti sisu- ja funktsioonisõnade tasakaal, mis erineb kõnes ja kirjas, monoloogis ja dialoogis. Isegi omavaheliste teismemeilide vaba keel on suulisest täiskasvanudialoogist veidi täpsem ja ilmekam, ehkki sidendi- ja deiktikuvaesem. (Vt Kerge, Pajupuu 2010) Kõik see avaldub oletamisi ka õpilaste kirjutistes.

Eesti õpilaste emakeele kirjutusoskust on algkooli ulatuses doktoritööna uurinud Anne Uusen (2006a, vt ka 2006b). Gümnaasiumilõpetajate kirjadeid käsitleb Kersti Lepajõe väitekirj (2012a, vt ka 2012b). Praegu täieneva emakeeleõppija korpuse märgendusvõimalusi vaeb Kadri Sõrmus (2008). Täiskasvanute emakeel-seid (L1) toimetamata kirjutisi ja sõnavara on uuritud ülikooli esmakursuslaste keeleoskuse osana (Ehala jt 2010), kirjaliike ka rööpselt teismelistega (vt Kerge 2010). Emakeele sõnavara arengut on peamiselt uuritud eelkoolieas, mil laps jõuab paari-kolme tuhande sõna valdamiseni (nt Urm 2012).

Enam tähelepanu on saanud eesti keel kui teine keel (L2). Näiteks on riikliku kõrgtasemeeksami kirjutised pakkunud võrdlust suulise keeleoskuse näitajatele (Pajupuu jt 2010), k.a sõnavara (Pajupuu jt 2009). Viimati on sõnavara uuritud eesti keele kui L2 eri taseme eksamite kirjutusosas (Alp jt 2013).

Allpool kajastub loomuliku kirjaliku keelekasutuse uurimuse üks tahk: eesti õppekeelega kooli loovkirjutiste individuaalsed ja ealised erinevused 5., 7., 9. ja 11. klassis. Artikli teema on teismeliste aktiivne sõnavara, mis lubab inimesel end reservatsioonideta väljendada (Corson 1995: 44–45), nagu seda siinkohal võimaldavad ülesanne ja teema (vt osa 2).

Uurimus on katseline ja alustav. Eesmärk on kirjeldada kirjaliku sõnavara arengut täiskasvanulikkuse suunas, andes lähte- ja võrdlusandmeid kognitiivse ja keelelise arengu ning teise keele omandamise uurimustele. Ühtlasi aitab tutvustatav statistiline ülevaade leida kirjutiste kvalitatiivse analüüsi suunad ja nendega sobivad kirjutised.

Tulemused toetavad ka emakeeleõpetuse praktikat. Oluline on teada, kui võrd kirjutise hindaja jälgib sõnavaramuutujaid, mida uurijal on võimalik mõõta, kuid õpetajal mitte, ehkki seda ametlikes juhendites nõutakse. Näiteks eeldab eesti keele riigieksami loovkirjutise maksimaalhinne muu kõrval, et sõnastus on täpne ja isikupärane ning sõnavara rikkalik. Hea hinde puhul on sõnavara mitmekülgne ja rahuldava puhul piisav; puuduliku ja nõrga hinde korral on kas tekst üldsõnaline või sõnavara piiratud. (Vt Riigieksam 2013: 6) Samas saab Hille Pajupuu jt (2009: 188) väitel juhendi järgi hinnata seda, kui ladus on jutt ja kui sidus tekst, kuid väga keeruline on hinnata sõnavara rikkust ja ulatust, mille objektiivne mõõtmine nõuab peale täpsete arvutuste ka sõnade keskmise kasutussageduse tundmist.

Sõnavara kui intuiitiivselt hinnatava õpisihi mõtestamiseks on valitud loovkirjutistes uuritavad sõnavaramuutujad, mis objektiivselt mõõdetuna näitavad sõnavara mitmekülgust ja ulatust (vt osa 3). Ainus hüpotees on, et laps areneb kogu aeg (vt Krull 2000) – küsimus on, kuidas.

2. Uuritud kirjutiste laad ja maht

Kirjutised on loodud 2011/12. kuni 2013/14. õa eakohaselt ühetaolise ülesande alusel ning kindla kirjutusaja jooksul üheksas Tallinna, Harjumaa, Hiiumaa ja Lääne-Virumaa üldhariduskoolis, mille hulgas ei ole nn eliitkoole. Eetilistel põhjustel vabatahtlikke osalejakoole ei avaldata – leitud tendentse ei soovita kindlate koolide, õpilaste ega õpetajatega siduda.

Kirjutisi on omal harjunud viisil (holistiliselt ehk tervikuna ja püstitatud ülesandest lähtudes) hinnanud eesti keele ja kirjanduse õpetajad, kes kirjutuskatse oma õpilastega läbi viisid. Nad ei olnud teadlikud, mida kirjutistes uurima hakatakse ega saanud hindamiseks erijuhtumiseid.¹ Erandina hindas ühe klassi töid n-ö võõras emakeeleõpetaja, sest katse viis läbi matemaatik. Kaasatud õpilased pandi kirjutama sotsiaalse suunitlusega arvamust väärtuste kohta, manades neile ülesandes ette kujuteldava olukorra ja rolli ning piiritledes žanri ja kirjutusaja. Ülesande sõnastus on kohandatud eri vanuses laste jaoks.

5. ja 7. klassi ülesanne on sõnastatud järgmiselt:

Kujuta ette järgmist OLUKORDA.

Teie kool tahab korraldada konkursi kõige õnnelikuma ja elus hästi hakkama saava inimese leidmiseks. Kooli ajaleht kogub ja avaldab nüüd võimalikult palju arvamuskirjutisi selle kohta, milline on õnnelik ja elus hästi hakkama saav inimene.

KIRJUTA koolilehele ka oma arvamus: milline inimene on sinu arvates õnnelik ja saab elus hästi hakkama?

Sinu arvamustlugu peaks olema u. 150 sõna pikk ja mahub parasjagu alla kasti – ruumi peaks ülegi jääma. Kirjutamiseks on aega üks koolitund.

9. ja 11. klassi ülesanne erineb seepoolt, et ülesandega on seotud loomuliku kirjutusolukorra tavaline kaastegevus, s.o lugemine (üldsõnaline tekst väärtuskasvatuse tähtsusest Eestis eri aegadel; 421 sõnet).² Ülesanne on sõnastatud järgmiselt:

Käisid hiljuti

ÕPILASKONVERENTSIL, MILLE TEEMA OLI “VÄÄRTUSKASVATUS”.

Lugesid ka ühe ettekande põhjal kirjutatud teksti. Sinu ülesanne on kirjutada üritusel kuuldust ja loetust artikkel koolilehe jaoks. Lisaks üldisele ülevaatele peatu lähemalt kahel kolmest konverentsil käsitletud teemast.

1. Mis on hariduse peamine eesmärk – kas teadmised või tervikliku isiksuse kasvatamine?
2. Millised iseloomuomadused tagavad õnneliku elu ja ühiskonnas hakkamasaamise?
3. Kas kõik algab kodust või saab kool samuti lapse iseloomuomadusi kujundada?

Teksti nõutav pikkus on umbes 250 sõna. Aega kirjutamiseks on 60 minutit.

¹ Kirjutistes ei uurita ainult sõnavara, muud tulemused on avaldamata.

² Lugemispõhist ülesannet ei saanud kasutada kahes nooremas klassis, kus ollakse harjunud ümberjutustusi kirjutama (vt osa 4.3).

Ülesanded esitati 12 p kirjas. Mõlemal juhul järgnes joonitud ala, kuhu optimaalse pikkusega tekst mahutada.

Kokku uuriti 159 õpilaskirjutist arvestusliku kogumahuga u 30 000 sõnet. Kirjutise keskmine pikkus oli 5.-7. klassis 142 sõnet (eeldatud 150) ja 9.-11. klassis 230 (eeldatud 250). Juhisest sõltumata kirjutasiid õpilased väga erineva pikkusega tekste. (Vt tabel 1) Edgar Krulli järgi erinebki laste õpitempo tavaliselt kuni kuus, harvem isegi 10 korda (Krull 2000), aga individuaalsete mõjurite üle saab ühetaolise katse korral vaid spekuloida.

Valimist jäeti välja ja tabelis 1 ei kajastu paar tööd, millel puudusid teksti tunnused (eriti lühike ja selgelt pooleli). Valimisse kaasati nende 16 õpilase tööd, kes deklareerisid end kakskeelsena (11 eesti ja vene, 2 eesti ja inglise) või kelle kakskeelsus tuleneb hariduskeelest (3 vene emakeelega last). Selle rühma osatähtsus (u 10%) vastab ligikaudu eesti koolide üldisele olukorrale.³

Sugude osatähtsus valimis lubab võrdlusi – tüdrukuid on u 53%. Nn kakskeelsete kirjutisi vähese arvu tõttu soo järgi ei vaadelda. Kooli asukoha järgi näitajaid ei võrrelda, kuid lapsi valiti igas vanuses nii Tallinnast kui väljastpoolt.

Valim näitab seega vaid tendentse, mille olulisust statistiliselt kontrolliti.

Tabel 1. Uuritud kirjutiste arv ja mahuandmed

Klass	Arv	Sh kakskeelsed	Maht sõnedes				
			Keskmine	Standardhälve	Mediaan	Lühim	Pikim
5. kl poisid	20	1	136	36,78	140	79	236
5. kl tüdrukud	23	3	157	32,69	151	102	269
5. klass	43	4	147	34,73	146	91	253
7. kl poisid	14	0	125	17,18	123	106	157
7. kl tüdrukud	17	1	147	28,10	147	100	191
7. klass	31	1	136	22,64	135	103	174
9. kl poisid	25	4	179	80,42	154	81	345
9. kl tüdrukud	24	1	241	77,62	244	99	448
9. klass	49	5	210	79,02	199	90	397
11. kl poisid	12	3	234	25,06	229	200	291
11. kl tüdrukud	19	3	289	59,02	282	203	430
11. kl sugu märkimata	5	0	205	36,72	219	150	243
11. klass	36	6	247	47,87	251	177	337
Kokku	159	16					

Võrdluseks on kasutatud peamiselt eesti mittefiloloogidest haritlaste emakeelseid toimetamata kirjutisi, mis on loodud kõrgtaseme ametlikku eksamiolukorda ja kirjutusülesandeid jäljendades (vt Pajupuu jt 2010). Täiskasvanute korpuse kirjutisi on vaadeldud kui õpilase arengu loomulikku sihti ja hea keekekasutuse etaloni. Teine võrdluse suund seob emakeele ja teise keele (vt Alp jt 2013).

³ Oletus toetub veebis leitud artiklitele (kõikumine aastati ja paikkonniti 7% ja 13% vahel). HTM-i statistikas (<http://www.hm.ee/index.php?048055;30.9.2013>) neid andmeid ei leidunud.

3. Uuritavad muutujad ja metoodika

Sõnavaramuutujate nimetamisel on pilt kirju. Samastatakse ja eristatakse sõnavara rikkust (ingl *vocabulary richness*), leksikaalset varieeruvust (*lexical variation*) ja mitmekesisust (*lexical diversity*), mõõtes neid mitmesugustel meetoditel, mis kõik suhestavad eri sõnade arvu (N) sõnede ehk sõnaeksemplaride koguarvuga (n) tekstis või korpuses (vt Laufer, Nation 1995: 310, Vermeer 2000, Malvern jt 2004: 3 jm). Terminite ebaühtlust analüüsib põhjalikult Guoxing Yu, kellele toetudes võiks sõnavara rikkuseks nimetada mitme uuritud muutuja kooskõla (Yu 2009: 238–241).

Uurisime kolme tüüpi muutujaid: 1) sõnavara üldine varieeruvus (kui palju eri sõnu kirjutaja kasutab); 2) sõnavara koostis ja ulatus (sagedaste ja harvade sõnade osatähtsus ja viimaste varieeruvus); 3) sisusõnade osatähtsus ehk leksikaalne tiheus. Allpool esitatakse iga tunnuse mõõdikud (vrd L2-ga Alp jt 2013).

Kõigepealt SÕNAVARA ÜLDISEST VARIEERUVUSEST. Sõnade ja sõnede lihtne protsendi-suhe (ingl *type-token ratio*) andis piloteerimisel kõige nõrgemaid korrelatsioone ega ole kõnekas (vt ka Vermeer 2000). Väiksemahuliste tekstide sõnavara varieeruvuse (rikkuse) mõõtmisel on Anne Vermeer (samas) pidanud sobivaks Uberi indeksit U , mis on osutunud heaks eristajaks nii eesti toimetamata täiskasvanutekstide (Pajupuu jt 2009) kui ka õpiteksti uurimisel (Puksand, Kerge 2012). Valem põhineb sõnede arvul n ja sõnade arvul N :

$$U = (\log n)^2 / (\log n - \log N)$$

Selle meetodi kõrvale valisime väga levinud Guiraud' indeksi G (Guiraud 1960, viidatud Mellor'i 2011 järgi), mis iseloomustab teksti sõnavara üldist varieeruvust sõnade arvu suhte järgi sõnearvu ruutjuuresse:

$$G = N / \sqrt{n}$$

Ehkki David Malvern jt (2004) samastavad sõnavara mitmekesisuse tema ulatusega (“the range of vocabulary and avoidance of repetition”), võib varieeruvus suur olla ka siis, kui kordamata kasutada sõnu paari tuhande kõige sagedama seast. Just nüüdistekstides harva esineva sõnavara valdamine on aga noortel kasin (vt Ehala jt 2010). Järelikult on oluline mõõta ka nn harva sõnavara (ingl *rare, advanced* või *sophisticated vocabulary*) rikkust (van Hout, Vermeer 2007), mida Soomes ja Eestis on nimetatud sõnavara ulatuseks (Tuokko 2007, Pajupuu jt 2009). Haritud sõnakasutus aitab süvenenumalt käsitleda teemat ning toetab kirjutise isikupära. Nimetame seda keerukaks sõnavaraks.

Näitamaks sõnavara ulatust, leiti iga osaleja LEKSIKAALNE PROFIL (LFP; vt Laufer, Nation 1995, Meara 2005), kus kirjutise sõnad on sagedussõnastiku (Kaalep, Muischnek 2002) alusel jagatud 1000 sõna kaupa sagedusjärgkudesse: 1. järgus on 1000 sagedasimat sõna, edasi sagedusvahemikud 1001–2000, 2001–3000 9001–10 000 ning 11. järguna sagedussõnastikust välja poole jäävad sõnad.

Üldistamiseks jagati iga kirjutise sõnastik tinglikult kaheks:

- *baassõnavara* ehk 4000 sagedasima sõna hulka kuuluvad sõnad (järgud 1–4), mis jagati jälgimise huvides omakorda kaheks: 1. järk (ülisagedad) ning koos 2.–4. järk;⁴
- *keerukas sõnavara* ehk kõik harvemad sõnad (järgud 5–11).⁵

KEERUKA SÕNAVARA OSATÄHTSUS (LS, ingl *lexical sophistication*; Laufer, Nation 1995: 309) näib eristavat olulisi arenguid, nt eesti keele kui L2 kirjutistes tasemeid A2 ja C1, olemata oluline lähemate tasemete võrdluses (vt Alp jt 2013: 281–282). Saamaks teada, mis toimub eesti koolis, mõõtsime seda näitajat teksti sõnede kohta:

$$LS = \text{keerukate sõnede arv} / n \times 100$$

KEERUKA SÕNAVARA VARIEERUVUST mõõtsime Advanced Guiraud' indeksiga AG (Daller jt 2003), mis suhestab sõnearvu n ruutjuurega n -ö harituma leksika:

$$AG = \text{keerukate sõnade arv} / \sqrt{n}$$

Laialt kasutatud LEKSIKAALSE TIHEDUSE indeks LD (ingl *lexical density*; Ure 1971) põhineb sisusõnade määral võrreldes funktsioonisõnadega, iseloomustades ühtaegu sõnavalikut ja süntaksit. Muutuja näitab teksti infotihedust ja võib olla teksti mõtelise keerukuse näitaja. (Vt arutlusi ja viiteid Kerge 2010: 45–48.) Lõtv tekst jääb paratamatult üldsõnaliseks. Ühtlasi räägib LD funktsioonisõnade määra muutumisest klassiti. Kuna arvutianalüüs (nt ka sagedussõnastik) ei erista funktsioonisõnu sisusõnade liigi piires (abimäärsõnu ja -verbe muudest), siis on leksikaalne tihedus arvatud järgmiselt:

$$LD = \text{nimisõnade} + \text{omadussõnade} + \text{verbide} + \text{adverbide sagedus}$$

Muutujat on mõõdetud ka õpitekstis kõrvu muud liiki tekstidega (Puksand, Kerge 2012: 189–191).

Hinnetega ja omavahel korreleeriti individuaalsete kirjutiste mediaannäitajad.

4. Tulemused ja diskussioon

Allpool peatutakse üldnäitajatel, hindamisse puutuval, sõnavara üksikmuutujatel ja lõpuks tervikpildil, diskuteerides pidevalt tulemuste üle. Arvutused, mida pole kommenteeritud, põhinevad tervikvalimil, sest õpetaja ei hinda töid kirjutaja emakeele järgi.

4.1. Üldised suundumused

TEKSTI PIKKUS. Nagu nähtus ülevaatest (tabel 1), on laste kirjutised erineva pikkusega, kuigi rühmiti oli seatud ühetaoline kirjutusaeg. Keskmiselt lühemad on igas vanuses poiste kirjutised – teksti pikkuse ja soo korrelatsioon on nõrk, kuid oluline ($r = 0,29$; $p < 0,001$). Teksti pikkus mõjutab oluliselt nii sõnavara üldist varieeruvust kui ka (küll nõrgalt) keeruka sõnavara näitajaid.

⁴ Ehkki Kaalepi-Muischneki (2002) sagedussõnastiku sissejuhatuse järgi katavad 3000 sagedasimat sõna 80% kirjakeele korpuse sõnedest ning nende määr katab ka haritud täiskasvanute L1 esseedes u 66% tekstist (Pajupuu jt 2009: 192), oleme arvestanud ka võimalikke muid kontekste kui korpuses esindatud. Selline järgujaotus lubab õpilaste LFP-d võrrelda samas grandis uuritud L2 kirjutiste omaga (vt Alp jt 2013), kuid raskendab võrdlust haritlaste L1-ga (vt 4.4).

⁵ Google'iga saab otsida viidatud sagedussõnastiku (eesti kirjakeele korpuse ressursside 1 mln sõna baasil loodud sagedussõnastik). Siin harvaks loetud sõnu näeb seal pealkirja "10 000 sagedasimat lemmat" all, kust lingitud tabeli veerus "Kokku" on nende esinemiskordi 20 või vähem.

Uuringudisaini usaldusvääruse huvides kontrollisime ka seda, kui tugev on teksti pikkuse korrelatsioon eaga ja see osutus keskmiseks ($r = 0,5$). Tugeva korrelatsiooni korral ($r > 0,7$) moodustaksid nad koos tulemusi kallutava faktori. Kuna teksti pikkus ei mõjuta valimis hinnet, ei korreleeru eaga 5.-7. kl rühmas ja seos on nõrk ka 9.-11. kl osas ($r = 0,31$; $p = 0,004$), on tema mõju võrdlustele üldiselt väike.

Kirjutamise kulgu ja seeläbi teksti pikkust peaks mõjutama emakeel. Ann Chenoweth ja John Hayes (2001) on leidnud, et L2 puhul ilmnevad erinevused L1-st kirjutamise sujuvuses (sõnade arv minutis) ja mahus (sõnade hulk tekstiosa kirjutamisel ajaühikus). Siinses katses puudus teksti pikkusel deklareeritud emakeeltega vähimigi korrelatsioon.

KIRJUTAJA IGA. Ülevaate kirjutiste otsestest sõnavaranäitajatest annab klassiti tabel 2. (Leksikaalset tihedust vaatame allpool.)

Tabel 2. Olulisemate sõnavaramuutujate erinevused klassiti

Klass	Möödik		
	Sõnavara varieeruvus G	Keeruka sõnavara varieeruvus AG	Keeruka sõnavara osatähtsus sõnedest LS
5. kl	5,73	0,40	6,80
7. kl	5,96	0,45	6,94
9. kl	8,06	1,10	14,1
11. kl	9,56	1,82	18,4

Vanusel on erinevate sõnavaranäitajatega alati oluline vastastikmõju (vt tabel 3). Tugevasti sõltuvad kirjutaja east nii üldine (G) kui ka keeruka sõnavara varieeruvus (AG) ja keerukate sõnade osatähtsus (LS). Nõrgemini seostub eaga sisusõnade määr (LD).

Tabel 3. Sõnavaramuutujate sõltuvus kirjutaja vanusest

Sõnavaramuutuja	Möödik	Vanuse ja muutuja korrelatsioonikordaja r
Üldine varieeruvus	G	0,83***
	U	0,59***
Keeruka sõnavara varieeruvus ja ulatus	AG	0,81***
	LS	0,74***
Sisusõnade osakaal	LD	0,52***

Märkus. Kordaja $r < 0,30$ puhul on korrelatsioon olematu, väga nõrk; $r < 0,70$ puhul keskmise tugevusega; $r > 0,70$ puhul tugev. Olulisuse kahesuunaline tõenäosus on siin alati maksimaalne: *** $p < 0,001$.

Selgelt erineb muutujate ja ea seose tugevus kahe vanuserühma sees. 5.-7. kl rühmas on näitajate erinevus väike, nende korrelatsioon eaga alati statistiliselt ebaoluline ja väga nõrk. Kahe vanema klassi erinevus on silmnähtav, muutujate korrelatsioon eaga keskmine/tugev ja alati väga oluline. Enim eristavad 9. ja 11. klassi sõnavara variatiivsuse näitajad ($r = 0,69$).

Sõnavara üldvariatiivsust mõõdavad nii Guiraud' kui ka Uberi indeks, kuid viimane seostub eaga nõrgemini. Kuna peamine huviobjekt on õpilase areng, siis viimase tulemusi (vt U tabelis 3) kõigi parameetritega ei seostata.

KIRJUTAJA SUGU ei mõjuta valimis sõnavara üksikmuutujaid praktiliselt üldse. Nõrgalt seostub sooga vaid 9. ja 11. kl sõnavara üldvarieeruvus ($r = 0,27$; $p = 0,015$). Üksikmuutujad sõltuvad kirjutajast, alates 7. klassist mõjutab neid oluliselt iga, mitte sugu. Küll aga muutub sooline pilt tugevasti, kui vaadata sõnavara ealisi näitajaid koos (vt osa 4.6).

4.2. Hinded ja sõnavara

Hinded on nooremas rühmas veidi paremad kui vanemas: 5. ja 7. kl keskmine ja mediaan on puhas 4 ning 9.-11. kl keskmine 4-, mediaan pigem 3+.

Teisalt mõjutab sõnavara hinnet harva oluliselt (vt tabel 4), enamasti juhuslikult (vrd korrelatsioone ja nende suunda klassiti ühe muutuja piires) ja vähe (vt kordajate tähendusi tabeli 3 alt).

Ainus keskmiselt tugev ja oluline seos on hindel 5. kl sõnavara üldvarieeruvusega (G). Väheoluliselt ja vähem mõjutab 5. kl hinnet kõik: hinnet tõstab keeruka leksika varieeruvus (AG), aga langetavad keerukate sõnade ja sisusõnade suurem määr (LS, LD). Arvestatav, kuid väheoluline on keeruka sõnavara osakaalu vastasuunaline seos hindegaga 11. klassis (LS). Muud seosed puuduvad või on statistiliselt ebaolulised. (Vt tabel 4)

Tabel 4. Hinde ja olulisemate sõnavaramuutujate korrelatsioonid eärühmiti

Sõnavaramuutuja	Möödik	Klass	Hinde ja muutuja korrelatsioonikordaja <i>r</i>
Sõnade üldine varieeruvus	G	5. kl	0,54***
		7. kl	-0,23
		9. kl	0,20
		11. kl	0,19
Keerukama sõnavara varieeruvus ja ulatus	AG	5. kl	0,34*
		7. kl	-0,15
		9. kl	-0,03
		11. kl	-0,14
	LS	5. kl	-0,35*
		7. kl	-0,24
		9. kl	-0,18
		11. kl	-0,34*
Sisusõnade osakaal	LD	5. kl	-0,32*
		7. kl	-0,06
		9. kl	-0,19
		11. kl	-0,16

Märkus. Olulisuse kahe-suunaline tõenäosus *p* on märgitud järgmiselt: **p* < 0,05, ***p* < 0,01, ****p* < 0,001.

Klasse rühmitades⁶ on pilt teistsugune. 5. ja 7. klassil rühmana ei korreleeru hindega enam ükski tunnus. Seevastu vanematel, s.o 9. ja 11. kl osas hinnatakse kõrgemini mitmekesisema sõnavalikuga teksti: hinnet mõjutab märgatavalt ja oluliselt ($p < 0,001$) sõnavara üldvarieeruvus G ($r = 0,40$); ebaoluliselt ja väga nõrgalt ka keeruka sõnavara varieeruvus AG ($r = 0,21$; $p = 0,060$). Muud muutujad hindega ei seostu.

Helmut Dalleri ja David Phelani (2007) ning Roeland van Houti ja Anne Vermeeri (2007) järgi peaks kirjutise hinnet tõstma sõnavara suurem üldvarieeruvus – seda kinnitasid selgesti 5. kl tulemus ja vanemad klassid rühmana. Samadel viidetal peaks hinnet enam tõstma keeruka sõnavara osakaal, mis 5. ja 11. klassis hinnet väheoluliselt hoopis langetas. Tegurite alamõju õpilaskirjutiste hindele võib põhjustada see, et võrreldavad uurimused põhinevad L2-l, kus sõnavara arengut hoolega jälgitakse (vrd mõjuga 5. klassis). Halvema hinde võivad põhjustada keeruka sõnavaraga kaasnevad vead ja õpetaja hindamisstiil. Piloteerimisel ilmnes nt hinnete oluline seos hindajaga 9. ja 11. klassi osas: $r = 0,54$; $p < 0,001$. Hindetegurid, sh korrektsus vajavad aga eraldi uurimist.

Õpilase emakeelega hinne praktiliselt ei seostu ($r = -0,13$; $p = 0,111$): noorematel puudub sõltuvus üldse, vanema rühma löikes on see väga nõrk ja ebaoluline ($r = -0,25$; $p = 0,255$), sest seost mõjutab nn kakskeelsete osalejate väike osatähtsus – juttu on 16 (10%) kirjutisest enamuskeelses koolis.

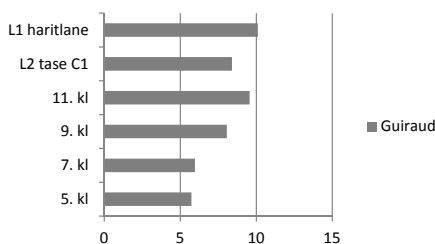
Hinde ja sõnavara nõrku seoseid seletab see, et õpetajal tuleb ka kirjutisi holistiliselt hinnates intuiitiivselt arvestada eri tüüpi sisu- ja vormiseiku, kuid kirjutiste hindamise tahud ja rubriigid on praeguseks omaette teadus. Barbara Moskal (2003) toob esile kümneid rühmitatud komponente, mida kirjutamise eeldusena arvesse võtta ning hiljem kirjutiste juures hinnata, ja need on selgunud pärast tuhandete tööde läbilugemist. Oma liigitusega rõhutab ta rubriikhindamise eeliseid holistilise ees. Vicky Spandel ja Rick. J. Stiggins (1990) pakuvad välja seitse efektiivsele kirjutamisele iseloomulikku tunnust: sisu, struktuur, sõnum, sõnavalik, lausestus, kokkulepped (grammatika ja õigekiri, k.a interpunktsioon) ning esitus (käekiri, paigutus, kujundus). Kõiki neid hiljem on arvestatud ja arendatud (nt McMackin, Siegel 2002, Tierney, Marielle 2004 jpt, Uusen 2006a). Kirjutusoskuse arendamise olulisima tahuna emakeeletunnis rõhutab Ellen Krogh (2012) kirjutaja oskust sõnumit oma vaatepunktist organiseerida ja põhjendada (ingl *voice*). Niisugusel taustal on ilmne, et kirjutiste rubriikhindamine vajab eraldi tähelepanu, kusjuures sõnavara puhul on ilmseid eeliseid elektroonilisel hindamisel (vt nt Enright, Quinlan 2010, Cushing Weigle 2013). Viimane vajab aga eraldi käsitlust. (Vt ka osa 4.6.)

Tuleb siiski veel kord rõhutada, et hindamisuurimus on eraldi teadusharu – siinne alustav analüüs ei räägi midagi hindamise kvaliteedist, vaid annab teada kvalitatiivanalüüsi vajadusest ning lubab valida suurest materjalist välja sellised tööd, mida kvalitatiivuurimuses käsitada kindla parameetri või nende kombinatsiooni tasememudelina (vrd nt Tillema jt 2012).

⁶ Rühmitamine on siin küsitav, sest õpetaja kasutab eri eas sama hindedkaalat.

4.3. Üldine leksikaalne varieeruvus

Nagu eespool näidati (vt tabel 2) ja võiski oletada, on sõnavara üldvarieeruvus tugevasti seotud eaga ja need seosed alati väga olulised. Haritud emakeelekõnelejate toimetamata esseedes uurisime sama muutujat võrdluseks mõõdikuga G (mediaan 10,1), sest varem oli kasutatud vaid Uberi indeksit (Pajupuu jt 2009). 11. klassis jõuab G (9,6) haritlastele väga lähedale. Sama L2-näitaja sarnaneb sõltumata täiskasvanu suuremast üldkogemusest veel C1-tasemel 9. klassiga, jõudes 8,4-ni (Alp jt 2013; 280).⁷ Ülevaate leksika üldise varieeruvuse arengust annab joonis 1.



Joonis 1. Sõnavara üldise varieeruvuse loomulik kasv kirjutistes

Guiraud' indeks G näitab seoseid eaga tugevamana kui Uberi indeks U (vt tabel 3). Seejuures ei eristu varieeruvusest 5. ja 7. kl sõnavara; ainult U näitab nõrka ja väheolulist pöördkorrelatsiooni ($r = -0,26$; $p = 0,025$). Küll erinevad taas keskmiselt tugevasti ja oluliselt ($p < 0,001$) 9. ja 11. klass – nii G järgi ($r = 0,68$) kui ka U järgi ($r = 0,49$). Seega leidis kinnitust, et Uberi indeksil on seda enam eeliseid, mida lühem on tekst (vt Vermeer 2000).

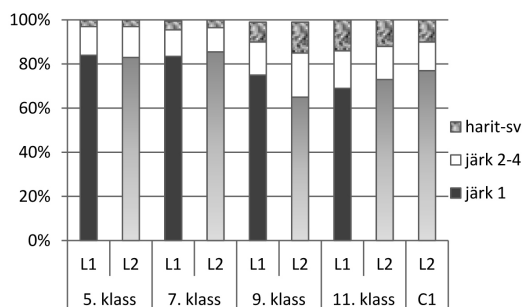
Kummatigi ei ole tulemused küsimärkideta. Näiteks võib leksika enam varieeruda neil, kes kasutavad ülesande täitmisel alusteksti sõnu (Baba 2009), nagu siinses katses 9. ja 11., kuid mitte 5. ja 7. klass. Niisugusel ülesandepüstitusel on aga koolieas selged põhjendused. 1) Täiskasvanu ei kirjuta asjalikku teksti kunagi ilma asjakohaseid varasemaid tekste lugemata ja arvesse võtmata. 2) Noorematel ei saa lugemistekste samas kasutada: eas, kus ümberjutustusega ollakse harjunud, kuid arvamuskirjutisega mitte, oleks niisugune katse tõenäoliselt lõppenud domineeriva ümberjutustamisega ja kirjutiste loomuliku sõnavara olnuks raske mõõta. Ka muutnuks lugemine kirjutuskatse noorematele väsitavaks. Küll viitab Kyoko Baba väide sellele, et loetud alusteksti ja ülesande sõnavarale tuleks kirjutise sõnastiku varieeruvust ja ulatust mõõtes anda selgelt vähem kaalu.

4.4. Leksikaalne profiil ja sõnavara keerukus

Võrdlesime sõnede sagedusjärkude jagunemist eraldi emakeelsete ja kakskeelsete laste individuaaltöodes, ehitades kolmel üldistataval sagedusjärgul põhineva LFP igas vanuserühmas. Profiilide võrdlus näitab hüpet kognitiivses arengus pärast 7. klassi: harituma-keerukama sõnavara osatähtsus sõnedes kasvab 5. kl kõrval põhikooli lõpus umbes kahekordseks ning gümnaasiumis kolmekordseks. (Vt ees

⁷ C1-eksamil ja õpilaskatses on kaalul ka hinne ja pingutus vahest suurem kui L1 haritlastel.

tabel 2 ja all joonis 2.) Joonisel 2 näeb ka 10% kakskeelsete laste eripära võrreldes kirjaliku täiskasvanutasemega C1 (Pajupuu jt 2010).

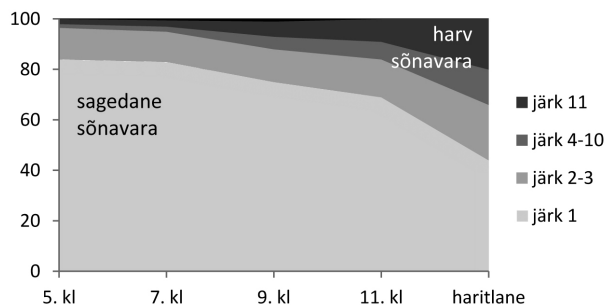


Joonis 2. Leksikaalsete profiilide võrdlus ea ja emakeelte järgi. Lühend *harit-sv* viitab keerukale sõnavarale (sagedusjärk 5–11). L2 (sagedusjärk 1 eristavana helehall) tähistab tinglikult ka kakskeelseid õpilasi

Keerukama sõnavara osatähtsus sõnedes (LS) on emakeelsetel õpilastel eraldi võetuna 5.-7. kl rühmas 6,8% ning 9.-11. kl rühmas 15,4%. Kakskeelsete laste rühmad teistest ei erine: nooremate mediaannäit kattub emakeelsetega (6,8%) ning vanemate rühmas, k.a nõrgema keskhindega venelased, on see teistest ebaoluliselt paremgi (17%).

Uuritud eestikeelse kooli vähesed kakskeelsed lapsed kasutavad keerukat sõnavara poole enam kui täisealised L2 kõnelejad, kelle haritud sõnavara kasvab A2-taseme 6,5%-lt 9,8%-ni C1-tasemel (Alp jt 2013). L2 leksikon on küll alati suhteliselt väike (vt Laufer, Nation 1995: 307), kuid sinne uurimus toob esile kakskeelsuse eelised, olgu siis tegu kakskeelse kodu või emakeelest erineva hariduskeelega. Muidugi vajab tulemus kontrollimist vähegi usaldusväärsel valimil.

Kuna varasemad andmed (Pajupuu jt 2009) ei kajasta eesti haritlaste leksikat sagedusjärgudes 1–4 ja 5–11, ehitasime üldistava võrdluse teistsugustel alustel. Jooniselt 3 on näha, et õpilased ei jõua 3000 sagedasima sõna osas (järgud 1–3) veel haritlaste kirjaliku arvamusalvuse tasemele: viimastel on nende osatähtsus u 65% sõnedest ja kahe vanimagi earühma õpilastel paarkümmend protsenti suurem.



Joonis 3. Õpilaskirjutiste sõnede sagedusjärgud (%) võrreldes eesti haritud täisealiste mittefiloloogidega (võrdluse allikas Pajupuu jt 2009: 192)

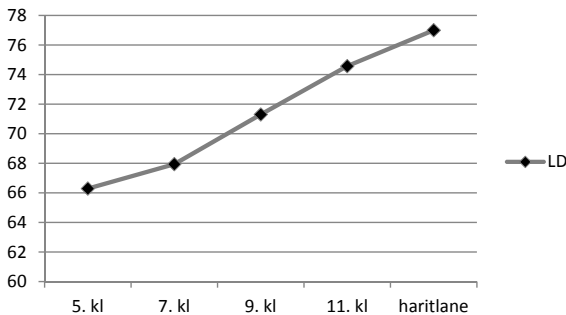
Kuigi Pilvi Alp jt (2013) 1. järku eraldi ei kaalu, vajavad sõnavarauurimustes ilmselt kvalitatiivset tähelepanu kaks äärmist sagedusrühma.

- 1) 1000 sagedasima sõna sõnesid on L1 haritlastel kolmandik vähem kui gümnasistidel (u 44% 11. kl 69% vastu). Põhikooli lõpus on nende määr u 74% sõnedest ning 5. ja 7. klassis võrdselt 83-84%. Ülisagedaste sõnade osatähtsus erineb noorematel (5. ja 7. klassi rühmal) haritlase keskmisest ligi kaks korda ja kahaneb teismeea jooksul u 14%.
- 2) Suur erinevus on sagedussõnastikust väljapoole jäävate sõnade osas – sellest sagedusrühmast, mida suuresti täidab sõnamoodustus, on haritlas-kirjutistes sõnesid 2,3 korda enam kui 11. kl õpilastel (haritlaste 20% gümnasistide 9% vastu) ning 8–10 korda enam kui 5. ja 7. klassis (2% ja 2,5%). Ka põhikooli lõpetajad (6%) erinevad siin oluliselt täiskasvanutest ja gümnaasiumist.

Kuna aga haritlasi on siinkohal mõõdetud korpusena, vajab see võrdlus kontrolli individuaalsete kirjutiste alusel.

4.5. Leksikaalne tihedus

Tekstide leksikaalne tihedus (LD) kasvab suhteliselt ühtlaselt (joonis 4) ja kõigub õpilaste mediaanväärtuse 69 lähedal, mis on omane ka teismeea meilidele. 5. klassi tekst on mediaanist u 4% lõdvem, sarnanedes enim täisealiste erameilidele, ja 11. klassil u 6% tihedam (75). (Vt meilide kohta Kerge 2010: 47.) Täiskasvanud L1-esseedes on LD näitaja 77, toimetatud ja toimetamata kirjalike žanrite mediaan 71 (vt Puksand, Kerge 2012: 189).



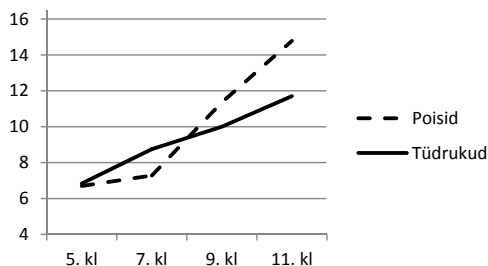
Joonis 4. Kirjutiste leksikaalne tihedus vanuserühmiti (haritlaste andmed: Pajupuu jt 2010)

Teksti tihedus sõltub vanusest üldiselt ($r = 0,52$; $p < 0,001$) ning kahe vanema rühma lõikes ($r = 0,34$; $p < 0,001$), kuid mitte noorematel. Leksikaalsel tihedusel on olulised seosed ($p < 0,001$) variatiivsuse näitajatega G ($r = 0,49$) ja AG ($r = 0,56$) ning keerukama sõnavara osatähtsusega LS ($r = 0,54$).

Lähemalt käsitleme seda teemat mujal kirjutiste süntaksi raames, jälgides sisu- ja funktsioonisõnu liigiti.

4.6. Sõnavara rikkuse kombineeritud indeksid ja hinne

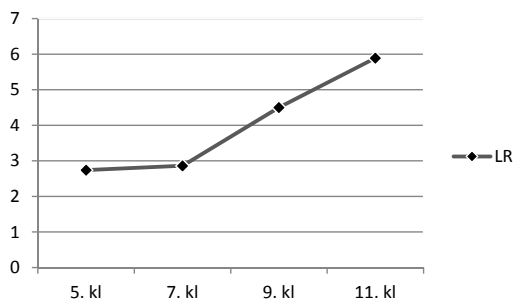
Kuivõrd sõnavara rikkuseks on soovitatud nimetada kõigi uuritud parameetrite kooskõla (Yu 2009), siis arvutasime kirjutiste sõnavara rikkuse tingliku kombineeritud indeksi LR esiteks kõigi 10-le keskmistatud mediaannäitude keskväärtusena klasside ja sugude kaupa. Poistel on sõnavara rikkuse tinglik indeks 10,03 ja tüdrukutel 7% madalam (9,32). Sugude LR-erinevus ei ole statistiliselt oluline 5. ja 7. klassil, kuigi tüdrukud lähevad 7. klassis tuntavalt ette, ega 7. ja 9. klassil, mille vahel poiste LR kasvab järsult. Küll aga on poiste sõnavara tüdrukute omast oluliselt paarkümmend protsenti rikkam 9. ja 11. klassis ($r = 0,64$; $p = 0,001$). Arengukõveraid võrdleb joonis 5.



Joonis 5. Sõnavara rikkuse üldistatud areng sooti

Ootamatuna mõjub poiste sõnavaraandmete taustal see, et poiste hinded on oluliselt nõrgemad nii 5.-7. klassis (hinde ja soo seos: $r = 0,47$; $p = 0,001$) kui ka 9.-11. klassis ($r = 0,30$; $p = 0,005$). Koguvalemis on soo mõju hindele märgatav ja statistiliselt väga oluline ($r = 0,36$; $p < 0,001$).

Teise meetodina kasutasime LR-indeksi ilmestamiseks sõnavara kõigi 1-le keskmistatud mediaannäitajate summeerimist klassiti. Earühmade sõnarikkust võrdleb keskmistatud näitajate summana joonis 6.



Joonis 6. Sõnavara rikkuse areng eanäitajate summaarse indeksi alusel

Hindamise soospekt sarnaneb mõneti riigieksamikirjanditega (vt Külli Habichti analüüsi 2007).⁸ Mediaanis oli poiste sisuhinne seal 11,1% võrra tüdrukutest nõrgem (sisuhinde kaal on 45% üldhindest). Keelehinne (kaal 55%), mida kujundanud õigekiri (35%), sõnastus ja stiil (15%) ning vorm ja kompositsioon (5%), jääb osi

⁸ Vt http://www.ekk.edu.ee/vvfiles/0/Kirjandi_analyyis_2007.pdf (28.12.2013). Hiljem on riik igakülgse eksamianalüüsi asendanud koolide võrdlusega.

kaaludes kokku tüdrukutele alla juba 17,3% võrra. Kuna sõnavalik seostub otseselt sisuga, siis võib see alateadlikult mõjutada poiste suhteliselt paremat sisuhinnet. Habicht analüüsib vaid passiivset sõnavara (teemade mõistmist), kuid soovib õpetada vältima kõnekeelsust ja üldsõnalisust. Analüüs puudutab aktiivset sõnavara 2-3 näitega tähenduseksimuste ja kõnekeele mõju kontekstis, kus 2/3 näiteid ilmestab üld- ja erislangi.

Praegu on riigieksamikirjutise sisul võrdne 25% kaal õigekirjaga, sõnavara ja stiiliga ning vormi ja ülesehitusega; iga tahu hinne 1–5 korrutatakse teguriga 5 (vt Riigieksam 2013). Ilmne on tarve arvutiekсамите järele, et sisu ja kompositsiooni saaks hinnata subjektiivselt, kuid keelekasutust objektiivselt.

5. Kokkuvõtte ja järeldused

Lapsed alustavad 5. klassis väga erineva sõnavaraga – individuaalselt erineb üldine varieeruvus pea kaks korda ja keeruka sõnavara osas isegi 10–12 korda.

5. ja 7. kl statistiliselt lähedased mediaannäitajad vajaks arengupsühholoogilist selgitust. Teisalt võis 7. klassi sõnavalikut kaudselt mõjutada ka uurimuse disain: noorimatega ühesugune ülesanne võis kahandada oma arvamuse võimetekohast argumentatsiooni. Katset tuleks pisteliselt korrata raskema ülesandega.

7. klassi järel toimub kirjaliku sõnavara arengus ilmne hüpe, mis enim väljendub keeruka sõnavara lisanemise näol. 9. kl sõnavara erineb 11. kl omast väga oluliselt, kuid tuleb arvestada, et nende kahe klassi vahel on juba toimunud valik gümnaasiumisse, nii et kognitiivse arengu kohta ei ole see võrdlus kõnekas – põhikooli lõpetajatega tuleks võrrelda ka täiskasvanugümnaasiumi ja kutsekooli õpilasi.

Nagu võiski oletada, kasvab teismeeas kirjaliku sõnavara varieeruvus ja ulatus: 5. klassist 11.-ni tõuseb variatiivsusindeks keskmiselt 1,7 korda, kuid üle kolme korra kasvab keerukam sõnavara. Arenevad ja õpivad kooli jooksul kõik – tundub, et nõrgemad kirjutajad enamgi: Guiraud' järgi on kõige üheülbalisema sõnavaraga kirjutis 11. klassis 39% parem kui 5. klassi nõrgimal; enim varieeruva sõnavaraga kirjutised neis klassides erinevad vähem (27%). Summeeritud LR-näitaja järgi kasvab leksikarikkus teismeeas üle kahe korra.

Eritähelepanu vajavad soolise arengu erinevus ja kirjutise hinde kujunemine.

Sõnavara eristab eri vanuses lapsi hästi, kuid seostub hindega vaid osati, suhteliselt juhuslikult ja nõrgalt – ainult 5. klassis tõstab sõnavara üldine varieeruvus hinnet; nõrk pöördmõju on hindede keerukal sõnavaral 5. ja 11. klassis. See on ka kõik.

On ilmne, et emakeeleomandaja korpusi on vaja arendada pidevalt ning kõiki vanuseid ja eri liiki tekste hõlmates, pluss tingimusi ja kirjutajaid igati (sh hindega) iseloomustades.

Järgmises artiklis tõlgendame õpilaskirjutiste juba uuritud tekstisüntaksi arengut. Siinsete kombineeritud ja üksiknäitajate järgi on võimalik leida ja kvalitatiivselt analüüsida n-ö mudelkirjutisi. Ajapikku tuleks uurida harvema sõnavara koostist ja arvestada ülesande ja lugemisteksti mõju sellele; suhestada õigekeelsust ja konstruktsioonide sagedust jpm.

Kõik see aitab ajakohastada ka loovkirjutiste eakohase hindamise juhendeid. Pole aga kindel, kas leitud andmed annavad juhenditele selge sisu, kui viimaseid

endistviisi rakendada vaid intuiitiivselt. Õigem on teha ja analüüsida õpilase elukäiku mõjutavaid töid (tasemetööd, lõpu- ja riigieksamid, k.a muude ainete esseed) üha enam arvutis ja varustada hindaja niiviisi objektiivsete lisaandmetega. Arvuti ei asenda kirjutiste subjektiivset hindamist, vaid täiendab seda objektiivsete näitajatega, k.a sõnavara omad.

Rakenduslikult aga on, mida siinsetest tulemustest õppida.

Ühelt poolt vajab laps veel 7. klassiski väga individuaalset lähenemist – alles pärast seda saab arvestada individuaalse sõnavara väiksema kõikumisega ja üldise loomuliku edasiminekinga. Teisalt juhivad tulemused taas tähelepanu asjaolule, et puudub ülevaade eri ainete õpetajate üldisest teadlikkusest kirjaoskuse, kirjutusoskuse ja sõnavara arengu või arendamise alal. Selgeks rääkimist vajab emakeeleõpetaja ja muude õpetajate vastutus kirjutusoskuse kujunemise eest.

Täheldasime, et kakskeelsetel ja õppekeelest erineva emakeelega lastel on sõnavaranäitajad omaealistega suhteliselt sarnased, ületades nt keeruka sõnavara osakaalus ligi kahekordselt C1-taseme näitajat (vt Alp jt 2013). Siinsed tulemused ei luba alahinnata eestikeelset haridust, nagu see gümnaasiumis on juba kehtestatud. Et aga sinne valim on üliväike ning teise keele täiskasvanud omandajal suurem üldine ja emakeelekogemus, tuleks võrrelda eesti- ja venekeelse kooli muukeelsete õpilaste eesti ja emakeele kirjutisi.

Earühmade eristajana tõestasid oma olulisust kõik kasutatud meetodid, kuid tulemuste kombineerimine vajab kokkuleppeid ja võrdlusi. Leksikaalses profiilis vajavad tähelepanu äärmised sagedusjärgud.

Batia Laufer ja Paul Nation (1995) on mõõtnud ka leksikaalset originaalsust ehk õppijale eriomaste sõnade osatähtsust, s.o leksikoni seda osa, mida teised kasutanud ei ole. Tulevikus väärriks see muutuja noorte sõnavara võrdluseks sobivana tähelepanu.

Tänu

Täname kõiki koole, õpetajaid ja õpilasi, kelle najal kogu uurimus teoks sai; Helin Puksandit, kes oma pika õpetajakogemuse toel aitas katseülesandeid koostada ja katseid korraldada, ning veel kord ETF-i (grant 8605).

Viidatud kirjandus

- Alp, Pilvi; Kerge, Krista; Pajupuu, Hille 2013. Measuring lexical proficiency in L2 creative writing. – Jozef Colpaert, Mathea Simons, Ann Aerts, Margret Oberhofer (Eds.). *Language Testing in Europe: Time for a New Framework?* Antwerpen: Linguapolis Universiteit Antwerpen, 274–286.
- Baba, Kyoko 2009. Aspects of lexical proficiency in writing summaries in a foreign language. – *Journal of Second Language Writing*, 18 (3), 191–208. <http://dx.doi.org/10.1016/j.jslw.2009.05.003>
- Berk, Laura E. 2003. *Child Development*. USA: Pearson Education Inc.
- Brügelmann, Hans 1990. Lapset ja kirjoittaminen. Kuinka lapset oppivat vähitellen kirjoittamaan oikein. – *Kielikukko*, 1, 21–25.
- Chenoweth, Ann N.; Hayes, John R. 2001. Fluency in writing: Generating text in L1 and L2. – *Written Communication*, 18 (1), 80–98. <http://dx.doi.org/10.1177/0741088301018001004>

- Corson, David 1995. *Using English Words*. Dordrecht: Kluwer Academic Publishers. <http://dx.doi.org/10.1007/978-94-011-0425-8>
- Cushing Weigle, Sara 2013. English language learners and automated scoring of essays: Critical considerations. – *Assessing Writing*, 18 (1), 85–99.
- Daller, Helmut; van Hout, Roeland; Treffers-Daller, Jeanine 2003. Lexical richness in the spontaneous speech of bilinguals. – *Applied Linguistics*, 24 (2), 197–222. <http://dx.doi.org/10.1093/applin/24.2.197>
- Daller, Helmut; Phelan, David 2007. What is in a teacher's mind? The relation between teacher ratings of EFL essays and different aspects of lexical richness. – Helmut Daller, James Milton, Jeanine Treffers-Daller (Eds.). *Modelling and Assessing Vocabulary Knowledge*. Cambridge Applied Linguistics. Cambridge: CUP, 234–244.
- Eesti keel 2011 = Öppekava. Põhikooli valdkonnaraamat: Eesti keel. Gümnaasiumi valdkonnaraamat: Eesti keel. Tallinn: REKK. <http://www.oppekava.ee> > Abiks õpetajale (25.9.2013).
- Ehala, Martin; Kerge, Krista; Lepajõe, Kersti; Sõrmus, Kadri 2010. Kõrgkoolide üliõpilaste eesti keele oskuse tase. Uuringukokkuvõte. Tartu: Tartu Ülikool.
- Enright, Mary K.; Quinlan, Thomas 2010. Complementing human judgment of essays written by English language learners with e-rater® scoring. – *Language Testing*, 27 (3), 317–334. <http://dx.doi.org/10.1177/0265532210363144>
- Goodman, Judith C.; Dale, Philip S.; Li, Ping 2008. Does frequency count? Parental input and the acquisition of vocabulary. – *Journal of Child Language*, 35 (3), 515–531. <http://dx.doi.org/10.1017/S0305000907008641>
- Guiraud, Pierre 1960. *Problèmes et méthodes de la statistique linguistique*. Dordrecht: D. Reidel.
- Habicht, Külli 2007. 2007. aasta eesti keele riigieksami analüüs. – Eesti keele eksamid. Tallinn: REKK, 1–26. http://www.ekk.edu.ee/vvfiles/o/Kirjandi_analyys_2007.pdf (28.12.2013).
- Hulstijn, Jan H. 2011. Language proficiency in native and nonnative speakers: An agenda for research and suggestions for second-language assessment. – *Language Assessment Quarterly*, 8 (3), 229–249. <http://dx.doi.org/10.1080/15434303.2011.565844>
- Kaalep, Heiki-Jaan; Muischnek, Kadri 2002. *Eesti kirjakeele sagedussõnastik*. Tartu: Tartu Ülikooli Kirjastus.
- Karlep, Karl 1998. *Psühholingvistika ja emakeeleõpetus*. Tartu: Tartu Ülikooli kirjastus.
- Kerge, Krista 2010. Kirjazaanrite keeleparameetrid mitme tekstiliigi taustal. – *Emakeele Seltsi aastaraamat*, 55, 32–62.
- Kerge, Krista; Pajupuu, Hille 2010. Text-types in speech technology and language teaching. – Jorge L. Bueno Alonso, Dolores González Álvarez, Úrsula Kirsten Torrado, Ana E. Martínez Insua, etc. (Eds.). *Analizar datos > Describir variación / Analyzing data > Describing variation*. Vigo: Universidade de Vigo (Servizo de Publicacións), 380–390.
- Kern, Richard 2000. *Literacy and Language Teaching*. Oxford: Oxford University Press.
- Krogh, Ellen 2012. Writing in the literacy era: Scandinavian teachers' notions of writing in mother tongue education. – *L1 – Educational Studies in Language and Literature*, 12 (special issue). Iris Pereira, Brenton Doecke (Eds.). *A contribution to the inescapability of language*, 1–28.
- Krull, Edgar 2000. *Pedagoogilise psühholoogia käsiraamat*. Tartu: Tartu Ülikooli kirjastus.
- Laufer, Batia; Nation, Paul 1995. Vocabulary size and use: Lexical richness in L2 written production. – *Applied Linguistics*, 16 (3), 307–322. <http://dx.doi.org/10.1093/applin/16.3.307>
- Leiwo, Matti 1993. *Lapse keeleline areng*. Tallinn: Tallinna Pedagoogikaülikool.
- Lepajõe, Kersti 2012a. Kirjand kui tekstiliik. Riigieksamikirjandite tekstuaalsed, retoorilised ja diskursiivsed omadused. *Dissertationes philologiae estonicae Universitas Tartuensis* 21. Tartu: Tartu Ülikooli Kirjastus.

- Lepajõe, Kersti 2012b. Argumenteerimisvõtted riigieksamikirjandites. – Eesti Rakenduslingvistika Ühingu aastaraamat, 8, 123–138. <http://dx.doi.org/10.5128/ERYa8.08>
- Malvern, David; Richards, Brian; Chipere, Ngoni; Durán, Pilar 2004. *Lexical Diversity and Language Development: Quantification and Assessment*. Hampshire, New York: Palgrave Macmillan. <http://dx.doi.org/10.1057/9780230511804>
- McMackin, Mary C.; Siegel, Barbara S. 2002. *Knowing How. Researching and Writing Non-fictions*, 3–8. Portland: Stenhouse Publishers.
- Meara, Paul 2005. Designing vocabulary tests for English, Spanish and other languages. – C. Butler, M. A. Gómez González, S. Doval Suárez (Eds.). *The Dynamics of Language Use: Functional and Contrastive Perspectives*. Amsterdam: John Benjamins, 271–285.
- Mellor, Andrew 2011. Essay length, lexical diversity and automatic essay scoring. – *Memoirs of the Osaka Institute of Technology, Series B*, 55 (2), 1–14.
- Moskal, Barbara M. 2003. Recommendations for developing classroom performance assessments and scoring rubrics. – *Practical Assessment Research & Evaluation*, 8 (14). PAREonline.net. <http://pareonline.net/getvn.asp?v=8&n=14> (30.10.2013).
- Nippold, Marilyn A. 2004. Research on later language development: International perspectives. – Ruth A. Berman (Ed.). *Language Development Across Childhood and Adolescence: Psycholinguistic and Crosslinguistic Perspectives*. Trends in Language Acquisition. Amsterdam: John Benjamins, 1–8.
- Pajupuu, Hille; Kerge, Krista; Alp, Pilvi 2009. Sõnavara loomulik rikkus haritud keeleteadajate tekstides. – Eesti Rakenduslingvistika Ühingu aastaraamat, 5, 187–196. <http://dx.doi.org/10.5128/ERYa5.12>
- Pajupuu, Hille; Kerge, Krista; Meister, Lya; Asu, Eva Liina; Alp, Pilvi 2010. Natural speaking and how to assess it. – *Trames: Journal of the Humanities and Social Sciences*, 59 (2), 120–140. <http://dx.doi.org/10.3176/tr.2010.2.02>
- Puksand, Helin; Kerge, Krista 2012. Õpikuteksti analüüs kirjaoskuse omandamise kontekstis. – *Emakeele Seltsi aastaraamat*, 57, 162–217. <http://dx.doi.org/10.3176/esa57.09>
- Riigieksam 2013 = Riigieksamite materjalid 2013. Eesti keele riigieksami eristuskiiri. Tallinn: Innove. http://www.innove.ee/UserFiles/Riigieksamid/2013/eesti_keelee_re_eristuskiiri_040413.pdf (20.9.2013).
- Schoon, Ingrid; Parsons, Samantha; Rush, Robert; Law, James 2010. Children's language ability and psychosocial development: A 29-year follow-up study. – *Pediatrics*, 126 (1), 73–80. <http://www.ifs.org.uk/caytpubs/schoon9.pdf> (2.7.2013).
- Spandel, Vicki; Stiggins, Richard. J. 1990. *Creating Writers: Linking Writing Assessment and Instruction*. New York: Longman.
- Sõrmus, Kadri 2008. *Emakeeleõppija korpus. Statistiline analüüs ja veamärgendusüsteem. Magistritöö*. Tartu: TÜ eesti ja üldkeeleteaduse instituut.
- Tierney, Robin; Marielle, Simon 2004. What's still wrong with rubrics: focusing on the consistency of performance criteria across scale levels. – *Practical Assessment, Research & Evaluation*, 9 (2). PAREonline.net. <http://PAREonline.net/getvn.asp?v=9&n=2> (15.8.2013).
- Tillema, Marion; van den Bergh, Huub; Rijlaarsdam, Gert; Sanders, Ted 2012. Quantifying the quality difference between L1 and L2 essays: A rating procedure with bilingual raters and L1 and L2 benchmark essays. – *Language Testing*, 30 (1), 71–97. <http://dx.doi.org/10.1177/0265532212442647>
- Tuokko, Eeva 2007. Mille tasolle perusopetuksen Englannin opiskelussa päästään? Perusopetuksen päästövaiheen kansallisen arvioinnin 1999 Eurooppalaisen viitekehysten taitotasoihin linkitetty tulokset. *Jyväskylä Studies in Humanities* 69. Jyväskylä: University of Jyväskylä.
- Ure, Jean 1971. Lexical density and register differentiation. – G. Perren, J. L. M. Trim (Eds.). *Applications of Linguistics. Selected Papers of the Second International Congress of Applied Linguistics, Cambridge 1969*. London: Cambridge University Press, 443–452.

- Urm, Ada 2012. Individuaalsete erinevuste stabiilsus eesti laste sõnavara arengus. Magistritöö. Tartu: Tartu Ülikooli sotsiaalteaduskond, psühholoogia instituut.
- Uusen, Anne 2006a. Põhikooli I ja II astme õpilaste kirjutamisoskus. Tallinna Ülikooli sotsiaalteaduste dissertatsioonid 19. Tallinn: Tallinna Ülikooli Kirjastus.
- Uusen, Anne 2006b. Kirjutamisest ja kirjutama õpetamisest. Tartu: Atlex.
- van Hout, Roeland; Vermeer, Anne 2007. Comparing measures of lexical richness. – H. Daller, J. Milton, J. Treffers Daller (Eds.). *Modelling and Assessing Vocabulary Knowledge*. Cambridge: CUP, 93–115.
- Vermeer, Anne 2000. Coming to grips with lexical richness in spontaneous speech data. – *Language Testing*, 17 (1), 65–83. <http://dx.doi.org/10.1177/026553220001700103>
- Yu, Guoxing 2009. Lexical diversity in writing and speaking task performances. – *Applied Linguistics*, 31 (2), 236–259. <http://dx.doi.org/10.1093/applin/amp024>

Krista Kerge (Tallinna Ülikool) huviala on eesti kirjakeele allkeelte, žanrite ja individuaalstiili lingvistiline analüüs, eesti keele omandamine esimese ja teise keelena.
Narva mnt 25, 10120 Tallinn, Estonia
krista.kerge@gmail.com

Anne Uuseni (Tallinna Ülikool) huviala on emakeele kirjaoskuse omandamine ja õpetamine algkoolieas.
Narva mnt 25, 10120 Tallinn, Estonia
uusen@tlu.ee

Halliki Põlda (Tallinna Ülikool) huviala on kvalitatiivanalüüsi meetodika, mõistete ja fenomenide analüüsi ja interpreteerimise võimalused avalikus kommunikatsioonis.
Narva mnt 25, 10120 Tallinn, Estonia
halliki.polda@gmail.com

TEENAGE VOCABULARY AND ITS ASSESSMENT IN CREATIVE WRITING

Krista Kerge, Anne Uusen, Halliki Põlda

Tallinn University

Within a study designed to find parameters of educated adult non-philologist L1-writings which give “ideal” benchmarks for assessing high-stakes L1 and L2 writing tests, students’ written vocabulary was targeted in grades 5, 7, 9, and 11 (age 12 to 18) to describe development towards those benchmarks (i.e., higher language cognition; see Hulstijn 2011). 159 students (90% of them native Estonians) were tasked to write an age-appropriate argumentative text on individual and social values. All writings (see table 1) were holistically assessed by the participants’ mother tongue teachers and then analyzed individually for their LFP (lexical profile), measuring the frequency range of all tokens (Laufer, Nation 1995, Meara 2005); their LS (lexical sophistication), measuring the percentage of advanced tokens (see Laufer, Nation 1995); the Uber, Guiraud, and Advanced Guiraud indices, measuring the diversity of general and advanced vocabulary (van Hout, Vermeer 2007, Guiraud 1960, Daller et al. 2003); and the LD (lexical density), showing the proportion of content words to function words (Ure 1971).

No significant difference was found between the vocabulary of grades 5 and 7, although individual values of parameters of the strongest writings may be up to ten times higher than those of the weakest ones. Differences between grades 9 and 11 were quite large and all statistically significant. From grade 5 to 11, variation of general vocabulary grows by 1.6 times and that of advanced vocabulary two times; the amount of sophisticated lexis grows by three times.

The writings of the few non-native participants (80% of them being bilingual also at home) were not typical of L2. The text length and vocabulary of this group did not differ from those of their age peers using L1. The lexical sophistication of this group was twice as big (17) as that of L2-speakers who had passed a C1-test (9.6; see Alp et al. 2013). There is an obvious need to test non-native writing in Estonian comparatively in two local school systems (with Russian or Estonian as the language of instruction).

As supposed, teachers assessing writing holistically were not able to take into account vocabulary parameters. In grade 5, ratings correlated significantly with Guiraud; in grade 11, there was a weak significant correlation between ratings and advanced vocabulary measures. Despite having better late vocabulary (Figure 5), boys’ ratings were significantly lower. This seems to argue in favour of partial computer-scoring.

Keywords: teenage, cognitive development, vocabulary variation and range, lexical sophistication and density, majority language in teaching

GRAMMATIKAKATEGOORiate SEKUNDAARSEST KASUTUSEST: VENE JA EESTI AJAVORMID

Sirje Kupp-Sazonov, Irina Külmoja

Ülevaade. Tänapäeva keeleteaduses üha aktuaalsemad tüpoloogija ja kontrastiivuuringud vajavad ühtlustatud grammatikakategooriate süsteeme ning üheselt mõistetavat terminoloogiat. Eesti keele grammatiliste vormide põhitähendusi on uuritud põhjalikult, kuid sekundaartähendused vajavad veel tähelepanu. Artiklis kõrvutatakse mõningate eesti ja vene keele grammatikakategooriate esinemist ülekantud tähenduses, et näidata selliste kasutusviiside olulisust keele uurimise, õpetamise ning samuti tõlketeooria ja -praktika seisukohast. Põhitähelepanu on ajavormidel. Artikli teises pooles üritatakse sekundaartähendusi tähistavaid oskuskeelendeid analüüsides leida sobivaimaid variante eesti keele grammatika jaoks.*

Võtmesõnad: metafoor, ajakategooria, terminoloogia, eesti keel, vene keel

1. Sissejuhatus

Keeletüpoloogia ja kontrastiivgrammatika on keeleteaduse harud, mis tegelevad keelte kõrvutamise, sisuliselt – võrdlemisega. Sõltumata sellest, kas tegu on mitmeid keeli koondavate tüpoloogiuuringutega (nt Hrakovskij 1985, 1989, 1992, 1998, 2004, Dahl, Koptjevskaja-Tamm 2001, Dahl 2007, 2010, Erelt 1996, 1997, 1999, 2000, 2001) või kontrastiivuuringutega, mis põhinevad enamjaolt kahel keelel, on väga oluline, et eri keelte nähtusi oleks võimalik võrrelda. Eri keelte grammatikakategooriaid saab võrrelda vaid lähtudes nende semantikast ja funktsioonidest. Kategooriate semantika on suuremal või vähemal määral esitatud keelte akadeemilistes grammatikates, mis võimaldab leida vasteid kahe keele kategooriate vahel. Eri keelte võrdlemist hõlbustab oluliselt grammatiliste kategooriate struktuuri analüüs ja sellele järgnev täpsem kirjeldamine.

* Uurimust on rahastanud Eesti Teadusfondi grant nr 9300 "Temporaalsus eesti ja vene keeles".

Vaatleme artiklis eesti keele grammatikakategooriate põhi- ja sekundaar-tähendusi eelkõige eesti-vene kontrastiivgrammatika seisukohast, mis aga ei tähenda, et saadud tulemused ei võiks olla kasutatavad ka eesti keele kõrvutamisel mõne teise keelega. Meie hinnangul vajavad selle teema puhul tähelepanu kaks omavahel tihedalt seotud vaatenurka:

- 1) grammatikakategooriate käsitlus, mis hõlbustaks eesti ja vene keele kõrvutamist kontrastiiv- ja tüpoloogiauuringutes. Näiteks ajakategooria ja selle funktsioneerimine kõnealustes keeltes ei kattu ning vajab detailsemat analüüsi;
- 2) võrdlemisel tekib paratamatult vajadus ka terminoloogia ühtlustamise järele. Näiteks puudub eesti keele grammatikas vaste venekeelsele oskusele *perenosnoe / metaforičeskoe upotreblenie vremennyh form* (ajavormide ülekantud e metafoorkasutus), samas on ühtne terminoloogia kontrastiiv- ja tüpoloogiauuringute jaoks oluline.

2. Põhitähendus vs. sekundaartähendus

Igal grammatikakategoorial on olemas põhitähendus (esmane, primaarne tähendus). “Eesti keele grammatikas” on öeldud: “grammatilist tähendust, mida morfoloogiline tunnus (või tunnuse puudumine) väljendab, nimetatakse selle liikme KATEGORIAALSEKS TÄHENDUSEKS. Kategoriaalseid tähendusi võib ühel tunnusel olla ka mitu, kuid enamasti on neist üks primaarne, teised aga sekundaarsed. Paljudel kategooria liikmetel on selgelt eristuv primaarne tähendus” (EKG I: 45). Eri keelte samanimeliste kategooriate esmane tähendus on sageli sarnane (Gak 1988: 8).

Mihhail Šeljagini (2001: 27) sõnul esineb grammatikakategooriate kasutamise ülekantud tähenduses kontekstides, mille semantika on vastuolus sõnavormi esmase tähendusega. Nii näiteks on *praesens historicum*’i puhul olevikuvorm minevikulises kontekstis ning seetõttu ei väljenda ajavorm enam tegevust olevikus, vaid omandab ülekantud tähenduse ja tähistab olukorda või tegevust minevikus:

- (1) Uus tõus Visnapuu loomingus, eriti ta isamaalauludes ja armastuslüürikas **algab jälle 1940. aastatel.** (EKL)

Sellist nähtust nimetab Aleksandr Bondarko, üks funktsionaalgrammatika loojaid, ajavormide **ülekantud ehk metafoorseks kasutuseks** (*perenosnoe / metaforičeskoe upotreblenie vremennyh form*), tuues välja ajavormide ülekantud tähenduses kasutamise peamise tunnuse: ajavormi grammatiline tähendus säilib, kuid konteksti vastupidise temporaalse semantikaga kokkupuutumise tulemusel muutub ajavormi tegelik tähendus tavaliselt metafoorseks. Sealjuures nii kontekst kui ka ajavorm viivad tegevuse vastavusse tegeliku ajaga, ainult et üks nendest ajalistest suhetest (see, mida väljendab kontekst) on reaalne, teine aga (mida tähistab verbivorm) on kujundlik, metafoorne. Bondarko rõhutab veel, et sellise vormikasutuse kontekstile on sageli iseloomulik eriline ekspressiivsus ja emotsionaalsus, mistõttu ei sobi metafoorne vormikasutus igasugustesse tekstidesse. Eelkõige esineb ülekantud tähenduses ajavorme ilukirjandustekstides ja kõnekeeles. (Bondarko 1971: 170–174)

Vägagi oluline on eristada grammatilist vormi, selle tähendust keele süsteemis ja kasutusvõimalusi tekstis. Šeljakin (2001: 27) põhjendab seda nii: “Sõna vormi

kasutamisel ülekantud tähenduses ei muutu selle grammatiline tähendus, see annab kontekstile kujundlik-grammatilise olemuse ja esineb kui sünonüümne vahend.”

Veel üheks sekundaarkasutuse tüübiks, millest räägitakse funktsionaalgrammatikas, on grammatikakategooria **neutralisatsioon** e üldistusfunktsioon, mis tähendab kategooria liikmete vahelise vastanduse muutumist ebaoluliseks. Näiteks arvukategooria puhul ilmneb see muuhulgas definitsioonide sõnastuses, kus vastandus ‘üks vs. rohkem kui üks’ pole oluline, kuna öeldu kehtib kogu asjade klassi kohta. Edastatav info ei muutu sellest, kas kasutame ainsust või mitmust (2a), (2b).

(2a) Hunt on kiskja.

(2b) Hundid on kiskjad.

3. Grammatikakategooriad ja nende tähendused

Järgnevalt vaadeldgem mõningaid eesti keele grammatikakategooriaid ning nende tähendusi lähemalt.

3.1. Arvukategooria ja pöördekategooria

Arvukategooria puhul võime väita, et selle eesmärk on väljendada eelkõige esemete, nähtuste, isikute jne arvu: ainsuses – üks, mitmuses – rohkem kui üks. “Tähenduslikult eristavad arvukategooria liikmed üht asja (ainsus) kahest või enamast asjast (mitmus).” (EKG I: 61)

(3a) Tal oli käes **üks lilleõis**.

(3b) Tal oli käes **mitu lilleõit**.

Samas on keeltes arvuvorme, mis ei tähenda tegelikult arvu ‘üks’ või ‘mitu’. Tuntud näiteks on *plurale tantum* ja *singulare tantum*-sõnad, kuid leidub muidki kasutusi ning kontekste. Kui näiteks tõlkida eestikeelne lause (4) vene keelde, kasutades samu arvuvorme, tekitame arusaamatuse: milleks panna kaks saabast ühte jalga? Ilmselt võib arusaamatusi tekitada ka vene keelest eesti keelde tõlkides (5a), (5b). Vaatamata vormilisele mitmusele (hüperbool!) on venekeelses lauses tõenäoliselt silmas peetud reaalselt ühte teatriskäiku. Eestikeelsest tõlkest jääb mulje, et tegevus on pigem korduv.

(4) Mul on saapad **jalas**.

(5a) Vy **po teatram** hodite, a my doma sidim.

(5b) Teie käite mööda **teatreid**, aga meie istuge kodus.

Kummalgi juhul ei saa tõlkida teise keelde lihtsalt sama grammatilise vormi abil, nagu näitavad mõned tõlketekstid (vt Kupp-Sazonov 2012: 113–128), kuna tegemist pole tavalise, primaartähendusega, vaid ülekantud tähenduse, metafooriga¹. Niisugune kasutus tekitab raskusi nii tõlkijale kui keeleõppijale, see ei ilmne pelgalt arvuvormide vastandamisel, vaid mingite täiendavate vahendite toel tekstis.

Tegusõnadel esineb arvuvormide ülekantud kasutust teatud pöördevormides. Näiteks on monarhidel ajast aega olnud komme endast rääkides kasutada ainsuse

¹ Metafoor on “sõna või väljendi kasutamine sarnasuse alusel uudses, ülekantud tähenduses; (laiemalt:) igasugune kujundliku sõnatarvituse juht.” (EKSS 2009)

1. pöörde asemel mitmuse 1. pööret (6). Mati Erelt (1990: 37) nimetab seda kasutusviisi *klassikaliseks majesteedimitmuseks*.

- (6) **Meie**, Erik Neljateistkümnnes, Jumala armust rootslaste, gootide ja vendide .. kuningas .. (EK)

Vene teaduskeeles on olnud pikka aega stiilivõtteks mitmuse 1. pöörde kasutamine (ka ühe autori puhul) selleks, et olla tagasihoidlikum ja tunnustada oma eelkäijate panust teadusesse. Ka eesti keeles kasutatakse seda võtet mõnikord, kuigi “keelekorraldajad taunivad autorimitmuse kasutamist kui ebasobivat enesekeergitamise viisi. Mitmuse 1. pööret peetakse loomulikuks vaid siis, kui ta esineb otsetähenduses, st osutab ühteaegu nii autorile kui ka kuulajale.” (Erelt 1990: 37) Seega on eesti teaduskeele seisukohast eelistatum umbisikulise tegumoe kasutamine või viitamine töö autorile ainsuse 3. pöördes ning *meie*-vormi tuleks kasutada vaid mitme autoriga kirjutise puhul.

- (7a) V pervoj glave **rassmotrim**...
(7b) Esimeses peatükis **vaadeldakse**...

“Grammatiline isiku- ja arvukategooria realiseerub morfoloogilise pöörde-kategooriana, millel on kolm ainsuslikku ja kolm mitmuslikku liiget.” (EKG II: 27). **Pöörde-kategooria** võimaldab nii eesti, kui vene keeles sekundaarkasutusi ning paljudel juhtudel need ka kattuvad. Näiteks võib ema oma lapsest rääkides kasutada ainsuse 3. pöörde asemel (tema) hoopis mitmuse 1. pööret (*meie*):

- (8) **Meil** tulevad praegu hambad ja seepärast **tahame** kõike suhu toppida.

Arst võib meilt küsida mitmuse 1. pöördes:

- (9) Ja kuidas **me** ennast täna **tunneme**?

3.2. Ajakategooria

Ajakategoorias on keskne vastandus olevik *vs.* minevik, samas võib leida kontekste, kus on võimalik olevikuvorm vahetada minevikuvormi vastu ilma mõtte sisu muutmata.

- (10a) Suur tükk **ajab** suu lõhki.
(10b) Suur tükk **ajas** suu lõhki: Levada Olümpia vastas asuvad kinnistud sundmüügis. (ST)

Kui arvukategooria puhul võib rääkida lihtsalt eelistusest, näiteks lause (2a) puhul valib kõneleja tõenäoliselt pigem ainsuse vormi, siis ajavormide neutralisatsiooni puhul on olukord pisut keerulisem. Näiteks olevikuvormi asendamine minevikuvormiga toob kaasa lisatähenduse – viite mingile konkreetsele juhtumile, kus tuntud vanasõna kinnitust leidis. Sellest hoolimata võib väita, et väljaöeldud mõte sellest ei muutu, st fakt, et suur tükk ajab suu lõhki, kehtis minevikus mingis konkreetsetes olukorras ja kehtib edasi olevikus. Ka neutralisatsioon vajaks kahtlemata süvendatud uurimist, kuid siinse artikli raames piirdume vaid grammatikakategooriate metafoorkasutuste analüüsimisega.

Sanimeliste grammatikakategooriate puhul kahes keeles ilmneb iga keele eripära just sekundaartähendustes, samas primaartähendused on sageli sarnased. Seetõttu esineb keeleõppijail emakeele häiriv mõju õpitavale keelele e interferents just grammatikakategooria liikme (*resp.* grammatilise tähenduse) sekundaarkasutuses. Siiski peab ütlema, et arvu- ja pöördekategooria kasutuses ei ole vaadeldavates keeltes väga suuri erinevusi; olulisemad on need tegusõna ajakategooria liikmetel.

Hoolimata eesti keele grammatika põhjalikust uuritusest ei ole vormide sekundaarsed kasutusviisid veel piisavalt tähelepanu pälvinud. Selle tõestuseks on kasvõi asjaolu, et siiani ei ole erialakirjanduses kujunenud välja ühte konkreetset oskuskeelendit vaadeldava nähtuse tähistamiseks. Olgu öeldud, et terminit *sekundaarne kasutus* kohtab "Eesti keele grammatika I. Morfoloogia. Sõnamoodustus" tekstis vaid ühe korra: "Imperfekti sekundaarseid kasutusi on a) nn meenutusminevik" (EKG I: 77). Samas kalduvad käesoleva artikli autorid arvama, et täpsem oskuskeelend üldnimetusena niisuguste kasutusjuhtude jaoks oleks *metafoorkasutus* (kuivõrd selle tulemusena tekib vormil metafoorne e kujundlik tähendus), mille alaliikideks oleks ka meenutusminevik, ajalooline olevik jt.

Üks sagedasemaid metafoorkasutusi ongi olevikuvormi paigutamine minevikulisse konteksti, mistõttu preesensivorm omandab metafoorse tähenduse (1). Kui üritada viia metafoori tähendust pelgalt sõna tasemelt ka teistele keeletasanditele, võib öelda, et *praesens historicum*'i korral kasutatakse olevikuvormi mineviku kontekstis kujundlikult, sest selle loomulik (primaarne, esmane) tähendus (väljendada tegevust mitteminevikulisel ajahetkel) muutub vastupidiseks.

Eesti keeles on kindlas kõneviisis neli ajavormi: olevik, lihtminevik, täisminevik ja enneminevik. Tulevikus asetleidva tegevuse väljendamiseks kasutatakse olevikuvormi koos tulevikule viitavate ajamääruste või kontekstiga. Leidub ka mõningaid verbe, mis ei märgi oleviku tegevust, vaid sisaldavad viiteid tulevikule oma leksiikaaltähenduses (*hakkama, algama, saama, tulema* jms). Teiste tegusõnade puhul on olevikuvormil vaja konteksti tuge, et väljendada tulevikku. Eraldi võib arutleda selle üle, kas olevikuvormi kasutamine tulevikulise tegevuse või olukorra väljendamiseks on samuti selle ajavormi ülekantud ehk metafoorkasutus või mitte, kuna futuurumi tähendus tekib konteksti toel.

Igal ajavormil on oma tähendus, keeleõppijale antakse samuti küllaltki konkreetsed juhised, millal millist ajavormi kasutada. Seejuures aga peetakse enamjaolt silmas vaid aegade esmaseid tähendusi ja neidki mitte kõiki. Nagu väidavad Raili Pool ja Helle Metslang (ilmumas), ei käsitleta eesti keele õpikutes tulevikku väljendamist (v.a Roman Kallase 2011. a õpik). Hoolimata eesti keele ajakategooria põhjalikust uurimisest (nt Metslang 1994, 1997), pole nn sekundaarsed kasutused piisavalt tähelepanu pälvinud.

Siiski võib erinevatest eesti keele grammatika käsitlustest leida vähemalt viis erinevat sekundaartähendust (kaks oleviku ja kolm minevikuvormidel), mis siiani ei ole leidnud kohta ühtses ajavormide süsteemis.

Olevikuvormil eristatakse kahte peamist sekundaartähendust.

- 1) ***Praesens historicum*** ehk *ajalooline olevik*: olevikus esinev öeldisverb "võib väljendada ka minevikus asetleidnud [...] tegevust või olemist" (Mihkla, Valmis 1979: 34), st olevikuvorm esineb minevikule viitavas kontekstis.

- (11) 1941. a **mängib** “Vanemuine” edutult Mart Raua draamat “Mõök väravas”, mille tegelasteks on kommunistidest poliitvangid. (EKL)

Keeletüpoloogia ja kontrastiivgrammatika, samuti keeleõppe ja tõlke seisukohast mängib see konkreetne metafoorkasutus eriti olulist rolli. Nimelt on selline olevikuvormi kasutusviis väga populaarne vene ilukirjanduses ja kõnekeeles, seetõttu on oluline välja selgitada, kas *praesens historicum* on mõlemas keeles kasutatav samadel tingimustel. Juba Valter Tauli on raamatus “Eesti keele grammatika II. Lauseõpetus” märkinud, et lihtmineviku vormi asemel ei saa olla olevikuvorm siis, kui verbi laiendab minevikku märkiv väljend nagu *eile, möödunud nädalal, kahe aasta eest* (Tauli 1980: 33). Siiani leitud näited nagu (12b) näivad kinnitavat seda väidet selgelt piiritletud ajalõiku tähistava ajamääruse kasutamisel (nt *möödunud nädalal* tähendab jooksva nädala esmaspäevale eelnevat seitset päeva).

(12a) *Včera v polnoč' prohožu* čerez stolovoju, a tam sveča **gorit**. (AČv)

(12b) *Eile läksin* südaööl läbi söögitoa, seal **põles** küünel. (AČe)

(12c) **Eile lähen* südaööl läbi söögitoa, seal **põleb** küünel.

Samas on eesti keeleuurijate väitel see suulises kõnes võimalik sõnajärje muutmisel (*Lähen eile südaööl...*).²

- 2) **Praesens referens** ehk *refereeriv olevik* – seda “kasutatakse peamiselt kirjandus- ja kunstiteoste sisu ja vormi kirjeldamisel ning seletamisel” (Mihkla, Valmis 1979: 35).

(13a) Tsvetajeva **ei mängi** lugejaga võrdsust, ta **võrdsustab** ennast lugejaga leksikaalselt, loogiliselt.

(13b) Cvetaeva **ne igraet** s čitatelem v ravenstvo: ona sebja k nemu **priravnivaet** leksičeski, logičeski. (JBr)

Vene grammatikates nimetatakse sellist kasutust *eksponeerivaks olevikuks* (*nastojasčee èkspozicionnoe*) (Petruhina 2009: 135). Peale selle on Bondarko toonud välja veel *nastojasčee èmocional'noj aktualizacii* (*emotsionaalselt aktualiseeritud olevik*), millel põhimõtteliselt on ajaloolise minevikuga sama funktsioon, kuid seda iseloomustab veelgi suurem emotsionaalsus ja ekspressiivsus.

(14a) My, sidja v Uzle, sovsem zaplesneveli, a *tut vdruk javljaetsja* soverenno svežij čelovek, s gromadnoj èrudiciej... (DMSv)

(14b) Kükitades siin Uzelis olime juba päris hallitanud, aga **äkki ilmub** meie hulka täiesti värske inimene, tohutu eruditsiooniga... (DMSe)

Šeljakin (2001: 104) kirjeldab ajaloolise oleviku ja emotsionaalselt aktualiseeritud oleviku erinevust: “Ajavormi kasutusviisi osas ei erine emotsionaalselt aktualiseeritud olevik mitte millegi poolest ajaloolisest olevikust, kuid sel on stilistiline eripära: see ei ole näitlik, vaid väljendab kõneleja subjektiivset hinnangulist suhtumist erinevatesse mineviku faktidesse.”

Praesens historicum’i on kirjeldatud ka näiteks saksa, hispaania jt keeltes.

Minevikuvormil võib välja tuua järgmised kasutusviisid.

- 1) **Meenutusminevik:** “tuletades meelde minevikus teadasaadut kasutatakse minevikku vaatamata sellele, et öeldu võib sisuliselt kehtida olevikuski” (EKG I: 77).

(15a) Kas Teie nimi **oli** Tamm?

(15b) *Vaša familija **byla** Tamm?

Vene keele minevikuvormil kirjakeeles seda sekundaartähendust ei ole, lauset (15b) saab mõista vaid ühtmoodi: kõnetatu perekonnanimi on küsimise hetkeks muutunud. Kui inimene pole sekundaartähendusest teadlik, siis võib selline küsimus eesti keeles põhjustada mitte-emakeelekõneleja jaoks segadust: küsija soovib kontrollida oma teadmist, vastajale aga võib jääda mulje, et küsimus puudutab tema eelmist perekonnanime.

Esimesed analüüsitulemused näitavad, et tõenäoliselt võimaldavad sellist kasutusviisi ka teised tegusõnad. Järgnev näide (16) on pärit ühest tavavestlusest, osapool A räägib eesti keelt emakeelena, teine osaleja B on kakskeelne (eesti ja vene).

(16) A: Kas Siim elab nüüd Prantsusmaal?

B: Jah, juba aasta.

A: Oot, mida ta seal **õppis**, õigusteadust?

B: Ta õpib praegu ka – jah, õigusteadust.

Olukord on sarnane näitelausestega (15a), (15b). Vestluses osaleja A soovib vaid Siimu puudutatavat informatsiooni kontrollida, samal ajal kui vestluspartner B mõistab tema küsimust hoopis teisiti ning seetõttu parandab A-d.

- 2) **Tuleviku minevik:** “täismineviku vormi kasutatakse ka siis, kui kõnehetkele järgnevat tegevust vaadeldakse mingi veel hilisema vaatlushetke seisukohalt (tuleviku minevikus)” (EKK: 279), st minevikuvormi kasutatakse tuleviku kontekstis.

(17a) Loodame, et pühapäeva õhtuks **oled** sa selle töö **lõpetanud**.

(17b) Budem nadejat'sja, čto k večeru ty **zakončis'** ètu rabotu.

Vene keeles on näites (17b) tegu perfektivse aspekti tulevikuvormiga selle primaartähenduses.

Kõrvutagem nüüd eesti keele võimalusi minevikuvormi kasutamiselega ülekantud tähenduses vene keeles.

- 3) **Kujutletav tulevik** (*buduščee voobražaemoe* Bondarko järgi): perfektivse minevikuvormi kasutamine tuleviku kontekstis (18a). Eestikeelses tõlkes ei ole vasteks minevikuvorm, vaid tulevikku väljendav olevikuvorm (18b). Siiski võiks ka eestikeelses lauses olla verb liht- või täisminevikus (18c).

(18a) Nu, položim, ja sdelaju usilie, sdelaju èto. Ili ja poluču oskorbitel'nyj otvet, ili soglasie. Horošo, **polučila** soglasie... [---] Ja poluču soglasie, a sy... syn? (LTv)

(18b) Noh, oletame, et ma sunnin ennast seda tegema, ma teen seda. Ma saan kas eitava või jaatava vastuse. Hästi, ma **saan** nõusoleku... [---] Ma saan nõusoleku, aga poeg... poeg? (LTE)

(18c) Ma saan kas eitava või jaatava vastuse. Hästi, ma **sain / olen saanud** nõusoleku...

Analoogiline metafoorkasutus on võimalik ka vene tingimuskõrvallausega lauses, kuid see pole stiililt neutraalne, vaid ekspressiivne (Bondarko 1971: 132).

(19a) Esli zavtra ne budet doždja – urožaj **pogib**. (ATv)

(19b) Kui homme vihma ei tule, **hävib** vili / on vili hävinud. (ATe)

Nii eesti kui vene kõnekeeles esineb ka mõningate verbide ekspressiivset lihtminevikuvormi kasutust, nt kavatsusena: *Noh, ma siis **läksin!** Nu, ja **pošël!***

Kuna artikli maht ei võimalda kõikide metafoorkasutuste käsitlemist, siis piirdume eelpool loetletuga. Käsitletud metafoorkasutused võtab kokku tabel 1, lisatud on vastavate näidete numbrid.

Tabel 1. Ajavormide sekundaartähendused eesti ja vene keeles

Sekundaartähendus	Eesti keel	Vene keel
<i>Praesens historicum</i> (olevikuvorm mineviku kontekstis)	+ (11)	+ (sh emotsionaalselt aktualiseeritud olevik) (12a)
<i>Praesens referens</i> / eksponeeriv olevik (olevikuvorm mineviku kontekstis)	+ (13a)	+ (13b)
Meenusminevik	+ (lihtminevikuvorm oleviku kontekstis) (15a)	–
Tuleviku minevik	+ (täisminevikuvorm tuleviku kontekstis) (17a)	–
Kujutletav tulevik	+ (täis- või lihtminevikuvorm tuleviku kontekstis) (18c)	+ (perfektiivse aspekti minevikuvorm tuleviku kontekstis) (18a)

4. Terminoloogia

Eraldi tähelepanu väärib ka see, kuidas kirjeldatud keelenähtusi eesti keeles nimetada, nii et terminite süsteem oleks piisav tüpoloogiliste ja kontrastiivsete uuringute jaoks ning ühtne muude keelte terminitega. Loomulikult tuleb enne iga termini loomist ja kasutuselevõttu küsida, kas keeles on üldse vajadust uue oskussõna järele.

Olukord oleks erinev, kui eesti keeles ajavormide kasutamist ülekantud tähenduses üldse ei esineks, sellisel juhul puuduks ka otsene vajadus oma oskuskeelendi

järele. Nii kasutavad Triin Lõbus ja Silvi Tenjes (2010: 157) terminit *narratiivne imperfekt* ning lisavad, et “erinevalt eesti grammatikatradsioonist, kus terminit *imperfekt* kasutatakse üldise lihtmineviku tähenduses, mõistame käesolevas artiklis selle all hispaania keele imperfektiivse aspektitähendusega minevikuvormi [---]. Narratiivseks imperfektiks nimetatakse romaani keeltes narratiivses diskursuses esinevat eripärast aspektikasutust, mille puhul imperfektiivset verbivormi kasutatakse selliste sündmuste kujutamiseks, mille ajalis-aspektiline iseloom eeldaks perfektiivset aspekti.” Analüüsiv nähtus esineb romaani keeltes, seega pole eestikeelse termini olemasolu ilmtingimata vajalik, kontrastiivuuringutes on võimalik kasutada hispaania terminile vastavat võõrsõna. Ajavormide kasutamine ülekantud tähenduses on aga igati omane ka eesti keelele, seega on vajadus omakeelse termini järele tunduvalt rohkem põhjendatud.

Tegelikult vajaksime koguni kahte oskuskeelendit: üks tähistaks ajavormi kasutamist selle esmases (primaarses, põhilises, prototüüpses) tähenduses (nt kui olevikuvorm märgib sündmuste toimumist kõnehetke suhtes mitteminevikulisel ajahetkel) ja teine sekundaartähendust (nt *praesens historicum*).

Vaadeldes, millised nimetused on sekundaarkasutuste kohta kasutusel eesti keelt käsitlevas grammatikakirjanduses (teosed on esitatud kronoloogilises järjestuses).

Venekeelne eesti ja vene keele kõrvutatav grammatika (1962): üldtermin puudub, juttu on *praesens historicum*'i tähendusest – olevikuvormi kasutamisest “v značeni nastojaščego istoričeskogo” (Päll jt 1962: 155). Minevikuvormide võimalikke metafoorkasutusi ei ole käsitletud.

“Eesti keele süntaksi küsimusi” (1963): Boriss Serebrennikov (1963: 444–449) kirjutab olevikuvormide tüüpidest, mille hulgas on eraldi välja toodud *praesens historicum*, *praesens tabulare* jms. Lisaks eristab ta veel variante: *osnovnoe značenie* (põhitähendus) ja *derivatnoe značenie* (tuletatud tähendus), pidades viimase all silmas näiteks olevikuvormiga mingi objekti alalise omaduse tähistamist.

“Eesti keele lauseõpetuse põhijooned I” (1974): kasutatakse sõna *põhitähendus*: “Peale oma põhitähenduse võib olevikus esinev öeldisverb väljendada olendile, esemele või nähtusele püsivalt iseloomulikku tegevust, seisundit või tunnust” (Mihkla jt 1974: 91). Kuigi autorid ei tähista olevikuvormi mittepõhilisi kasutusviise ühegi oskussõnaga, võib eeldada, et terminile *põhitähendus* võiks vastandada nt *sekundaartähenduse*, *kõrvaltähenduse* jms.

“Eesti keele süntaks kõrgkoolidele” (1979): sarnaselt eelmisele käsitlusele on ka siin kasutusel *põhitähendus*: “Peale põhitähenduse võib olevikus öeldisverb väljendada ka minevikus asetleidnud või tulevikus toimuvat tegevust või olemist” (Mihkla, Valmis 1979: 34).

Teistest erinev ja teatud määral keerulisem on olukord Tauli grammatikaraamatus “Eesti grammatika II. Lauseõpetus” (1980): autor kasutab läbivalt, ka muudes kontekstides kui verbist rääkides, sõna *esinemine*. Seega ei saa väita, et tegu oleks terminiga, pigem tähendab see Tauli jaoks sama mis kasutamine, tähistades lihtsalt erinevaid kasutusnäiteid: “Esinemine: (1) kui tegevusaega ei väljendata või tegevus toimub alati [---]” (Tauli 1980: 32).

“Eesti keele grammatika I. Morfologia. Sõnamoodustus” (1995) annab eelmainitud *sekundaarse kasutuse* (EKG I: 77). Tähelepanu väärib asjaolu, et kõnealust terminit on kasutatud vaid üks kord (imperfekti puhul), kuigi ka olevikuvormil on eesti keeles oma sekundaartähendused täiesti olemas.

Uuematest käsitlustest tuleks nimetada veel näiteks Pille Esloni, Katre Öimu, Annekatrin Kaivapalu, Reili Arguse ja Erika Matsaki ühisartiklit “Kuidas uurida esimese ja teise keele omandamist?”, kus autorid viitavad kõnealusele nähtusele järgneval viisil:

Sõnade ja väljendite mõistestamisel tugineb inimene suuresti oma kogemusele ning sellest tulenevalt metafoorsele, metonüümsele vm laadi analoogiapõhisele mõtlemisviisile, mis motiveerib meie keelekasutust oluliselt suuremal määral, kui arvatud. See ilmneb sõnade, vormide ja süntaksistruktuuride kasutamises ülekantud tähenduses (fraseologismid, registrierinevustega seotud vormivariatiivsus jne) või neile **mitteomases funktsioonis** [siinne esiletõst] (nt *praesens historicum* või verbi 2. pöörde üldistatud kasutamine kelle tahes suhtes, kes situatsiooni kogevad: *See kordumatu igaviku tunne, kui istud niisugustel sügisõhtutel kodusoojas ja kuulad raugeva tormi müha!*). (Eslon jt 2010: 18)

Siinkohal tahaksime juhtida tähelepanu väljendile *mitteomases funktsioonis*, sest seda pole kõnealuses artiklis pikemalt lahti seletatud. Kuna autorid lisavad, et “Neid nähtusi ei saa pidada anomaaliaks: tegu on keelele omase iseorganiseerumisvõime ilmnemisega, tänu millele on omavahel tasakaalus keele säilimine ja pidev tsükliline uuenemine” (Eslon jt 2010: 18), siis tekib küsimus: kui keel võimaldab sellist kasutusviisi, kas seda saab nimetada *mitteomaseks* või on tegu pigem lihtsalt tavakasutusest erineva nähtusega?

Vene keele grammatikas ongi kasutusel keelend *perenosnoe / metaforičeskoe upotreblenie vremennyh form* ehk ‘ajavormide ülekantud / metafoorne kasutus’ (vt lähemalt Bondarko 1971: 129–237, Bondarko 2002: 191, Šeljakin 2001: 98–111, Knjazev 2007: 79). Eelmainitud artiklis kasutavad autorid sõnaühendit *ülekantud tähenduses kasutamine* fraseologismide ja vormivariatiivsuse kohta. Ühes varasemas Esloni artiklis (2006: 365) esineb termin *ülekantud tähenduses kasutamine* seoses aspektisituatsioonidega vene keeles.

Vaadeldagem lähemalt ka ingliskeelsetes käsitlustes leiduvaid nimetusi.

Östen Dahl vastandab omavahel kahte nähtust: *prototypical use* (prototüüpne, põhiline kasutamine) ja *secondary use* (sekundaarne kasutamine): “The prototypical uses of FUT that we have cited have clear future time reference. It is well known that many categories called ‘Future’ can also be used for ‘predictions’ or ‘inferences’ about non-future states of affair. It is reasonable, however, to regard this as a secondary use of FUT categories, which preserves one element of the prototypical uses, viz. that of prediction (in the wide sense of that word) although leaving out the others.” (Dahl 1985: 108) Niisiis on tuleviku prototüüpsetel kasutusviisidel selge viide tulevikule, aga futuurumi sekundaarne kasutamine tähendab, et säilib üks prototüüpse kasutuse elementidest, ning teised jäävad välja.

Bernard Comrie (1995: 43) eelistab sõnaühendit *normal usage*: “The usages noted so far in this section have typically shifted the tense (e.g. present for past), and either shifted aspect in accordance with normal usage in the present, or retained aspect in accordance with the distinction between description and narration.” Comrie puudub sõna tähistamiseks ajaloolise oleviku laadset ajavormi kasutamist, ta viitab vaid sellele, et tegu on teatud nihkega tavalisest kasutamisest.

Renaat Declerck (2006: 27) kasutab kahte väljendit: *special uses of tenses* (ajavormide erikasutused) ja *metaphorical uses* (metafoorsed kasutused): “Some tenses have one or more metaphorical uses: they are used to represent a particular time as if it were another time. The present tense, for example, can be used to refer not only to the present, but also to the past and to the future. For example, in narration we often find a switch from the past tense to the ‘historic present’.” Seega väidab Declerck, et mõnedel ajavormidel on üks või mitu metafoorset kasutusvõimalust, mille käigus need peavad väljendama mingit aega nii, nagu oleks tegu hoopis teise ajaga, ehk siis olevikuvormiga saab väljendada tegevust minevikus või tulevikus.

Tähelepanuväärne on see, et näiteks James Forsyth ei pea *praesens historicum*’i sugugi olevikuvormi sekundaarseks ega ka mitte metafoorseks kasutamiseks, vaid liigitab ajaloolise oleviku preesensivormi põhitähenduste (*basic meanings*) hulka: “The basic meanings are: (a) the ‘real’ or actual present [---] (c) the historic present, used in narrating past events” (Forsyth 1970: 147). Kõike, mis ei mahu põhitähenduste alla, nimetab ta *eccentric usage* (iseäralikud kasutusviisid): “‘Eccentric’ usages of present-future forms in the ‘real’ mode are few. As in many languages, including English, the (imperfective) present tense can be used to refer to a future action [---]” (Forsyth 1970: 149). Forsythi hinnangul on vene olevikuvormi kasutamine tuleviku tähenduses just selline iseäralik kasutusvariant.

Suzanne Fleischman, kes on põhjalikult uurinud ajavormide kasutamist narratiivis, eelistab rääkida sellistest nähtustest nagu *basic tense function* (ajavormi põhifunktsioon) ja *not basic tense function*: “What I propose to demonstrate here is that the function of tense forms IN NARRATIVE is frequently not the basic tense function of locating events relative to the moment of speech (or an alternative reference point)” (Fleischman 1985: 852). Jutustavat olevikku nimetab ta oleviku variandiks (*variety*): “the term narrative present (= NP) will be used to designate that variety of present” (Fleischman 1985: 867). Sünonüümiks ‘põhifunktsioonile’ on tema töödes veel ‘põhitähendus’ (*basic meaning*): “grammatical meanings, including the so-called ‘basic’ meanings or functions of tense and aspect” (Fleischman 1991: 77).

Loomulikult ei ole siin loetletud võimalike terminite nimekiri ammendav, kindlasti võib grammatikakäsitlustest leida veel teisigi variante. Kuid kuna artikli eesmärk ei ole esitada lõplikku loetelu kõikidest võimalikest oskuskeelenditest, siis piirdume eeltooduga. Juba see küllaltki lai terminite valik kinnitab veelkord veendumust, et vajadus ühe konkreetse termini järele on eesti keele grammatika kirjeldamisel olemas.

5. Kokkuvõte

Kuigi ajavormide kasutamine (sh ülekantud tähenduses) eesti ja vene keeles sajabprotsendiliselt ei kattu, näib siiski, et ka eesti keeles on rohkem aegade metafoorkasutusi, kui võiks esmapilgul arvata. Grammatiline metafoor on süsteemne nähtus, mis tugineb grammatilise vormi semantilisel potentsiaalil ning vajaks eesti grammatikas veel uurimist. Meie arvates võimaldaks ja hõlbustaks grammatiliste tähenduste struktuuri põhjalikum vaatlus kahe keele kõrvutamist ning tüpologiauuringuid. Ühtlasi võimaldab ülekantud kasutuste tundmine ka sellist keele õpetamist, kus vaadeldakse esmalt põhitähendusi ning seejärel keeleõppe

kõrgtasemel ja tõlkeõpetuses käsitletakse juba ka grammatikakategooriate metafoorseid funktsioone (vt näited 15a, 15b ja 16).

Terminoloogia seisukohast on meil vaja leida vastus kahele põhiküsimusele.

- 1) Kas eesti keele grammatikas on vaja kõnealuse nähtuse tähistamiseks terminit? Vastamisel tasuks lähtuda Tiiu Ereli (2007: 69) püstitatud küsimusest: “kas mõiste on erialal nii oluline, käibiv või iseseisev, et üldse vajab nimetamist, või piisab kirjeldusest?” Meie hinnagul on tegu piisavalt olulise ja iseseisva nähtusega. Selle argumendi kasuks räägib juba seegi, et oleme siin tutvustanud vähemalt viit ajavormide kasutamisevõimalust ülekantud tähenduses (kaks oleviku- ja kolm minevikuvormide puhul), seega ei saa väita, et tegu oleks marginaalse nähtusega, mille tähistamiseks me terminit ei vaja.
- 2) Kui leiame, et vajadus termini järele selle grammatilise nähtuse tähistamise jaoks on olemas, samas aga kinnistunud oskuskeelend eesti keele grammatikas puudub, siis on vaja jõuda selgusele, kas peaksime välja mõtlema täiesti uue termini, võtma kasutusele mõne juba kirjanduses esinenud variandi või võime laenata selle ka mõne teise keele grammatikast.

Põhimõtteliselt võiks kasutada “Eesti keele grammatikas” esinevat sõna *sekundaarkasutus*, kuid võib tekkida küsimus, kas see termin annaks täies mahus edasi seda, mis tegelikult ajavormi ja selle tähendusega juhtub, kui näiteks olevikuvorm satub mineviku konteksti vms. Kui kasutada olevikku minevikulises kontekstis, siis võib öelda, et preesensivorm omandab metafoorse tähenduse. Metafoor võib esineda mitte pelgalt sõna tasemel, vaid ka teistel keele tasanditel, nimelt grammatikas (Metslang 2001: 222–231), nt kui ajavormi kasutatakse talle mitteoomases kontekstis.

Veel ühe argumendina selle poolt, et *metafoorkasutus* oleks veidi täpsem termin kõnealuse nähtuse tähistamiseks, võib võrrelda muid olukordi, kus räägitakse samuti sekundaarkasutustest. Näiteks võib väita, et küsilause sekundaarkasutuseks on retooriline küsilause (*Kes siis tänapäeval enam postkaarte saadab?*), millele selle esitaja ei ootagi vastust, kuid ometi ei kaota retooriline küsimus täies ulatuses küsimuse esmast funktsiooni ja tähendust, sest soovi korral võib ka retoorilisele küsimusele siiski vastata. Olevikuvorm minevikulises kontekstis või minevikuvorm oleviku kontekstis kaotab aga täielikult oma nn esmase funktsiooni ja omandab sõltuvalt ümbritsevast kontekstist sootuks vastupidise tähenduse.

Kahtlemata üheks võimaluseks, mida tasub kaaluda, on ka sobiva termini n-ö laenamine. Heaks näiteks on perfektitüübid, millele juhtis tähelepanu Comrie (1976: 56–61) ning mille Helle Metslang ja Mati Ereli edukalt ka eesti keele grammatikas kasutusele võtsid (Metslang 1997: 102–106, Ereli, Metslang 1998: 665–667). Artikli autoritele tundub, et kõige paremini annaks ajaloolise oleviku, meenusmineviku jms olemust edasi termin, mis sisaldaks viidet kasutusele ülekantud tähenduses, seejuures ülekande tüübi täpsem eritlus oleks juba edasise uurimise ülesanne. Seetõttu tuleks meie analüüsist lähtudes kõne alla venekeelsele terminile analoogiline oskuskeelend *metafoorkasutus*, mis võimaldaks mõista ajavormide ülekantud tähenduses kasutamise olemust.

Materjali allikad

- AČe = Tšehhov, Anton 2006. Näidendid. Tõlk. Ernst Raudsepp, Otto Samma. Tallinn: Avita.
- AČv = Tšehhov, Anton 1982. P'jesy. Moskva.
- ATe = Tolstoi, Aleksei 1974. Nikita lapsepõlv. Tõlk. Jüri Piik. Tallinn: Eesti Raamat.
- ATv = Tolstoj, Aleksej 1983. Detstvo Nikity. Minsk: Junactva.
- DMSe = Mamin-Sibirjak, Dmitri 1952. Privalovite miljonid. Tõlk. Nelly Toiger. Tallinn: Eesti Riiklik Kirjastus.
- DMSv = Mamin-Sibirjak, Dmitrij 1958. Privalovskie milliony. Orel.
- EK = Küng, Enn. 2011. "Paludes kristlikku päästmist ja kaitset": Eestimaa seisuste alistumine Rootsi võimule 1561. aasta suvel. – Tuna, 3, 88–98.
- EKL = Annus, Epp; Epner, Luule; Järv, Ants; Olesk, Sirje; Süvalep, Ele; Velsker Mart 2001. Eesti kirjanduslugu. Tallinn: Koolibri.
- JBr = Brodski, Jossif 1992. Stihotvorenija, èsse, p'jesy. Minsk: Èridan.
- LTe = Tolstoi, Lev 1978. Anna Karenina. Tõlk. Selma Holberg. Tallinn: Eesti Raamat.
- LTV = Tolstoj, Lev 1976. Anna Karenina. Moskva: Hudožestvennaja literatura.
- ST = Suur tükk ajas suu lõhki: Levada Olümpia vastas asuvad kinnistud sundmüügis. – Ärileht, 7.11.2012. <http://arileht.delfi.ee/news/uudised/suur-tuk-ajas-suu-lohki-levada-olumpia-vastas-asuvad-kinnistud-sundmuugis.d?id=65226732> (4.10.2013).

Viidatud kirjandus

- Bondarko, Aleksandr 1971. Vid i vremja russkogo glagola. [Aspect and Tense of the Russian Verb.] Moskva.
- Bondarko, Aleksandr 2002. Teorija značenija v sisteme funkcional'noj grammatiki. [The Theory of Meaning in the System of Functional Grammar.] Moskva.
- Comrie, Bernard 1976. Aspect. An Introduction to the Study of Verbal Aspect and Related Problems. Cambridge: Cambridge University Press.
- Comrie, Bernard 1995. Aspect. An Introduction to the Study of Verbal Aspect and Related Problems. Cambridge: Cambridge University Press.
- Dahl, Östen 1985. Tense and Aspect Systems. Blackwell Publishers.
- Dahl, Östen 2007. From Questionnaires to Parallel Corpora in Typology. – Sprachtypologie und Universalienforschung, 60, 172–181.
- Dahl, Östen 2010. Typology of negation. – Laurence R. Horn (Ed.). The Expression of Negation. Berlin: De Gruyter Mouton, 9–38.
- Östen Dahl; Koptjevskaja-Tamm, Maria (Eds.) 2001. Circum-Baltic Languages. Vol. 1: Past and Present. Studies in Language Companion Series 54. Amsterdam: Benjamins.
- Declerck, Renaat 2006. The Grammar of the English Verb Phrase. Vol. 1: The Grammar of the English Tense System. Berlin, New York: Mouton de Gruyter.
- 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. [The Grammar of the Estonian Language I: Morphology. Word-formation.] Eesti Teaduste Akadeemia Eesti Keele Instituut. Tallinn.
- EKG II = Erelt, Mati; Kasik, Reet; Metslang, Helle; Rajandi, Henno; Ross, Kristiina; Saari, Henn; Tael, Kaja; Vare, Silvi 1993. Eesti keele grammatika II. Süntaks. [The Grammar of the Estonian Language II: Syntax.] Eesti Teaduste Akadeemia Keele ja Kirjanduse Instituut. Tallinn.
- EKK = Erelt, Mati; Erelt, Tiiu; Ross, Kristiina 2007. Eesti keele käsiraamat. [The Handbook of Estonian.] Tallinn: Eesti Keele Sihtasutus.
- EKSS = Langemets, Margit; Tiits, Mai; Valdre, Tiia; Veski, Leidi; Viks, Ülle; Voll, Piret (Toim.) 2009. Eesti keele seletav sõnaraamat 1–6. [The Explanatory Dictionary of Estonian.] Tallinn: Eesti Keele Sihtasutus.

- Erelt, Mati 1990. Kõneleja ja kuulaja kaudse väljendamise võimalusi eesti keeles. [The possibilities of expressing the speaker and the listener in an indirect way in Estonian.] – Keel ja Kirjandus, 1, 35–39.
- Erelt, Mati (Ed.) 1996. Estonian: typological studies I. Publications of the Department of Estonian of the University of Tartu 4. Tartu: Tartu University Press.
- Erelt, Mati (Ed.) 1997. Estonian: typological studies II. Publications of the Department of Estonian of the University of Tartu 8. Tartu: Tartu University Press.
- Erelt, Mati (Ed.) 1999. Estonian: typological studies III. Publications of the Department of Estonian of the University of Tartu 11. Tartu: Tartu University Press.
- Erelt, Mati (Ed.) 2000. Estonian: typological studies IV. Publications of the Department of Estonian of the University of Tartu 14. Tartu: Tartu University Press.
- Erelt, Mati (Ed.) 2001. Estonian: typological studies I. Publications of the Department of Estonian of the University of Tartu 18. Tartu: Tartu University Press.
- Erelt, Mati; Metslang, Helle 1998. Oma või võõras? [Our own or foreign?] – Keel ja Kirjandus, 10, 657–668.
- Erelt, Tiiu 2007. Terminõpetus. [Terminology.] Tartu: Tartu Ülikooli Kirjastus.
- Eslon, Pille 2006. Deiktikud ja aspektisituatsioonid. – Regards multidisciplinares sur la deixis: de l'énoncé à l'énonciation et vice-versa = Multidisciplinary Views on Deixis: from Utterance to Uttering and Vice Versa = Multidistsiplinaarsed vaated deiksi-sele: lausungist lausumiseni ja vastupidi. Kõide 4b. Tartu: Tartu Ülikooli Kirjastus, 363–382.
- Eslon, Pille; Õim, Katre; Kaivapalu, Annekatrin; Argus, Reili; Matsak, Erika 2010. Kuidas uurida esimese ja teise keele omandamist? [Perspectives of the research of the first and of the second language acquisition.] – Lähivõrdlusi. Lähivertailuja, 20, 11–48. <http://dx.doi.org/10.5128/LV20.01>
- Fleischman, Suzanne 1985. Discourse functions of tense-aspect oppositions in narrative: Towards a theory of grounding. – Linguistics, 23 (6), 851–882. <http://dx.doi.org/10.1515/ling.1985.23.6.851>
- Fleischman, Suzanne 1991. Toward a theory of tense-aspect in narrative discourse. – Jadranka Gvozdanovic, Theo A. J. M. Janssen (Eds.). The Function of Tense in Texts. Amsterdam etc.: North-Holland, 75–97.
- Forsyth, James 1970. A Grammar of Aspect: Usage and Meaning in the Russian Verb. Cambridge: Cambridge University Press.
- Gak, Vladimir 1988. Russkij jazyk v sopostavlenii s francuzskim. [Russian Language in Comparison with the French.] Moskva.
- Hrakovskij, Viktor 1985. Tipologija konstrukcij s predikatnymi aktantami. [Typology of Constructions with Clause Arguments.] Leningrad.
- Hrakovskij, Viktor 1989. Tipologia iterativnyh konstrukcij. [Typology of Iterative Constructions.] Leningrad.
- Hrakovskij, Viktor 1992. Tipologija imperativnyh konstrukcij. [Typology of Imperative Constructions.] Sankt-Peterburg.
- Hrakovskij, Viktor 1998. Tipologija uslovnyh konstrukcij. [Typology of Conditional Constructions.] Sankt-Peterburg.
- Hrakovskij, Viktor 2004. Tipologija ustupitel'nyh konstrukcij. [Typology of Concessive Constructions.] Sankt-Peterburg.
- Knjazev, Jurij 2007. Grammatičeskaja semantika. Russkij jazyk v tipologičeskoj perspektive. [Grammatical Semantics. Russian in the Typological Perspective.] Moskva.
- Kupp-Sazonov, Sirje 2012. Vene-eesti tõlgete mõjust eesti keele minevikuvormide kasutamisele B. Akunini mõningate teoste näitel. [Exploring the influence of Russian-Estonian translations on the use of Estonian past tense forms in Boris Akunin's works.] – Eesti ja soome-ugri keeleteaduse ajakiri. The Journal of Estonian and Finno-Ugric Linguistics, ESUKA / JEFUL, 3 (2), 113–128.

- Kupp-Sazonov, Sirje; Külmoja, Irina 2013. Eesti keele grammatikakategooriatest keeletüpo-
loogia ja kontrastiivgrammatika aspektist. – Ettekanne 12. rakenduslingvistika kevad-
konverentsil 18.-19.4.2013 Tallinnas.
- Lõbus, Triin; Tenjes, Silvi 2010. Narratiivi aeg ja tõlke aeg: diskursiivne käsitlus. [Time of
narrative and time of translation: A discursive approach.] – Eesti Rakenduslingvistika
Ühingu aastaraamat, 6, 157–174. <http://dx.doi.org/10.5128/ERYa6.10>
- Metslang, Helle 1994. Temporal Relations in the Predicate and the Grammatical System of
Estonian and Finnish. Oulu: Oulun Ylipisto.
- Metslang, Helle 1997. On the use of the Estonian past tense forms through the last one hundred
years. – Mati Ereht (Ed.). Estonian: typological studies II. Publications of the Depart-
ment of Estonian of the University of Tartu 8. Tartu: Tartu University Press, 98–145.
- Metslang, Helle 2001. Metafooridest grammatikas. [About grammatical metaphors.] – Reet
Kasik (Toim.). Keele kannul: pühendusteos Mati Erehti 60. sünnipäevaks. Tartu
Ülikooli eesti keele õppetooli toimetised 17. Tartu: Tartu Ülikooli Kirjastus, 222–231.
- Mihkla, Karl; Rannut, Lehte; Riikoja, Elli; Admann, Aino 1974. Eesti keele lauseõpetuse
põhijooned I. Lihtlause. [Basic Characteristics of Estonian Syntax I. Simple Clause.]
Tallinn: Valgus.
- Mihkla, Karl; Valmis, Aavo 1979. Eesti keele süntaks kõrgkoolidele. [Estonian Syntax for
Universities.] Tallinn: Valgus.
- Petruhina, Elena 2009. Russkij glagol: kategorija vida i vremeni (v kontekste sovremennyh
lingvističeskikh issledovanij). [Russian Verb: Categories of Aspect and Tense (in the
Context of Contemporary Linguistic Studies).] Moskva.
- Pool, Raili; Metslang, Helle (ilmumas 2014). Jazykovye sredstva peredači vremennogo plana
buduščego v rabotah russkikh studentov na èstonskom jazyke. [Ways of Expressing
Future Tense in Russian-Speaking Students' Written Papers.] – Acta Slavica Estonica.
(Trudy po russkoj i slavjanskoj filologii. Lingvistika, V (XVI).
- Päll jt 1962 = Pjall', Èduard; Totcel', Èliza; Tukumcev, Gvidon. Sopostavitel'naja grammatika
èstonskogo i russkogo jazyka. [Comparative Grammar of Estonian and Russian.] Tal-
lin: Èstonskoe gosudarstvennoe izdatel'stvo.
- Russkaja korpusnaja grammatika. [Russian Corpus Grammar.] <http://rusgram.ru/Лицо>
(4.10.2013).
- Serebrennikov, Boris 1963. Kategorija vremeni v pribaltijsko-finskikh jazykah. – Eesti keele
süntaksi küsimusi. [Questions about Estonian Syntax.] Keele ja Kirjanduse Instituudi
uurimused 8. Tallinn: Eesti Riiklik Kirjastus, 426–511.
- Šeljakin, Mihail 2001. Funkcional'naja grammatika russkogo jazyka. [Functional Grammar
of Russian.] Moskva.
- Tauli, Valter 1980. Eesti grammatika II. Lauseõpetus. [Estonian Grammar II. Syntax.]
Uppsala: Finsk-ugriska institutionen.

Sirje Kupp-Sazonovi (Tartu Ülikool) peamised uurimisvaldkonnad on vene-eesti kontrastiivne
grammatika, vene-eesti-vene tõlkeküsimused.

Ülikooli 17-120, 51014 Tartu, Estonia

Sirje.Kupp-Sazonov@ut.ee

Irina Külmoja (Tartu Ülikool) peamised uurimisvaldkonnad: vene keele funktsionaalgrammatika,
vene-eesti kontrastiivne grammatika, eesti keele tüpoloogiuuringud, keelekontaktid, vene keel Eestis,
Eesti vanausuliste keel ja kultuur.

Lossi 3-228, 50090 Tartu, Estonia

irina.kulmoja@ut.ee

ON SECONDARY USE OF GRAMMATICAL CATEGORIES: RUSSIAN AND ESTONIAN TENSE FORMS

Sirje Kupp-Sazonov, Irina Külmoja

University of Tartu

As typological and contrastive studies of languages are becoming increasingly more important in linguistics, the need for generalised systems of grammatical categories applicable to different languages and described by a unified unambiguous terminology is acutely felt. The article focuses on secondary use of grammatical categories in Estonian and Russian. It is noted that while primary meanings of Estonian grammatical categories have already received thorough treatment, their secondary use requires further elaboration. Numerous examples of secondary use of both Estonian and Russian grammatical categories (such as number and, in particular, tense) are presented and analysed. It is shown that secondary use of grammatical categories cannot be considered marginal in either language, nor does it completely overlap in them, thus being a substantial potential source of negative transfer in language learning as well as translation and interpretation. Further contrastive linguistic study of the phenomenon is clearly required. An attempt is therefore made, proceeding from an examination of the varied terms used to designate the phenomenon in pertinent linguistic literature, to suggest a term most suitable for Estonian grammar.

Keywords: metaphor, tense, terminology, Estonian, Russian

KÜSIMUSED EESTI LAPSELE SUUNATUD KÕNES LEEDU JA VENE KEELE TAUSTAL

Helen Kõrgesaar

Ülevaade. Eesti hoidjakeele uurimine on valdkond, millega on põhjalikumalt tegelema hakatud alles viimasel kümnendil, seetõttu leidub hoidjakeele tasandeid, mille kohta info on napp või puudub üldse. Siinne artikkel annab esmalt ülevaate uurimusest, mis vaatleb eesti hoidjakeeles kasutatavaid küsilausungeid teiste keelte (vene, leedu, aga ka prantsuse ja Austria saksa) taustal. Seejärel kirjeldatakse kaht tulemusi täpsustavat kontrolluuringut. Artiklist selgub ka, kuidas jagunevad küsilausungid eesti, vene ja leedu hoidjakeeles nende positsioonist, pragmaatilisest rollist, funktsioonist ja struktuurist lähtuvalt.

Võtmesõnad: keeleomandamine, pragmaatika, eesti keel, vene keel, leedu keel

1. Sissejuhatus

Viimase kümne aasta jooksul on eesti laste keeleomandamise uurimise juures hakatud üha suuremat tähelepanu pöörama ka suhtluse teisele poolele – lastele suunatud kõnele ehk hoidjakeelele. Hoidjakeel on dünaamiline kõneregister, mida kasutab täiskasvanu lapsega kõneldes. Hoidjakeele sõnavara ja suhtlemisstrateegiad on individuaalsed ning muutuvad lapse kasvamisega. Varasema perioodi hoidjakeelt iseloomustavad arvukad kordused, täiskasvanule suunatud kõnega võrreldes lühemad laused, mis on oma suhtluseesmärgi tõttu pigem küsi- või väitlused, ning onomatopoeetiliste sõnade ja deminutiivide rohkus. Siinse artikli kontekstis on hoidjakeele all silmas peetud varajase perioodi hoidjakeelt (ingl *baby talk*), mida vanem kõneleb alla kolmeaastase lapsega.

Kuigi esimesi vihjeid lastele suunatud kõne kohta võib leida juba ligi saja aasta tagusest kirjutisest (Mägiste 1924), jäävad esimesed tänapäevased hoidjakeele käsitlused tõepoolest veidi enam kui kümne aasta tagusesse aega (Tulviste 2000, Pajusalu 2001, Argus 2003). Nüüdseks on eesti lastekeele uurijad jõudnud üksmeelsele seisukohale, et laste- ja hoidjakeel on teineteisega tihedalt seotud ning

paremaid tulemusi on võimalik saada, kui üht uurides vaadeldakse paralleelselt ka teise andmeid.

Ühena esimestest hakkas lapsele suunatud kõne uurimisega tegelema Tiia Tulviste, kes on uurinud vanema (enamasti ema) vestlusstrateegiaid lapsega kõnel-des. Tulviste paigutas eesti hoidjakeele kohe ka teiste keelte konteksti, kõrvutades kohalikku keelematerjali USA, Soome ja Rootsi omaga (Tulviste 2000, Tulviste, Kants 2001, Tulviste jt 2002), samuti on ta vaadelnud vestlusstrateegiaid eri situatsioonides, nagu söögilauavestlused või suhtlemine puslede kokkupanemisel (Tulviste 2000, Tulviste, Mizera 2002).

Eesti hoidjakeeles on palju uurimata valdkondi. On kirjeldatud väikelastele suunatud kõnet ning seda pigem morfoloogia ja sõnavara koostise seisukohast, kuid täiesti käsitlemata on jäänud näiteks teismeliste suunatud kõne, samuti on väga vähe tegeletud hoidjakeele pragmaatikaga. Siinse artikli eesmärk on kirjeldada eesti lapsele suunatud kõnes esinevaid küsimusi ja nende tüüpe teiste keelte taustal, esitada eesti küsimuste hulga ja liikide kohta andmeid võrdluses teiste keeltega (vene, leedu) ning seejärel vaadelda edasiarendatud liigituse alusel lähemalt eesti küsimusi. Lõpuks kõrvutatakse andmeid teiste eesti laste ja nende vanemate kõne kontrollmaterjaliga.

Analüüs võimaldab esiteks asetada eesti hoidjakeele laiemasse konteksti ning vaadelda eesti vanemate keekekasutust lapsega rääkides, pidades silmas rahvuslikke ja kultuurilisi eripärasid. Kuna esmase keelematerjali analüüsist selgus, et eesti hoidjate keekekasutus erineb vene ja leedu hoidjate omast, ongi siinse artikli eesmärk selgitada välja, kas eesti hoidjakeel sarnaneb oma vestlusstrateegiatelt pigem germaani ja romaani või meile geograafiliselt lähedasemate balti ja slaavi keelte omaga. Teiseks annab keelematerjali uurimine andmeid selle kohta, mis funktsiooni küsimused vanema kõnes täidavad ja kuidas küsimuste eri liigid vanema kõnes jaotuvad.

2. Teoreetiline lähtepunkt ja kõneandmed

Eesti hoidjakeele vestlusstrateegiad on ainekultuurist pakkunud ka teistele uurijatele. 2012. aastal viidi läbi ühisuurimus, kus võrreldi eesti, vene ja leedu hoidjakeele andmeid ja analüüsiti eesti, vene ja leedu vanemate küsilausungeid 1;8–2;8 vanustele lastele suunatud kõnes (Kazakovskaja jt 2012). Uurimuse varasemas etapis (Kazakovskaja, Balčiūnienė 2012b) olid selgunud teatavad sarnasused vene ja leedu hoidjate keekekasutuses (näiteks kasutab vanem rohkesti küsimusi, millest enam kui 85% on vestlusküsimused)¹; eesti hoidjakeel kaasati uurimusse eesmärgiga vaadata, kas võiks väita, et kõigi kolme keele hoidjakeel kannab teatud kultuurilist ühismõju ehk on mõjutatud meie ühisest nõukogude minevikust. Siinse artikli aluseks olnud ühisuurimuse tarbeks liigitas küsimusi vene keele dialoogides Viktoria Kazakovskaja, leedu keele dialoogides Ingrida Balčiūnienė ja eesti keele materjalis siinse artikli autor Helen Kõrgesaar.

Laste ja vanemate vahelise spontaanse kõne dialoogi oli uuritavas vanusevahemikus lindistatud iga kuu (ainult eesti lapse kohta puudus lindistus vanuses 2;2). Eesmärk oli vaatluse alla võtta iga lindistuse 300 esimest küsimust, mis võimaldaks

¹ Küsimuste liigitus on toodud osas 2.1.

määrata ka nn initsiatiivküsimused e vestlust algatavad küsimused. Vene ja leedu lindistustes nii ka tehti, kuid eesti dialoogides ei olnud niisugune valik võimalik, kuna ainult kahes lindistuses esines üle 300 küsimuse. Seetõttu võeti vaatluse alla kõik eesti dialoogides esinenud küsimused. Tabel 1 annab ülevaate sellest, kuidas kõneandmed eri keelte vahel jagunesid. Samuti on tabelis esile toodud, kui suur oli küsimuste osakaal kõigist täiskasvanu lausungitest.

Nii vene kui ka leedu keelematerjal hõlmas 13 tunniajast lindistust, millest oli vaatluse all niisiis 300 esimest küsimust (kokku 3900 küsimust mõlema keele kohta). Iga keele vestlusmaterjal oli pärit ühe lapse ja vanema vahelistest dialoogidest. Kõik kolm last olid ükskeelsest perest, olid normikohase arenguga, elasid oma kodumaa suuruselt teises linnas (Peterburi, Kaunas, Tartu) ning olid pere ainsad lapsed. Vaatlusaluse vene lapse vestluspartneriks dialoogides oli tema vanaema, leedu lapsega vestles ema ning eesti lapse lindistustes oli vestluse teiseks osapooleks enamasti ema, kolmes lindistuses ka isa. Eesti keelematerjal koosnes umbes 10 dialoogitunnist, kuna kõik lindistused ei olnud päris tunniajased. Kokku esines eesti täiskasvanu kõnes 957 küsimust.²

Tabel 1. Analüüsitud küsimuste hulk lindistustes ning küsimuste osakaal kõigi vanema lausungite hulgas (Kazakovskaja jt 2012)

Lapse vanus	Vene keel		Leedu keel		Eesti keel	
	Küsimusi hoidjakeeles	Küsimuste osakaal kõigist vanema lausungitest	Küsimusi hoidjakeeles	Küsimuste osakaal kõigist vanema lausungitest	Küsimusi hoidjakeeles	Küsimuste osakaal kõigist vanema lausungitest
1;8	300	45,4%	300	69,7%	256	34,7%
1;9	300	43,9%	300	65,8%	66	24,2%
1;10	300	56%	300	70%	221	31%
1;11	300	50,4%	300	91,4%	61	43,9%
2;0	300	46,5%	300	62,5%	394	22,1%
2;1	300	59,6%	300	85,7%	674	15,8%
2;2	300	55%	300	90,7%	–	–
2;3	300	55,5%	300	42,5%	43	33,3%
2;4	300	58,9%	300	79%	47	25,8%
2;5	300	53,7%	300	69,7%	88	25,6%
2;6	300	58,5%	300	65,5%	53	23,1%
2;7	300	47,9%	300	70,1%	89	31,7%
2;8	300	42,5%	300	72%	66	34,2%
Keskmine	–	51,8%	–	71,9%	–	28,8%

Uurimusest ilmnes, et kui vene hoidjakeeles esines tunniajases lindistuses keskmiselt 51,8% küsimusi ja leedu omas koguni 71,9% kõigist vanema lausungitest, siis eesti hoidjakeel koosnes küsimustest vaid 28,8% ulatuses. Kui vene ja eriti leedu hoidjakeel on üles ehitatud küsimustele, siis eesti täiskasvanu kasutab lapsega kõneldes peamiselt väiteid ja käskluseid.

Wolfgang U. Dressleri tööühma varasemast uurimusest (Kilani-Schoch jt 2009) selgus, et nii prantsuse kui ka Austria saksa hoidjakeeles kasutatakse samuti

² Nii eesti, vene kui ka leedu keelematerjal on pärit CHILDESi lastekeele andmepangast. Eesti materjal on andmepangas Vija alamkorpusest (kogunud Maigi Vija), vene keelematerjal oli kogutud Natalja Gagarina juhatusel ning leedu pikiuuringu tarbeks oli materjali kogunud Ingrida Balčiūniene.

rohkesti mitteküsimivaid kordusi ja vähem küsimusi, mistõttu eesti keel paigutus mainitud ühisuurimuse valguses pigem meist geograafiliselt kaugemate germaani-romaani keelte konteksti. Ka siinkirjutaja kogutud esmased andmed eesti hoidjakeele kohta kinnitavad, et täiskasvanu kasutab 4-aastase lapsega kõneldes suurel hulgal mitteküsimivaid kordusi, korrates peamiselt üle lapse vastuseid küsimustele, aga ka küsimusi endid, seevastu vestluses teise täiskasvanuga esines vaid mõni üksik kordus (1).

- (1) LAPS: mis see on siis?
EMA: no mis see on?
LAPS: kaubaauto.
EMA: kaubaauto.

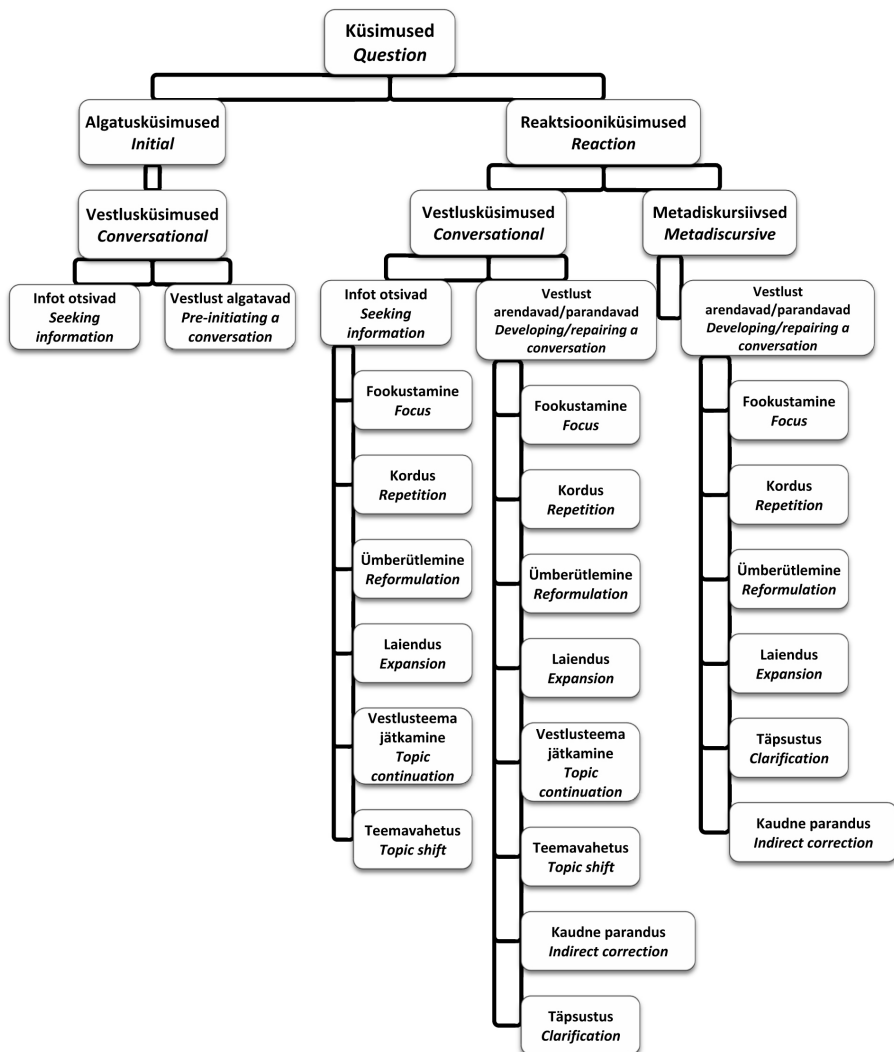
Lapsega kõneldes kordas täiskasvanu ühest küljest üle sõnu, mida pidas vajalikuks lapsel kinnistada, teisest küljest korrati aga ka neid sõnu, mis lapsel kindlasti selged olid, andmaks tagasisidet, et vastus on õige ja laps on õigel teel (Kõrgesaar 2007). Tuginedes nendele andmetele kui ainsatele selles eesti hoidjakeele valdkonnas, võib tõepoolest väita, et eesti lastele suunatud kõnes on kordused väga sagedad.

2.1. Hoidjakeele küsimuste liigitus

Analüüsi aluseks olnud küsimuste liigituse on varasemate, vene ja leedu hoidjakeelt käsitleva uurimuse tarbeks³ välja arendanud Kazakovskaja ja Balčiūnienė (2011, 2012a, 2012b). Sama liigitust kasutasime ühiselt ka vene, leedu ja eesti hoidjakeeli võrdlevas uurimuses ning hiljem on seda mugandanud ning eesti lastekeele uurimiseks kohandanud siinse artikli autor. Liigitus koosneb viiest eri tasemest, jagasime küsimused esmalt positsiooni, seejärel pragmaatilise rolli, funktsiooni, struktuuri ja viimaks vormi järgi. Siinkohal pean oluliseks lugejale selgitada, mida täpselt on liigituse iga jaotuse all (vt joonis 1) silmas peetud, mistõttu olen täiendanud autorite kirjeldust (Kazakovskaja, Balčiūnienė 2012b: 201–203) omapoolsete selgitavate kommentaaride ning eesti keele näidetega:

- 1) positsioon vestluses – kas tegemist on algatus- (ingl *initial*) või reaktsiooniküsimusega (*reaction*). Algatusküsimus on vestlust/lindistust alustav küsimus ning enamasti saab neid lindistuse jooksul esineda vaid ühel korral. Samas on mõnes lindistuses tehtud paus ning hiljem vestlust jätkates alustatud taas algatusküsimusega, nt *Noh, mida sa siis nüüd teha tahaksid?* Reaktsiooniküsimus on küsimus, mille vanem esitab reaktsioonina lapse eelnevale lausungile;
- 2) pragmaatiline roll – algatusküsimused saavad oma pragmaatilise rolli järgi ollagi ainult vestlusküsimused (*conversational*), samas kui reaktsiooniküsimused võivad lisaks sellele olla ka metadiskursiivsed (*metadiscursive*), mille all mõistetakse küsimusi, mis täpsustavad lapse öeldut. Vanem küsib üle, kui ta ei saanud päris täpselt aru, mida laps ütles, või parandab öeldu sisu, nt *Mida-mida, kuidas sa ütlesid?*;

³ Uurimus oli osa keeltevahelisi mõjusid uurivast projektist "Pre- and Protomorphology in Language Acquisition", mille juhiks oli W. U. Dressler Austria Teaduste Akadeemiast (Kazakovskaja, Balčiūnienė 2012: 70).



Joonis 1. Küsimuste liigitus (Kazakovskaja, Balčiūnienė 2012b: 202)

- 3) funktsioon – algatusküsimused jagunevad oma funktsiooni järgi kaheks: infot otsivateks (*seeking information*) ja vestlust algatavateks küsimusteks (*pre-initiating a conversation*). Reaktsiooniküsimustest jagunevad vestlusküsimused samuti kaheks: infot otsivateks ning vestlusteemat arendavateks ja parandavateks küsimusteks (*developing/repairing a conversation*). Infot otsiva küsimuse puhul täiskasvanu tõepoolest ei tea küsimusele vastust, nt *Kus te siis täna vanaemaga käisite?*, vestlust arendava küsimuse puhul on vastus vanemale teada ja tema küsimus kannab muud eesmärki, nt *Noh, mis värvi see pall on?* (Viimasel juhul ei ole tegemist info otsimisega, vaid vestlusteemat suunava küsimusega.) Metadiskursiivsed küsimused saavad oma struktuurilt olla vaid vestlusteemat arendavad ja parandavad;

- 4) struktuur – viis, kuidas küsimuste eri struktuurid eri funktsiooni täitvate reaktsiooniküsimuste puhul jagunevad, on näidatud joonisel 1. Selle järel on lühidalt ära toodud selgitus eri struktuuride kohta.

Kazakovskaja ja Balčiūnienė liigitus (2012a: 75–78) jagab vestlusteemat arendavad ja parandavad ning infot otsivad reaktsiooniküsimused nende struktuuri järgi kaheksaks. Igale struktuuritüübile olen lisanud ka üks-kaks iseloomulikku näitelauset.

Fookustamine (ingl *focus*) – vanem kordab ainult osa lapse lausest:

- (2) EMA: pime on jah, õhtu on käes.
LAPS: must.
EMA: must või?

Kordus (ingl *repetition*) – vanem kordab üle kogu lapse lausungi. Korduseks on siin loetud ainult vanema reaktsioon lapse korrektsele lausungile, muul juhul on tegu parandusega:

- (3) ISA: mis sa nägid?
LAPS: mis sa otsid?
ISA: mis sa otsid?
- (4) tegevus: LAPS matkib söömist, matsutab.
LAPS: oli magus.
EMA: oli magus?

Ümberütlemine (ingl *reformulation*) – vanem ütleb lapse lausungi ümber nii, nagu oleks parem olnud, nt muudab sõnajärge, muudab murdekeele sõna kirjakeele sõnaks, parandab onomatopoeetilise või pudikeele sõna õigeks jne. Mingit viga iseenesest ei pruugigi lapse lausungis olla. Samuti võib vanem öelda lause ümber nii, nagu ta arvas, et laps tahtis öelda:

- (5) LAPS: seda maitsesid.
EMA: maitsesid seda?

Laiendus (ingl *expansion*) – vanem lisab midagi lapse lausele, laiendab lapse öeldut leksikaalselt või grammatiliselt:

- (6) LAPS: kassi ka.
EMA: kassiraamatut ka vaatame?

Vestlusteema jätkamine (ingl *topic continuation*) – vanem ei korda midagi, ei paranda midagi, ei vasta otseselt lapse fraasile, vaid jätkab vestluse loomulikku kulgu:

- (7) EMA: päkapikud kõnnivad õues või?
LAPS: kõnni.
EMA: mis päkapikud teevad?

Teemavahetus (ingl *topic shift*) – vanem ei vasta otse lapse lausungile, vaid muudab lihtsalt vestluse teemat:

- (8) LAPS: Atu seda.
ISA: kas sul on piim otsas?

- (9) EMA: siia klotside juurde pissile.
 LAPS: pissin passin pissi pa.
 LAPS: pisin pasin.
 LAPS: xxx. (Laps ütleb midagi arusaamatut.)
 EMA: kus su telefon on?

Kaudne parandus (ingl *indirect correction*) – vanem parandab lapse eelmise lausungi vale vormi või sisu:

- (10) LAPS: Ee:sti: immast.
 EMA: Eesti ilmast või?
 (11) LAPS: telekas oli palju kutsat.
 EMA: palju kutsasid?

Täpsustus (ingl *clarification*) – vanem palub lapsel korrata või täpsustada eelmist lausungit:

- (12) EMA: see on diktofon.
 LAPS: xxx. (Laps vastab midagi arusaamatut.)
 EMA: mida?
 (13) LAPS: see on, mis see on?
 ISA: kuskohas?

Jaotuse viies tase puudutab küsimuste jagunemist nende vormist lähtuvalt ning selle põhjal jagunevad küsimused kolmeks: üldküsimused (ingl *general (yes/no) questions*), mis ootavad jah/ei-vastust, alternatiivküsimused (... või/or ...) ja avatud ehk eriküsimused (*wh-questions*), millele vastates tuleb valida lausekomponendi etteandmata variantide vahel.⁴ Siinse artikli aluseks olnud 2012. aasta vene-leedu-eesti võrdlevas uurimuses küsimuste vormilist jaotumist ei vaadeldud, kuid kindlasti võiks seda uurida edaspidi.

2.2. Tulemuste võrdlus

Kui välja arvata küsimuste hulga eelmainitud erinevus eri rahvusest vanemate kõnes, olid näitajad eri keelte materjalis üldjoontes üsna sarnased. Algatusküsimusi esines vene hoidjakeeles 0,3%, leedu keeles 0,2% ja eesti keeles 0,6%, millest järeldub, et reaktsiooniküsimusi oli vastavalt vene keeles 99,7%, leedu keeles 99,8% ja eesti keeles 99,4%. Sama vähe (0,3%) esines eesti hoidjakeeles ka metadiskursiivseid küsimusi, mida leedu vanem kasutas 11,1% ja vene vanem 7,7% kõigist juhtudest.

Infot otsivate ehk n-ö tõeliste küsimuste hulk oli eesti hoidjakeeles oluliselt väiksem kui teistes keeltes – eesti keeles 1,3%, vene keeles 8,7% ja leedu keeles koguni 13,3%. Kõrvutades neid andmeid eesti keele suulise kõne dialogide uurimise

⁴ Eesti keele küsilause jagunemist vormi järgi on käsitlenud Helle Metslang ja Tiit Hennoste tööühm. Metslangi (1981, 2013) järgi jagunevad eesti keele küsimused laiemalt samuti üld-, alternatiiv- ja eriküsimusteks, kõigi kolme küsimusevormi mõistest vastavad Kazakovskaja ja Balčiūnienė (2012b: 206) omadele. Põhitüüpeide all toob Metslang lisaks välja alltüüpe.

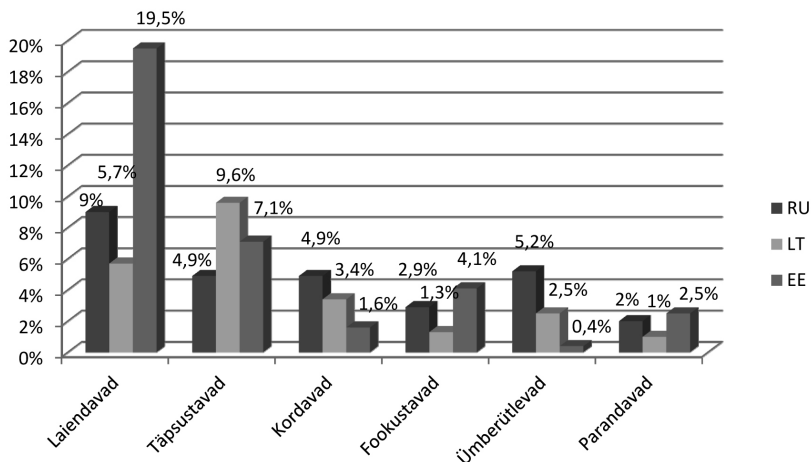
Hennoste jt (2009: 341, 344–345) esitavad jaotuse, mis koosneb kuuest küsimusetüübist. Esmalt koosnes liigitus kolmest rühmast, mis samuti vastavad Kazakovskaja ja Balčiūnienė liigitusele (*kas*-küsimused, alternatiivküsimused ja avatud küsimused), kuid dialooge ja neis sisaldunud küsimusi analüüsides tuli esmast jaotust laiendada ning tulemuseks saadi järgmine jaotus: 1) avatud küsimused, 2) alternatiivküsimused, 3) suletud *kas*-küsimused (vastuseks jaatus-eitus), 4) vastust pakkuvad küsimused (küsimus sisaldab küsija arusaama või oletust võimaliku õige või sobiva vastuse kohta), 5) jutustavad *kas*-küsimused (tavapärase jaatus-eituse asemel info otsimine), 6) muud (nt analüüsimatud näited ja lausungid, kus on koos kaks küsimust jne).

andmetega (Hennoste jt 2013: 23) selgub, et viimastes esineb infot otsivaid küsimusi 37% kõigist juhtudest. Kuivõrd Hennoste jt uurimus põhineb aga nimelt infodialoogidel (lindistatud kõned infotelefonile, reisibüroosse, polikliinikusse jne, kuhu inimene helistab tavaliselt selleks, et vajalikku infot saada), jääb üle imestada koos autoritega, mis võib olla põhjuseks, et isegi infodialoogides esineb nii vähe infot otsivaid suhtlustegevusi (Hennoste jt 2013: 24). Selliste tulemuste taustal ei tundugi ehk infot otsivate küsimuste nii vähene hulk hoidjakeeles kummaline. Igal juhul lubavad tulemused järeldada, et eesti täiskasvanu lapsega rääkides pigem suunab vestlust ega soovi infot hankida, samuti ei kipu ta last ka ülemäära parandama.

Struktuuritüübi järgi jagunesid küsimused eri rahvusest vanemate keelekasutuses enam-vähem võrdselt (vt joonis 2), esiletoomist vajaks ehk vaid see, et eesti hoidjakeeles kasutati võrreldes teiste keeltega erakordselt suurel hulgal (19,5%) laiendavaid küsimusi (14), (15), (16) ning märkimisväärselt väikesel hulgal (0,4%) ümberütlevaid küsimusi (17), (18), (19).

- (14) LAPS: lomme. /loeme/
EMA: missugust juttu loeme?
(Andreas, 1;08.25, emaga)
- (15) LAPS: see kingutus.
EMA: pakid kingituse ära vä?
(Andreas, 2;00.16, emaga)
- (16) LAPS: Atsu taab veel seda.
EMA: veel sousti ja kartulit?
(Andreas, 2;01.08, emaga)
- (17) LAPS: notsu limpsis notsu limpsis kaks selme.
EMA: kaks sõrme limpsis?
(Andreas, 2;01.01, emaga)
- (18) LAPS: seda vanat albumit, seal on väikesed Goorgid.
EMA: väikesed Georgid on seal?
(Andreas, 2;06.12, emaga)
- (19) LAPS: sõidame bussiga.
EMA: bussiga sõidame?
(Andreas, 2;07.12, emaga)

Analüüsist (vt joonis 2) selgus, et kui leedu vanem on väga aldis lapse öeldut täpsustama (kõige sagedasem küsimusestruktuur leedu keele andmete põhjal), siis nii eesti kui ka vene täiskasvanu kasutab pigem vestlusteemat laiendavaid küsimusi, mida esineb mõlema keele materjalis kõige rohkem. Vahe on aga selles, et eesti täiskasvanu laiendab vestlusteemat kaugelt rohkem kui vene täiskasvanu (eesti 19,5% vs. vene 9%), kuid kasutab samas silmatorkavalt vähem kordavaid ja ümberütlevaid küsimusi.



Joonis 2. Küsimuste jagunemine vene (RU), leedu (LT) ja eesti (EE) hoidjakeeles nende struktuuri järgi

3. Eesti küsimuste analüüs täpsustatud liigitamispõhimõtete alusel

Juureldes siiski eesti hoidjakeele küsimuste vähesuse üle ning analüüsid andmeid üha uuesti, jäi silma võimalik põhjus, miks 2012. aasta ühisuurimuse analüüsis osas kasutatud küsimuste liigitus eesti keele küsimuste määramiseks sugugi sobida ei pruugi. Küsimuste liigitamisel kehtis nõue, et analüüsimisele kuuluvad ainult need täiskasvanu küsimused, mis järgnevad vahetult lapse lausungile. Eesti hoidjakeele andmete hilisemal vaatlusel selgus aga, et täiskasvanul on sageli kombeks hoopis iseenda varasemat lausungit/küsimust küsimusega ümber öelda, täpsustada, korrata vm.

- (20) EMA: mis on must?
LAPS: ake.
EMA: miks aken on must?
EMA: miks aken must on?
- (21) LAPS: metsa.
EMA: metsas.
EMA: mis metsa sees on?
- (22) LAPS: kepsega.
EMA: küpsetavad.
EMA: mis päkapikud küpsetavad?
EMA: mida päkapikk küpsetab?

Sellele leiule tuginedes tekkiski mõte jagada küsimused nende positsiooni järgi mitte kaheks, vaid kolmeks: algatusküsimusteks ning reaktsiooniks a) lapse eelmisele lausungile ning b) vanema enda eelmisele lausungile. Kuna vestlust analüüsid selgus, et vanem ütleb sageli küsimusega ümber hoopis iseenda öeldut või kordab

iseenda eelmist lausungit, võis oletada, et kui võtta arvesse ka need küsimused, kasvab küsimuste üldine hulk hoidjakeeles märgatavalt. Sellest lähtuvalt otsustasin eesti hoidjakeele küsimuste määramiseks täiendada Kazakovskaja ja Balčiūnienė liigitust lisajaotisega ning viia läbi kontrolluurimuse, mis võimaldaks arvesse võtta ka vanema enda eelmisele lausungile järgnevaid küsimusi.

Selleks lisasin määramistabelisse pragmaatilise rolli reale reaktsiooniküsimuste kõrvale veel teisegi võimaluse: *self reaction* ehk reaktsioon iseenda kõnevoorule. Kogu muu jaotus, mis puudutas küsimuste edasist jagunemist nende funktsiooni, struktuuri ja vormi järgi, jäi samaks. Kontrolluurimuses analüüsisin 15 lindistust lapse vanuses 1;7–2;8⁵ (sh 13 lindistust emaga ja kaks isaga) ning võtsin vaatluse alla liigituse kolm esimest tasandit: küsimuste positsiooni, pragmaatilise rolli ja funktsiooni.

Võrreldes uusi tulemusi eelmistega selgus, et kui varem esines vanema kõnes 957 küsimust, siis uue jaotuse alusel oli võimalik määrata kõik 1201 küsimust (ülevaade tulemuste kohta on toodud tabelis 2). Küsimused, mis on reaktsiooniks vanema enda eelmisele lausungile, moodustavad u 20% kõigist küsimustest, seega oli eelmise jaotuse alusel jäänud määramata ja analüüsimata viiendik vanema küsimustest. Uus liigitus ei muutnud algatusküsimuste arvu, infot otsivate küsimuste arv oli kasvanud kahe ja metadiskursiivsete küsimuste hulk kolme võrra. Kõige rohkem muutus (lisandus 244 küsimust), nagu võiski arvata, just vestlusteemat arendavate ja parandavate reaktsiooniküsimuste hulk.

(23) LAPS: panen kinni selle.

EMA: ära pane, sa pidid oma arvutiga numbreid tegema.

EMA: kus su arvuti on?

(Andreas, 2;03.26, emaga)

(24) LAPS: anna sealt ülevalt.

EMA: siit kapist, näed sul on ka sinikas siin.

EMA: oled ära löönud?

(Andreas, 2;05.30, emaga)

(25) LAPS: tead, mis kondi sees on?

EMA: ei tea.

EMA: mis seal sees on?

(Andreas, 2;07.12, emaga)

Tabel 2. Tulemuste jagunemine eesti keele materjalis varasema ja täpsustatud liigituse põhjal

Küsimuste jagunemine	VARASEM LIIGITUS		TÄPSUSTATUD LIIGITUS	
	Küsimuste arv	% kõigist küsilauseingitest	Küsimuste arv	% kõigist küsilauseingitest
Küsimusi kokku	957	100%	1201	100%
Algatusküsimusi	5	0,5%	5	0,4%
Reaktsioone lapsele	952	99,5%	952	79,3%
Reaktsioone iseendale	–	–	244	20,3%

⁵ Endiselt puudusid keeleandmed lapse vanuses 2;2.

Kokkuvõtteks võib öelda, et uus küsimuste liigitus võimaldab määrata kõik küsimused ning saada paremat ettekujutust selle kohta, kuidas, miks ning mida vanem küsib. Küsimuste hulk vanema kõnes suureneb sellise liigituse alusel viiendiku võrra. Uurimise tulemustest selgus, et eesti vanem kordab ja sõnastab sageli ümber just enda öeldut. Samas tuleb siiski tõdeda, et ehkki uue liigituse järgi sai ära määrata veel 244 küsimust ning küsimuste osakaal vanema kõnes tõusis seetõttu 36,2 protsendini (kasvas 7,4% võrra), on seda ikkagi märkimisväärselt vähe võrreldes vene (51,8%) ja leedu (71,9%) hoidjakeelega.

4. Eesti tulemuste kontroll võrdluses laia valimi meetodil saadud keelematerjaliga

Eeltoodud tulemuste valguses tekkis siiski küsimus, miks on vene-leedu ning eesti tulemuste vahel nii suur erinevus. Kuna 2012. aasta ühisuurimuse keeleandmed pärinesid vaid ühe ema ja tema lapse vahelisest vestlusest, jäi kahtlus, et küsimuste vähesus hoidjakeeles on tingitud selle konkreetse ema keelekasutusest. Et seda välistada, oli tarvis läbi viia teinegi kontrolluuring võrdleva materjali põhjal, mis kinnitaks või lükkaks ümber, kas eesti hoidjakeel on tõepoolest küsimuste vaene. Selle tarbeks otsustasin võrrelda olemasolevaid andmeid veel nelja sama vana lapse ja nende vanemate keeleandmetega. Kuna varem kasutatud materjali mõnes lindistuses kõneles lapsega ema, mõnes aga isa, valisin ka võrdlevat materjali selle alusel, et vestluse täiskasvanud osapoolena oleksid esindatud nii isad kui ka emad. Võrdleva materjali põhjal tehtud uurimuses ei vaadeldud ma küsimuste jagunemist nende positsiooni, funktsiooni, pragmaatilise rolli ega struktuuri seisukohast, vaid keskendusin küsimuste hulga täpsustamisele.

Teise katse valim hõlmas ajaperioodilt laste teise eluaasta esimest poolt ning jagunes kolmeks vanusevahemikuks, igast vahemikust kuulus valimisse 2–4 vestlust.

I vanusevahemik:

Andreas 2;01.03 (vestlus toimus isaga)

Martina 2;01.17 (vestlus toimus emaga)

Ruuben 2;02.03 (vestlus toimus isaga)

II vanusevahemik:

Andreas 2;03.26 (vestlus toimus isaga)

Mia 2;03.17 (vestlus toimus emaga)

III vanusevahemik:

Andreas 2;04.13 (vestlus toimus isaga)

Andreas 2;05.02–2;05.30 (vestlus toimus emaga)

Jaana 2;05.12 (vestlus toimus emaga)

Kõigis kolmes vanusegrupis olen aluseks võtnud esmalt algses uurimuses kasutatud lapse, Andree kõneandmed, millele olen võrdlusmaterjalina lisanud teiste sama vanade laste ja nende vanemate kõneandmed lähtuvalt sellest, mis vanuses laste ja vanemate spontaankõne lindistuste litereeringuid pakub CHILDESi lastekeele andmebaasi eesti lastekeele korpus. Andmete analüüsimisel kasutasin enda täiendatud küsimuste liigitust (vt eespool 3. osa). Ülevaate tulemustest annab tabel 3.

Tabel 3. Küsimuste hulk hoidjakeeles laia valimi meetodi põhjal

Vanuse- vahemik	Lapse nimi ja vanus	Lausungite arv	Küsimuste arv	Küsimuste osakaal
I	Andreas 2;01.03	393	84	21,4%
	Martina 2;01.17	418	111	26,6%
	Ruuben 2;02.03	447	140	31,3%
II	Andreas 2;03.26	130	48	36,9%
	Mia 2;03.17	539	193	35,8%
III	Andreas 2;04.13	182	55	30,2%
	Jaana 2;05.12	357	103	28,9%
	Andreas 2;05.02–.30	343	89	26%
Kokku / keskmine		2809	823	29,3%

Analüüsitulemused näitavad, et teised valimisse kuulunud vanemad kasutavad tõepoolest rohkem küsimusi, kuid see ei muuda siiski märkimisväärselt küsimuste hulga osakaalu kõigi lausungite hulgas (st teistes lindistustes oli proportsionaalselt suurem ka täiskasvanu lausungite koguhulk). Niisiis kinnitab laia valimi põhjal tehtud kontrolluuring samuti, et 2-aastase lapsega kõneldes kasutab eesti täiskasvanu vaid keskmiselt 29,3% küsimusi. Seega ei saa küsimuste vähesust hoidjakeeles pidada vaid ühe konkreetse vanema keelekasutuse eripäraks. Niisugune tulemus kinnitab veel kord, et eesti hoidjakeel on tõepoolest küsimustevaene ning selle vestlusstrateegiad ei ole üles ehitatud peamiselt ega ka mitte suuremalt jaolt küsimustele, nagu võime näha vene ja leedu hoidjakeele tulemustest.

5. Kokkuvõtteks

Küsimuste kasutamise sageduse põhjal võib väita, et eesti hoidjakeel sarnaneb pigem meist geograafiliselt kaugemate germaani ja romaani keeltega, mis on üles ehitatud peamiselt väit- ja käsklausetele ning kordustele, mitte niivõrd küsimustele nagu meile geograafiliselt lähemad balti ja slaavi keeled. Samas võib öelda, et küsimuste vähesest kasutamisest hoolimata sarnaneb nende üldine jaotumine vastavalt küsimuse positsioonile (enam kui 99% reaktsiooniküsimusi), funktsioonile (ligi 90% vestlusküsimusi), pragmaatilisele rollile (ligi 90% vestlusteemat arendavaid ja parandavaid küsimusi) ja struktuurile (kõige rohkem on kõigis kolmes keeles kasutatud vestlusteemat laiendavaid ja jätkavaid küsimusi) eesti hoidjakeeles vene ja leedu keelega.

Siiski ei võimalda keelematerjali piiratus praegu teha põhjanevaid järeldusi kultuuriliste ühisjoonte või meie ühise nõukogude tausta võimaliku mõju kohta eesti, vene ja leedu hoidjakeelele. Selleks tuleks kindlasti uurida ka vene ja leedu hoidjakeelt laia valimi abil ning välistada sel moel neiski keeleandmetes võimalus, et küsimusterohke suhtlusstiil iseloomustab tegelikult ainult ühte konkreetset vanemat. Vene materjali puhul võib olulist rolli mängida fakt, et lapse vestluspartner on vanaema, mitte ema.

Veel oleks edaspidi huvitav vaadata, kuidas muutub vanema kõnes küsimuste hulk, nende funktsioon ja pragmaatiline roll lapse kasvades, aga ka seda, kas ja kuidas see peegeldub lapse kõnes. Samuti oleks huvitav võrrelda neid aspekte vene ja leedu lapsele suunatud kõne andmetega.

Viidatud kirjandus

- Argus, Reili 2003. Lastekeelest ja selle uurimisest. – *Oma Keel*, 1, 26–32.
- Hennoste, Tiit; Gerassimenko, Olga; Kasterpalu, Riina; Koit, Mare; Rääbis, Andriela; Strandson, Krista 2009. Küsimused eestikeelses infodialoogis I: küsimuste vorm. – *Keel ja Kirjandus*, 5, 341–359.
- Hennoste, Tiit; Andriela Rääbis; Kirsi Laanesoo 2013. Küsimused eestikeelses infodialoogis II: küsimused ja tegevused. – *Keel ja Kirjandus*, 1, 7–28.
- Kazakovskaja, Viktoria; Balčiūnienė, Ingrida 2011. Ossobennosti retši vzroslovo, obraššenni k rebjonku: voprossitelnoje repliki v russkom i litovskom jazōkah. – *Isljedovanija po slavjanskim jazōkam*, 16 (1), 259–292.
- Kazakovskaja, Viktoria; Balčiūnienė, Ingrida 2012a. Lithuanian and Russian child-directed speech: Why do we ask young children so many questions? – *Eesti Rakenduslingvistika Ühingu aastaraamat*, 8, 69–89. <http://dx.doi.org/10.5128/ERYa8.05>
- Kazakovskaja, Viktoria; Balčiūnienė, Ingrida 2012b. Interrogatives in Russian and Lithuanian child-directed speech: Do we communicate with our children in the same way? – *Journal of Baltic Studies*, 43 (2), 197–218
- Kazakovskaja, Viktoria; Balčiūnienė, Ingrida; Kõrgesaar, Helen 2012. Conversational strategy in early childhood: Insights into Estonian, Lithuanian and Russian child-directed speech. – *Ettekanne 16.–18.5.2012 Tallinna Ülikoolis toimunud rahvusvahelisel konverentsil “Keeleteaduse päevad 2012”*.
- Kilani-Schoch, Marianne; Balčiūnienė, Ingrida; Korecky-Kröll, Katharina; Laaha, Sabine; Dressler, Wolfgang U. 2009. On the role of pragmatics in child-directed speech for the acquisition of verb morphology. – *Journal of Pragmatics*, 41 (2), 219–239. <http://dx.doi.org/10.1016/j.pragma.2008.10.001>
- Kõrgesaar, Helen 2007. Lapsele suunatud kõne võrdlus lapse ja täiskasvanule suunatud keeikasutusega. Seminaritöö. Käsikiri Tallinna Ülikoolis.
- Metslang, Helle 1981. Küsilause eesti keeles. [Interrogative sentence in Estonian.] Tallinn: Valgus.
- Metslang, Helle 2013. “Kes palju küsib, saab targemaks.” Küsimusest ja selle keelelisest vormist. [“Who asks a lot, can be smarter.” Question and its linguistic form.] – *Oma Keel*, 2, 5–12.
- Mägiste, Julius 1924. Paar sõna lastekeelest. – *Eesti keel*, 1, 1–9.
- Pajusalu, Karl 2001. Baby Talk as a Sophisticated Register: A Phonological Analysis of South Estonian. – *Psychology of Language and Communication*, 5 (2), 81–92.
- Tulviste, Tiia 2000. Socialization at meals: A comparison of American and Estonian mother-adolescent interaction. – *Journal of Cross-Cultural Psychology*, 31 (5), 537–556. <http://dx.doi.org/10.1177/0022022100031005001>
- Tulviste, Tiia; Kants, Luule 2001. Conversational styles of mothers with different value priorities: Comparing Estonian mothers in Estonia and Sweden. – *European Journal of Psychology of Education*, 16 (2), 223–231. <http://dx.doi.org/10.1007/BF03173027>
- Tulviste, Tiia; Mizera, Luule 2002. Eesti emad, Rootsi lapsed: võrdlev uurimus sotsialisatsioonist eesti, rootsi ja soome üks- ja kakskeelsetes kodudes. – Aune Valk (Toim.). *Eesti ja eestlased võrdlevas perspektiivis: kultuuridevahelisi uurimusi 20. sajandi lõpust*. Tartu Ülikool: Tartu Ülikooli Kirjastus, 143–155.

Tulviste, Tiia; Mizera, Luule; De Geer, Boel; Tryggvason, Marja-Terttu 2002. Regulatory comments as tools of family socialization: A comparison of Estonian, Swedish and Finnish mealtime interaction. – *Language in Society*, 31 (5), 655–678. <http://dx.doi.org/10.1017/S0047404502315045>

Helen Kõrgesaar (Tallinna Ülikool) uurib eesti hoidjakeelt, selle dünaamikat ja mõju lapse kõne arengule.
Narva mnt 25, 10120 Tallinn, Estonia
helen.korgesaar@gmail.com

INTERROGATIVES IN ESTONIAN CHILD-DIRECTED SPEECH

Helen Kõrgesaar

Tallinn University

It has been established from previous studies that Estonian child-directed speech includes fewer questions than in Russian and Lithuanian and adults' interaction with children is based on declarative and command sentences, which aligns Estonian child-directed speech more closely to Germanic and Romance languages than to the geographically closer Baltic and Slavic languages. The aim of the present article is to find out whether Estonian child-directed speech is indeed as poor in questions as it appeared by initial research.

Analysing the Estonian data from previous research, it appeared that Estonian adults do not use questions only in reaction to child speech but also to reformulate, specify and repair their own utterances. For that reason the original classification of questions used in this research has been refined. To the previous classification that only took into account questions that the parent posed in immediate reaction to the child's speech, a new category has been added in order to consider the questions that parents pose in reaction to their own prior utterances.

The Estonian-Russian-Lithuanian joint study was based on speech data from one Estonian, one Russian and one Lithuanian child-parent couple. About 12 hours of audio material for each language was observed, all three children were recorded at age 1;8–2;8. First, for a control study, 15 recordings of one-hour spontaneous speech of children aged 1;7–2;8 were analysed. Then, the results were verified by a wide selection method and speech data from another four children was included, in order to exclude the possibility that the small number of questions characterizes the phrasing of a specific mother rather than that of Estonian child-directed speech in general.

The results of the study show that although the new classification of questions allowed the identification of one-fifth more questions, Estonian child-directed speech compared to Russian and Lithuanian child-directed speech remains relatively poor in questions and according to its conversation strategy is more similar to Germanic and Romance languages.

However, it can be said that according to the position, function, pragmatic role and structure, the results concerning the overall distribution of questions in Estonian child-directed speech are similar to Russian and Lithuanian results.

Keywords: language acquisition, pragmatics, Estonian, Russian, Lithuanian

PÕHIMÕTTELISELT VÕI PRAKTILISELT PAIDES – KAHE SARNASE FUNKTSIOONIGA SÕNA KÄEKÄIGUST

Annika Küngas

Ülevaade. Artikkel analüüsib sõnade *põhimõtteliselt* ja *praktiliselt* kasutust eesti kirjakeele korpuse andmete põhjal, lähtudes ennekõike semantikast ja pragmaatikast, kuid ka süntaksist ja foneetikast. Keskendutakse viimase kahe sajandi materjalile, kuna oma tänapäevasel kujul ilmnesid *põhimõtteliselt* ja *praktiliselt* esmakordselt 20. sajandi esimesel poolel. Sõltuvalt kontekstist võivad need sõnad funktsioneerida mitmel viisil. Adverbilises kasutuses on *põhimõtteliselt* ja *praktiliselt* üksused, mille leksikaalne väli ei kattu. Siiski on nendele lisandunud uusi tarvitusi, mille sarnasus on hõlpsasti tajutav, nt kui neid pruugitakse umbmäärastavas funktsioonis või diskursuseülese üksusena.

Võtmesõnad: pragmaatilised markerid, mäarsõnad, kirjalik keel, korpusuuringud, eesti keel

1. Sissejuhatus

Artiklis võetakse vaatluse alla kahe sõna, *põhimõtteliselt* ja *praktiliselt* kasutusmuster, eesmärgiks on jälgida korpuseandmete põhjal nende üksuste funktsioonide arengut. Näidatakse, et *põhimõtteliselt* ja *praktiliselt* esindavad polüseemiat, mille uurimine annab aimu keelest ja keele muutumise seaduspäradest üldisemalt. Seega on uuritavate sõnade puhul tähtis, et neid on kasutatud erinevates funktsioonides. Üksused esinevad tänapäeval nii adverbide kui ka pragmaatiliste markeritena, jagunedes pragmaatilises kasutuses omakorda modaalpartikliteks ja diskursusemarkeriteks.¹ Ulatuslik ülevaade eri autorite pragmaatiliste markerite käsitlestest on esitatud siinkirjutaja varasemas artiklis (Valdmets 2013: 110–112). Selle kirjatöö sisu siin dubleerima ei hakata, kuid tuuakse välja tähtsamad pidepunktid adverbide

¹ Mäarsõna, adverb ja täistähenduslik viitavad artiklis autosemantilisele üksusele ning (pragmaatiline) marker, partikkel, mittetäistähenduslik sünsemantilisele üksusele.

grammatisatsiooni kohta. Artiklis toetutakse grammatisatsiooni seletamisel mudelele (Hopper 1991), mille järgi keelendi A grammatiseerumine mingis kontekstis algab tähendusnihkega. Seejärel saab keelend uue kasutusmustriga B. Kolmandas (eeldatavas) staadiumis jääb kasutusse vaid B, A taandub. Küll aga on tõdetud, et kõik keelendid ei jõuagi tsükli kolmandale astmele.

Kuigi siin vaadeldavate sõnade esmane kasutus eesti keeles on täistähenduslik (vt nt EKSS), on neile grammatiseerudes lisandunud pragmaatilisi nüansse, mis on selgesti märgatavad just viimaste aastakümnete keeletarvituses. Nii on adverbidest kujunemas pragmaatilised markerid (vt nt Fraser 1996, Traugott, Dasher 2002, Brinton, Traugott 2005). Erinevalt adverbidest ei osale pragmaatilised markerid propositsiooni väljendamisel, pigem ilmneb nende funktsioon kindlas kontekstis. Pragmaatilisi markereid on keeleteaduslikus kirjanduses defineeritud eri vaatepunktidest. Siinkohal on peamiselt tuginetud Laurel J. Brintoni (1996) käsitlusele: need on raskesti määratletavad üksused, mida kasutatakse keeles erinevates pragmaatilistes funktsioonides, kusjuures nende algne adverbiline ülesanne on ajapikku hääbumas. Pragmaatilisi markereid võib omakorda jagada modaalspartikliteks ning diskursusemarkeriteks (Traugott, Dasher 2002, Valdmets 2013), kusjuures kahe kategooria eristamine võib olla problemaatiline (Traugott 2007: 141). Modaalpartikliteks loetakse üldjuhul sõnad, mis funktsioneerivad fraasi- või lauselaiendina (Valdmets, Habicht 2013). Diskursusemarkerite rühma kuuluvad üksused, mis on vaid marginaalselt modaalsed (vt analüüsi Traugott, Dasher 2002: 152 jj), pigem on need suurema mõjualaga kui lause, sidudes teksti osi ja luues tervikut (vt täpsemalt ka Valdmets 2013). Diskursusemarkerid asetuvad enamasti lausealgulisse positsiooni, kuid ei seostu muu lausega. Nii on nende ülesandeks tihtipeale kõnes vooruvahetuse hõlbustamine ja kirjas lõikude markeerimine. (Brinton 1996: 6)

Toetudes nii üldistele kui ka partikleid puudutavatele grammatisatsiooni käsitlustele (Brinton 2010, Traugott, Dasher 2002, Heine 2002, Heine, Narrog 2010 jt), võib märgata, kuidas täistähenduslikud adverbid grammatiseeruvad ajapikku pragmaatilisteks üksusteks. Eespool esitatud kolmeastmelise grammatisatsioonimudeli (Hopper 1991) mõistetes on *lt*-liitelised sõnad jõudnud protsessi teise järku. Uues funktsioonis keelendid ei ole enam sisult nii läbipaistvad kui nende adverbiline lähe. Üldist muutumismustrit arvesse võttes võime eeldada, et *põhimõtteliselt* ja *praktiliselt* on kujunemisejärgus üksused, mille kasutus varieerub, ja seega on võimalik küsida, kas ja missugustes kontekstides nad on muutumas adverbist pragmaatiliseks markeriks.

2. Analüüsi lähtekohti

Eeltoodust nähtub, et sellistel üksustel võib tänapäeval olla nii adverbiline (1a), (1b) kui ka pragmaatiline (2a), (2b) kasutus.

- (1a) Ta on juba **põhimõtteliselt** igasuguse propaganda vastane.
- (1b) Ma tunnen oma tööd nii teoreetiliselt kui ka **praktiliselt**.
- (2a) Tal **põhimõtteliselt** puudub kontakt välismaailmaga.
- (2b) **Praktiliselt** pool saalisviibijatest on mehed.

Nimetatud üksuste ühe kasutusvõimalusena tõuseb esile võrdsussuhte neutraalse ligikaudsuse väljendamine (vt ka Erelt 1986: 105). Näiteis (2a), (2b) on sõnad *põhimõtteliselt* ja *praktiliselt* oma täistähenduslikust kasutusest nihkunud ning nende sünonüümideks on saanud umbmäärased hulgasõnad *umbes*, *peaaegu*, *enam-vähem*, *üsna* jm. Samasse rühma kuuluvad veel mõned teisedki kirjakeeles laialdaselt kasutatavad sõnad (nagu *suhteliselt*). *Põhimõtteliselt* ja *praktiliselt* on siinsesse artiklisse valitud kui tüüpnäited, mis esindavad ka teisi samalaadseid rõhusõnu. Randolph Quirk jt (1985: 445) on välja toonud intensiivsussõnad (ingl *intensifiers*), jagades need kaheks: võimendajad (*amplifiers*; eesti keeles nt *kohutavalt hägune*, *täielikult naeruväärne*) ja pehmendajaid (*downtoners*; nt *peaaegu võimatu*, *üsna normaalne*, *natuke igav*). Douglas Biber jt (1999: 554–555) on neid grappe nimetanud vastavalt *intensifiers* ja *diminishers*. Nagu eesti keeleski, võib ingliskeelsetes näidetes leida võimendajate ja pehmendajatega samakujulisi üksusi, mida täistähenduslikust kasutusest pruugitakse viisimäärsõnana, kuid edasise arengu käigus on märgatavad nihked määratähenduse poole, seda eriti modaalse fraasilaiendite puhul. Pehmendajad on mõnes kontekstis kasutusel ka selleks, et mahendada lausega väljendatut, et ei tekiks konflikti kõneleja ning vastuvõtja vahel (vt ka Aijmer 2013). Siinkohal on mh võimalik leida paralleelse lauselaiendilise kasutusega.

It-sõnade pragmaatilist kasutust jälgides jäävad silma nii mõnedki keelendid, mis laiendavad omadussõnu, paiknedes omadussõnafraasis vahetult enne peasõna. Siin käsitletavat kaks üksust on valitud just seetõttu, et mõlemal on selline kasutus võimalik. Nende kollokatsioonide näol on enamasti tegu fraasidega, mille alistusseoses olev liige iseloomustab mingit omadust täpsemalt. Laiendav funktsioon on seotud n-ö tavaliste määralaienditega, ometi erinevad uuritavad muutumatud sõnad selle poolest, et mõnel juhul on nende tähendus nihkunud. Teatud kontekstis on küll võimalik tõlgendada nende (määrata/täis)tähendust, kuid enamasti on lisandunud pragmaatiline varjund, mida alg tähenduses polnud. Sageli on määratletav sildkontekst (Heine 2002: 86), mis kajastab vaheetappi teel rohkem grammatiseerunud üksuse poole.

Lisaks siin käsitletavatele sõnadele *praktiliselt* ja *põhimõtteliselt* käituvad sarnaselt veel näiteks partiklid *suhteliselt* (*suhteliselt tavaline päev*), *hullult* (*hullult hea film*), *hübematult* (*hübematult kallis korter*), *sügavalt* (*mul on sügavalt savi*). Toimetamata ning kõnekeelelähedases kirjakeeles leidub ka juba lühenenud üksusi, näiteks *absull* (*absoluutselt*), *suht* (*suhteliselt*): *absull null*, *suht igav*. Lähtutakse eeldusest, et üksus on teisenenud, kui tüve algne tähendus (nt *hübematult* 'jultunult, häbitult', *hullult* 'vaimuhaigele kohaselt' või *sügavalt* 'pinnast allapoole ulatuvalt') ei ole enam täielikult läbipaistev. Kõigil neil sõnadel on lisaks määratähendusele muidugi ka muid funktsioone (rõhutav, kinnitav, pehmendav vms).

Kõike eelnevat arvesse võttes võib modaalpartikleid defineerida kui lauses üldlaiendina toimivaid sõnu, mis lisavad väljendatule modaalse varjundi. Need saavad olla nt episteemilised (vahel ka evidentsiaalsed) või emfaatilised üksused, kusjuures siinses artiklis käsitletud sõnadel võib modaalpartiklina ilmnedagi veel pehmendav lisafunktsioon. Lauses võrdlemisi vabalt liikuva fakultatiivsed üksused lisavad öeldule subjektiivse varjundi ja tihtipeale on neile raske omistada mingit konkreetset tähendust. Seega on fraasi- ja lauselaienditeks määratletavad keelendid mitmekülgsed, kuna nende funktsioon ilmneb tihtipeale alles konkreetsetes kontekstis.

Näidetes (2a), (2b) on vaadeldavate sõnade semantika juba nihkunud, kuid süntaktiliselt on nad veel lausega seotud. Sellised kasutused üldlaiendina võiks lugeda vähem grammatiseerunuks kui mõned muud pragmaatiliste sõnade näited (vt allpool), millel on võime tekstis vabalt ümber paikneda. Kaugemale grammatiseerunud üksuste alla kuuluvad mh diskursusemarkerid, mis ei muuda lause tavaehitust (adverbi ja markeri eristamise testi kohta vt Aijmer 1997: 4 jj). Siinkohal on paslik vaadata eristamise võimalust eesti keele puhul. Teadupärast paikneb eesti kirjakeele iseseisvas jaatavas väitlauseis verb teisel kohal. Mõõndustega on võimalik verbi asetada ka muusse positsiooni, kuid enamasti on sellisel juhul märgatav kohmakam lauseehitus. Nimelt on Kaja Taela (1988) andmeil SVX- ja XVS-sõnajärg eesti kirjakeeles sagedaim, moodustades umbes poole kõigist juhtudest. Suulises kõnes (ja tõenäoliselt ka toimetamata kirjakeeles) võib olukord olla teistsugune, Liina Lindström (2005) on näidanud, kuidas suulises kõnes on sagedaim SVX- ja SV-sõnajärg. Seega on mõlema töö põhjal võimalik väita, et eesti keele põhisõnajärjes on enamasti verb teisel kohal. Samale järeldusele on jõutud teisteski uurimustes, vt nt Annekatrin Kaivapalu psühholingvistilise tajutesti põhjal (2010) ning Tuomas Huumo (1995) eesti ja soome keelt võrreldes. On välja toodud mitmesuguseid sõnajärje valiku tegureid infostruktuurist kuni lausetüübini. SVX- ja XVS-sõnajärjes esineb adverb X-positsioonil. Siinses uurimuses tuuakse muuhulgas välja, et XSV-järg on adverbide puhul ebaloomulikum. Kui laiendada siinset sõnajärjemalli perifeeriatega² käsitlusega, kus vasak perifeeria omistatakse tihtipeale just lauset alustavatele diskursusemarkeritele (Hansen 1997: 156, Schourup 1999: 233), siis järjend *lt*-sõna-subjekt-verb (sõna topikaliseeritakse ilma subjekti ja verbi inversioonita) võiks mõnes kontekstis viidata just pragmaatilisele kasutusele. Diskursusemarkereid käsitletakse kui grammatiseerunud üksusi, mille lisamisel lause algusesse ülejäänud lause ehitus ei muutu.

Võrdleme kaht lauset, kus sõna *põhimõtteliselt* võib lugeda pragmaatiliseks markeriks. Mõlemas lauses on sõna semantilis-pragmaatiline tähendus muutunud, kuid lausete erinevus seisneb sõnajärjes. Nimelt näites (3) toimub sõna *põhimõtteliselt* järel aluse ja öeldise inversioon, kuid näites (4) mitte. Siinkohal on *põhimõtteliselt* näites (4) loetud rohkem grammatiseerunud diskursusemarkeriks, sest see on lisatud konteksti nii, et lause põhistruktuur ei muutu ning sõna *põhimõtteliselt* väljajätt ei muudaks lauset (4) ebaharilikuks.

- (3) Minul küll hämaras esinemisega probleeme pole. **Põhimõtteliselt** oskan ma klaverit mängida ka kinnisilmi.
- (4) Minul küll hämaras esinemisega probleeme pole. **Põhimõtteliselt** ma oskan klaverit mängida ka kinnisilmi.

Diskursusemarkerid on niisiis sõnad, mis struktureerivad diskursust. Modaalpartiklitega sarnaselt on need semantiliselt raskesti määratlevad üksused. Diskursusemarkerid toimivad aga pigem lauseüleselt kui lausesiseselt. Nad paiknevad tavaliselt lause alguses (vahel ka lõpus), seostades eri lauseid omavahel, et tekiks koherentne diskursus. On väidetud, et sellised ülejäänud lausega nõrgalt seotud üksused, mille tähendust on keeruline defineerida, pole samas kontekstis vahetatavad (Schourup 1999: 242–243): nende kommunikatiivne funktsioon võib olla sootuks erinev. Seega on need kontekstisidusad sõnad. Siinkohal on märkimisväärt, et

² Lähemalt vasaku ja parema perifeeria (ingl *left and right periphery*) kohta pragmaatiliste markerite käsitluses vt Traugott 2012, Beeching, Detges (ilmumas), Van der Wouden, Foolen 2011.

et *põhimõtteliselt* ja *praktiliselt* on vähemalt oma modaalpartiklilises tarvituses sarnased, st vahetatavad.

Järgnevalt ongi vaadatud uuritavate sõnade semantilis-pragmaatilist külge ning põgusalt ka süntaksit (adverbide ja pragmaatiliste markerite tunnusoonte loetelu vt Valdmets 2013: 126). Keeleüksuste arengu liikumapanevaks jõuks on tähendusnihe, seetõttu on esmalt kirjeldatud sõnade tähendust ning vaadeldud kasutusvõimalusi. Nii on täheldatud (Valdmets 2013), et adverbid kannavad konkreetset sisu ja neid on võimalik küsimustada, samal ajal kui markerid on abstraktsed küsimustamatud üksused. Määrsõnad on kontekstist sõltumatud, kuid pragmaatilised markerid võivad olla lauseümbrusega seotud (omandadeski tihtipeale tähenduse konkreetse ümbruses). *lt*-sõnad võivad lause propositsioonis osaleda (adverbid) või sellest välja jääda (markerid). Lisaks semantikale ja pragmaatikale on uuritavate sõnade puhul määrav ka süntaktiline käitumine. Adverbid on seotud üksused, samal ajal kui pragmaatilised markerid on süntaktiliselt enamasti nõrgalt seotud. Adverbidele, kuid mitte pragmaatilistele markeritele, on omane võime siduda laiendeid ja esineda lauseliikmena. Lisades siia ka foneetilise tasandi, võib märkida, et ennekõike pragmaatilise markeri funktsioonis on võimalik selliste üksuste lühenemine.

Artikli materjal pärineb Tartu Ülikooli eesti kirjakeele korpusest ja tasakaalus korpusest. Peamiselt on vaatluse all toimetatud ilukirjanduslikud ja ajakirjanduslikud tekstid, kuid kasutatud on ka toimetamata kirjakeele näiteid (uue meedia allkorpusest). Uuritav ajavahemik ulatub sõnade esimestest kasutusjuhtudest korpuses kuni praeguseni. *Põhimõtteliselt* ja *praktiliselt* oma tänapäevasel kujul vana kirjakeele korpuses ei esine. Küll on aga vanematest tekstidest võimalik leida nende sõnade tüvesid (sellekohased viited esitatakse edaspidi vastavate alapeatükkide juures). Siinse uurimuse eesmärgiks on ennekõike kvalitatiivne vaatlus nende sõnade kujunemisest ja kasutusest. Kvantitatiivsed andmed kirjakeele korpuse aja- ja ilukirjanduslike tekstide kohta esitatakse vaid taustinformatsioonina, kasutades andmeid siinkirjutaja magistritööst (Valdmets 2010).

3. *Põhimõtteliselt*

Põhimõtteliselt on sõna, mis adjektiivist *põhimõtteline* tuletatuna kannab tähendust 'põhimõttekindlalt', 'põhimõtteisse puutuvalt', 'põhimõtte pärast' ja 'printsipiaalselt'. Vana kirjakeele korpuses nimetatud sõnakujusid ei esine, küll aga annavad vanad tekstid aimu selle liitsõna arengust. Nii leidub Heinrich Gösekeni grammatikas ja sõnastikus liitsõna *pohhi* (aastast 1660, siin kasutatud Kingisepa jt 2010. aasta uustrukki) ning Anton Thor Hellel *pohhi* ja *mõtlik*, *mõtte* (aastast 1732, siin kasutatud Kilgi ja Rossi 2006. aasta uustrukki).

Põhimõtteliselt ilmneb oma tänapäevasel kujul esmakordselt 1930. aastate tekstides. Vanemast keelest (1890.–1910. aastate korpusest) on leitavad ka tänapäevasest erineva täiendsõnakujuga vormid nagu *põhjusmõttelikult* ja *põhjusmõtteliselt* (5). Hiljem saab valdavaks tänapäeval tuntud vorm.

- (5) **Põhjusmõttelißelt** ei olnud wiimaßel küll mitte eebtwõtja koha wastu midagi, aga tegelißelt beda enam. (ILU1910)

Sõna *põhimõtteliselt* esinemissagedus on kuni 1990. aastateni väga madal, umbes kümmekond tarvitajuhtu ühe aastakümne tekstides (täpseid arvandmeid vt Valdmets 2010: 44). Viimastel aastakümnetel on sõna kasutus suurenenud, 1990. aastate tekstides ulatub kasutus üle poolesaja juhu, 2000. aastatel on see muuhulgas korpuse mahu suurenemise tõttu sagedam kui kogu 20. sajandi peale kokku. Ka normaliseeritud andmeid vaadates on võimalik märgata ühtlast kasutussageduse kasvu (Valdmets 2010: 75).³ Tõdedes, et keelendi sage kasutamine annab eelduse selle grammatisatsiooniks (Bybee 2003, Hopper 1998: 156–159), on lähtekoht uurimuseks olemas. Konkreetseid arvandmeid on siin kasutatud vaid taustinformatsioonina, keskendudes pigem sõnade kvalitatiivsele analüüsile.

Näites (6) on näidatud adverbiline kasutus, millele viitavad nii tähendus ('põhimõtte poolest'), positsioon lauses kui ka rõhusõna (*juba*).

- (6) Kärusid oli nelja liiki – eesli-, härja-, traktori- ja inimkärud, kusjuures viimased erinesid esimesest *juba* **põhimõtteliselt**. (OR2002)

Olgu siinkohal toodud ka teine näide adverbilisest kasutusest (7), kus konkreetne tähendus ilmneb ümbritsevast kontekstist. Siinses näites kannab *põhimõtteliselt* tähendust 'põhimõtte pärast', kuna autor kirjeldab inimesi, kes on oma põhimõtetelt vägivallavastased.

Teades, et uuritava üksusel on eesti keeles ka teine, laiendav funktsioon, võiks *põhimõtteliselt* siinses lauses olla analüüsiv ka *iga* laiendina, kollokatsioonis *põhimõtteliselt iga* 'peaaegu iga'. Selline tõlgendus pole aga väga tõenäoline ümbritsevat konteksti arvesse võttes. Siinne näide on huvitav veel teiselgi põhjusel: nimelt alustab lauset pragmaatiline marker *muidugi*, mis on iseseisva üksusena ülejäänud tekstist komaga eraldatud.

- (7) Muidugi, need, kes **põhimõtteliselt** iga vägivalda vastu räägivad, ei tea, mis nad räägivad. (ILU1938)

Edaspidi on toodud näited, kus uuritava sõna funktsioon liigub adverbilisest kasutusest pragmaatilise suunas. Funktsioonimuutus on enim märgatav kontekstides, kus sõna *põhimõtteliselt* algne tähendus ei ole enam võimalik, selle kui pragmaatilise markeri sünonüümiks võivad olla *peaaegu*, *üldiselt* jmt. *Põhimõtteliselt* pragmaatiline tähendus ilmneb selgelt siis, kui tekib semantiline vastuolu, nagu on märgata lauses (8).

- (8) Peab kohe ütlema, et Vahemere kultuuriruumis pole selles **põhimõtteliselt** midagi skandaalset (peaasi on säilitada macho hoiak), kuid tema balansseeris siiski skandaali piiril. (ET2002)

Siin modifitseerib *põhimõtteliselt* pronoomenit *miski*, kuid sage on see ka koos näitava omadussõnalise asesõnaga, nt *sama* või *samasugune* (*põhimõtteliselt sama* esines nt Eesti Päevalehe 2007. aasta kogus viies lauses, kokku kasutati sõna *põhimõtteliselt* selle aastakäigu tekstides 101 korral). Fraas *põhimõtteliselt midagi* on siinkohal välja toodud, kuna ilmneb huvitav asümmeetriline kollokatsioon: seda fraasi kasutatakse eitavas lauses sagedasti, kuid juhtumit, kus pronoomen *miski* oleks nominatiivis ja kasutusel jaatavas lauses (**põhimõtteliselt miski*), korpuses ei leidu. Lauses (8) on *põhimõtteliselt* fraasilaiend ning see on süntaktiliselt seotud

³ Sõna *põhimõtteliselt* kasutussagedus 100 000 korpuse sõna kohta on esimesel esinemisel kirjakeele korpuses (1930. aastatel) 1,08; 2000. aastatel on see tõusnud 4,54 sõnani (Valdmets 2010: 120).

üksus. Samal ajal on aga sõna semantiline kese nihkunud uue kasutuse suunas ning üksus on kaotanud võime paikneda lauses rõhulises positsioonis. Sellise olukorra tingib keeles toimuv pidev protsess, kus konkreetse üksusega liituvad lisatähendused. Lõpuks pääseb mõjule lisatähendus, millest saab mingis kontekstis sõna uus semantiline kese. *Põhimõtteliselt* näites (8) on loetud modaalpartiklite hulka, mis pole küll täielikult oma funktsiooni muutnud, kuid on grammatiseerumas.

Lauses (9) esineb pragmaatiline marker dialoogis vastusena koos sõnaga *ei*. See funktsioneerib episteemilise partiklina, mis pehmendab öeldut ning väljendab kõneleja ebakindlust, vältides konkreetset eitavat vastust küsimusele. Toodud on ilmekas näide selle kohta, et teatud olukordades võib ka modaalpartikkel esineda lausealguselises positsioonis. Kõneleja on otsustanud otsekohesuse vältimiseks kasutada modaalsõna *põhimõtteliselt*, selle asemel et vastata konkreetse eitusega. Sama partiklit oleks saanud kasutada ka jaatava konstruktsiooni puhul (*põhimõtteliselt jah / põhimõtteliselt küll*). *Põhimõtteliselt* võib olla partiklina ka üksikliikmena üldküsimuse vastuseks. Kui aga sõna kasutatakse eriküsilausele vastamiseks (nt kui küsimus algab sõnaga *kuidas*), siis funktsioneerib *põhimõtteliselt* eranditult adverbina. Näitega (9) sarnast kasutust on võimalik leida ka *praktiliselt* puhul, vt lauset (21) allpool.

- (9) – Kas te ei ole ühessegi oma töösse siiamaani armunud?
– **Põhimõtteliselt** ei. Ma tahan ikka ja jälle uut, uut ja uut. (AJA1990)

Lauses (10) on sõnal pehmendav funktsioon ning *põhimõtteliselt* liigub modaalpartikli ja diskursusemarkeri äärealal, kuid kuulub sellegipoolest pragmaatiliste markerite hulka.

- (10) fun|tegeleb: a lincy|s2tib millal sa liikuma saad hakata? [---]
lincy|s2tib: kle kohe **põhimõtteliselt** (JT)

Kasutades sõna *põhimõtteliselt*, võtab kõneleja tagasi öeldu sirgjoonelisust. Jututoa kasutaja lincy|s2tib ütleb alguses, et ta hakkab kohe liikuma, kuid seejärel lisab partikli *põhimõtteliselt* näitamaks, et ta sel konkreetset hetkel siiski ei stardi. Modaalpartikkel *põhimõtteliselt* on lauses (10) järelasendis, mis reaajas toimivas suhtluses on tihtipeale kasutusel seisukohavõtu markeerimiseks (Keevallik 2003: 74–76, sh arutus markeri *ma arvan* kohta). Reaajas toimivas suhtluses, nt telefoni- või jututoavestluses, lisatakse väitele tihtipeale pehmendus, mis vähendab väite jõulisust ning muudab fakti vähem kindlaks. Näites (10) väldib jututoa küllaline lincy|s2tib partikli *põhimõtteliselt* abil otsekohesust ja seetõttu on võimalik tuua funktsionaalne paralleel ka järelasendis kasutatud episteemilise üksusega *ma arvan*.

Lauses (11) on aga esile kerkimas uus, diskursusefunktsioon: *põhimõtteliselt* märgib muutust diskursuses. Nimelt on see analüüsiv sõnana, mis ei funktsioneerid üldlaiendina, vaid pigem on sellel lauseteülene haare. Seega on üksus interpreteeritav diskursusemarkeriks, mida iseloomustab vaid marginaalne modaalsus ja esile on kerkinud uus, infot siduv või laiendav funktsioon. Täielikult väljakujunenud juhul on diskursusemarker lisatud lause algusesse nii, et see ei muuda järgneva lause ülesehitust. Diskursusemarker on üksus, mis ühildab eelnevalt väljendatud järgnevaga.

- (11) Esiteks on neis riietes palav. Teiseks on need kleidid nii pikad, võimatu on graatsiliselt kõndida. **Põhimõtteliselt** koperdan ja komistan ma iga sekundi tagant. (KR2003)

Näites (11) on *põhimõtteliselt* lausega süntaktiliselt seotud, kuna uuritavale sõnale järgneb verb ning sellele subjektpronoomen *ma* (eeldades tavapärasest jaatava lause SV-sõnajärge; sõnajärje kohta vt ka Aijmer 1997). Seega võib öelda, et uuritaval sõnal on veel ühiseid jooni adverbilise kasutusega, kuna sõna *põhimõtteliselt* oleks raske tekstist välja jätta: see on seotud üksus, mille eemaldamine tekitaks lauses vastuvõetamatu sõnajärje (*Teiseks on need kleidid nii pikad, võimatu on graatsiliselt kõndida. *Koperdan ja komistan ma iga sekundi tagant.*). Sõna alg tähendus on aga kadunud ning esile on kerkinud uus interpreteerimisvõimalus: sõnaga *põhimõtteliselt* näitlikustatakse seda, mis on ebagraatsiline kõndimine. Seda võib pidada nii põhjusseose kui ka selgitava-laiendava funktsiooni kandjaks. Kuigi diskursusemarkerid üldjuhul modaalsust ei kanna, on nendel siiski aeg-ajalt marginaalne modaalsus hoomatav. Nii ka ülaltoodud näites, kuna iga sekundi tagant komistamine on ilmne liialdus.

Järgnevas näites esineb *põhimõtteliselt* samuti diskursusemarkerina. See ühendab kaht lauset, sidudes eelneva järgnevaga. Sõnal puudub rõhutav funktsioon, kuid märgatav on siduv rakendus.

- (12) 15. koha saanud Kirsipuule on tänu võlgu McEwen, kes tema tuulest esimesena finišijoone ületas. “Üritasin oma taktikaga läbi ajada. **Põhimõtteliselt** see ka õnnestus, kuid kahjuks sattusin grupi ette liiga vara ja viiskümmend meetrit enne lõppu sain kangi. [---]” (EPL2005)

Näites (12) pole *põhimõtteliselt* lausega süntaktiliselt seotud ning see ei põhjusta ka subjekti ja verbi inversiooni. Seega leidub korpusel kasutusjuhte, kus sõna on juba selgelt oma algsest kasutusest eemaldunud, st arenguteel adverbist pragmaatiliseks markeriks on siin uuritavad üksused partiklistunud. Sõnad on minetanud oma algse ülesande ning neid kasutatakse uutest funktsioonides, nii ka lauses (12), kus autor püüab anda oma tegevusele hinnangut. Alustades teist lauset sõnaga *põhimõtteliselt*, püüab autor näidata, et tema õnnestumine teatud ettevõtmises ei olnud täielik, kuid siiski valdav. Kasutades sõna *põhimõtteliselt*, selgitab ja laiendab autor eelöeldut ehk täpsemalt seostab sellega kaht lauset, kuid see, kas tegemist on laiendusega, selgub alles järgnevast hinnangust õnnestumise kohta. *Põhimõtteliselt* toimib ka siinses näites diskursuse tasandil ning see üksus on võimalik lausest välja jätta. Võrreldes diskursusemarkereid näidetes (11) ja (12) võib märkida, et semantika põhjal otsustades on *põhimõtteliselt* selgemini markerina kasutusel näites (11), süntaksi põhjal otsustades aga näites (12), kus markeri olemasolu või puudumine ei muuda lause üldist struktuuri. Erinevalt näitest (11) oleks diskursus näites (12) ka ilma sõnata *põhimõtteliselt* grammatiliselt korrektne: *Üritasin oma taktikaga läbi ajada. See ka õnnestus, kuid kahjuks sattusin grupi ette liiga vara ja viiskümmend meetrit enne lõppu sain kangi.* Selline sõnajärg mõjuks aga diskursuses ebaloomulikumalt ning kaoks ka hinnangu marginaalne pehmendav interpretatsioon.

Kõnelähedases kirjakeeles on levinud ka lühenenud vormid, mis funktsioneerivad vaid pragmaatiliste markeritena ega esine kunagi täistähenduslikena. Näiteks võib jututubadest, foorumitest jm toimetamata tekstidest leida vorme nagu *põhimõtselt*, *põmst*, *pmst*, *pms* jne. Järgnevalt on välja toodud kaks näidet korpusel

jututubadest, kus on kasutatud diskursusemarkereid, mis haakuvad kirjakeele analüüsiga. *Põhimõtteliselt* (õigemini selle lühenenud vorm) korraldab siinkohal diskursust. Lausetes (13) ja (14) alustavad kasutajad truggy ja modegirl voo ru sõnaga *põhimõtteliselt*, sidudes selle abil eelnevat järgnevaga.

Näites (13) funktsioneerib *põhimõtteliselt* üksusena, mis alustab summeeriva sõnana vastust eelnevatele voo ru dele. Siinse *pmst* kasutuse võib lugeda veel marginaalseks modaalsuse väljendajaks, kuid esmajoones on siiski tegemist diskursusemarkeriga. Kasutajad snow ja kagu küsivad väikese vahega üldjoontes sama küsimuse, millele kasutaja truggy mõlemal korral reageerib. truggy ei vasta küsimusele otse, vaid formuleerib selle ümber, kasutades küsijatega sama verbi. Markeriga *pmst* annab truggy mõista, et ta ei vasta küsimusele nii, nagu see on esitatud, vaid annab olukorrale oma hinnangu, koondades selle kokkuvõtva sõna *pmst* järele.

- (13) kagu: ainult vaikus
snow: tore
snow: seegi
kagu: :)
kagu: jah tore
truggy: suht ei ole vaikus hetkel
truggy: on vaiksemaidki hetki siin olnd
snow: jaja mida iganes
snow: mis te siis r22gite
snow: ka
kagu: ei tea pole sin enne käinud
truggy: rääkida võib kõigest, enam vähem :P
kagu: tore
truggy: jaaa
truggy: mis teoksil teil ka?
kagu: millest te siin räägite ka?
truggy: **pmst** siin räägitakse kõikidel teemadel üldiselt :P (JT)

Lauses (14) tekitatakse selge vastandus SScripti kirjutatule ja *põhimõtteliselt* kasutatakse siduva sõnana. modegirl alustab voo ru diskursusemarkeriga selleks, et pöörduda tagasi eelneva vaidluse juurde. Laiemast kontekstist jääb selgusetuks, kas modegirl on kohtumõistja või vaidluses osaleja. modegirl ründab SScripti, kasutades selleks käibefraasi (*vanematega ei vaielda*). Tegemist on pöördega diskursuses, kus eelnev kontekst seostatakse järgnevaga ehk modegirl seob diskursusemarkeri *põhimõtteliselt* abil teise inimese voo ru enda omaga.

- (14) plika_tartust: script palun suundu istmikusse
j22karupoeg: tubli tubli :P
citty: ei ei ei sul asjast täitsa vale arusaam ikka sina oled
modegirl: <plika_tartust> script palun suundu istmikusse <<vapsee viisakas
SScript: hahaaaaa targem annab järele järelikult oled sina
SScript: hihiiii
SScript: =)
SScript: vot nii

Henry: tre
modegirl: **põhimõtselt**, SScript, vanematega ei vaielda, seega ole tasa
(JT)

Viimastes näidetes on toodud juhtumid, kus *põhimõtteliselt* on võimalik lausest välja jätta, ilma et üksiklause tasandil miski muutuks, s.h lause propositsioon teiseks. Kui lausetest (10), (12), (13), (14) jätta *põhimõtteliselt* välja, siis lause üldine struktuur ei muutuks, kuid lause kaotaks nüansse pragmaatilisest vaatepunktist. *Põhimõtteliselt* on aga lausetesse lisatud põhjusega ning sõna täidab nendes kontekstides kindlat funktsiooni.

Eelnevas on näidatud sõna kujunemislugu viimase sajandi jooksul, kusjuures välja on toodud areng adverbist pragmaatiliseks markeriks. Nende üksuste viisimääruslikule kasutusele on lisandunud määramääruslik nüanss. Sellist üksust on võimalik interpreteerida samal ajal modaalpartikliks. Samas on välja kujunenud ka diskursuseülene funktsioon, muutes üksuse veelgi abstraktsemaks. Muutus konkreetsest adverbiks abstraktseks markeriks on tüüpiline kujunemisahel, mida on välja toodud ka teiste keelte näitel (vt Traugott, Dasher 2002).

4. *Praktiliselt*

Praktiliselt, nagu ka *põhimõtteliselt*, on sõna, mis pärineb tuletatud adjektiivist ehk sõnast *praktiline*. F. J. Wiedemanni eesti-saksa sõnaraamatus leidub adjektiiv kujul *praktiline* : *praktilise* (Wiedemann 1923 [1869]). *Praktiliselt* tähendab 'tegelikkuses, praktikas' ja võib olla sünonüümne sõnaga *otstarbekalt*. Korpuses on sõna esmakordselt 1900. aastate tekstides. Kuigi esinemissagedus on väike, näitab materjal, et 20. sajandi esimesel poolel kasutatakse seda vaid viisiadverbina (Valdmets 2010: 32). Nagu ka sõna *põhimõtteliselt* puhul, on konkreetseid arvandmeid *praktiliselt* kohta kasutatud vaid taustinformatsioonina ja keskendutakse pigem kvalitatiivsele analüüsile. Olgu aga siinkohal välja toodud, et alles 20. sajandi teisel poolel on võimalik sõna *praktiliselt* leida uues ehk pragmaatilises funktsioonis. 1960. aastatel esineb see esimest korda ka ilukirjanduses, sinnamaani tuli see välja vaid korpuse ajakirjandustekstides.

Praktiliselt esineb esmakordselt 1900. aastatel vaid ühes näitelasuses, kus see on kasutusel viisiadverbina (15). 1930. aastate korpuses leidub teinegi üksik ajakirjandusest pärit täistähenduslik kasutus (16). Sajandi esimesel poolel on sõna kasutusel vaid adverbina, esitatud näide (17) on pärit 1950. aastatest.

- (15) Rahwaid näemegi selle kallal nägeliselt (teorialiselt) ja tegeliselt (**praktiliselt**) tööol olema. (AJA1900)
- (16) Sinagi lahkelt tunnustad, et mu selged teoreetilised vahetegemised olevat kenad lugeda – aitäh sellegi eest! – kuid väidad, et ma **praktiliselt**, laval ajan asjad untsu. (AJA1930)
- (17) 32 valla- ja külanõukogu täitevkomitee sekretäri Viru-, Harju-, Lääne- ja Hiiu maalt sõitis tagasi oma kodukohtadesse, et **praktiliselt** rakendada omandatud teadmisi täitevkomitee igapäevases töös. (AJA1950)

Edaspidi kirjeldatakse sõna modaalseid kasutusjuhte. Modaalne tarvitus ilmneb 20. sajandi teisel poolel. Arvestades aga vähest esinemissagedust korpuses (1900. ja 1930. aastatel ühes lauses, 1950. aastatel viies lauses ja 1960. aastatel 14 lauses), ei saa kauguleulatuvaid järeldusi teha. Kuigi üldine kasutussagedus korpuses suureneb aasta-aastalt, ei esine see sõna korpuses kuni 21. sajandi alguseni kuigi tihti (sarnane tendents keelendiga *põhimõtteliselt*, täpseid arvandmeid vt Valdmets 2010: 32).⁴ Kõnelähedasemas kirjakeeles ehk jututubade materjalis on *praktiliselt* siiski sagedamini esindatud. Samas on toodud taas uuema aja andmed.

Alustame paari näitega modaalpartiklilisest kasutusest. Lausetes (18), (19) ja (20) on *praktiliselt* loetud partiklite hulka, kuid modaalsus ei ole siinkohal alati läbinähtav. Pigem märgib selline kasutus nihet algsest täistähenduslikust kasutusest eemale, nii et *praktiliselt* funktsioneerib uues ümbruses fraasi- või lauselaiendina. On öeldud, et *praktiliselt* kui modaalne üldlaiend kipub tänapäevases keelekasutuses välja tõrjuma modaalsõnu *tegelikult*, *õigupoolest*, *peaaegu*, *enam-vähem* (Erelt 2006: 74).

Lausetes (18) ja (19) on sõna kasutatud fraasilaiendina. Näites (18) asetub *praktiliselt* küll lause algusesse, kuid siin ei seostu see diskursusega (täpsemalt vt Valdmets 2013), üksus funktsioneerib sõna *kõik* laiendina,⁵ näidates, et enam-vähem kõigis sadamates on töö katkenud. Samamoodi nagu lausetes (19) ja (20), on ka näites (18) esile tõstetud sõna, mille tähendus on nihkunud pragmaatilisema kasutuse suunas ning mida enam ei saa interpreteerida kui üksust, mis näitab asja olemust tegelikkuses või praktikas nagu näidete (15), (16) ja (17) korral. Tegemist on pigem metafoorse kasutusega, mis ilmselt tundub keelekasutajale ilmekam. *Praktiliselt* omandab lõpuks sama funktsiooni, nagu on sõnadel *peaaegu* või *enam-vähem*. Ka näitelause (19) asetub modaalpartikkel fraasi algusesse (*praktiliselt ainukene koht*). Lauses (20) kuulub *praktiliselt* pigem verbifraasi, asetudes verbi ette. Siiski leidub korpuses ka selliseid juhtumeid, kus *praktiliselt* paikneb pärast verbi. Enamasti asetuvad partiklid aga peasõna vahetusse lähedusse. See on tüüpiline ka teiste pehmendajate puhul. Lauses (21) on *praktiliselt* episteemiline partikkel, vt arutlust näite (9) juures.

- (18) **Praktiliselt** kõigis Inglise sadamates on töö katkenud. (AJA1970)
- (19) Talvistel aegadel on trepikoda **praktiliselt** ainukene koht, kus saab magada rahu ja vaikus. (ILU1990)
- (20) Ometi üle rahvuspiiride minevad kontaktid **praktiliselt** puuduvad. (AJA1990)
- (21) – Kas sa linna vahel ka kuskil liigud?
– **Praktiliselt** ei. (EE1996)

Lisaks modaalpartiklilaadsetele juhtudele leiab sõna *praktiliselt* ka diskursusemarkeri kasutust, millele näites (22) viitab lausealguline positsioon ning funktsioneerimine diskursuse korraldajana ehk lauseteülese üksusena. Sõna tähenduski on juba nihkunud, kuid süntaktiliselt ei ole see veel vabalt liikuv liige. Küsimus käib 11. septembril 2001 toimunud terrorirünnaku kohta. Lugeja EditX esitab küsimuse ning laiendab seda kohe järgneva lausega. Siinkohal funktsioneerib *praktiliselt* õigustusena eelnevalt öeldule. See täpsustab eelnevalt öeldut, samal ajal küsimust

⁴ Sõna *praktiliselt* kasutussagedus 100 000 sõna kohta on esimesel esinemisel kirjakeele korpuses (1900. aastatel) 0,43 sõna korpuse 100 000 sõna kohta, 2000. aastatel on see tõusnud 3,97 sõnani (Valdmets 2010: 120).

⁵ Fraas *praktiliselt kõik* (sealjures on *kõik* erinevates käänetes) jääb korpuse kõigis žanrites silma kõrge esinemissagedusega.

lahti mõtestades või parandades. Siiski pole see täielikult väljakujunenud diskursusemarker, kuna sõna süntaktiline käitumine on veel muutumata, st normaal-lause ehitust silmas pidades poleks võimalik *praktiliselt* siinsest lausest välja jätta (*Kuidas on võimalik, et USA meedia on nii halvasti informeeritud sündmuste käigust, ohvrite arvust jms? *Ketras USA meedia eile terve päeva jooksul samu kaadreid ja kordas samu tsitaate...*). Nimelt on *praktiliselt* lauses esimesel kohal, seejärel tuleb verb ning alles siis järgneb subjekt. Iseseisva diskursusemarkeri puhul oleks tavapärasem teises lauses sõnade järjekord *USA meedia ketras...* Seega on *praktiliselt* näites (22) kujunemisjärgus diskursusemarker.

- (22) EditX, 15:34 12.09.2001: Kuidas on võimalik, et USA meedia on nii halvasti informeeritud sündmuste käigust, ohvrite arvust jms? **Praktiliselt** ketras USA meedia eile terve päeva jooksul samu kaadreid ja kordas samu tsitaate, mis olid kõigil juba pähe kulunud? (EPL2001)

Näidetes (23) ja (24) funktsioneerib *praktiliselt* samuti diskursusemarkerina, kuid siin on kasutusel juba väljakujunenud üksus. Nendes lausetes on sõna järgnevaga vaid nõrgalt seotud ning üksus asub väljaspool lause süntaktilist struktuuri. Kui see lause algusest välja jätta, oleks süntaktiliselt määratav siiski tavapärane lauseehitus.

- (23) “Ega sellisest sitasest mersust rohkem välja ei pigistagi!”
“Võibolla ütlete, et seisite lausa paigal?” küsib GAI-mees.
“**Praktiliselt** ma seisingi paigal,” kinnitab mersujuht.
“Sellisel juhul viiskümmend dollarit keelatud kohas parkimise eest!”
(EE1997)

- (24) “Innar Mändoja kannatas tänase etapi poolhaigena lihtsalt lõpuni. Järgmise etapist jääb meid alles ainult viis meest, nii et liidrisärgi hoidmine saab raske olema. **Praktiliselt** suurt sõnaõigust meil enam ei ole.” (EPL2001)

Praktiliselt on lauses (23) kui abivahend, mis aitab voo alustada. Kuna eelnev on üldküsimus, siis eeldatav vastus on jaatus või eitus. *Praktiliselt* näitab kohe voo alguses, et vastus on modifitseeritud. Seega võib diskursusemarkeri kasutuses märgata marginaalset modaalsust, kuna mersujuhi vastus dialoogis näitab, et vastaja ei aktsepteeri küsimust tingimusteta ja vaidleb selle proportsioonile vastu (küsimus on selgelt irooniline ning esitab paigalseisu versiooni kui ilmvoimatut).

Näites (24) on tegu sama autori kahe järjestikuse lausega, kusjuures *praktiliselt* on siin lausete liigendaja diskursuses. See aitab luua suuremaid infolõike, kusjuures *praktiliselt* ei seostu siinkohal ainult vahetult eelneva lausega, vaid kogu tsitaadiga, ja selge on ka sõna kokkuvõttev funktsioon (vt ka näidet (13) *põhimõtteliselt* juures). Tegemist on kas autori tehtud pöördega monoloogis, kus püütakse eelnevalt öeldut leevendada, või siis jäetakse väide veel mõneti lahtiseks.

Nagu *põhimõtteliselt*, võib ka *praktiliselt* esineda toimetamata kirjakeeles lühenenult, kui kirjutaja soovib matkida kõnekeelset vormi. Nii saab *praktiliselt* olla kujul *praktse*, *prakse* vms. Sellised vormid on oma funktsioonilt sarnased eelnevalt kirjeldatud pragmaatiliste markeritega.

Järgnevates näidetes jututubade korpusest (25) ja (26) on esindatud mõlemad artiklis uuritavad üksused küll erinevates kontekstides ja erinevate inimeste kirjutatuna, kuid sarnases funktsioonis ja positsioonis. Sellised pragmaatilised markerid näitavad tõenäoliselt, et nende kasutajad määravad oma elukoha või

päritolu geograafiliselt mingi linna tasemel, kuid siis umbmäärastavad selle väite modaalpartikli abil. Teine seletus võiks olla ajalise määratluse umbmäärastamine. Mõlema seletuse järgi on sõna semantiliselt analüüsiv modaalpartikliks, mida saaks tõlgendada tähenduses 'enam-vähem' või 'enamasti, enamjaolt'.

(25) La_Femme_Nikita: ++heigo18+- kus elad?????

[---]

Heigo18: M;a elan Paides **praktiliselt!** (JT)

(26) msn_muzy_13: sa paidest we

kissy: sirtz ma trtust

murru: ma ka **põhimõtteliselt** (JT)

Siinkohal on partiklitel samasugune funktsioon nagu voorulõpulisel üksusel *põhimõtteliselt* näites (10) eespool. Nii kasutaja Heigo18 kui ka murru vastavad endale esitatud küsimusele. Näide (25) on vaid täislause, mis on vastuseks konkreetsele eriküsilausele. Näites (26) lõpetatakse sõnaga *põhimõtteliselt* voor, mis ilmselt vastab valikküsilausele. Siinses näites on täheldatav lünklause (*ma ka põhimõtteliselt*), kust paraku ei selgu, kas tegemist on reageeringuga kasutajale msn_muzy_13 või kissy. Ümbritsev laiem kontekst ei loo samuti selgust, kuna rohkete osalejatega vestluses tekib konkreetse vestluse raames palju ebakõla, mis omakorda viib jututoa kasutajate ärrituse ning ropendamiseni. Sellegipoolest saab siinkohal määrata sõna *põhimõtteliselt* modaalseks (vt ka sõna *põhimõtteliselt* analüüsi näites (9)).

Selliseid otseseid kattumisi *põhimõtteliselt* ja *praktiliselt* funktsioonide vahel võib korpusest leida veelgi. Seetõttu võib märkida, et mõlemad sõnad on teatud kontekstides võrdväärsed ja seetõttu ka vahetatavad.

Võttes arvesse semantikat-pragmaatikat, süntaksit ja foneetikat, on võimalik esitada mõned tähelepanekud sõnade kujunemise kohta. Algsest adverbilisest funktsioonist on esmajärgus välja arenenud üksuste modaalpartikliiline kasutus, kus fraasi- ja lauselaiendid on lausungisiselt tekstiga seotud. See kajastab sõnade umbmäärastavat funktsiooni. Kuna aga semantiliselt on nihe juba toimunud, siis edasine areng puudutabki tõenäoliselt sõnu, mis esmajärgus on küll oma täistähenduse juba minetanud, kuid süntaktiliselt võivad veel olla lausega seotud. Järgmine arenguaste puudutab markereid, mis funktsioneerivad diskursuseüleselt ega ole konkreetse lausega, mida nad alustavad, otseselt seotud. Sõna lühenenud kuju viitab kindlasti pragmaatilisele funktsioonile, kus adverbiline interpretatsioon enam võimalik ei ole.

Põhimõtteliselt ja *praktiliselt* on sõnad, mis on liikumas otsese tähendusega kasutusest pragmaatilise suunas. Kui *praktiliselt* erinevad tähendused on seletavas sõnaraamatus fikseeritud, siis sõnartikli *põhimõtteliselt* juures on seal nimetatud vaid üks tähendus ('põhimõtetelt, põhimõtte poolest; põhimõtte pärast'). Samas tuleb tõdeda, et mõned kasutusnäited paistavad siiski illustreerivat pragmaatilist kasutust, nt *Meie olukorrad on põhimõtteliselt sarnased* (EKSS). Eeldades, et nende kahe sõna arengusuund on sarnane teiste grammatiseeruvate adverbidega maailma keeltes, võime näha, kuidas neid kasutatakse tänapäeval sagedasti just pragmaatilises funktsioonis. Arvestades grammatiseerumise mehhanisme, võib eeldada, et aja jooksul üksuse lähtefunktsioon kaob ja alles jääb vaid sihtfunktsioon. Samas on tõdetud (Heine 1993: 48–49, Hopper, Traugott 2003: 49, 124–125), et paralleelne eksisteerimine võib kesta sadu aastaid ning kõik grammatiseeruvad üksused ei jõua

selleni, et mingi funktsioon täielikult kaoks. Ennustavalt võib arutleda selle üle, mis saab siis, kui artiklis uuritud sõnade lähtefunktsioon peaks ühel hetkel kaduma. Tõenäoline on, et sellisel juhul asendab keelekasutaja selle teise vormiga, mõne muu üksusega või kaob sootuks vajadus selle konkreetse sõna järele.

5. Kokkuvõte

Põhimõtteliselt ja *praktiliselt* on kujunemisjärgus üksused, mille algne kasutus hakkas mingil ajaperioodil (kirjakeele korpuse andmetel vastavalt 1930. ja 1960. aastatel) muutuma ning täistähenduslikule adverbilisele tarvitusele on lisandunud uusi, pragmaatilisi funktsioone. Seega ilmneb korpusematerjali põhjal, et keeles eksisteerivad kõrvuti nii adverbiline kui ka pragmaatiline tarvitus. Selle tõestuseks on artiklis esitatud nii süntaktilisi, semantilis-pragmaatilisi kui ka foneetilisi fakte.

Artiklis näidati, kuidas nimetatud üksused on kirjakeeles kasutusel kahes erinevas põhifunktsioonis: adverbilises ja pragmaatilises. Korpusematerjalile tuginedes kirjeldati pragmaatiliste üksuste edasist jagunemist modaalpartikliteks ja diskursusemarkeriteks. Kui enamjaolt arenevad *lt*-liitelised adverbid eesti keeles episteemilisteks partikliteks (vt nt Valdmets 2013 sõnade *lihtsalt*, *tegelikult*, *ilmselt* ja *loomulikult* kohta), siis *põhimõtteliselt* ja *praktiliselt* on erilised, kuna nende üheks funktsiooniks partiklina on umbmäärastamine (*põhimõtteliselt kõik*, *praktiliselt pool*). Selliseid fraasilaiendeid kasutatakse sageli kirjalikus diskursuses, kuid eeldatud on, et kirjalik eesti keel on mõjutatud suulisest eesti keelest. Seetõttu kaasati artiklisse näiteid ka kõnelähedasest kirjakeelest (interneti jututubadest).

Põhimõtteliselt ja *praktiliselt* on oma adverbilises kasutuses kaks erinevat leksikaalset üksust, mille semantiline väli ei kattu. Eesti kirjakeele korpuse põhjal on aga võimalik leida näiteid pragmaatilise kasutuse kohta, kus uuritavad üksused funktsioneerivad sarnaselt. Foneetiliselt on võimalik välja tuua mõlema üksuse lühenenud kuju kõnelähedases kirjakeeles. Süntaktiliselt on määratletavad üksused, mis adverbina on kindlalt lausega (verbiga) seotud. Adverbiaalina kasutatakse neid sõnu rõhulises positsioonis, kusjuures neile on võimalik lisada laiendeid. Kui sõna omandab uusi funktsioone pragmaatilise markerina, siis esmajärgus (süntaksist lähtuvalt) on see veel mõnevõrra seotud ülejäänud lausega, kuid sellisel juhul esineb see enamasti rõhutus positsioonis. Siinkohal on mõeldud sildkontekste, kus modaalpartiklid käituvad üldlaienditena, mh ka pehmendajatena (kindel positsioon peasõna vahetus läheduses) ja diskursusemarkerid paiknevad lause alguses, neile järgneb verb ning seejärel subjekt. Lauses vabalt liikuvad üksused on siinse käsitluse järgi sõnad, mida on võimalik diskursusest hõlpsalt välja jätta ilma propositsiooni muutuseta. Tegu on üksustega, millel on peamiselt diskursuseülene kasutus ja teksti liigendav funktsioon.

Semantiliselt on adverbid täistähenduslikud ning konkreetse sisuga, mida on kerge ka kontekstiväliselt määratleda. Samuti on adverbid küsimustatavad. Semantilisest-pragmaatilisest vaatepunktist on *põhimõtteliselt* ja *praktiliselt* järgmine arenguaste vaheetapp, kus algne adverbiline tähendus on veel mõneti tajutav, kuid tegelikult on võimalik näha ka uut, pragmaatilist funktsiooni. Edasise arengu käigus “täieõiguslikuks” modaalpartiklikuks või diskursusemarkeriks täistähendus taandub. Keelendite muutumise tulemuseks on abstraktsed hägusa tähendusega üksused,

mida pole võimalik küsimustada ega leksikaalse üksusena tõlgendada. Kõike eelnevalt arvesse võttes võib esitada grammatiseerimiskäiguhela, kus pragmaatilised markerid on pärit adverbilisest lähtest, kusjuures diskursusemarkeri funktsioon on rohkem grammatiseerunud kui modaalpartikli oma.

Kokkuvõtvalt võib öelda, et *põhimõtteliselt* ja *praktiliselt* on oma leksikaalselt lähtelt erinevad, kuid grammatiliselt arengukäigult sarnased üksused. Kuniks inimestel on vajadus oma väiteid ähmastada, selline pragmaatiliste üksuste areng ka jätkub. Seda tüüpi tugevdavatel ja pehmedavatel üksustel on eri keeltes kalduvus kiiresti vahetuda ning muutuda, mida ka artiklis käsitletud moesõnad ilmekalt näitavad.

Lühendid

EE	Eesti Ekspress
EPL	Eesti Päevaleht
ET2002	Emil Tode 2002, Raadio
ILU	ilukirjandus
JT	jutuoad
KR	seltskonnaajakiri Kroonika
OR2002	Olev Remsu 2002, Kevad Hiinas
S	subjekt
V	verb
X	objekt, adverbiaal, predikatiiv

Viidatud kirjandus

- Aijmer, Karin 1997. I think – An English Modal Particle. – Toril Swan, Olaf Jansen Westvik (Eds.). *Modality in Germanic Languages. Historical and Comparative Perspectives*. New York: Mouton De Gruyter, 1–47. <http://dx.doi.org/10.1515/9783110889932.1>
- Aijmer, Karin 2013. *Understanding Pragmatic Markers. A Variational Pragmatic Approach*. Edinburgh: Edinburgh University Press Ltd.
- Beeching, Kate; Detges, Ulrich (Eds.) (ilmumas). *The Role of the Left and Right Periphery in Semantic Change*. Amsterdam: Benjamins.
- Biber, Douglas; Johansson, Stig; Leech, Geoffrey; Conrad, Susan; Finegan, Edward 2007 [1999]. *Longman Grammar of Spoken and Written English*. Kuues parandatud trükk. Pearson Education Limited.
- Brinton, Laurel J. 1996. *Pragmatic Markers in English. Grammaticalization and Discourse Functions*. Berlin/New York: Mouton de Gruyter. <http://dx.doi.org/10.1515/9783110907582>
- Brinton, Laurel J. 2010. *Discourse Markers*. – Andreas H. Jucker, Irma Taavitsainen (Eds.). *Historical Pragmatics. Handbooks of Pragmatics 8*. Berlin/New York: Mouton de Gruyter, 285–314.
- Brinton, Laurel J.; Traugott, Elizabeth Closs 2005. *Lexicalization and Language Change*. Cambridge: Cambridge University Press. <http://dx.doi.org/10.1017/CBO9780511615962>
- Bybee, Joan L. 2003. *Mechanisms of change in grammaticization: The role of frequency*. – Brian D. Joseph, Richard D. Janda (Eds.). *The Handbook of Historical Linguistics*. Oxford: Blackwell, 602–623. <http://dx.doi.org/10.1002/9780470756393.ch19>
- Erelt, Mati 1986. *Eesti adjektiivisüntaks*. [The Syntax of Estonian Adjectives.] Tallinn: Valgus.
- Erelt, Mati 2006. *Lause õigekeelsus. Juhatused ja harjutused*. [Sentence Norms. Instructions and Exercises.] Tartu.
- Fraser, Bruce 1996. *Pragmatic Markers*. – *Pragmatics*, 6 (2), 167–190.

- Hansen, Maj-Britt Mosegaard 1997. Alors and donc in Spoken French: A reanalysis. – *Journal of Pragmatics*, 28 (2), 153–187. [http://dx.doi.org/10.1016/S0378-2166\(96\)00086-0](http://dx.doi.org/10.1016/S0378-2166(96)00086-0)
- Heine, Bernd 1993. *Auxiliaries. Cognitive Forces and Grammaticalization*. New York/Oxford: Oxford University Press.
- Heine, Bernd 2002. On the role of context in grammaticalization. – Ilse Wischer, Gabriele Diewald (Eds.). *New Reflections on Grammaticalization. Typological Studies in Language* 49. Amsterdam/Philadelphia: Benjamins, 83–101. <http://dx.doi.org/10.1075/tsl.49>
- Heine, Bernd; Narrog, Heiko 2010. Grammaticalization and linguistic analysis. – Bernd Heine, Heiko Narrog (Eds.). *The Oxford Handbook of Linguistic Analysis*. New York: Oxford University Press, 401–423.
- Hopper, Paul J. 1991. On some principles of grammaticization. – Elizabeth Closs Traugott, Bernd Heine (Eds.). *Approaches to Grammaticalization, Vol. I. Typological Studies in Language* 19:1. Amsterdam: John Benjamins, 17–36. <http://dx.doi.org/10.1075/tsl.19.1>
- Hopper, Paul J. 1998. Emergent grammar. – Michael Tomasello (Ed.). *The New Psychology of Language. Cognitive and Functional Approaches to Language Structure*. Mahwah, NJ: Lawrence Erlbaum Associates, 155–175.
- Hopper, Paul J.; Traugott, Elizabeth Closs 2003. *Grammaticalization. Second edition*. Cambridge: Cambridge University Press. <http://dx.doi.org/10.1017/CBO9781139165525>
- Huumo, Tuomas 1995. On the position of subject in Finnish and Estonian. – R. Kataja, K. Suikkari (Toim.). *XXI kielitieteen päivät Oulussa 6.–7.5.1994. Acta Universitatis Ouluensis, series B Humaniora* 19. Oulu, 79–87.
- Kaivapalu, Annekatrin 2010. Mõnede eesti sõnajäremallide psühholingvistilisest reaalsusest. [On psycholinguistic reality of some word order patterns of Estonian.] – *Eesti Rakenduslingvistika Ühingu aastaraamat*, 6, 103–120. <http://dx.doi.org/10.5128/ERYa6.07>
- Keevallik, Leelo 2003. From Interaction to Grammar: Estonian Finite Verb Forms in Conversation. *Studia Uralica Upsaliensia* 34. Uppsala.
- Kilgi, Annika; Ross, Kristiina 2006. Lühike sissejuhatus eesti keelde 1732. Saksa keelest tõlkinud ja järelsõnad kirjutanud Annika Kilgi ja Kristiina Ross. [A short introduction to Estonian 1732. Translation from German and epilogue by Annika Kilgi and Kristiina Ross.] Tallinn: Eesti Keele Sihtasutus.
- Kingisepp, Valve-Liivi; Ress, Kristel; Tafenau, Kai 2010. Heinrich Gösekeni grammatika ja sõnastik 350. [The grammar and dictionary of Heinrich Göseken 350.] Tartu: Tartu Ülikooli eesti ja üldkeeleteaduse instituut.
- Lindström, Liina 2005. Finiitverbi asend lauses. Sõnajärg ja seda mõjutavad tegurid suulises eesti keeles. [The position of the finite verb in a clause: word order and the factors affecting it in Spoken Estonian.] *Dissertationes philologiae estonicae Universitatis Tartuensis* 16. Tartu: Tartu Ülikooli Kirjastus.
- Quirk, Randolph; Greenbaum, Sidney; Leech, Geoffrey; Svartvik, Jan 1985. *A Comprehensive Grammar of the English Language*. London, NY: Longman.
- Schourup, Lawrence 1999. Discourse markers. – *Lingua*, 107 (3–4), 227–265. [http://dx.doi.org/10.1016/S0024-3841\(96\)90026-1](http://dx.doi.org/10.1016/S0024-3841(96)90026-1)
- Tael, Kaja 1988. Sõnajäremallid eesti keeles (võrrelduna soome keelega). [Estonian word order patterns (compared to Finnish).] Preprint KKI-56. Tallinn: Eesti NSV Teaduste Akadeemia Keele ja Kirjanduse Instituut.
- Traugott, Elizabeth Closs 2007. Discourse markers, modal particles, and contrastive analysis, synchronic and diachronic. – *Catalan Journal of Linguistics*, 6, 139–157.

- Traugott, Elizabeth Closs 2012. Intersubjectification and clause periphery. *Intersections of Intersubjectivity*. – Lieselotte Brems, Lobke Ghesquière, Freek Van de Velde (Eds.). Special Issue of *English Text Construction*, 5 (1), 7–28. <http://dx.doi.org/10.1075/etc.5.1>
- Traugott, Elizabeth Closs; Dasher, Richard B. 2002. *Regularity in Semantic Change*. Cambridge: Cambridge University Press.
- Valdmets, Annika 2010. Mõne modaalpartikli kujunemine ja kasutamine eesti kirjakeeles alates 1890. aastatest. [The development and use of some of the modal particles of written Estonian since the 1890's.] *Magistritöö. Käsikiri Tartu Ülikooli eesti keele õppetoolis*.
- Valdmets, Annika 2013. Modal particles, discourse markers, and adverbs with *lt*-suffix in Estonian. – Liesbeth Degand, Bert Cornillie, Paola Pietrandrea (Eds.). *Discourse Markers and Modal Particles. Categorization and Description. Pragmatics & Beyond New Series 234*. Amsterdam: John Benjamins, 107–132. <http://dx.doi.org/10.1075/pbns.234>
- Valdmets, Annika; Habicht, Külli 2013. Episteemilistest modaalpartiklitest eesti kirjakeeles. [About epistemic modal particles in Estonian.] – *Teoreetilises keeleteaduses Eestis III*. Ilona Tragel, Ann Veismann, Piret Piiraja (Toim.). *Eesti ja soome-ugri keeleteaduse ajakiri ESUKA*, 4 (1), 205–222.
- Van der Wouden, Ton; Foolen, Ad 2011. Dutch Particles in the Right Periphery. http://www.tonvanderwouden.nl/index_files/papers/fipa-2011-05b.pdf (16.12.2013).
- Wiedemann, Ferdinand Johann 1923 [1869]. *Estnisch-deutsches Wörterbuch: mit einer Karte Eestis*. Kolmas, muutmata trükk. Tartu: Eesti Kirjanduse Selts.

Võrgumaterjalid

- EKSS = Eesti keele seletav sõnaraamat. <http://www.eki.ee/dict/ekss/> (9.12.2013).
- Eesti kirjakeele korpus. <http://www.cl.ut.ee/korpused/kasutajaliides/> (27.7.2013).
- Tasakaalus korpus. <http://www.cl.ut.ee/korpused/grammatikaliides/> (27.7.2013).

Annika Küngas (Tartu Ülikool), peamine uurimisvaldkond on eesti kirjakeele pragmaatiliste markerite kujunemine ja kasutamine.
 Jakobi 2, 51014 Tartu, Eesti
annika.kungas@ut.ee

PÕHIMÕTTELISELT OR PRAKTILISELT IN PAIDE – TWO WORDS WITH SIMILAR FUNCTIONS

Annika Küngas

University of Tartu

This article offers a corpus-based overview of two Estonian words, namely *põhimõtteliselt* ‘in principle; basically’ and *praktiliselt* ‘practically; virtually’, illustrating their development and use mostly during the last hundred years. These two items may have various functions in the language. Used as adverbs, *põhimõtteliselt* and *praktiliselt* are words with different lexical fields. However, both words have undergone grammaticalization, as a result of which their functions have become similar. They are used more and more as pragmatic markers and therefore have developed from full-content adverbs into abstract items. As pragmatic words, they occur as modal particles or discourse markers. Therefore, they are mainly seen as downtoners or words that make the discourse more coherent. The data show that there have been changes in the way that these two words are used semantically, as well as syntactically and phonetically. Phonetically, these items are losing their lexical content; but this kind of erosion is only present with pragmatical uses. Syntactically, when used as adverbs, these words are bound to the sentence. They are stressed items that may get a modifier on their own. As pragmatic markers, these items are only loosely attached to the nearby discourse or not at all. Semantically, as grammaticalized words, *põhimõtteliselt* and *praktiliselt* have become fuzzier: now they are seen as words that have an abstract meaning while they are used as sentence or phrase modifiers. Alternatively, they may function on the discourse level to structure the text.

Keywords: pragmatic markers, adverbs, written language, corpus research, Estonian

ÜHENDVERBID EESTI KEELE PINDSÜNTAKTILISES ANALÜÜSIS

Kadri Muischnek, Kaili Müürisep, Tiina Puolakainen

Ülevaade. Artikkel käsitleb ühendverbide tuvastamist eesti keele automaatse pindsüntaktilise analüüsi käigus. Ühendverbide äratundmine on vajalik lause täpsemaks süntaktiliseks analüüsiks, sest lause osaliste süntaktilised funktsioonid, semantilised rollid ja nende keelendamine sõltub sellest, milline on lause keskmeks olev predikaatverb, sh sellest, kas predikaatverb on lihtverb või ühendverb.

Ühendverbide tuvastamiseks rakendatakse kahte strateegiat: leksikonipõhist ja reeglipõhist. Viimane tähendab seda, et osa korrapäraseid produktiivselt kombineeruvaid ühendverbe pannakse kokku reeglite abil. Artiklis kirjeldatakse kahte eksperimenti: esialgse ja täiustatud ühendverbide tuvastamise käiku. Täiustatud süsteemi tulemus on päris hea, saavutades saagise 97,4% ja täpsuse 96,6%.*

Võtmesõnad: arvutilingvistika, püsiühendite tuvastamine, eesti keel

1. Sissejuhatus

Meie igapäevane suhtlus toimub üha sagedamini elektroonsete kanalite kaudu, kogu loodav tekstiline info säilitatakse digitaalsel kujul ning seetõttu kasvab ka vajadus seda infot töödelda, parandada, tõlkida või tekstimassiivist infot otsida. Loomuliku keele töötleses mängib kesksel rolli süntaktiline analüüs, olles vahelülilis morfoloogilise analüüsi ja semantilise töötleses vahel.

Eesti keele automaatse süntaksianalüüsiga kitsenduste grammatika (ka piirangute grammatika, ingl *Constraint Grammar*) abil on kõige põhjalikumalt tegelenud Kaili Müürisep ja tulemused esitanud oma doktoriväitekirjas (Müürisep 2000). Süntaksianalüüsiaatori sealse variandi väljundis ei analüüsitud ühend- ja väljendverbe kui kokkukuuluvaid ja tervikuna lause keskmeks olevaid ühendeid. Ühendverbi afiksaaladverbiline komponent analüüsiti adverbiaaliks ja väljendverbi substantiivne komponent vastavalt tema käändevormile kas objektiks või adverbiaaliks. Selline lahendus oli esiteks predikaadi kui lause süntaktilise keskme tuvastamise

* Artikli valmimist on toetanud Euroopa Regionaalarengufond Eesti Arvutiteaduse Tippkeskuse kaudu ning Haridus- ja Teadusministeerium institutsionaalse uurimistoetuse IUT 20-56 "Eesti keele arvutimudelid" kaudu.

seisukohalt ebatäpne ning teiseks raskendas objekti tuvastamist. Sõltub ju eesti keeles lause aspekt (perfektiivne vs. imperfektiivne) ja seega ka objekti kääne sageli sellest, kas predikaadiks on lihtverb või ühendverb: *luges raamatut*, *luges raamatu läbi*, **luges raamatut läbi*, **luges raamatu*. Automaatses süntaksianalüüsis sõltub seega sellest, kas predikaadiks on liht- või ühendverb, nende sõnavormide hulk, mis võiksid saada selles lauses objekti märgendi. Ühend- ja väljendverbide probleemi lahendamise vajadus on välja toodud süntaksianalüsaatori edasiste arengusuundade all (Müürisep 2000: 96).

Selle probleemi ühe osa – ühendverbide analüüsi – lahendamisega tegelebki käesolev artikkel, mis on üles ehitatud järgmiselt. Kõigepealt, osas 2, käsitletakse ühendverbi kui keelenähtust; tegeldakse tema defineerimise ja piiritlemisega. Osas 3 antakse lühike ülevaade eesti keele morfoloogilisest ühestamisest ja pindsüntaktilisest analüüsist kitsenduste grammatika formalismi abil. Ühendverbide automaatsele tuvastamisele on pühendatud osa 4, mille alaosas 4.1 selgitatakse leksikoni ja reeglite rolli ühendverbide tuvastamisel, osas 4.2 esitatakse esialgse, lihtsama tuvastamisviisi töökäik ja tulemused ning alaosas 4.3 käsitletakse täiustatud tuvastamisviisi ning selle tulemusi. Osa 5 annab lühiülevaate sarnasest uurimistööst ning artikkel lõpeb kokkuvõtte ja tulevikuplaanidega.

2. Ühendverbi määratlemisest ja piiritlemisest

Ühendverbi on eesti keeleteaduslikus kirjanduses defineeritud kui verbist ja afiksaaladverbist koosnevat, lauses liitöeldisena toimivat ühendit, mille sisuliseks tuumaks on verb, mille tähendusele afiksaaladverb lisab oma nüansi (nt EKG II: 20, Erelt 2013: 62).

Alates Huno Rätsepast (1978: 28) jagatakse eesti keele ühendverbid kahte rühma: ainukordseteks ja korrapärasteks (nt Erelt 2013: 62 jj). Ainukordsed ühendverbid on idiomaatilised ühendid, mille tähendus ei moodustu nende osade tähendustest, vaid on omane ühendile tervikuna (nt *alla vanduma*, *üles ütleva*). Korrapäraste ühendverbide kõige tüüpilisemateks esindajateks on liikumisverbide ja suunda väljendavate afiksaaladverbide kombinatsioonid, mille tähendus moodustub komponentide tähendustest kompositsionaalsel moel, ühendverbi osad on säilitanud oma tähendusliku iseseisvuse (nt *alla / ette / juurde / järele / kokku / kõrvale / laiali / ligi / läbi / mööda / otsa / peale / sisse / välja / üle / üles / vahele astuma / jahutama / jooksmata / kõndima / lendama / lonkima*).

Ülalkirjeldatud rühmad on tüüpilised keeleteadusliku klassifikatsiooni näited selles mõttes, et nad esitavad iga rühma tüüpilisi esindajaid iseloomustavad tunnused. Loomulik keelekasutus on teadagi hästi kirjeldatav prototüübi mudeli abil, kus igas rühmas on kesksed eksemplarid, mis vastavad kõigile esitatud tunnustele, ja hajusatele äärealadele kuuluvad eksemplarid, mis ei vasta kõigile selle rühma tunnustele või “halbemal” juhul ühendavad endas kahe rühma tunnuseid. Keelesüsteemi kirjeldamiseks sobivad hajusate piiride ja kattuvate äärealadega kategooriad hästi, kuid automaatanalüüsil ja korpuste märgendamisel tuleb anda analüüs igale tekstisõnale; siin kirjeldataval juhul siis ära tunda kõik tekstis esinevad ühendverbid ning täpselt eristada ühendverbid verbi ja adverbi vabadest ühenditest. Nagu automaatanalüüs on ikka teatud määral lihtsustus, nii on ka piir ühendverbide ja mitte-ühendverbide vahel automaatsel analüüsil mõnevõrra suvaline.

Järgnevalt peatume lühidalt mõnel ühendverbide piiritlemise probleemil. Alustame *olema*-ühenditest osas 2.1, siis käsitleme korrapäraste ühendverbide kombineerumist osas 2.2. ning lõpuks arutleme afiksaaladverbi ja täistähendusliku adverbi piiri üle osas 2.3.

2.1. Kas verbi *olema* ja (afiksaal)adverbi ühendid on ühendverbid?

Üheks seni järjekindla lahenduseta probleemiks ühendverbide hulga piiritlemisel on *olema*-ühendid. Eesti keele grammatika (EKG II: 22) nimetab modaalsete ühendverbidena ühendeid *vaja olema* ja *tarvis olema* ning kuigi seda otse pole öeldud, jääb mulje, et muid *olema*-ühendeid ühendverbide hulka kuuluvateks ei peeta. Ka eesti keele seletussõnaraamatus (EKSS 2009) ei esitata verbi *olema* kirje juures ühtegi ühendverbi.

Seega käsitletakse tüüpiliselt ühendverbidena selliseid ühendeid nagu *alles hoidma*, *kinni panema*, *üle saama*, kuid mitte selliseid ühendeid nagu *alles olema* (*see paber on mul õnneks alles*), *kinni olema* (*pood on kinni*, *kõik on peas kinni*), *üle olema* (*meeskond on oma vastasest selgelt üle*), mis tundub *olema*-ühendite suhtes n-õ pisut ebaõiglane.

Teiselt poolt, kui ühendverb on defineeritud kui sõnaühend, mille põhitähendus tuleb verbilt, siis kas *kinni olema* põhitähendus tuleb ikka selliselt suhteliselt sisutihjalt verbilt nagu *olema*? Ning enamik muutumatu sõna ja verbi *olema* püsivaid ühendeid on sellised, kus muutumatu sõna ei klassifitseeru hästi afiksaaladverbiks kui pelgalt verbi tähendust modifitseerivaks sõnaks – nt *käsil olema*, *nõus olema*, *päri olema*, *rahul olema*, vt ka osa 2.3.

Praegu on automaatsel süntaksianalüüsil see probleem lahendatud nii, et ühendverbidena käsitletakse järgmisi *olema*-ühendeid: *alles olema*, *ära olema*, *kinni olema*, *kohal olema*, *käsil olema*, *lahti olema*, *nõus olema*, *päri olema*, *rahul olema*, *tarvis olema*, *vaja olema*, *valmis olema*, *üle olema*. Loend on ebajärjekindel ega ole kindlasti lõplik.

2.2. Korrapärased ühendverbid

Korrapäraste ühendverbide näidetena on ikka esitatud liikumisverbide kombinatsioon suunda väljendavate afiksaaladverbidega (Rätsep 1978: 28, EKG II: 21, Erelt 2013: 62–63). Nii moodustatud ühendverbide näitel saab aga ka veenduda, et teatud semantiliselt välja kuuluvate verbide ja afiksaaladverbide kombineerumine korrapäraste ühendverbide moodustamiseks pole siiski täisproduktiivne, olles seotud semantilist laadi kitsendustega, mida on aga raske formaliseerida.

Kui jääda liikumisverbide näite juurde, siis need jagunevad selgelt kaheks rühmaks. Enamik neist kombineerub tõesti vabalt suunda väljendavate afiksaaladverbidega, moodustades korrapäraste ühendverbe, nt *astus / hüppas / jalutas / kõndis / lendas / läks / marssis / roomas / sõitis / traavis / tormas / veeres / vonkles jne (kodust) välja / (teistest) ette / (trepist) üles või alla, (sillast) üle / (teistele) järele*. Ent väiksem osa liikumisverbe suunda näitavate partiklitega ei kombineeru, nt **hulkus / luusis / käis / seikles (kodust) välja / (teistest) ette / (trepist) üles*

või *alla* / (*sillast*) *üle* / (*teistele*) *järele* pole võimalikud kombinatsioonid. Varem mainitud verbide näol on tegemist direksionaalidega, teise loetelu verbid ei ole aga mitte direksionaalid, vaid liikumisverbid, mis ei spetsifitseeri liikumise algus- ega lõpp-punkti ega ka liikumisteed ja nii nad ei kombineerugi suunda väljendavate afiksaaladverbidega.

Ka afiksaaladverbide puhul eksisteerivad tüüpilised mõistemetafooriteooriale vastavad ülekanded konkreetsematest tähendusvaldkondadest (allikvaldkondadest) abstraktsematesse tähendusvaldkondadesse (sihtvaldkondadesse). Ruum ja selles liikumine on üks populaarsemaid allikvaldkondi, kohasuhteid saab kasutada ka näiteks ajast rääkides (Pajusalu 2009: 122 jj). Nii väljendab afiksaaladverb *ette* nii ruumi- kui ajasuhet, olemas on nii ühendverbid *ette astuma* / *hüppama* / *jooksma* / *jõudma* kui ka *ette aimama* / *helistama* / *kuulutama* / *teadma*. Ajasuhet väljendades võib *ette* samuti produktiivselt kombineeruda teatud semantilisse välja kuuluvate verbidega, nt teadmis- ja teatamisverbidega: *ette aimama*, *ette teadma*, *ette teatama*, *ette helistama*, *ette hoiatama*, *ette kuulutama* jne. See tähendab, et sagedased suunda väljendavad afiksaaladverbid võivad kombineeruda ka teistesse tähendusgruppidesse kuuluvate verbidega peale liikumisverbide. Sellise kombineerumise produktiivsuse kohta pole meil praegu aga piisavalt teavet.

Hetkel on süntaksianalüsaatori reeglistikus produktiivselt kombineeruvatena käsitletud lisaks liikumisverbidest (direksionaalidest) ja suunda väljendavatest afiksaaladverbidest moodustuvatele ühendverbidele veel mitut väiksemat hulka, täpsemalt vt osa 4.2.

2.3. Afiksaaladverb vs. täistähenduslik adverb

Afiksaaladverbi defineeritakse kui muutumatut sõna, mis lauses verbi juurde kuuludes annab sellele uue tähendusvarjundi või konkretiseerib verbi tähendust (EKG I: 33, Ereht 2013: 21). Huno Rätsep eristab väljendverbi ühendverbist ka selle põhjal, et väljendverbi nominaalsetel komponentidel “puudub ühendverbide adverbidele (*peale*, *sisse*, *alla* jne.) omane üldistatud tähendus.” (Rätsep 1978: 21)

Eksisteerib aga hulk verbi ja adverbi püsivaid ühendeid, mis ei ole “päris” ühendverbid selles mõttes, et nende adverbiline komponent ei ole redutseerunud mitteiseseisvaks ja mittetäistähenduslikuks (EKG I: 18) afiksaaladverbiks, vaid on täistähenduslik sõna, nt *hukka mõistma*, *silmitsi seisma*, *kõhuli* / *selili keerama* jms). Nende näidete kohta ei saa väita, et adverbiline komponent ainult täpsustaks või modifitseeriks verbiga väljendatud tähendust või et talle oleks omane üldistatud tähendus, pigem tuleneb adverbist põhiline osa ühendi kui terviku tähendusest.

Selliseid ühendeid märgendatakse süntaksianalüsaatori väljundis praegu ühendverbidena, sest kindlasti on tegemist püsiühendiga ning me ei näe hetkel põhjust postuleerida veel ühte adverbist ja verbist koosneva püsiühendi liiki lisaks ühendverbile.

3. Morfoloogiline ühestamine ja süntaksianalüüs kitsenduste grammatika abil

Eesti keele pindsüntaksi analüsaator põhineb kitsenduste grammatika formalismil (Karlsson jt 1995). Analüsaator on reeglipõhine, see tähendab, et tuvastamisreeglid on koostatud inimese poolt masinõppimist kasutamata. Analüsaator koosneb analüüsimootorist, mis on keelest sõltumatu, ja reeglite baasist. Eesti keele analüüsiks kasutatakse Lõuna-Taani Ülikoolis (*Syddansk Universitet*) väljatöötatud VISL-parserit.¹ Reeglite baas on jaotatud kaheks: morfoloogilise ühestamise reeglistik, mis valib morfoloogiliselt mitmestele sõnavormidele konteksti sobiva analüüsi, ning pindsüntaksi reeglid, mis lisavad igale sõnavormile tema süntaktilist funktsiooni näitava märgendi (vt Roosmaa jt 2003).

Analüsaator saab sisendiks morfoloogiliselt analüüsitud teksti, milles on sõnavormidele antud kõik võimalikud morfoloogilised tõlgendused (nt sõnavorm *või* võib olla sidesõna, nimisõna nominatiivis ja genitiivis või hoopis verbivorm). Morfoloogilise ühestamise reeglid eemaldavad konteksti suhtes sobimatud tõlgendused. Samal etapil toimub ka osalausepiiride esialgne määramine. Kui konteksti põhjal pole võimalik ühest analüüsi leida, jäetakse alles kõik sobivad märgendid.

Pindsüntaktilisel analüüsil lisatakse kõigepealt kõik sõnavormi grammatiliste tunnustega sobivad süntaktilised märgendid ning seejärel hakatakse iteratiivselt eemaldama neid, mis ikkagi lausesse ei sobitu (näiteks objekti märgend eemaldatakse partitiivis nimisõnalt, kui verbivorm eeldab totaalobjekti olemasolu või ei leidu osalause üldse transitiiivset verbi või on juba objekt tuvastatud ning antud sõnavorm ei ole selle tuvastatud objektiga koordineeritud jmt).

Analüsaatori antav märgendus on väga pindmine: tuvastatakse osalause subjektid, objektid, predikatiivid, aga ei eristata pea- ega kõrvallauseid, objekt ning verb ei ole omavahel ühendatud, ka atribuudid ei ole seotud põhjadega, vaid lihtsalt eristatakse ees- ja järeltribuute. Samuti puudub eri tüüpi adverbiaalide eristus. Sügavam süntaktiline analüüs on järgmise, sõltuvuspuid ehitava grammatikamooduli ülesanne.

Morfoloogilise ühestamise ja pindsüntaktilise analüüsi reeglite baasis oli 2013. aasta detsembri seisuga umbes 4500 reeglit. Eksperimendid 95 000-sõnalisel käsitsi märgendatud korpusega² andsid kogu pindsüntaktilise analüüsi saagiseks (kõigi automaatselt korrektselt leitud märgendite arvu suhe tegelikult korrektselt märgendite arvu) oli analüsaatori sellel versioonil 92,9% ning täpsuseks (kõigi automaatselt korrektselt leitud märgendite arvu suhe kõigi leitud märgendite arvu) 69,3%; vigu oli 7,1%

See tähendab et 7% sõnadest jäävad ilma õigest märgendist ning 30–31% pakutud märgenditest on kas üleliigsed või vigased. Oluline hulk mitmesustest tekib adverbiaali ning adverbiaalse atribuudi eristamatusest, vead tekivad aga eelkõige objekti, subjekti ja predikatiivi määramisel.

Näiteks lauses *Ta on neis linnades aastaid tsaariajal elanud* on sõnavorm *linnades* jäänud mitmeseks eestäiendi (vrd *turvistes mehi*) ja adverbiaali tõlgenduse vahel ning sõnavorm *tsaariajal* on jäänud mitmeseks järeltäiendi (vrd *mehi pildil*) ja adverbiaali vahel.

¹ <http://beta.visl.sdu.dk/> (30.9.2013).

² Testkorpus on osa käsitsi märgendatud korpusest veebiaadressil <http://www.keeletehnoloogia.ee/ekt-projektid/vahendid-teksti-mitmekihiliseks-margendamiseks/soltuvussyntaktiliselt-kasitsi-analuusitud-korpus> (16.2.2014).

Vigu subjekti, objekti ja predikatiivi määramisel põhjustab asjaolu, et nende süntaktiliste funktsioonide kodeerimiseks kasutatakse samu käändeid (v.a ainsuse genitiiv), lisakeerukuse toob see, et need käändevormid on paljudel sõnadel homonüümsed. Nii näiteks on lauses *Aastaid on ta korraldanud jõulukontserti*. sõnavorm *ta* saanud morfoloogilisel ühestamisel ekslikult genitiivi tõlgenduse ja süntaksianalüüsil seetõttu objekti analüüsi ning partitiivis nimisõna *jõulukontserti* on süntaktilisel analüüsil saanud subjekti tõlgenduse, st subjekt ja objekt on ära vahetatud.

Et tulemusi parandada, on vaja reeglitesse lisada rohkem leksikaalset informatsiooni ja lausemustreid, eelkõige vajab jätkuvalt täiendamist verbirektsioonide sõnastik.

4. Ühendverbide automaatne tuvastamine

Varasem lähenemine (Müürisep 2000) ei üritanudki tuvastada ühendverbe, kuid analüsaatori efektiivsuse parandamiseks ja puustruktuuri analüüsiks on nende leidmine muutunud väga oluliseks.

Selles peatükis vaadeldakse kõigepealt kahte strateegiat ühendverbide tuvastamisel: leksikoni ning kombineerimisreeglite kasutamist; kirjeldatakse lühidalt leksikoni koostamise protsessi, seejärel esitatakse kahe ühendverbide tuvastamise katse töökäik ja tulemused.

4.1. Leksikoni ning reeglite roll ühendverbide tuvastamisel

Ühendverbide tuvastamisel rakendatakse kahte strateegiat: leksikoni ning kombineerimisreegleid. Ühendverbide esialgne nimekiri loodi Paul Saagpaku eesti-inglise sõnaraamatu (Saagpakk 1992) põhjal ning selle nimekirja alusel koostati esialgne ühendverbide tuvastamise reeglistik. Kuna paljud ühendverbide koosseisu kuuluvad afiksaaladverbid võivad olla morfoloogiliselt mitmesed kaassõna või mõne nimisõnavormi suhtes (nt *juurde võtma*, kus afiksaaladverb *juurde* saab morfoloogilise analüüsi käigus ka kaassõna ning nimisõna aditiivi e lühikese sisseütleva käändevormi tõlgenduse), siis tuli esialgsete reeglite hulka lisada täiendavaid kitsendusi. Samuti tuli arvestada ühendverbide nominalisatsioonivõimalusega ning sellega, et sama afiksaaladverb võib kombineeruda erinevate verbidega, moodustades erinevaid ühendverbe.

Ühendverbide esialgset leksikoni on täiendatud eesti keele seletussõnaraamatu (EKSS 2009) andmetega ja korpuse käsitsi märgendamise käigus n-õ avastatud ühendverbidega, nt *olema*-ühenditega, täpsemalt vt osa 2.1. Leksikon kindlasti veel muutub.

Lisaks ühendverbide leksikonile on loodud ühendverbide tuvastamise grammatikas esialgu kaks suuremat gruppi reegleid korrapäraste ühendverbide tuvastamiseks: reeglid liikumisverbide ja suunda väljendavate afiksaaladverbide tuvastamiseks ning perfektivse afiksaaladverbiga *ära* seotud verbide reeglid.

Kuigi mitte kõik liikumisverbid ja suunda väljendavad afiksaaladverbid ei saa omavahel kombineeruda, võime arvutigrammatika koostamisel eeldada, et kui nad

juba samas osalauses koos paiknevad, siis nad ilmselt ka ühendverbi moodustavad. Erandiks on juhud, kus muutumatu sõna on morfoloogiliselt mitmene afiksaaladverbi ja adpositsiooni tõlgenduste vahel ning toimib konkreetsetes osalauses adpositsioonina. Nagu vigade analüüsi osast näha, on need juhud ka põhiline vigade allikas.

Perfektiveeriva afiksaaladverbi *ära* puhul tuleb arvestada kontekstiga, kus *ära* võib esineda ka imperatiivi eitava vormi koosseisus (nt *ära tee*) ning sel juhul ühendverbi ei moodustu.

Ühendverbide tuvastamise protsess toimub iteratiivselt. Esimestel etappidel üritatakse afiksaaladverbikandidaate ühestada afiksaaladverbi, pre- ja postpositiooni ning substantiivi tõlgenduste vahel, järgmisel sammul leitakse, millised selles osalauses esinevad afiksaaladverbid ja verbid omavahel kombineeruvad. Näiteks lauses *Koju minnes tuli saapal tald ära* võib afiksaaladverb *ära* moodustada ühendverbi nii verbiga *minema* kui ka *tulema*.

Ühendverbide tuvastamise reegleid kontrolliti 95 000-sõnalisel käsitsi süntaktiliselt märgendatud korpusel.³ Korpus esindab normeeritud kirjakeelt, koosnedes tasakaalus korpuse⁴ ilukirjandus-, ajakirjandus- ja teadustekstidest.

4.2. Tuvastamise esialgsed tulemused

Ühendverbide tuvastamise protsess algab morfoloogilise ühestamise etapis, kus afiksaaladverbi võimaliku tõlgendusega sõnal võib olla ka mitmeid teisi tõlgendusi: kas kaassõna (nt *alla*, *järele*, *vastu*, *üle*, *ümber*), nimisõna (nt sõnavorm *lahti* võib olla nimisõna *laht* vorm, sõnavorm *maha* võib olla nimisõna *maa* vorm), võimalikud on ka nii kaassõna kui ka nimisõna tõlgendused ühel sõnavormil (nt *ilma* (*ilm*), *juurde* (*juur*), *külge* (*külg*), *kõrval* (*kõrv*), *peale* (*pea*)); võimalik on ka omadussõna käändevormi tõlgendus (nt *heaks* (*hea*), *üleväl* (*ülev*)), verbivormi tõlgendus (nt *lahku* (*lahkuma*), *taha* (*tahtma*)), nii kaassõna kui ka verbivormi tõlgendused (nt *läbi* (*läbima*), *taga* (*tagama*)), nii nimisõna kui ka verbivormi tõlgendused (nt võib sõnavorm *välja* lisaks afiksaaladverbile olla nimisõna *väli* vorm või verbi *väljama* vorm). Mõnel sõnavormil on võimalik mitmesus afiksaaladverbi, omadussõna, nimisõna ja verbivormi tõlgenduste vahel (nt sõnavorm *täis* võib olla afiksaaladverb, omadussõna, nimisõna *täi* vorm ning verbi *täima* vorm; sõnavorm *valmis* võib olla afiksaaladverb, nimisõna *valm* ning verbi *valmima* vorm). Võimalikud mitmesused on veel afiksaaladverbi, kaassõna, nimisõna ning verbivormi tõlgenduste vahel (nt *ringi* (*ring*, *ringima*)), afiksaaladverbi, kaassõna, arvsõna, asesõna tõlgenduste vahel (nt sõnavorm *ühes* võib olla afiksaaladverb, kaassõna, arvsõna *üks* vorm või asesõna *üks* vorm) ning afiksaaladverbi, kaassõna, arvsõna, asesõna ja nimisõna tõlgenduste vahel (nt sõnavorm *ühte* võib olla afiksaaladverb, arvsõna *üks* vorm, nimisõna *üks* vorm või nimisõna *ühe* vorm).

Afiksaaladverbi ning muude tõlgenduste vahel ühestamisele pandi alus kit-senduste grammatika süntaksianalüsaatori morfoloogilise ühestamise moodulis (Puolakainen 2001), mida on käesolevas töös oluliselt täiendatud.

Kuna enamasti on ka võimalikku afiksaaladverbi tekstis ümbritsevad sõnavormid sel hetkel veel ühestamata, st kindlaks määramata võib olla kas kääne (eriti sage on mitmesus nominatiiv / genitiiv / partitiiv) või ka sõnaliik (adverb või

³ Testkorpus on osa käsitsi märgendatud korpusest veebiaadressil <http://www.keeletehnoloogia.ee/ekt-projektid/vahendid-teksti-mitmekihiliseks-margendamiseks/soltuvussuntaktiliselt-kasitsi-analuusitud-korpus> (16.2.14).

⁴ <http://www.cl.ut.ee/korpused/grammatikakorpus/> (16.2.14).

nimisõna, verb või nimisõna), siis alati ei õnnestu afiksaaladverbi sõnaliiki õigesti määrata. Näiteks, lauses *Gripiviirus liigub kah ringi* ühestatakse sõnavorm *ringi* ekslikult nimisõna *ring* aditiivi e lühikese sisseütleva vormiks.

Ühendverbide leksikoni kasutatakse abileksikonina morfoloogilise ühestamise etapil ja põhileksikonina sellele järgneval ühendverbide osade kokkuviimisel ning süntaktiliste funktsioonide määramisel. Sellise analüüsijada tulemuse saagis ühendverbide tuvastamise osas oli 79,3%, st 20,7% ühendverbidest jäi tekstis ära tundmata. 5,2% afiksaaladverbidest tuvastati ühendverbi osana, kuid pandi seejuures kokku vale verbiga. 1,7% pakutud afiksaaladverbidest ei osutunud tegelikult ühendverbi koostisosaks, st nad ei olnud tegelikult afiksaaladverbid ja sõnaliigi määramisel oli tehtud viga.

Enamik vigu tulenes puudulikust ühendverbide leksikonist ning morfoloogilise ühestamise etapil tehtud valedest valikutest afiksaaladverbi ja adpositsiooni tõlgenduste vahel, millele liitusid veel küllaltki sagedased ühestamisvead käändsõna nominatiivi, genitiivi ja partitiivi tõlgenduste vahel.

Puuduliku leksikoni näideteks on laused ..*jäi üle ära mõõta kulminatsiooni-kõrguste erinevusele vastav kaare pikkus maapinnal ning Järjest tugevamini kiirgava Päikese kiirgus hakkab eemale puhuma vesinikust ja heeliumist koosnevat gaasi*, kus esimeses lauses tuvastati ühendverb *ära mõõtma*, kuid ei tuntud ära ühendverbi *üle jääma* ja teises lauses jäi tuvastamata ühendverb *eemale puhuma*.

4.3. Ühendverbide täiustatud tuvastamine

Tulemuse parandamiseks võeti ette järgmised sammud. Kuna enamik vigu tulenes morfoloogilise ühestamise etapil tehtud ekslikest valikutest afiksaaladverbi ja adpositsiooni tõlgenduste vahel, siis selliste olukordade minimeerimiseks toodi sisse reeglid, mis pärast morfoloogilist ühestamist tegelevad just kaassõna ja afiksaaladverbi vahekorra korrigeerimisega. Morfoloogilise ühestamise käigus on selline valik raskendatud, kuna sel ajal on üldine mitmesus väga kõrge ning sageli pole veel teada lähimas naabruses olevate noomenite käänded ega mõnikord ka sõnaliik.

Näiteks lauses *Ameerika alustas ränkadest vastasseisudest rahvuslikul pinnal ja töötas läbi vere, pisarate ja seadusloome tänase rassiliselt, usuliselt ja rahvuslikult tolerantse riigini* .. osutus võimalikuks parandada sõna *läbi* analüüsi afiksaaladverbist (enne moodustati sellega ekslikult ühendverb *töötas läbi*) kaassõnaks, mis moodustab kaassõnafraasi *läbi vere, pisarate ja seadusloome*. Samuti lauses *Erinevalt veest võib jää tagant pealpressiva jäämassi survele liikuda isegi üle kõrgustike* õnnestub kaassõna *üle* õigesti seostada kaassõnafrasiks *üle kõrgustike*, kuigi lauses esinev verb *liikuda* võimaldab teoreetiliselt moodustada ühendverbi *liikuda üle*. Ning vastupidiselt lauses .. *selgus, et osooni lagunemise reaktsioonide efektiivsus jääb tekkimise omale alla* taastati *alla* õige afiksaaladverbi tõlgendus, mis moodustab ühendverbi *jääb alla*.

Teiseks lisati võimalus reeglite abil vabalt kombineerida ühendverbe liikumisverbist ja suunda väljendavast afiksaaladverbist. Sel viisil kombineeritakse näiteks ühendverbid *ette minema/tulema/hüppama/jooksma/kargama/kiirustama/käänama/langema* jne.

Samuti loodi eraldi reegel moodustamaks produktiivselt ühendverbe perfektiveeriva afiksaaladverbiga *ära*.

Produktiivselt moodustatavate ühendverbide gruppidega esitati veel kehaasendi muutmist väljendavad ühendverbid (afiksaal)adverbidega *püsti, pikali, istuli, põlvili, selili, külili, kõhuli* ning afiksaaladverbidega *kinni ja lahti* kombineeruvad verbid, näiteks *ajama, heitma, hüppama, laskma, laskuma, jääma, kargama* jne.

Omaette kombineerumisgrupina lisati väga sageli ühendverbi moodustavate verbidega *jääma, saama* ning *minema* (mis ka üksikverbidena on väga sagedased) kombineeruvad afiksaaladverbid.

Veel ühe täiendusena lisati reeglid, mis arvestavad nominaliseerunud ühendverbidega. Näiteks lauses *Siis tuleb harjuda lühikesest basseinist pikka üle minemisega* suudetakse nende abil ära tunda ühendverbi *üle minema* nominalisatsioon.

Selle tuvastamismeetodi saagis on praegu 97,4% ehk kõikidest afiksaaladverbidest jääb tuvastamata 2,6%, seega eelmises osas kirjeldatud lihtsa meetodiga võrreldes tõusis saagis 18% võrra, mis on väga hea tulemus. 2,3% afiksaaladverbidest küll tuvastati afiksaaladverbina, kuid antud lause kontekstis pandi kokku vale verbiga. Ülegenereerimise protsent on 3,4, st lisaks igale sajale nende reeglite abil tuvastatud ühendverbile pakkus analüsaator veel 3,4 ühendverbi analüüsi, mis ei olnud tegelikult ühendverbid.

Osa vigadest on tingitud endiselt vales valikust kaassõna ja afiksaaladverbi morfoloogiliste tõlgenduste vahel. Näiteks lauses *Paneme suurde tuppä riiuli peale ..* analüüsitakse muutumatu sõna *peale* ekslikult afiksaaladverbiks ning koos verbivormiga *paneme* ühendverbiks *paneme peale*, kuigi õige on selles kontekstis kaassõnafraas *riiuli peale*. Aga tuleks tähele panna, et selles lauses on teoreetiliselt ka semantiliselt võimalik selline analüüs, kus *riiuli* on süntaktiline objekt ning *peale* on tõesti ühendverbi osa.

Lauses *Nüüd lasti neid kastist välja jalutama* on näide olukorrast, kus afiksaaladverb küll tuvastatakse, kuid antud lause kontekstis pannakse kokku vale verbiga: afiksaaladverb *välja* kombineeriti ekslikult verbiga *jalutama* ühendverbiks *välja jalutama*, kuigi tegelikult esineb lauses ühendverb *lasti välja*.

Positiivse näitena on lauses *Tulin neli päeva tagasi mägedest alla* ära tuntud õige ühendverb *alla tulema*, kuigi lauses on olemas veel teise potentsiaalse produktiivselt kombineeruva ühendverbi *tagasi tulema* komponendid.

Üheks veapõhjuseks on kõrvallausega poolitatud ja üksikeisest eraldatud lauseosad, kus üks osa sisaldab põhiverbi ja teine afiksaaladverbi, nt lauses *.. lõppkokkuvõttes tasub hoonete soojustamine, milleks majaomanikel kulub arvutuste järgi 15,2 miljardit eurot, ennast mitmekordselt ära ei tuntud ära ühendverbi ära tasuma*; analüüsi raskendava asjaoluna esineb siin pealauset poolitavas kõrvallauses verbivorm *kulub*, mis samuti saaks afiksaaladverbiga *ära* ühendverbi moodustada.

5. Sarnane uurimistöö eesti keele ja ka teiste keelte alal

2002. aastal avaldas Ivan A. Sag koos kolleegidega artikli “Multiword Expressions: Pain in the Neck for NLP”, kus tõdetakse, et püsiühendid on võtmeprobleemiks tegelikke tekste lingvistilise põhjendatusega analüüsivate keeletöötlusvahendite loomisel (Sag jt 2002). Sellest teedrajavaks tunnistatud artiklist alates on püsiühendite, sh ühendverbidega seonduv probleemistik olnud arvutilingvistikas aktuaalne uurimisteema.

Eestikeelse ülevaate püsiühendite probleemistikust arvutilingvistikas annab Kadri Muischneki ja Heiki-Jaan Kaalepi artikkel “Eesti keele püsiühendid arvutilingvistikas: miks ja kuidas” (Muischnek, Kaalep 2009).

Käesolevas artiklis kirjeldatud tööle kõige sarnasem on Heiki-Jaan Kaalepi eesti keele püsiühendite märgendamise programm, mida on kirjeldatud Eesti keele- tehnoloogia teisel konverentsil⁵ ettekandes “Mitmesõnalised verbid” (Kaalep 2009). Selle programmi ülesandeks on tekstis märgendada kõik püsiühendite andmebaasis sisalduvad verbikesksed püsiühendid, sh ka ühendverbid. Sisendina kasutatakse morfoloogiliselt ühestatud teksti, kus pole aga märgendatud osalausepiire, st osalausepiiride määramine on püsiühendite märgendamise programmi osa, samas aga on valesti märgendatud, õigemini märkimata jäänud osalausepiir kõige tavalisem vigade põhjus selle programmi väljundis. Verbikesksete püsiühendite tuvastamise saagis on 92% ja täpsus 90%, ühendverbide tuvastamise saagist ja täpsust eraldi mõdetud ei ole.

Eesti keele ühendverbide korpusest ekstraheerimise probleemistikku on käsitletud ka murdematerjali näitel. Kristel Uihoaed on otsinud vastust küsimusele, milline statistik sobib kõige paremini ühendverbide loendi koostamiseks murdekorpuse põhjal (Uihoaed 2010).

Üldisemalt palju käsitletud teemad on olnud tekstikorpuse baasil ühendverbide loendi koostamine (nt Baldwin, Villavicencio 2002, Ramisch jt 2008, Kaalep, Muischnek 2002), sellise loendi põhjal mitmesuguse lisainfoga varustatud leksikoni koostamine (nt Villavicencio 2003, Kaalep, Muischnek 2008) ja püsiühendite kohtlemine automaatsel süntaktilisel analüüsil.

Mis puudutab viimast allteemat, siis näiteks Aline Villavicencio ja Ann Copestake tutvustavad oma 2002. aastal avaldatud artiklis (Villavicencio, Copestake 2002) partikkelverbide käsitlemist HPSG-põhise inglise keele formaalse grammatika LinGO ERG raamistikus. Inglise keele partikkelverb (ingl *particle verb*, ka *phrasal verb*) koosneb verbist ja adverbiaalsest partiklist, *particle verb* on ka eestikeelse termini *ühendverb* ingliskeelseks vasteks (Erelt 2003: 101). LinGo ERG süsteemis esitatakse partikkelverbid leksikonis ning nad on jagatud üheteistkümneks tüübiks vastavalt sellele, kas ja milline seotud laiend neil saab olla ja milline võib olla partikli ja seotud laiendi sõnajärg. Tulevikuplaanina mainitakse teatud semantilist laadi partikkelverbide esitamist verbide ja partiklite omavahel kombineeruvate hulkadena, näitena tuuakse ikka liikumisverbide ja suunda või kohta väljendavate partiklite hulgad.

Martin Forst jt (2010) kirjeldavad võrdlevalt ühendverbide töötlemist inglise, saksa ja ungari keelt analüüsivates leksikaalfunktsionaalse grammatika formalismi

põhistes süntaksianalüsaatorites (ParGram LFG). Autorid väidavad esiteks seda, et nende kolme keele ühendverbe saab arvutigrammatikas esitada ühesugusel moel ja teiseks seda, et kompositsioonilistele (eesti traditsioonis nimetatud korrapärasteks) ja idiomaatilistele ühendverbidele tuleb läheneda erineval moel: idiomaatilised tuleb esitada leksikonis ja kompositsioonilised tuleb “kokku panna” süntaktilise analüüsi käigus. Kui partikli lisamine korrapärase ühendverbi puhul lisab verbi argumenti-struktuuri uue argumenti, siis luuakse spetsiaalse reegli abil partikkelverbi jaoks uus formaalne kirjeldus, mis erineb vastava lihtverbi formaalsest kirjeldusest ainult selle lisatud argumenti poolest.

6. Kokkuvõtte ja tulevikuplaanid

Artiklis käsitleti ühendverbide töötlemise probleeme eesti keele automaatse reegli-põhise süntaksianalüüsi raames. Ühendverbide tuvastamisel kasutatakse kahte strateegiat: leksikonipõhist ja reeglipõhist, viimasel juhul pannakse mõned korrapäraste ühendverbide hulga kokku teatud semantiliselt hulka kuuluvate verbide ja afiksaaladverbide kombineerimisel. Tuvastamine ise on iteratiivne protsess. Esimestel etappidel üritatakse ühestada afiksaaladverbina esineda võivaid sõnavorme afiksaaladverbi, pre- ja postpositsiooni ning substantiivi tõlgenduste vahel, järgmisel sammul leitakse, millised afiksaaladverbid ja millised verbid omavahel kombineeruvad. Ühendverbide tuvastamise esialgsed tulemused olid keskpärased, saagiseks saavutati 79,3%, mis jäi alla ka varasemale lihtmeetodil tehtud eksperimentidele. Täiendades reeglistikku morfoloogilise järelühendamise reeglitega ning luues eraldi reeglid produktiivselt kombineeruvate korrapäraste ühendverbide tarbeks, saadi tulemuseks saagis 97,4% ning täpsus 96,6%.

Tabelis 1 on toodud erinevate eksperimentide tulemused.

Tabel 1. Ühendverbide tuvastamise eksperimentide saagised ja täpsused

	Kaalep 2009	Esialgne	Täiustatud
Saagis	92%	79,3%	97,4%
Täpsus	90%	97,9%	96,6%

Kaalepi (2009) ja siinses artiklis kirjeldatud eksperimentide tulemuste võrdlemine on natuke meelevaldne, sest tekstide hulga, žanrid, aga ka märgendatavad ühendid on erinevad: Kaalepi tulemused on pigem üldisemalt verbikesksete püsiühendite kohta. Samuti hinnati Kaalepi eksperimentis tulemusi morfoloogilise märgenduse suhtes, kasutades sisendina perfektselt ühestatud teksti. Samas on selle katse aluseks olnud püsiühendite andmebaas väga oluliseks allikaks meie leksikoni edasiseks täiustamiseks.

Tuvastamata jäänud ühendverbide veel üheks allikaks on ebatäielik leksikon, mida on plaanis täiendada, esmajärjekorras võtta kasutusele püsiühendite andmebaasis⁶ sisalduvatel ühendverbidel need, mis meie leksikonis puuduvad. Samuti tuleks täiendada korrapäraste ühendverbide kombineerimiseks kasutatavaid reegleid ning tuvastada uusi omavahel kombineeruvate verbide ja afiksaaladverbide rühmi.

⁶ <http://www.cl.ut.ee/ressursid/pysiyhendid/> (30.9.2013).

Järgmisena tuleb luua samasugune leksikon ja reeglid ka väljendverbide jaoks ning siis jõuda kõigi levinumate püsiühendite analüüsimiseni.

Loodud ja loodav reeglistik on väga vajalik süvasüntaktiliseks analüüsiks, ilma püsiühendite tuvastamiseta oleks analüüsipuu poolik, et mitte öelda vigane.

Viidatud kirjandus

- Baldwin, Timothy; Villavicencio, Aline 2002. Extracting the unextractable: a case study on verb particles. – Proceedings of the Conference on Computational Natural Language Learning (CoNLL 2002), Taipei, Taiwan, 31 August – 1 September 2002. Association for Computational Linguistics, 17.
- EKG I = Erelt, Mati; Reet Kasik; Helle Metslang; Henno Rajandi; Kristiina Ross; Henn Saari; Kaja Tael; Silvi Vare 1995. Eesti keele grammatika I. Morfoloogia. Sõnamoodustus. [The Grammar of the Estonian Language I: Morphology. Word-formation.] Eesti Teaduste Akadeemia Eesti Keele Instituut. Tallinn.
- EKG II = Erelt, Mati; Reet Kasik; Helle Metslang; Henno Rajandi; Kristiina Ross; Henn Saari; Kaja Tael; Silvi Vare 1993. Eesti keele grammatika II. Süntaks. Lisa: kiri. [The Grammar of the Estonian Language II: Syntax.] Eesti Teaduste Akadeemia Keele ja Kirjanduse Instituut. Tallinn.
- EKSS = Langemets, Margit; Tiits, Mai; Valdre, Tiia; Veski, Leidi; Viks, Ülle; Voll, Piret (Toim.) 2009. Eesti keele seletav sõnaraamat 1–6. [The Explanatory Dictionary of Estonian.] Tallinn: Eesti Keele Sihtasutus.
- Erelt, Mati (Toim.) 2003. Estonian Language. *Linguistica Uralica Supplementary Series Vol. 1*. Tallinn: Estonian Academy Publishers.
- Erelt, Mati 2013. Eesti keele lauseõpetus. Sissejuhatus. Õeldis. [Estonian Syntax. Introduction.] Tartu ülikooli eesti keele osakonna preprintid 4. Tartu Ülikool.
- Forst, Martin; Holloway King, Tracy; Laczko, Tibor 2010. Particle verbs in computational LFGs: Issues from English, German, and Hungarian. – Proceedings of LFG 10. CSLI Publications, 228–248.
- Kaalep, Heiki-Jaan 2009. Mitmesõnalised verbid. [Multiword verbs.] – Ettekande slaidid internetiaadressil <http://www.keeletehnoloogia.ee/konverentsid/ekkt-teine-konverents/slaidid-2009/mitmesonaliste-verbide-aratundmine-tekstides> (30.9.2013).
- Kaalep, Heiki-Jaan; Muischnek, Kadri 2002. Using the text corpus to create a comprehensive list of phrasal verbs. – Proceedings of Language Resources and Evaluations: Third International Conference on Language Resources and Evaluation, Las Palmas, May 29–31, LREC 2002. European Language Resources Association, 101–105.
- Kaalep, Heiki-Jaan; Muischnek, Kadri 2008. Multi-word verbs of Estonian: a database and a corpus. – Proceedings of the LREC Workshop “Towards a Shared Task for Multiword Expressions”, Marrakech, Morocco, June 1, 2008. European Language Resources Association, 23–26.
- Karlsson, Fred; Voutilainen, Atro; Heikkilä, Juha; Anttila, Arto (Eds.) 1995. Constraint Grammar: A Language-Independent System for Parsing Running Text. Natural Language Processing 4. Berlin, New York: Mouton de Gruyter.
- Muischnek, Kadri; Kaalep, Heiki-Jaan 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>
- Müürisep, Kaili 2000. Eesti keele arvutigrammatika: süntaks. [Computer Grammar of Estonian: Syntax.] *Dissertationes Mathematicae Universitatis Tartuensis* 22. Tartu: Tartu Ülikooli Kirjastus.

- Pajusalu, Renate 2009. Sõna ja tähendus. [Word and Meaning.] Tallinn: Eesti Keele Sihtasutus.
- Puolakainen, Tiina 2001. Eesti keele arvutigrammatika: morfoloogiline ühestamine. [Computer Grammar of Estonian: Morphological Disambiguation.] *Dissertationes Mathematicae Universitatis Tartuensis* 27. Tartu: Tartu Ülikooli Kirjastus.
- Ramisch, Carlos; Villavicencio, Aline; Moura, Leonardo; Idiart, Marco 2008. Verb-particle constructions, noise and idiomaticity. – Proceedings of the Twelfth Conference on Computational Natural Language Learning (CoNLL 2008), Manchester, UK, August 2008. Association for Computational Linguistics, 49–56
- Roosmaa, Tiit; Koit, Mare; Muischnek, Kadri; Müürisep, Kaili; Puolakainen, Tiina; Uiibo, Heli 2003. Eesti keele arvutigrammatika: mis on tehtud ja kuidas edasi? [A formal grammar of Estonian: Experience and prospects.] – *Keel ja Kirjandus*, 46 (3), 192–209.
- Rätsep, Huno 1978. Eesti keele lihtlausete tüübid. [Types of Simple Sentences in Estonian.] Eesti NSV TA Emakeele Seltsi toimetised 12. Tallinn: Valgus.
- Saagpakk, Paul F. 1992. Eesti-inglise sõnaraamat. [Estonian-English Dictionary.] 2. trükk. Tallinn: Koolibri.
- Sag, Ivan A.; Baldwin, Timothy; Bond, Francis; Copestake, Ann; Flickinger, Dan 2002. Multiword Expressions: A Pain in the Neck for NLP. – Alexander Gelbukh (Ed.). *Computational Linguistics and Intelligent Text Processing*. Third International Conference, CILCling 2002, Mexico City, Mexico, February 17–23, 2002. Proceedings. Lecture Notes in Computer Science 2276. Springer Verlag, 1–15.
- Uiiboed, Kristel 2010. Statistilised meetodid murdekorpuse ühendverbide tuvastamisel. [Statistical methods for phrasal verb detection in Estonian dialects.] – *Eesti Rakenduslingvistika Ühingu aastaraamat*, 6, 307–326. <http://dx.doi.org/10.5128/ERYa6.19>
- Villavicencio, Aline 2003. Verb-particle constructions and lexical resources. – Proceedings of the Meeting of the Association for Computational Linguistics: 2003 workshop on Multiword Expressions, Sapporo, Japan, July 2003. Association for Computational Linguistics, 57–64
- Villavicencio, Aline; Copestake, Ann 2002. Verb-particle constructions in a computational grammar of English. – Jong-Bok Kim, Stephen Wechsler (Eds.). *Proceedings of the 9th International Conference on Head-Driven Phrase Structure Grammar*, Kyung Hee University, Seoul 5–7 August, 2002. CSLI Publications, 357–371

Võrgumaterjalid

- Eesti keele verbikesksete püsiühendite andmebaas. <http://www.cl.ut.ee/ressursid/pysiyhendid/> (30.9.2013).
- Sõltuvussüntaktiliselt käsitsi analüüsitud korpus. <http://www.keeletehnoloogia.ee/ekt-projektid/vahendid-teksti-mitmekihiliseks-margendamiseks/soltuvussuntaktiliselt-kasitsi-analuusitud-korpus> (16.2.2014).
- Tasakaalus korpus. <http://www.cl.ut.ee/korpused/grammatikakorpus/> (16.2.2014).
- World of VISL. Visual Interactive Syntax Learning. <http://beta.visl.sdu.dk/> (30.9.2013).

PARTICLE VERBS IN ESTONIAN SHALLOW SYNTACTIC ANALYSIS

Kadri Muischnek, Kaili Müürisep, Tiina Puolakainen

University of Tartu

Particle verbs are a frequent phenomenon in Estonian and they need special attention during automatic syntactic analysis. This article deals with the computational treatment of particle verbs in the course of automatic syntactic analysis using the Constraint Grammar framework.

The recognition of particle verbs is needed for further deep syntactic analysis, as often the particle verb as a predicate introduces a different argument structure compared to the corresponding simplex verb.

The processing of particle verbs in the CG framework is mainly lexicon-based, although some groups of particle verbs are also combined by the rules, e.g. combinations of verbs of movement and directional particles.

The Constraint Grammar analyser of Estonian consists of separate modules for morphological disambiguation and syntactic analysis. The recognition of particle verbs starts during morphological disambiguation, but at that stage there is still a lot of general morphological ambiguity in text and that ambiguity hinders the recognition of particle verbs.

In the reported project a special post-disambiguating module was introduced in order to achieve more correct analysis of the uninflecting particle, which could be a part of a particle verb, function as an adposition or a declensional form of a noun or even a verb.

Recognizing particle verbs using the aforementioned extra module achieved recall 97.4% and precision 96.6%.

Keywords: computer linguistics, detection of fixed word combinations, Estonian

Kadri Muischneki (Tartu Ülikool) teaduslikud huvialad on korpuslingvistika, eesti keele süntaktiline struktuur ning automaatne süntaktiline analüüs.

Tartu Ülikool, arvutiteaduse instituut, Liivi 2, 50090 Tartu, Estonia

kadri.muischnek@ut.ee

Kaili Müürisepa (Tartu Ülikool) peamisteks uurimisuundadeks on eesti keele automaatne süntaktiline analüüs ja sellega seonduvad teemad.

Tartu Ülikool, arvutiteaduse instituut, Liivi 2, 50090 Tartu, Estonia

kaili.muurisepp@ut.ee

Tiina Puolakaineni (Tartu Ülikool) põhilisteks uurimishuvideks on eesti keele automaatne süntaktiline analüüs ja morfoloogiline ühestamine.

Tartu Ülikool, arvutiteaduse instituut, Liivi 2, 50090 Tartu, Estonia

Tiina.Puolakainen@ut.ee

PALVED EESTI, SOOME JA VENE KEELES: GRAMMATIKA PRAGMAATIKA TEENISTUSES

Renate Pajusalu

Ülevaade. Artikkel käsitleb palvete ja küsimuste vormistamise stereotüüpe eesti, vene ja soome keeles. Materjal on kogutud küsitluste abil, milles vastajad pidid kirjutama, mida nad kirjeldatud situatsioonis ütleksid või kirjutaksid (nn diskursuse täiendamise ülesanne). Iga vaadeldava keele kohta loodi andmebaas, milles on 200 palvet. Nende alusel vaadeldakse esitatud kümme situatsiooni võrdlevalt kolmes keeles, tuues välja nii sarnasused kui erinevused. Kokkuvõttes tehakse üldistused palvete kodeerimisel esinenud sagedamate grammatiliste vahendite (lausetüüp, verbi morfoloogiline vorm, eituse olemasolu, pöördumine *sina-* või *teie-*vormiga jm) esinemise kohta eesti, soome ja vene keeles.*

Võtmesõnad: küsimus, pöördumine, konditsionaal, eitus, eesti keel, vene keel, soome keel

1. Sissejuhatus

Artikkel käsitleb palvete vormistamist eesti, vene ja soome keeles. Eesmärgiks on välja selgitada, mille poolest erinevad nende keelte stereotüüpsed palved omavahel, sest kuigi vaadeldavate keelte strateegiad on pealtnäha üsna sarnased, ei ole need siiski täielikult identsed. Tegemist on suhtluses küllalt tundliku kõneaktitüübiga, mistõttu võivad normist erinevad palve vormistused (näiteks teise keelena kõneleja poolt) partnerit solvata.

Palved (ingl *request*) ja nende lingvistiline vormistamine kuuluvad kõneaktiteooria (Searle 1969) ja klassikalise viisakusteooria (Brown, Levinson 1987) keskmesse. Tegemist on selge kõneaktitüübiga – palve on sotsiaalne tegevus, mille tagajärjel teate vastuvõtja peaks sooritama palvega esitatud teo, millest on kasu kõnelejal või kellelgi kolmandal (vt Taleghi-Nikazm 2006: 1). Palve on seega tüüpiline direktiiv¹ ja siinkohal ei tehta vahet erineva tugevusastmega palvetel, st ka käsud, anumised jm seda tüüpi suhtlustegevused on liigitatud palvete hulka.

* Autor tänab anonüümseid retsensente väga kasulike märkuste eest.

¹ Lähtun siin Searle'i (1969) kõneaktide liigitusest, mis paigutab direktiivide hulka ka küsimused. Palved on siis laias laastus see osa direktiividest, mis ei ole küsimused.

Vastavalt näo mõiste ümber rajatud viisakuskäsitlusele (Brown, Levinson 1987) on palve puhul tegemist partneri nägu ohustava aktiga, sest paludes kellelgi midagi teha seatakse ohtu tema iseseisvus ja sõltumatus. Selle ohu pehmemdamiseks kasutatakse leksikaalseid ja grammatilisi vahendeid (nn negatiivset viisakust). Loomulikult on palvete pehmemdamise vajadus kultuuriti erinev, sõltudes eelkõige kõnelejatevahelistest suhetest nii sotsiaalses plaanis kui käimasoleva suhtluse raames, olukorra formaalsusastmest ja palve enda n-õ raskusest (kuivõrd tehakse teisele inimesele tüli, kuivõrd lihtne on palvet täita jms). Keeliti erinevad on ka pehmemdusvahendid.

Palveid on uuritud mitmesuguste meetoditega ja mitmetes keeltes, varasemast perioodist on enim tuntud rahvusvaheline projekt CCSARP, mis võrdles küsitluste põhjal (nn *diskursuse täiendamise ülesanne*) eri keelte palveid ja vabandusi (Blum-Kulka jt 1989), kuid meetodit kasutatakse palju ka tänapäeval (nt Ogiermann 2009, Peterson 2010, Yu 2011 jpt). Samas on ka tegeliku suhtluse palvete uurimisel juba pikk traditsioon, üldiselt peetakse valdkonna teerajajaks Susan Ervin-Trippi (1976) käsitlust Ameerika inglise keele direktiividest. Ka hilisemad palvete uurimused on tihti pigem rõhutanud autentsete situatsioonide analüüsi olulisust, meetodina on kasutatud konversatsioonianalüüsi (Taleghi-Nikazm 2006, Curl, Drew 2008, Sorjonen jt 2009, Tanner 2012, Rääbis 2012 jm). Ühte ja ainuõiget meetodit keeleteaduses siiski ei eksisteeri. Kuigi reaalsete vestluste uurimine annab meile kõige olulisemat infot keelekasutuse kohta, on see väga ajamahukas ja eri keelte kohta võrreldavate tulemuste saamine nõuab suuri ja mitmekülgseid korpuseid. Siinse artikli eesmärgiks on kõneaktide uurimine eri kultuurides, ja selleks annab küsitlusmeetod hõlpsamini võrreldavat informatsiooni (vt ka Jucker 2009, Yu 2011). Siinjuures ei tohi muidugi unustada, et küsitluste abil saame ainult ligilähedase pildi keelekasutuse nüanssidest, sest ühelt poolt ei tea inimene kunagi päris täpselt, kuidas ta tegelikus olukorras käitub, ja teiselt poolt on suhtlus alati mitme suhtleja poolt üles ehitatud protsess, küsitlusele vastaja aga üldjuhul ei suuda arvestada hüpoteetilise partneriga.

Grammatiline vahend palve edastamiseks võiks eesti keeles primaarselt olla imperatiivi verbivormi sisaldav käsklause (*Anna mulle soola!*), kuid pehmemdatud kujul peaks palve ilmselt esitama pigem küsimuse vormis ja tingivas kõneviisis modaalverbiga (*Kas te ei võiks mulle soola ulatada?*, vt Metslang 2004: 252). Tegelikus suhtluses kasutatakse siiski ka mitmesuguseid muid palvekonstruktsioone. Eesti keele direktiive on varem käsitletud infodialoogide materjali põhjal (Hennoste, Rääbis 2004). Direktiivide klass on selles liigituses jagatud soovideks ja ettepanekuteks, piir nende vahel on suhteliselt hägus; põhiliste vormivahenditena esinevad väitlause ja verbi tingiv kõneviis (nt *ma tahaksin...*), kuid võimalikke konstruktsioone on muidugi palju (*op. cit.*: 72–79). On käsitletud ka palveid ja käske eesti keele suulises materjalis viisakustestooriast lähtuvalt (Keevallik 2004) ja direktiive isa ja lapse suhtluses ühe salvestuse põhjal (Rääbis 2012). Nendes uurimustes on välja toodud muuhulgas eestlaste suhteliselt otsene suhtlusstiil võrreldes teiste rahvastega ja sarnasus pigem saksa kui inglise keeleruumiga (Keevallik 2004: 214).

Siinne käsitlus lähtub oletusest, et iga keele emakeelena kõnelejal on mingi stereotüüpne ettekujutus olukorraga sobivast keelelisest käitumisest. See ei pruugi kokku langeda tegeliku käitumisega, kuid selle teadmine on oluline nt keeleõppes. Kuna õppematerjal peab esitama suhtluse kõige olulisemad aspektid, on keele õpetamisel vajalik teada, milline on “tüüpiline” palve eri olukordades ja millest

sõltub keeleliste vahendite valik. Samuti on emakeelelt homogeensete õpirühmade puhul kasulik teada, mille poolest erineb õpitava keele palvete moodustamine nende emakeele omast. Soome keele kohta on teada, ja ilmselgelt kehtib see ka teiste keelte kohta, et õpiku palved ei vasta tegelikele (Tanner 2012), kuid see on paratamatu, vähemalt niikaua, kui õppematerjalina kasutatakse enamasti valmistekste.

Palvete vormistamine sõltub ühiskonnas kehtivatest viisakusnormidest. Ka nii lähedaste kultuuride puhul nagu siin käsitletud eesti, soome ja vene, eksisteerivad pealtnäha sarnaste palvete kodeerimise vahendite kasutuses ometi erinevused, mida paljud teise keele kõnelejad on suhtluses ebamugavusena tundnud.² Siinne uurimus võrdlebki eesti, soome ja vene keele vahendeid, mida stereotüüpsete palvete moodustamisel kasutatakse. Eesmärgiks on anda võrdlev ülevaade väga laiaast valdkonnast, mille süvitsi uurimine seisab alles ees.

2. Meetod ja materjal

Meetodina kasutatakse siinses uurimuses küsitlusi, täpsemalt kümnest olukorrast koosnevat diskursuse täiendamise ülesannet (ingl *discourse completion task*; Blum-Kulka jt 1989). Küsitluse koostamisel püüti esitada võimalikult erineva formaalsus- ja raskusastmega olukordi, millest enamik nõudis palve sõnastamist, mõned aga võrdluseks ka küsimuse esitamist (nt S7, S8, S9). Situatsioonide loomisel oli eesmärgiks võimalikult suur variatiivsus mitmes aspektis: suhtluskanalitena on esitatud meilisuhklus ja mobiiltelefoni sõnum (S1, S8, S10), telefonisuhklus (S4, S6) ja näost-näku suhtlus (S2, S3, S5, S7, S9); tuttavusastmelt varieerusid situatsioonid sõpradevahelisest suhtlusest (S4, S6, S8) suhtluseni võõraste inimestega argikeskkonnas (S3, S7), teiselt poolt oli osa situatsioonidest institutsionaalsed (S1 õppejõud, S2 raamatukoguhoidja, S5 ja S9 poemüüja, S10 ametnik). Samuti oli erinev palvete “raskusaste”, varieerudes partneri töökohustuste hulka kuuluvatest tegudest (S5 kauba müümine poes ja S10 tõendi väljastamine kindlustusfirmas), mis ei peaks palve olemuse poolest pehmemdamist vajama, kuni võõralt inimeselt suure teene (S3) või õppejõult vastutuleku (S1) palumiseni. Situatsioonide loomisel piiras mahtu see, et küsitletud ei jaksa tavaliselt väga pikalt keskenduda.

Sama küsitlus tõlgiti vene ja soome keelde. Iga keele kohta on siinses pilootuurimuses analüüsitud 20 vastaja tekstid, mille nad kirjutasid kujutletava reaktsioonina esitatud kümnele situatsioonile. Nimetan neid tekste vastustekstideks, mida kokku oli 600 (200 iga keele ja 20 igas keeles iga situatsiooni kohta) ja mis võisid olla pikemad või lühemad. Kõik vastajad olid üliõpilased: eesti ja vene keele puhul Tartu ülikoolist, soome keele puhul Helsingi ülikoolist.³ Grupid ei olnud ühiskonna kui terviku suhtes representatiivsed (enamasti oli tegemist naistega, kes õppisid ülikoolis keelega seotud erialal), kuid just gruppide sarnasuse tõttu on tulemused omavahel võrreldavad. Tegemist on pilootuurimusega, mida on plaanis hiljem jätkata teisi keeli arvesse võttes.

Kõikide küsitluste vastustekstidest kodeerisin palve peaaktid (ingl *head acts*, vt Blum-Kulka jt 1989: 17–18), st vastustekstis kõige rohkem palvet esitavad süntaktilised üksused. Selle identifitseerimine ei olnud alati üheselt selge, sest mõnikord esines vastustekstides ka pöördumisi, vabanduse palumist, eelpalveid

² Väga huvitavat materjali on eesti ja soome suhtluse erinevuste kohta kogunud Ninni Jalli, näiteks on tema materjalis kirjeldus sellest, kuidas soomlase arvates on eesti arstid Soomes ebaviisakad, kuna ei vormista korraldusi küsivas vormis. (Ninni Jalli, isiklik vestlus.)

³ Soomekeelse küsitluse viis Helsingi ülikoolis läbi Pire Teras, venekeelse Tartu ülikoolis Oksana Palikova. Tänan neid südamest abi eest!

jm kõneaktitüüpe, mida on varasemateski uurimustes peetud palve pehmenamise vahenditeks (eesti keele kohta vt Keevallik 2004: 211). Näiteks järgnevas vastustekstis on palve pealauseks (ja seega analüüsitavaks lauseks) loetud *ehk võtad siis kaasa selle?*

S4: Raamat

Ma tahaks lugeda seda raamatut, äkki saad mulle seda laenata? Homme me näeme niikuinii, ehk võtad siis kaasa selle?

Peaakti identifitseerimisele järgnes keeleliste nähtuste kodeerimine, määrati iga peaakti lausetüüp (väitlause, käsklause või küsilause), verbi vorm (indikatiiv, imperatiiv või konditsionaal), modaalkonstruksiooni olemasolu ja kõneliik (jaatav või eitav). Kogu vastusteksti ulatuses (st ka siis, kui see esines ainult väljaspool peaakti) kodeeriti veel pöördumisvorm (kas *sina* või *teie*).⁴

3. Eesti, soome ja vene palved küsitluses esitatud situatsioonide kaupa

Selles peatükis esitatakse küsitluses kasutatud situatsioonid ja analüüsitakse näiteid tüüpilistest vastustekstide peaaktidest. Näited on otsesed tsitaadid, kuid valitud on nende hulgas kõige paremini keskmist esindav (või vajadusel kaks kõige tüüpilisemat).⁵

Igas situatsioonis vaadeldakse eelkõige peaakti lausetüüpi, verbi kõneviisi, modaalkonstruksiooni ja eituse olemasolu. Eraldi kommenteeritakse pöördumisvormi ja vajadusel ka muid palvete vormistamisel relevantseid keelelisi vahendeid. Näitelause järel esitatakse selle konstruksiooni relevantseid osad ja selliste näidete arv, mis materjalis seda konstruksiooni esindas, nt lause (1a) *Kas mul oleks ehk võimalik eksam siiski sooritada?* järel on esitatud konstruksioonimall KÜSIL(ause); JAATAV KOND(itsionaal); MOD(aalkonstruksioon) ja number 14 näitab, et kõiki neid elemente sisaldas 20 peaaktist 14. Muudes aspektides (nt vabandused, pehmedaja *ehk* jms) võivad need 14 lauset üksteisest erineda, ja samu elemente (nt konditsionaalivormi või eitust) võib eraldi esineda ka teistes peaaktides. Seega mida enam läheneb number 20-le, seda üksmeelsemad on olnud selle keele osas vastajad. Lisaks on toodud ka muid näitelauseid, mille kohta ei ole konstruksioonimalli ja esinemiskordade arvu esitatud.

3.1. Situatsioon 1: Eksam

Magasid eksamipäeval sisse. Sul on siiski väike lootus, et õppejõud lubab sul eksami sooritada mingil muul ajal. Kirjutad talle meili. Kuidas sõnastad oma soovi?

Eesti materjalis oli enamasti tegemist küsilausega (1a). Grammatilistest pehmedajatest esines enamasti verbi konditsionaalivorm, selle puudumisel nt partikkel *ehk*. Kõige rohkem oli esimese isiku verbivorme, mistõttu jäi selgusetuks, kas õppejõudu sinatatakse või teietatakse. Vähesed esinenud pöördumised olid siiski *teie*-vormis.

⁴ Kodeeriti ka muid nähtusi (nt vabandussõna olemasolu, pehmedavaid partikleid jms), kuid siinse artikli maht ei võimalda neid põhjalikult käsitleda, mistõttu nende olemasolu on kommenteeritud ainult siis, kui see hakkab materjalis eriti silma.

⁵ Soome ja vene näited on varustatud võimalikult sõnasõnalise tõlkega. Modaalverbi vene keelest tõlkimisel on alati antud verb *võima*, soome keelest tõlkimisel sama tüvi, ka siis, kui see ei ole kõige loomulikum tõlkevaste. Kuna soome partikkel *-han* ei saa viisakuskonstruksioonides eesti keelde tõlkida, on see märgitud verbi juurde sulgudes.

Soome vastused olid eesti omadega väga sarnased (1b): tüüpilisim predikaat oli *olisiko mahdollista* ‘kas oleks võimalik’ ja ka pöördumised puudusid enamasti. Ühel juhtumil, kus verb oli teises isikus, oli erinevalt eesti keelest tegemist *sina*-vormiga.

Ka vene vastustes oli sama tüüpi palveid (1c). Siiski olid vene vastused oluliselt rohkem kirjutatud nii, et oli näha *teie*-vormis pöördumist õppejõu poole, samuti esines verbi konditsionaali eitust (1d).

- (1) a Kas mul oleks ehk võimalik eksam siiski sooritada?
KÜSIL, JAATAV KOND, MOD – 14
- b Olisiko mahdollista tehdä tentti jonakin toisena päivänä?
‘Kas oleks võimalik teha eksam mingil muul päeval?’
KÜSIL, JAATAV KOND, MOD – 18
- c Est’ li u menja vozmožnost’ sdat’ ego v drugoe vremja?
‘Kas mul on võimalus sooritada see muul ajal?’
KÜSIL, JAATAV IND, MOD – 9
- d Ne mogli by vy razrešit’ mne perezdat’ ego?
‘Kas te ei võiks lubada mul see uuesti sooritada?’

3.2. Situatsioon 2: Raamatukogu

Sa ei suuda raamatukogus leida vajalikku raamatut. Lähed laua juurde, kus istub raamatukogutöötaja ja kirjutab midagi arvutis. Kuidas palud temalt abi?

Eesti materjalis oli kõige enam küsilauseid, millega küsiti: *kas te saaksite aidata* (2a). Kõik pöördumised olid *teie*-vormis, verbid enamasti konditsionaalis. Lisaks esines vastuses enamasti vabandus, mille tõenäoliselt tingis see, et töötaja pidi katkestama oma tegevuse.

Soome materjalis oli samuti kõige tavalisem *voisit(te)ko auttaa*, kuid see võis olla nii *sina*- kui *teie*-vormis. Suhteliselt rohkem kui eesti ja vene keeles oli ka esimeses isikus predikaate (nt *tarvitsisin tätä kirjaa* ‘mul oleks seda raamatut vaja’). Viis korda esines partikkel *-hAn* (2b), mida soome keeles kasutatakse palvete pehmemdamiseks. Vabandust paluti samuti väga paljudel juhtudel (2b).

Vene materjalis oli selles situatsioonis valdavalt eitava konditsionaaliga küsilausekonstruktsioone *ne mogli by vy* (2c). *teie*-vormis pöördumine oli eksplitsiitne kõigis vastustes, peaaegu kõik sisaldasid ka vabandust.

- (2) a Vabandage, kas te saaksite aidata mul üht raamatut leida?
KÜSIL, JAATAV KOND, MOD – 12
- b Anteeksi, olisikohan mahdollista saada apua?
‘Vabandust, kas oleks (*-hAn*) võimalik saada abi?’
KÜSIL, JAATAV KOND, MOD – 8
- c Izvinite, ne mogli by vy pomoč’ mne najti knigu?
‘Vabandage, kas te ei võiks aidata mul leida raamatut?’
KÜSIL, EITAV KOND, MOD – 13

3.3. Situtatsioon 3: Lennujaam

Oled hilinemas lennukile. Taksot ei ole kiiresti võimalik saada. Sinu naabril on auto ja sa just nägid aknast, et ta tuli koju. Sul tekib mõte, et ta võiks sind lennujaama ära viia. Te ei tunne küll hästi, olete ainult tervitanud, aga ta on noor inimene ja paistab väga lahke. Lähed naabri ukse taha, helistad uksekella. Naaber avab. Mida ütled?

Kolmas situatsioon oli vastajatele kõige ootamatum. Paar vene üliõpilast väitsid isegi, et olukord on täiesti võimatu. Vastustekstid olid kõigis kolmes keeles kõige pikemad ja sisaldasid situatsiooni kirjeldust ja hulgaliselt eelpalveid. Eesti materjalis olid kõik palvete peaaktid vormistatud küsimusena, valdavalt kujul *kas saaksite* või *kas teil oleks aega* (3a). Konditsionaal esines vaid natuke rohkem kui pooltel juhtudest (12 korda), mis ilmselt näitab, et nii ootamatu palve puhul ei ole konditsionaal eesti keeles automaatne pehmedaja, olulisem on kogu situatsiooni suhteliselt pikk kirjeldus ja palve põhjuste üksikasjalik lahtiseletamine. *sina* ja *teie*-vormi kasutati vastavalt 6 ja 14 korda, selles osas vastajatel seega üksmeel puudus. Näites (3b) on näha ebatüüpilisem juhtum: küsilause on vormistatud partikli *äkki* abil (esines selles situatsioonis kolm korda), verb on indikatiivis ja pöördumisvorm *sina*.

Soome materjalis olid kõik palved esitatud konditsionaalis modaalkonstruksiooniga. Samas naabrit sinatati 19 juhul 20-st. Viiel korral esines partikkel *-hAn* (3c). Selles situatsioonis näib siis soome keeles hoolimata solidaarsest pöördumisviisist (*sina* mitte *teie*) valitsevat üksmeel palve pehmedamise vajaduses grammatiliste vahenditega, sest kõikides vastustekstides kasutati konditsionaalset modaalverbi (*voisitko*) või modaalkonstruksiooni (*olisiko mahdollista* 'kas oleks võimalik').

Vene materjalis oli rohkem varieerumist. Ühelt poolt esines tüüpkonstruksioon *ne mogli by vy* (*ne mog by ty*) (st küsilause ja eitav konditsionaal, näide 3d), kuid leidis ka käsklauseid (nagu näide 3e).

- (3) a Kas saaksite mind, palun, lennujaama viia?
KÜSIL, JAATAV KOND, MOD – 11
- b Äkki sa saad mind ära viia?
- c Anteeksi kauheasti, mutta voisitkohan mitenkään viedä minut lentokentälle?
'Kohutavalt vabandan, aga kas võiksid (-hAn) viia mind lennuväljale?'
KÜSIL, JAATAV KOND, MOD – 20
- d Ne mog by ty podvezti do aëroporta?
'Kas sa ei võiks sõidutada lennujaamani?'
KÜSIL, EITAV KOND, MOD – 11
- e Opazdyvaju v aëroport, podvezite požalujsta!
'Hilinen lennujaama, sõidutage palun!'

3.4. Situatsioon 4: Raamat

Tahaksid lugeda raamatut, mis on olemas sinu sõbral. Kohtud temaga homme niikuinii. Helistad talle ja palud raamat homme kaasa võtta. Mida ütled?

See on kogu komplektis üks kõige kergemini täidetavaid palveid – sõbral pole ju raske raamatut kaasa võtta, eriti kui kohtumine on juba niikuinii kokku lepitud. Vastustekstid olid suhteliselt lühikesed, kõigis kolmes keeles pöörduiti ainult *sina*-vormis.

Eesti materjalis esines konditsionaali vaid pooltes palvetes (4a), lisaks modaalverbidele *võima* ja *saama* esines ka verbi *viitsima*. Soome materjalis on mõnevõrra rohkem konditsionaali (4b), muus osas oli aga olukord eesti keelega sarnane, sest leidus ka muid lausetüüpe (mitte ainult küsilauseid), üks kord ka *jos*-lause (*jos ottaisit...*). Vene materjalis oli konditsionaali ja eitust vaid üksikutel juhtudel, tüüpilisem on küsilause *možeš'...?* (4c).

- (4) a Kas sa võiksid homme selle raamatu ka kaasa võtta?
KÜSIL, JAATAV KOND, MOD – 8
- b Huomenna kun nähdään, niin voisitko ottaa kirjan mukaan?
'Homme kui kohtume, kas võiksid võtta raamatu kaasa?'
KÜSIL, JAATAV KOND, MOD – 12
- c Можеš' завтра захватит' с собой?
'Kas sa võid homme kaasa võtta?'
KÜSIL, JAATAV IND, MOD – 8

3.5. Situatsioon 5: Pood

Soovid osta poe kulinaarialetist kolm grillitud kanakoiba. Mida ütled müüjale?

Poesituatsiooni ostupalved erinevad kõigis kolmes keeles oluliselt muudest palvetest. Eesti ja vene keeles olid need enamasti nimisõnafraasid, mida täiendas ainult palvepartikkel *palun*⁶ (5a) või *požalujsta* (5d), vene materjalis esines mõnikord verb *dajte* 'andke'. Soome materjalis aga esines päris palju pikemaid palveid, milles oli kasutatud ka konditsionaalivormi, põhiliselt *mina*-vormis, nt *ottaisin* 'võtaksin' (5c). Soome palvete lõpus esines tihti *kiitos* 'aitäh' (5b).

- (5) a Palun kolm grillkanakoiba.
VERBITA KÄSKL – 15
- b Kolme grillatua kanankoipea, kiitos.
'Kolm grillitud kanakoiba, aitäh.'
VERBITA KÄSKL – 7
- c Ottaisin kolme noita grillatuja kanankoipia.
'Võtaksin kolm neid grillitud kanakoibi.'
VÄITL, JAATAV KOND – 10
- d Mne požalujsta tri kurinye nožki.
'Mulle palun kolm kanakoiba.'
VERBITA KÄSKL – 12

⁶ Eesti *palun* võib olla analüüsitud ka 1. isiku verbivormiks, kuid materjalis on tegemist pigem pragmaatilise partikliga.

3.6. Situatsioon 6: Teesid

Sul on vaja saata teesid rahvusvahelisele üliõpilaskonverentsile. Tähtaeg on täna õhtul, teesid peaks olema inglise keeles. Kahtled oma keeleoskuses ja tahad paluda head sõpra, kes inglise keelt hästi oskab, teesid üle vaadata. Aga tead ka, et tal on homme tähtis eksam ja arvatavasti väga vähe aega. Helistad siiski. Mida ütled?

Situatsioon esindab sõpradevahelist suhtlust, kuid erinevalt situatsioonist S4 (Raamat) on siin tegemist raskemini täidetava palvega, sest on teada, et sõbral on niigi väga kiire. Seetõttu on vastustekstid pikemad ja seletavad kogu olukorda rohkem. Kõigis kolmes keeles sõpra sinatatakse.

Eesti materjalis on verb konditsionaalis vaid pisut vähem kui pooltel juhtudel (6a), tavalisem on indikatiivis (6b) või imperatiivis verb. Soome materjalis on kõik palved esitatud konditsionaali abil, kuigi kõnekeelsust on markeeritud rohkem kui eesti materjalis (6c). Vene materjalis on lausestruktuurid küllaltki erinevad, mistõttu ükski konstruktsioon ei ole sage; konditsionaalivorme on vaid üksikuid, kuid modaalverbi *moč'* on kasutatud rohkem kui pooltel juhtudel (nagu nt 6d). Vene materjalis esineb erinevalt eesti ja soome omast emotsionaalseid hüüatusi abi vajalikkusest (nt *vopros žizni i smerti* 'elu ja surma küsimus').

- (6) a Ma tean et sul väga kiire, aga kas sa saaksid mind natuke aidata?
KÜSIL, JAATAV KOND, MOD – 7
- b Tean, et sul on niigi ajapuudus, aga ehk lased ühest tekstist silmad üle?
- c Mä tiedän, että oot kiireinen, mutta voisitko sä auttaa mua vähän tän englannin abstraktin kanssa?
'Ma tean, et sul on kiire, aga kas sa võiksid aidata natuke selle inglise keele abstraktiga?'
KÜSIL, JAATAV KOND, MOD – 20
- d Smožeš' proverit' moi tezisy?
'Kas saad kontrollida mu teese?'
KÜSIL, JAATAV IND, MOD – 4

3.7. Situatsioon 7: Bussijaam

Oled võõras linnas eksinud. Tahad küsida esimeselt vastutulijalt, kus asub bussijaam. Vastu tulebki üks vanem proua. Kuidas küsid?

Teeküsimine on väga stereotüüpne situatsioon, vastustekstid olid siin lühikesed ja omavahel sarnased. Kõigis kolmes keeles oli praktiliselt kõikides vastustes näha eksplitsiitne *teie*-vorm ja vabandussõna (*vabandust, anteeksi, izvinite požalujsta*). Eesti materjalis oli kõige rohkem indikatiivis verbi *oskama* (7a), soome keeles jagunesid laused aga enam-vähem võrdselt verbide *osata* 'osata', *voida* 'võida' ja *tietää* 'teada' vahel. Eesti ja soome materjalis oli muude situatsioonidega võrreldes suhteliselt vähe konditsionaalivorme (vastavalt ainult viis ja kaheksa 20-st), põhjuseks tõenäoliselt see, et tegemist oli pigem küsimuse kui palvega. Vene materjalis oli valdavalt palveid *ütelda* (*(pod)skažite*), kus bussijaam on, pehmen-dava

grammatilise vahendina kasutati erinevalt eesti ja soome keelest pealause verbi eitust, nt *ne podskazete* 'ei ütle' (7c).

- (7) a Vabandage, kas te oskate öelda, kus asub bussijaam?
KÜSIL, JAATAV IND, MOD – 12
b Anteksi, osaatteko sanoa, missä on bussiasema?
'Vabandage, kas oskate ütelda, kus on bussijaam?'
KÜSIL, JAATAV IND, MOD – 7
c Izvinite, vy ne podskazete, kak dojti do avtobusnoj stancii?
'Vabandage, ega te ei ütle, kuidas minna bussijaama?'
KÜSIL, EITAV IND – 16

3.8. Situatsioon 8: Pealkiri

Oled päeval rääkinud sõbraga ühest raamatust, mille pealkiri on miskipärast meelest läinud. Saadad sõbrale SMS-i, et ta selle meelde tulets. Mida kirjutad?

Kõikidest küsitluse situatsioonidest on see kõige vähem palve, pigem on tegemist info küsimisega. Tõenäoliselt seetõttu ei esinenud üheski keeles ei modaalverbe ega konditsionaali. Siiski esines igas keeles ka mõni peaakt, mis ei olnud vormistatud eriküsilauseks, vaid oli palve ütelda või küsimus *kas sa mäletad...?*. See näitab veel kord, et kõneakti eesmärk ja vorm ei ole üksüheselt seotud ka "puhaste" küsimuste puhul, mis pealegi on vastajale kerged ega vaja mingit erilist pehmendust. Kõigis kolmes keeles olid kõik eksplitsiitsed pöördumised *sina*-vormis. Venekeelses materjalis esines kolm korda eitava verbiga küsimus (*ne napomniš'* '(sa) ei meenuta').

- (8) a Mis oli selle raamatu nimi, millest täna rääkisime?
KÜSIL, JAATAV IND – 17
b Mikä se oli se kirja josta puhuttiin tänään?
'Mis see oli see raamat, millest me täna rääkisime?'
KÜSIL, JAATAV IND – 19
c Kak nazyvalas' ta kniga o kotoroj my dnem govorili?
'Mis oli selle raamatu nimi, millest me päeval rääkisime?'
KÜSIL, JAATAV IND – 11

3.9. Situatsioon 9: Graveering

Sul on vaja lasta kingiks ostetud lusikale graveerida nimi. Lähedal on üks kellassepaäri, aga sa ei tea, kas nad teevad sellist teenust või mitte. Otsustad siiski läbi astuda ja küsida. Mida ütled?

See situatsioon on küsimuste ja palvete vahealalt. Ühelt poolt võiks kirjutada lihtsalt küsimuse graveerimisvõimaluse kohta, kuid see võiks olla ka eelpalve, sest jaatava vastuse korral järgneks ilmselt tegelik palve graveerida lusikale nimi. Kõigis kolmes keeles kasutatakse pöördumises *teie*-vormi, kuid siin ei pruugi see olla viisakusest lähtuv teietamine, vaid kollektiivse subjekti (kellassepatöökoda kui institutsionaalne kollektiiv) poole pöördumine.

Eesti materjalis on kõige tavalisem küsimus *kas saab graveerida* (9a), konditsionaali esines vaid mõnel üksikul juhul. Ka soome vastustes on kõige rohkem otsest küsimust (9b), kuid pisut rohkem kui eesti materjalis esineb pehmendamist (siiski alla poole juhtudest) konditsionaali ja partikli *-hAn* abil (9c). Ka vene materjalis esineb valdavalt otseseid küsimusi (9d), mõnes üksikus lauses on ka pehmendavaid vahendeid (näiteks kolmes lauses eitut).

- (9) a Kas teil siin graveerida ka saab?
KÜSIL, JAATAV IND, MOD – 11
- b Kaiverratteko te lusikoihin?
'Kas te graveerite lusikatele?'
KÜSIL, JAATAV IND – 11
- c Kysyisin vain, että mahtaakohan teillä olla kaiverruspalvelua?
'Küsiaksin ainult, et kas teil võib (*-hAn*) olla graveerimisteenus?'
- d Izvinite, vy zanimaetes' gravirovaniem?
'Vabandage, kas te tegelete graveerimisega?'
- KÜSIL, JAATAV IND – 12

3.10. Situatsioon 10: Kindlustus

Sul on vaja saada oma kindlustusfirmalt tõend panga jaoks, et sinu korter on kindlustatud (kõik on korras ja korter tõesti on kindlustatud, aga pank muretseb laenu tagatise pärast). Leiad internetist ühe oma kindlustusfirma ametniku nime ja meiliaadressi. Mida kirjutad?

See situatsioon esindab formaalset kirjalikku suhtlust võõra vastuvõtjaga. Eesti materjalis oli kõige rohkem konditsionaalis verbiga küsilauseid adressaadi kohta (10a), kuid leidus ka *mina*-vormis konditsionaali, nt *ma sooviksin, mul oleks vaja* (10b). Valdavalt oli tegemist eksplitsiitse pöördumisega, alati *teie*-vormis.

Soome materjalis on samuti enamasti konditsionaalne verb (10c), kusjuures selles situatsioonis on eesti ja soome keeles konditsionaalsete verbide osahulk enam-vähem võrdne (kõigis teistes siinse küsitluse situatsioonides oli eesti materjalis konditsionaali vähem kui soome omas). Erinevalt eesti ja vene keelest esineb soome keeles *sina*-vormis pöördumist.

Vene materjalis esines nii eitava konditsionaaliga konstruktsioone (10e) kui ka otseseid palveid, mis on vormistatud verbiga *prošu* (10d) või mõne muu palvekonstruktsiooniga. Kõik pöördumised olid *teie*-vormis.

- (10) a Kas te võiksite kirjutada mulle tõendi selle kohta, et korter on kindlustatud?
KÜSIL, JAATAV KOND, MOD – 12
- b Mul oleks vaja saada tõend panga jaoks, et mu korter on kindlustatud.
- c Voisitteko lähettää todistuksen minulle?
'Kas te võiksite saata mulle tõendi?'
- KÜSIL, JAATAV KOND, MOD – 11
- d Prošu vydat' spravku o tom, čto...
'Palun väljastada mulle tõend selle kohta, et...'
- VÄITL, JAATAV IND – 7

- e Ne mogli by vy vyslat' mne spravku o tom čto...
'Kas te ei võiks saata mulle tõendi selle kohta, et...'

4. Palvete keeleline vormistus

Eelpool kirjeldatud tulemused annavad võimaluse teha esialgseid üldistusi selle kohta, mille poolest eesti, soome ja vene keele palvete stereotüübid erinevad ja sarnanevad.

4.1. Palve lausetüüp

Nii eesti, soome kui ka vene keeles on käsklausena vormistatud vaid väike osa küsimustikus formuleeritud lause peaaktidest. Kõige rohkem on neid vene (27 palvet/küsimust 200-st, nt 11c), kõige vähem soome materjalis (2 palvet/küsimust 200-st, nt 11b), eesti materjali käsklausete arv jääb nende kahe vahele (11 palvet/küsimust 200-st, nt 11a). Kõigi kolme keele materjalis esinevad käsklused vaid sõprade poole pöördumises (S4 ja S6), vt näited (11a)–(11c), ainult vene materjalis esineb paar käsklauset koos *teie*-vormis pöördumisega (11d)⁷. Eesti ja vene keeles sisaldavad need laused vastavalt palvesõnu *palun* ja *пожалуйста*. Vene keeles näib imperatiiv koos palvesõnaga seega üksikjuhtumitel olevat piisav ka formaalses situatsioonis palve pehmemdamiseks (vt ka Ogiermann 2009).

- (11) Käsklused materjalis
- a Vaata palun ingliskeelne dokument üle, viitsid? (S6 Teesid)
 - b Ota se kirja huomenna mukaan, voisin nyt lukea sen. (S4 Raamat)
'Võta see raamat homme kaasa, võiksin nüüd selle läbi lugeda.'
 - c Prover', požalujsta, tekst. (S6 Teesid)
'Kontrolli, palun, teksti.'
 - d Pomogite mne, požalujsta. (S2 Raamatukogu)
'Aidake mind, palun.'

Eraldi konstruktsioonina oli situatsioonis S5 (Pood) sage ilma verbita nimi-sõnafraasi kasutus (eesti materjalis nt 15, soome materjalis 7 ja vene materjalis 9 korda 20-st), vt näiteid (5a)–(5c) eespool.

Väitlauseid esines kõigi kolme keele materjalis enam-vähem võrdselt (eesti 20, soome 28, vene 21), neid on kõigis situatsioonides v.a S7 (Bussijaam), S8 (Pealkiri) ja S9 (Graveerimine), mis ongi pigem küsimused kui palved. Eesti materjalis esineb väiteid situatsiooniti enam-vähem võrdselt ja väite sisu on enamasti seotud *minakõneleja* vajaduste või soovidega⁸ (verbivormid *vajan*, *vajaksin*, *tahaks* jne, nt 12a). Vene materjalis on näha, et peaaegu pooled väitlausestest on esitatud situatsioonis S10 (Kindlustus), nt (12c), ja tüüpiline väitekujuiline palve sisaldab sõna *pros'ba* või *prošu*, olles seega performatiivlause. Soome materjalis on aga väitlause tüüpiline hoopis S5 (Pood) puhul, nt (12b), ja sisaldab esimeses isikus ja konditsionaalis verbi (*ottaisin* 'võtaksin'). Seega on väitlauseid uuritud materjalis küll keeliti enam-vähem võrdselt, kuid tüüpilised konstruktsioonid ja nende kasutuskontekstid siiski väga erinevad.

⁷ Venekeelset *teie*-pöördumisega käskivas kõneviisis ostupalvet kirjeldab ka Krista Vogelberg (2002: 306), tehes järelduse, et vene keeles piisabki paljudel juhtudel *teie*-vormis verbist, et palve piisavalt viisakalt vormistada.

⁸ Ervin-Tripp (1976) on seda tüüpi nimetanud vajadusväiteks (ingl *need statement*).

- (12) Väitlased materjalis
- a Vabandage, et segan, aga vajan pisut teie abi. (S2 Raamatukogu)
 - b Ottaisin kolme noita grillatuja kanankoipia. (S5 Pood)
'Võtaksin kolm noid grillitud kanakoibi.'
 - c Obračajus' k vam s pros'boj o predostavlenii dokumenta. (S10 Kindlustus)
'Pöördun teie poole palvega dokumendi saamiseks.'

Kõige rohkem peakte oli kõigis kolmes keeles esitatud küsilause, eelkõige üldküsilause vormis. Arvuliselt oli küsilauseid vene keeles küll kõige vähem (143 palvet/küsimust 200-st) ja soome keeles kõige rohkem (163 palvet/küsimust 200-st), kuid võib siiski öelda, et kõigis kolmes keeles on küsilause kõige tüüpilisem palve esitamise viis ja silma hakkavad ka väga kindlad küsilausekonstruktsioonid, mida selleks kasutatakse (vt 3. osa näiteid).

Kõige vähem esines küsilauseid poesituatsioonis, mis on ilmselgelt kõige rutiinsem siin käsitletavatest palvetest. Küsilauseid kasutatakse nende palvete puhul, mis nägu ohustavad, mitte aga teenindussituatsioonides, kus partneri ülesanne ongi kaupa kätte anda ja vastav palve teenindaja nägu ei ohusta.

4.2. Palve peakti verbi vorm küsi- ja väitlausetes

Verbi vorm pakub huvi eelkõige sellepärast, et üheks tüüpiliseks palve esitamise viisiks kõigis kolmes keeles on konditsionaali preesensivorm. Eesti ja soome konditsionaalivormid on sünteetilised ja kasutuselt mõnevõrra erinevad (Metslang 1999). Vene konditsionaalivormid on analüütilised, koosnedes partiklist *by* ja verbi minevikuvormist (vene keele praeguse süsteemi seisukohast).

Üldarvult esineb konditsionaali kõige rohkem soome materjali palvetes, neid on kokku 134, mis tähendab, et 67% juhtudest oli peakt vormistatud konditsionaalivormi kasutades. Kuna enamik lauseid, mis konditsionaali ei sisaldanud, kuulusid situatsioonidesse S7 (Bussijaam), S8 (Pealkiri) ja S9 (Graveerimine), mis on vaadeldud situatsioonidest eesmärgilt kõige rohkem küsimused, võib öelda, et palve puhul on konditsionaal väga sagedane.

Eesti keeles esines konditsionaalivormis verbe vähem, kokku 86 korda (43% kõigist juhtudest). Konditsionaalivormide jagunemine situatsioonitüübiti on üldiselt sarnane soome konditsionaalivormidele, kuid igas situatsioonis on proportsionaalselt vähem eesti konditsionaali. Siiski võib märgata, et kõige formaalsemas situatsioonis S10 (Kindlustus) on eesti ja soome keeles konditsionaale võrdselt, kõige suurem on aga erinevus sõpradevahelise keeruka palve situatsioonis S6 (Teesid), kus kõik soome peaktid on konditsionaalis, eesti materjalis aga vaid 8.

Vene keeles on konditsionaalivorme kõige vähem, ainult 48 korda (24% kõikidest juhtudest). Rohkem esineb neid situatsioonides S2 (Raamatukogu), S3 (Lennujaam) ja S10 (Kindlustus).

Kolme keele materjali konditsionaalivormide esinemise võib kokku võtta järgmiselt: kõige rohkem soome, kõige vähem vene materjalis, eesti materjalis vahepealsel hulgal. Erinevust on näha ka kasutusviisis. Soome materjalis näib konditsionaalis palve olevat vaadeldud keeltest kõige automaatsem, seda kasutatakse enamasti ka palves sõbrale. Samuti esineb soome materjalis päris tihti konditsionaalivormis

verbe, mis ei kuulu modaalkonstruksioonidesse. Eesti materjalis on konditsionaal eelkõige formaalsemate olukordade marker ja esineb põhiliselt modaalverbides. Vene konditsionaal on palvetes peaaegu alati konstruksiooni *ne mogli by vy* osa ja esineb samuti formaalsetes situatsioonides, kuid harvemini kui eesti ja soome keeles.

4.3. Eitus palvetes

Eitus esineb palvete moodustamisel eelkõige vene keeles: 200 peaktist 72 olid vene materjalis eitavad. Kõige suurem osa sellest on konstruksioonil *ne mogli by vy / ne mog(la) by ty* 'kas te/sa ei võiks' (13b), mida esines 41 korda. Kuid 31 korda esines eitus ka muudes konstruksioonides, mis näitab, et see on vene keeles oluline palve pehmemdamise vahend. Situatsioonides S2 (Raamatukogu) ja S7 (Bussijaam) esines ka teine sage konstruksioon *ne (pod)skazete* 'ega te ei ütle' (13c, vt ka 7c), mis ilmselt on samuti väljakujunenud vormel. Üksikuid näiteid eitavatest verbidest oli aga muidki (13d).

Eesti materjalis esines eituslikku peakti kuuel juhul 200-st, soome keeles mitte ühtegi. Kuigi eesti materjalis on eitavaid lauseid ilmselgelt väga vähe, on võrdluses soome materjaliga huvitav see, et neid siiski leidub. Kuuest kolm esines situatsioonis S3 (Lennujaam), mis on kõikidest vaadeldud olukordadest kõige ootatum ja partneri nägu kõige rohkem ohustav palve (13a). Võib siis oletada, et kui vajadust palvet pehmemdada on väga suur, siis tehakse seda eesti keeles mõnikord eitava vormiga.

- (13) a Ega teil pole äkki võimalik mind autoga ära viia?
b *Ne mogli by vy vyslat' mne spravku o tom, čto...*
'Kas te ei võiks saata mulle tõendi selle kohta, et...'
c *Prostitute, vy ne podskazete, gde ja mogu najti...*
'Vabandage, ega te ei ütle, kust ma võin leida...'
d *Ne pomniš' zaglavie?*
'Ei mäleta seda pealkirja?'

4.4. Partneri poole pöördumine: *sina* või *teie*

Kuigi siinse uurimuse eesmärgiks ei olnud otseselt pöördumisvormide uurimine, annab materjal siiski võimaluse teha mõningaid järeldusi ka selles valdkonnas. Eesti, soome ja vene pöördumisi ei ole teadaolevalt varem diskursuse täiendamise ülesande abil võrdlevalt uuritud. Eesti ja vene keele kohta on teada, et vene keeles kasutatakse *teie*-vormi rohkem kui eesti keeles ja see on rohkem kui eesti keeles seotud vanuselise distantsiga (Pajusalu jt 2010a, 2010b). Eesti ja soome keele võrdlusest on teada, et eesti keeles teietatakse omakorda rohkem kui soome keeles, kuid sarnaselt vene keelega on soome keeles hoolimata harvemast esinemisest samuti oluline just ealine distants, vt Jalli, Pajusalu (ilmumas).

Kõigis kolmes keeles kasutati *sina*-vormi sõprade poole pöördumiseks (S4, S6, S8) ja *teie*-vormi võõra vanaproua poole pöördumiseks (S7). Samuti pöörduti *teie*-vormis klassespatöökoja töötaja poole (S9), selles olukorras on tegemist nn kollektiivse *teie*-ga. Erinevused kolme keele vahel ilmsid võõraste poole pöördumises:

raamatukogutöötajat (S2) ja pangaametnikku (S10) vene ja eesti materjalis teietati, soome materjalis leidus nii *sina-* kui *teie-*vormi. Naabrit (S3) eesti ja vene materjalis mõnikord teietati ja mõnikord sinatati, soome materjalis oli siin vaid üks *teie-*vorm. Õppejõu (S1) ja poemüüja (S5) poole eesti ja soome keeles tavaliselt ei pöördutudki, vene keeles pöörduti *teie-*vormis.

5. Kokkuvõte

Artiklis vaadeldi võrdlevalt eesti, soome ja vene stereotüüpsete palvete moodustamist. Palvete vormistus sõltub kultuurilistest praktikatest, st sellest, mida inimesed peavad sobivaks või viisakaks. Kultuurilised käitumisnormid on suures osas ühiskondlikud ja võivad olla sarnased ka üle keelepiiride. Vaadeldud kolm keelt on seotud mitmekeelsusega Eestis: jättes kõrvale globaalse inglise keele, mõjutavad just soome ja vene käitumisnormid kõige rohkem tänapäeva eestlaste suhtlemistavasid.

Vaadeldud keelte materjalis kasutati palvete esitamiseks sarnaseid grammatilisi ressursse: küsilauseid ja verbi konditsionaalivormi, eesti ja vene materjalis ka eitust. Samuti oli kasutusel mitmeid leksikaalseid vahendeid: palve ja vabanduse pragmaatilisi markereid ja modaalkonstruksioone. Sotsiaalse distantsi osutamiseks on kõigis kolmes keeles ka mitmuslik pöördumisvorm *teie*. Samu ressursse aga kasutatakse mõnevõrra erinevalt, vähemalt siinses uurimuses kasutatud materjali põhjal otsustades.

Põhilised erinevused kolme keele materjali vahel olid järgmised:

- vene materjalis kasutati kõige vähem küsilauseid, suhteliselt rohkem oli käsklauseid; konditsionaalivorme oli kõige vähem ja need olid enamasti seotud ühe kindla konstruktsiooniga (*ne mogli by vy / me mog(la) by ty*); see-eest eitust oli suhteliselt sage;
- soome materjalis kasutati kõige rohkem küsilauseid ja konditsionaali, sealjuures esines muude kui modaalsete verbide konditsionaalivorme rohkem kui teistes vaadeldud keeltes; eitust ei esinenud soome palvetes kordagi;
- eesti materjali peaaktid olid kõigis vaadeldud parameetrites (küsilause, osakaal, konditsionaalis verbivormide hulk, modaalkonstruksioonide hulk, eitust, pöördumisviis) sageduslikult soome ja vene materjali palvete vahel.

Küsimus, kas eesti keele n-ö vahepealset olukorda saab seletada ainult nende kolme keele areaalse kontiinumiga või on oluline ka muude keelte mõju (eelkõige rootsi-soome, saksa-eesti ja prantsuse-vene mõjusid silmas pidades), ootab selgitamist järgmistes uurimustes.

Lühendid

KOND	konditsionaal
IND	indikatiiv
KÜSIL	küsilause
KÄSKL	käsklause
VÄITL	väitlause
MOD	modaalkonstruksioon
S1–S10	situatsioonid 1–10 peatükist 3

Viidatud kirjandus

- Blum-Kulka, Shoshana; House, Juliana; Kasper, Gabriele 1989. Investigating cross-cultural pragmatics: An introductory overview. – Shoshana Blum-Kulka, Juliana House, Gabriele Kasper (Eds.). *Cross-Cultural Pragmatics: Requests and Apologies*. Ablex Publishing Corporation.
- Brown, Penelope; Levinson, Stephen C. 1987. Politeness. Some Universals in Language Use. *Studies in Interactional Sociolinguistics* 4. Cambridge: Cambridge University Press.
- Curl, Traci S.; Drew, Paul 2008. Contingency and action: A comparison of two forms of requesting. – *Research on Language and Social Interaction*, 41 (2), 129–153. <http://dx.doi.org/10.1080/08351810802028613>
- 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>
- Hennoste, Tiit; Rääbis, Andriela 2004. Dialoogiaktid eesti infodialoogides: tüpologia ja analüüs. Tartu Ülikooli Kirjastus.
- Jalli, Ninni; Pajusalu, Renate (ilmumas). Sa uss ära sinata mind siin midagi, ega me koos seakarjas pole käind. Puhuttelu Virossa. – Hanna Lappalainen, Johanna Isosävi (Toim.). *Eurooppalaiset puhuttelukäytänteet. Suomalaisen Kirjallisuuden Seura*.
- Jucker, Andreas H. 2009. Speech act research between armchair, field and laboratory. The case of compliments. – *Journal of Pragmatics*, 41 (8), 1611–1635. <http://dx.doi.org/10.1016/j.pragma.2009.02.004>
- Keevallik, Leelo 2004. Politeness in Estonian: A matter of fact style. – Leo Hickey, Miranda Stewart (Eds.). *Politeness in Europe*. Clevedon: Multilingual Matters, 203–217.
- Metslang, Helle 1999. Is the Estonian and Finnish conditional actually a conditional? – Mati Ereht (Ed.). *Estonian Typological Studies III*. Publications of the Department of Estonian of the University of Tartu 11. Tartu: Tartu University Press, 97–127.
- Metslang, Helle 2004. Imperative and related matters in everyday Estonian. – *Linguistica Uralica*, 4, 243–256.
- Ogiermann, Eva 2009. Politeness and in-directness across cultures: A comparison of English, German, Polish and Russian requests. – *Journal of Politeness Research: Language, Behavior, Culture*, 5 (2), 189–216. <http://dx.doi.org/10.1515/JPLR.2009.011>
- Pajusalu, Renate; Vihman, Virve; Klaas, Birute; Pajusalu, Karl 2010a. Eestlaste ja venelaste suhtluskäitumine: sina, teie ja keegi veel. [Politeness behavior among Estonians and Russians: formal and informal second person and the third person.] – *Eesti Rakenduslingvistika Ühingu aastaraamat*, 6, 207–224. <http://dx.doi.org/10.5128/ERYa6.13>
- Pajusalu, Renate; Vihman, Virve; Klaas, Birute; Pajusalu, Karl 2010b. Forms of address across languages: Formal and informal second person pronoun usage among Estonia's linguistic communities. – *Intercultural Pragmatics*, 7 (1), 75–101. <http://dx.doi.org/10.1515/iprg.2010.004>
- Peterson, Elizabet 2010. Perspective and politeness in Finnish requests. – *Pragmatics*, 20 (3), 401–423.
- Rääbis, Andriela 2012. Direktiivisekventsid 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 1969. *Speech Acts: An Essay in the Philosophy of Language*. Cambridge: Cambridge University Press. <http://dx.doi.org/10.1017/CBO9781139173438>
- Sorjonen, Marja-Liisa; Raevaara, Liisa; Lappalainen, Hanna 2009. Mä otan tän. Käynnin syyn esittämisen tavat R-kioskillä. – Hanna Lappalainen, Liisa Raevaara (Toim.). *Kieli kioskillä. Tutkimuksia kioskiasioinnin rutiineista*. Helsinki: Suomalaisen Kirjallisuuden Seura, 90–119.
- Taleghani-Nikazm, Carmen 2006. *Request Sequences: The Intersection of Grammar, Interaction and Social Context*. Philadelphia: John Benjamins Publishing Company. <http://dx.doi.org/10.1075/sidag.19>

- Tanner, Johanna 2012. Rakenne, tilanne ja kohteliaisuus. Pyynnöt S2 oppikirjoissa ja autenttisisissa keskusteluissa. Helsingin yliopisto.
- Vogelberg, Krista 2002. The politeness debate continued: notes on some key controversial issues in Brown and Levinson's theory. – Renate Pajusalu, Tiit Hennoste (Toim.). Tähdendusepüüdja. Tartu Ülikooli üldkeeleteaduse õppetooli toimetised 3. Tartu: Tartu Ülikooli Kirjastus, 340–354.
- Yu, Kyong-Ae 2011. Culture-specific concepts of politeness: Indirectness and politeness in English, Hebrew and Korean requests. – Intercultural Pragmatics, 8 (3), 385–409. <http://dx.doi.org/10.1515/iprg.2011.018>

Renate Pajusalu (Tartu Ülikool) põhilised uurimisvaldkonnad on eesti keele semantika ja pragmaatika, ka võrdlevas perspektiivis.
Tartu Ülikool, eesti ja üldkeeleteaduse instituut, Ülikooli 18, 50090 Tartu, Estonia
renate.pajusalu@ut.ee

REQUESTS IN ESTONIAN, FINNISH AND RUSSIAN: GRAMMAR AND PRAGMATICS

Renate Pajusalu

University of Tartu

The article discusses the formulation of requests in Estonian, Russian and Finnish. The data was collected by a discourse completion task: participants had to write what they would say or write in a given situation. 200 requests in each of the languages were collected. The article compares ten of the presented situations in three languages based on this data, bringing out the similarities and differences. The analysis focuses on grammatical categories that are pragmatically essential to the formulation of requests (sentence type, morphological form of the verb, presence of negation, address by singular or plural pronoun). The main unit of analysis is the head utterance (request proper) of the reply. However, in some categories the whole replies are considered.

In the summary, generalizations are made about the more frequent grammatical devices used in the formulation of requests. All three languages typically used interrogative sentences to formulate requests. However, in Russian the number of imperative sentences was somewhat greater than in Estonian and Finnish. Another grammatical device common to all three languages was the conditional verb form. The conditional was most frequent in Finnish and least frequent in Russian. Negation was characteristic of requests only in Russian, where it occurred both with and without conditional forms. In Finnish there were no occurrences of negation, while in Estonian it was observed on rare occasions. Address by plural form was most frequent in Russian, less frequent in Estonian and least frequent in Finnish.

Estonian is placed between Finnish and Russian in the frequency of all grammatical categories studied, which can probably be explained by the fact that Estonian is influenced by both Finnish and Russian.

Keywords: interrogative, address forms, conditional mood, negation, Estonian, Russian, Finnish

AUTHORSHIP VERIFICATION OF OPINION PIECES IN ESTONIAN

Timo Petmanson

Abstract. Authorship verification is an important subproblem in authorship attribution and plagiarism detection tasks. We present a novel approach for extracting stylistic features unique to individual authors. We use the correlations of important textual features as a way to learn the style. The goal of our proposed method is to answer the following question: given a set of documents known to be written by the same person and an unknown document, is the unknown document also written by that individual. We present the first study of this problem conducted on opinion pieces written in Estonian. Our method achieves 74% precision, which is comparable with current state-of-the-art systems tested in other languages, whereas the recall level is still something to be improved on.

Keywords: natural language processing, text analysis, linguistic expertise, machine learning, pattern mining, feature correlations, Estonian

1. Introduction

The question of whether any two documents are written by the same author was proposed as a fundamental aspect of every authorship attribution problem by Koppel, Winter (2014). Authorship verification arises in many case studies such as police investigations requiring detection of the ownership of defamatory letters and ransom notes (De Vel et al. 2001, Abbasi, Chen 2005). It could also be used to learn the general profile of a group of persons such as sexual predators (Inches, Crestani 2012).

We address an extended version of the problem: given several documents known to be written by the same author, is that particular person the author of one or more unknown documents. This problem has been also addressed in PAN Workshop and Competition: Uncovering Plagiarism, Authorship and Social Software Misuse (Argamon, Juola 2011, Juola, Stamos 2013). Although authorship verification is

also important in other domains such as source code analysis (Frantzeskou et al. 2004), we concentrate on texts.

What makes this problem particularly difficult is that we are given only documents from a single author, not from a closed set of candidates. Although authorship attribution and identification are relatively well-studied areas, most of the literature addresses case studies with a closed set of authors, which might not be sufficient in real-world scenarios (Koppel et al. 2009). The task formalized in open-set fashion has gained more interest recently.

There have been a few studies related to authorship verification in Estonian, such as plagiarism detection in dictionaries (Langemets, Voll 2008). However, this study is the first addressing the issue directly, setting the baseline for future studies.

2. Data description

Our Corpus of Opinion Pieces of Estonian covers 1474 news articles from 318 authors, where 295 of them have ten articles or less. Figure 1 depicts a more detailed distribution of the number of documents per author. As many as 123 authors have only two articles in our corpus. We did not use authors that have only a single article, because we need at least one document for training and one for evaluation. The corpus is a subset of opinion pieces (editorials, columns, etc.) published by Postimees Online news portal from January 2008 till September 2013.

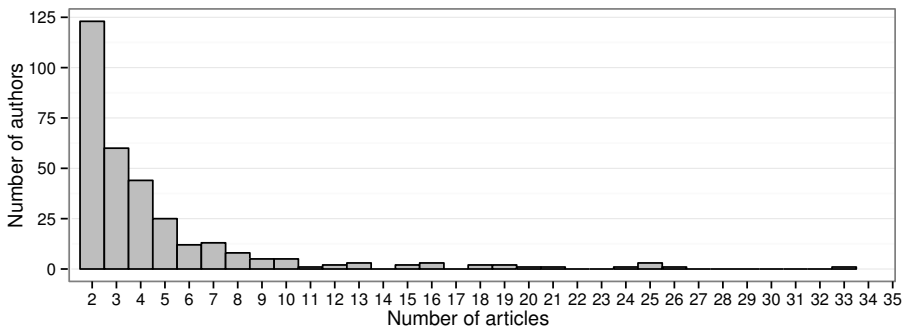


Figure 1. Histogram depicting the number of documents per author in the corpus

Our computational methods, which we describe in the following sections, require a reference corpus. For this task, we decided to use a subset of opinion pieces from the Balanced Corpus of Estonian¹ which were not written by journalists presented in our main corpus.

Authorship identification of news articles is particularly difficult as the articles are edited before publishing. The style and language use is more normalized compared to more informal texts such as emails and blogs. This makes it harder to exploit certain features like usage of emoticons, number of grammar errors, etc. which are often representative of people of a certain age and gender (Koppel, Argamon, Shimon 2002, Stamatatos 2009). Although news articles hide many stylistic aspects the authors would use informally, the news corpus makes an interesting example of

¹ The Balanced Corpus of Estonian <http://www.cl.ut.ee/korpused/grammatikakorpus/index.php?lang=en> (31.1.2014).

detecting less obvious patterns. This might be important in cases of someone trying to impersonate somebody else or intentionally hiding his/her own writing style.

3. Feature extraction

Common features used in authorship verification are traditional in natural language processing: word and character n-grams, word prefixes and suffixes of various sizes, exploiting external sources such as Wordnet and Wikipedia (Houvardas, Stamatatos 2006, Brocardo et al. 2013). We experimented with various combinations of features and decided to use the most easily interpretable ones. We used the following formal and lexical features: is the word uppercase, does the word start with an uppercase letter, does the word contain a digit. From morphological features we considered word lemma, part-of-speech (POS), case and verb type. Note that not all possible features exist for all words. For example, verbs do not have cases.

Another reason we did not include traditional n-grams is that typical information about the words in our corpus is already encoded by the lemma and other morphological features. N-grams would be valuable for capturing the use of emoticons, excessive punctuation and significant amount of grammar and typing errors. This would be important in more informal texts like e-mails, but not in our corpus.

Table 1. Example of important features used in the study

word	lemma	POS	case	verb type	uppercase
Usjas	usjas	adjective	nominative	–	yes
kaslane	kaslane	noun	nominative	–	no
jookseb	jooksma	verb	–	b	no
künklikul	künklik	adjective	adessive	–	no
maastikul.	maastik	noun	adessive	–	no

As an example, consider the following sentence: “Usjas kaslane jookseb künklikul maastikul” (‘The sinuous feline is running on a hilly terrain’). Table 1 shows the extracted features for the sentence. The morphological information was extracted using the morphological analyzer *t3mesta* (Kaalep 1997, Kaalep, Vaino 1998).

3.1. Pattern mining

Simple features described in Table 1 are good for many classic tasks like clustering the documents by topic. However, we might get more descriptive stylistic features if we combine several of them. For example, an author may have a unique way of ordering certain types of words in some phrases. Some of this information can be captured, if we encode each feature as a pattern

(feature, offset, value)

where *offset* determines the relative position of the feature/value combination. For instance, a pattern (*case, -1, nominative*) would say that the case of the previous word is nominative. Using the offset makes sense, when we want to combine two

simpler patterns. For example, a composite pattern identifying all words in the partitive case followed by a verb is (*case, 0, partitive*) & (*pos-tag, 1, verb*). The maximal absolute offset was two.

Although we extracted one set of patterns from all words, we also extracted sets of them specifically related to noun, verb and adjective phrases. We achieved this by separately considering contexts of nouns, verbs and adjectives with a radius of two words. A final set of patterns were extracted from contexts surrounding punctuation. Punctuation patterns may give insight to unique formatting preferences.

We used the PatNLP library (Petmanson, Laur 2012) to mine the frequent patterns for each author. We instructed the library to detect patterns covering at least five percent of the tokens. Smaller thresholds were not computationally feasible and larger thresholds yielded too generic patterns.

Not all the frequent patterns describe the writing style of the author, but can be just specific to the Estonian language or the news article style in general. Such patterns are not discriminative enough to capture the style of a journalist. To tell the relevance of a pattern, we estimated the p-value using 500 randomly selected articles from the Balanced Corpus of Estonian as a reference. The p-value was computed as the percentage of documents where the pattern was at least as frequent as in the articles of the author. The smaller the p-value, the more relevant and surprising the pattern is. We kept only patterns having p-value less than 2.5 percent. The threshold was chosen such that it would yield about a few hundred statistically significant patterns.

3.2. Classification model

As previously explained, we mined the patterns separately for five groups of tokens: all words, nouns, verbs, adjectives and punctuation. The simplest way to use them as features for a machine learning classifier is to encode them as a vector of frequencies – a single vector for a single document. However, such patterns only capture document level co-occurrence of features. Information on whether two frequent and statistically significant patterns are never used together in the same phrase or sentence is discarded. However, this kind of knowledge might be useful for classification.

To address the issue, we encoded the matches of the patterns as bitvectors, where a *true* bit indicates a match. Next, we computed the *Matthews correlation coefficient* (MCC) between all bitvectors. Note that we used MCC instead of alternatives such as the *Pearson correlation coefficient* because the matches are binary events. The Pearson correlation assumes that the events follow a Gaussian distribution, whereas MCC is specifically designed for bitvectors. Now, by encoding the correlations as a single vector, we can use it as an input for the classifier.

As our problem statement assumed an open set of authors, we can only learn from known documents, i.e. positive examples. Such a problem is known as a one-class learning problem in machine learning and has many solutions. In this work, we used a 1-class *support vector machine* (Cortes, Vapnik 1995) to perform the classification task. We employed the Python *scikit-learn* library (Pedregosa et al. 2011) implementation using the RBF kernel with default parameters.

4. Results

To evaluate our proposed method, we created training and testing collections for each author in the Corpus of Opinion Pieces of Estonian (see Chp. 2). In order to build the model, we could only use the documents written by a single person. As our corpus also contained authors with only two articles, the testing corpora were designed so that they contained one randomly selected document from the same author and another from a different author. The training set contained the rest of the known articles by the author. The total number of testing documents in all collections was 636 with 318 articles for both positive and negative examples respectively.

We built a separate model for each author using the respective training set and evaluated that particular model on the testing set of the same author. Next, we computed the overall number of true positive (TP), true negative (TN), false positive (FP) and false negative (FN) hits (see Table 2). Using these counts, we computed relevant statistical measures for the method (see Table 3).

We see that our method has an acceptable precision of 74%, low recall of 14%, which combined as an F1-score is 23%. Current state-of-the art authorship identification systems achieve an F1-score of 75% on the PAN'13 corpus containing various documents in Spanish, English and Greek (Juola, Stamatos 2013). We need to note that the evaluation setup used in (Juola, Stamatos 2013) also included correctly classified negative examples in precision and recall computation. The reason was that they did not require all the questions to be answered. Thus, resulting precision, recall and F1-scores were all equal if a system answered all the questions. But we cannot directly compare the performances, since the datasets on which the algorithms were evaluated are different.

Table 2. Number of true positive (TP), true negative (TN), false positive (FP), false negative (FN) hits for authors with more than 10 training documents ($N > 10$), with less or equal to ten documents ($N \leq 10$) and all authors

Datasets	TP	TN	FP	FN
$N \leq 10$	10	286	10	261
$N > 10$	35	16	6	12
All	45	302	16	273

Table 3. Precision, recall, accuracy and F1-scores computed from Table 2

Metric	$N \leq 10$	$N > 10$	All
precision	78%	63%	74%
recall	12%	45%	14%
accuracy	54%	59%	55%
F1-score	20%	52%	23%

Our method is very conservative for authors having ten training articles or less. We achieve good precision of 78%, but low recall. Authors with more than ten training documents achieve better recall of 45%, albeit worse precision of 63%.

As the topic and word usage in documents are very important features, we decided to study them and see if we could find any link between misclassified

examples. In Table 4, we have applied the Latent Semantic Indexing (Deerwester et al. 1990) method to extract the most influential keywords of the central topic in our document collections.

Table 4. Main keywords identifying the central topic in all documents, true positive and false positive examples

Document collection	LSI main topic keywords
All documents	0.5*"eesti", 0.3*"aasta", 0.2*"riik", 0.2*"inimene", 0.1*"laps"
True positive documents	0.4*"eesti", 0.3*"aasta", 0.2*"inimene", 0.2*"riik", 0.2*"laps"
False positive documents	0.5*"eesti", 0.2*"aasta", 0.2*"riik", 0.2*"inimene", 0.1*"euroopa"
False negative documents	-0.4*"laps", -0.3*"eesti", -0.2*"aasta", -0.2*"inimene", -0.2*"riik"

We fitted the LSI model separately on all true positive, true negative, false positive and false negative documents and extracted the most influential keywords. Coefficients show the weights of the keywords attributing to the central topic. A negative coefficient means that a particular keyword is statistically underrepresented in the document collection.

We see that for all documents the main keywords describing the central topic are, in order of importance, *eesti* ('Estonia', 'Estonian'), *aasta* ('year'), *riik* ('state'), *inimene* ('person') and *laps* ('child'). Almost the same keywords with similar coefficients are also present in true and false positive document corpora. Although the number of false positives was relatively small, it seems they were misclassified as the content was very similar, albeit by a different author. False negative documents seem in general not to be about the main topic in other collections (note the negative coefficients of main keywords in Table 4), indicating that the profile learned by the classifier included too many topic-related features.

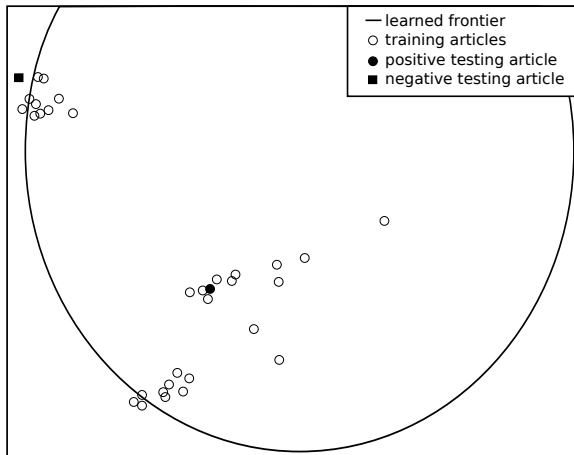


Figure 2. Frontier learned by support vector machine for columnist Ahto Lobjakas. The full coordinate space is transformed to 2D-coordinate space by applying principal component analysis. Everything inside the frontier would be classified as from the same author and everything outside as from a different author

Another possible reason for misclassification is that some authors do not in fact have a single profile or style, but may have several. In Figure 2, we have plotted document representations from a single training and testing corpora pair. We see at least three distinctive clusters of documents, which correspond to different styles or topics of the author. Although in this particular case both unknown documents were classified correctly, the model representation could have been better. Instead of a single large frontier, we might have tried to learn three smaller frontiers instead, each capturing the specifics of representative clusters. How to exploit this will be of particular interest of our future studies.

5. Summary

In this work, we described a novel method using pattern mining and feature correlations for the task of authorship verification. We achieved acceptable precision of 74%, albeit not very satisfying recall. The main reason for misclassification seemed to be the small set of training samples and the large impact of the topic on the final authorship model. Although the topic of the documents can be helpful in the case of niche authors, our future studies should concentrate on how to separate topics from the unique stylistic features of the authors.

References

- Abbasi, Ahmed; Chen, Hsinchun 2005. Applying authorship analysis to extremist-group web forum messages. – *Intelligent Systems, IEEE*, 20 (5), 67–75. <http://dx.doi.org/10.1109/MIS.2005.81>
- Argamon, Shlomo; Juola, Patrick 2011. Overview of the International Authorship Identification Competition at PAN-2011. – V. Petras, P. Forner, P. D. Clough (Eds.). *CLEF 2011 Labs and Workshop, Notebook Papers*, 19–22 September 2011, Amsterdam, The Netherlands. <http://www.informatik.uni-trier.de/~ley/db/conf/clef/clef2011w.html> (1.2.2014).
- Brocardo, Marcelo Luiz; Traore, Issa; Saad, Sherif; Woungang, Isaac 2013. Authorship verification for short messages using stylometry. – *Proceedings of the IEEE 2013 International Conference on Computer, Information and Telecommunication Systems (CITS)*, 1–6. <http://dx.doi.org/10.1109/CITS.2013.6705711>
- Cortes, Corinna; Vapnik, Vladimir 1995. Support-vector networks. – *Machine Learning*, 20 (3), 273–297. <http://dx.doi.org/10.1007/BF00994018>
- De Vel, Olivier; Anderson, Alison; Corney, Malcolm; Mohay, George 2001. Mining e-mail content for author identification forensics. – *ACM Sigmod Record*, 30 (4), 55–64. <http://dx.doi.org/10.1145/604264.604272>
- Deerwester, Scott; Dumais, Susan; Landauer, Thomas; Furnas, George; Harshman, Richard 1990. Indexing by latent semantic analysis. – *Journal of the American Society for Information Science (JASIS)*, 41 (6), 391–407. <http://www.informatik.uni-trier.de/~ley/db/journals/jasis/jasis41.html> (1.2.2014).
- Frantzeskou, Georgia; Gritzalis, Stefanos; MacDonell, Stephen 2004. Source code authorship analysis for supporting the cybercrime investigation process. – *Proceedings of the 1st International Conference on e-business and Telecommunications Networks (ICETEO4)*, Setúbal, Portugal, 85–92. <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.104.6476> (1.2.2014).

- Houvardas, John; Stamatatos, Efstathios 2006. N-gram feature selection for authorship identification. – Jérôme Euzenat, John Domingue (Eds.). *Artificial Intelligence: Methodology, Systems, and Applications*. 12th International Conference, AIMSA 2006, Varna, Bulgaria, September 12-15, 2006. Proceedings. Lecture Notes in Computer Science 4183. Berlin, Heidelberg: Springer, 77–86.
- Inches, Giacomol; Crestani, Fabio 2012. Overview of the International Sexual Predator Identification Competition at PAN-2012. – P. Forner, J. Karlgren, C. Womser-Hacker (Eds.). *CLEF 2012 Evaluation Labs and Workshop, Online Working Notes*, Rome, Italy, September 17-20. <http://dblp.uni-trier.de/db/conf/clef/clef2012w.html#InchesC12> (1.2.2014).
- Juola, Patrick; Stamatatos, Efstathios 2013. Overview of the Author Identification Task at PAN 2013. – P. Forner, H. Müller, R. Paredes, P. Rosso, B. Stein (Eds.). *Information Access Evaluation. Multilinguality, Multimodality, and Visualization*. 4th International Conference of the CLEF Initiative, CLEF 2013, Valencia, Spain, September 23-26, 2013. Proceedings. Springer. <http://www.clef-initiative.eu/documents/71612/3095ffc3-376b-40eb-af10-8251c5f107f6> (29.9.2013).
- Kaalep, Heiki-Jaan 1997. An Estonian morphological analyser and the impact of a corpus on its development. – *Computers and the Humanities*, 31 (2), 115–133. <http://dx.doi.org/10.1023/A:1000668108369>
- Kaalep, Heiki-Jaan; Vaino, Tarmo 1998. Kas vale meetodiga õiged tulemused? Eesti keele morfoloogiline ühestamine statistika abil. – *Keel ja Kirjandus*, 1, 30–38.
- Koppel, Moshe; Argamon, Shlomo; Shimoni, Anat Rachel 2002. Automatically categorizing written texts by author gender. – *Literary and Linguistic Computing*, 17 (4), 401–412. <http://dx.doi.org/10.1093/lc/17.4.401>
- Koppel, Moshe; Schler, Jonathan; Argamon, Shlomo 2009. Computational methods in authorship attribution. – *Journal of the American Society for information Science and Technology*, 60 (1), 9–26. <http://dx.doi.org/10.1002/asi.20961>
- Koppel, Moshe; Winter, Yaron 2014. Determining if Two Documents are by the Same Author. – *Journal of the American Society for Information Science and Technology*. *Journal of the Association for Information Science and Technology*, 65 (1), 178–187. <http://dx.doi.org/10.1002/asi.22954>
- Langemets, Margit; Voll, Piret 2008. Sõnaraamatu kohtulingvistiline analüüs: Eesti pretsetent. [Linguistic forensic analysis of a dictionary: an Estonian precedent.] – *Eesti Rakenduslingvistika Ühingu aastaraamat*, 4, 67–86. <http://dx.doi.org/10.5128/ERYa4.05>
- Pedregosa, Fabian; Varoquaux, Gaël; Gramfort, Alexandre; Michel, Vincent; Thirion, Bertrand; Grisel, Olivier; Blondel, Mathieu; Prettenhofer, Peter; Weiss, Ron; Dubourg, Vincent; Vanderplas, Jake; Passos, Alexandre; Cornapeau, David 2011. Scikit-learn: Machine learning in Python. – *The Journal of Machine Learning Research*, 12, 2825–2830.
- Petmanson, Timo; Laur, Sven 2012. Pattern based fact extraction from Estonian texts. National Programme for Estonian Language Technology. Project Report. University of Tartu. February 17, 2012. <http://www.keeletehnoloogia.ee/ekt-projektid/mallipohine-faktituletus-tekstikorpukest/projekti-tulemusi-kajastavad-uurimisraportid/esimene-vaheraport> (1.2.2014).
- Stamatatos, Efstathios 2009. A survey of modern authorship attribution methods. – *Journal of the American Society for Information Science and Technology*, 60 (3), 538–556. <http://dx.doi.org/10.1002/asi.21001>

AUTORITUVASTUS EESTIKEELSETES ARVAMUSARTIKLITES

Timo Petmanson

Tartu Ülikool

Autorituvastus on üks plagiaarismituvastuse olulisi alamprobleeme. Käesolevas töös pakume välja autorituvastuse algoritmi eestikeelsete arvamusartiklite jaoks. Ajakirjaniku tuvastamiseks otsime tema tekstidest mustreid, mis kirjeldavad kõige paremini just talle omaseid lause- ja sõnakonstruksioone.

Autorituvastuse jaoks on aegade jooksul pakutud välja mitmeid lahendusi, kuid reeglina on alati eeldatud kindlat kandidaatautorite hulka. Käesolevas töös vaatame selle probleemi keerukamat kuju, kus võimalike autorite hulk on täpsustamata ning tuleb vastata vaid küsimusele, kas etteantud tekst on konkreetse autori kirjutatud või mitte. Selle ülesande piisavalt hea lahendus omab mitmeid praktilisi rakendusi, nagu näiteks anonüümsete laimukirjade autorite tuvastus.

Siinne artikkel on esimene, mis käsitleb probleemi eestikeelsetel tekstidel. Kirjeldatud meetod saavutab Postimees Online korpusel 74% täpsuse. Selline täpsus on võrreldav parimate süsteemidega teiste keelte jaoks, kuid samas on 14% saagis siiski liiga madal.

Võtmesõnad: keeletöötlus, tekstianalüüs, keeleekspertiis, masinõpe, mustrikaeve, tunnuste korrelatsioon, eesti keel

Timo Petmanson is a second year PhD student at the Institute of Computer Science of the University of Tartu and works on topics related to machine learning and language technology.

Liivi 2, 50409 Tartu, Estonia

timo_p@ut.ee

INVESTIGATING CULTURAL VARIABILITY IN RATER JUDGEMENTS OF ORAL PROFICIENCY INTERVIEWS

Irina Stassenko, Liljana Skopinskaja, Suliko Liiv

Abstract. The article is concerned if the cultural background of non-native raters could lead to substantial differences in the assessment of students' oral performances during the national examination in the English language in Estonia. The discussion involves the analysis of the ratings of twenty video-recordings of oral proficiency interviews by two rater groups of the Estonian and Russian origin, and a questionnaire study investigating rater perceptions of the national examination. Despite the lack of any marked cultural bias being displayed in the raters' behaviour, the results of the study reveal a number of significant differences in their perceptions of various aspects of the national examination as well as their own behaviour during the rating process.*

Keywords: assessment, rater variation, rater reliability, cultural validity, oral proficiency interview

1. Introduction

The assessment of oral proficiency interviews can rarely be error-free as scores assigned to test takers are based on a reflection, not only of the quality of a candidate's performance, but also of the qualities of the rater (McNamara 2000: 37). Extensive research into rater behaviour indicates different factors affecting the variability in raters' judgements, such as severity, i.e., how severe or lenient a rater is (Lumley, McNamara 1995: 56, Bachman et al. 1995: 238–257), consistency in assessment (Wigglesworth 1993: 308–309, Luoma 2004), variety of perceptions of what constitutes speaking proficiency (Pollitt, Murray 1996, Fulcher 2003: 144), educational level, i.e., whether the rater is a teacher or a professional rater (Brown 1995, Fulcher 2003: 142, Pajupuu 2007) and linguistic background, i.e., whether the rater is a native or a non-native speaker (Brown 1995, Winke et al. 2012).

* This research has partly been supported by the Estonian Science Foundation Grant No 9037 "Assessing Variables: Creating Benchmarks in High-Stakes Proficiency Tests".

While most of the studies on raters' behaviour concentrate either on their linguistic background or professional experience, little attention has been paid to non-native rater variation, although some cultural variability has been detected in interlocutors' handling of oral proficiency interviews in E. Alas' doctoral thesis (2010). In addition, cultural variation may be observed in the diversity of communication styles, talk distribution, or turn-taking patterns (Tannen 1984, Trompenaars 1994). In other words, communication styles reflect cultural values and the different ways in which interpersonal relations are believed to be achieved best (FitzGerald 2003: 79). Hence, the concept of cultural validity, as a form of test validity, has recently been introduced to link cultural and linguistic factors to test takers' assessment outcomes (Abedi 2011). Based on the differences in value dimensions as well as communication styles between the two cultures, where Estonian culture is characterised as an individualist, small power distance and low uncertainty avoidance culture, as contrasted to Russian culture with its collectivist, large power distance and high uncertainty avoidance orientations (Hofstede 1997, Pajupuu 2001), it is possible to assume that the raters' ways of thinking that permeate their culture and predisposed notions of what good language proficiency constitutes may also influence their rating behaviour, thus calling into question the reliability of test results. Moreover, serious doubts have lately been raised by INNOVE, the former National Examinations and Qualification Centre of Estonia, as to rater validity and reliability in oral proficiency interview results of Russian-based school learners, as compared to their much lower scores in the other parts of the national examination in English (Kriisa 2012: 27).

In view of the absence of relevant studies conducted in Estonia the following research hypothesis was formulated: the conduct of the two non-native groups of Estonian and Russian raters in the oral part of the national examination in English in Estonia will display culture-related differences in their rating process that may in turn affect the candidates' scores.

2. Method and participants

As a rule, raters involved in the speaking test of the national English language examination in Estonia are practising teachers who have undergone standardised assessor training. The speaking test to be marked comprises an introduction (not rated), a monologue and follow-up questions (task 1) and a role play (task 2). The marking scale for speaking consists of four categories (communication, vocabulary, grammar, and fluency & pronunciation) in terms of which the language produced by a candidate is described (REKK 2008). The marking scale has six levels of language proficiency with the descriptors of what the candidate is supposed to be able to express under each of the abovementioned four categories. The minimum score for each of the criteria is 0, and the maximum score is 5 points. The maximum total score that can be assigned is 20 points. The candidate's final mark represents a sum total of all the aforementioned categories. However, for the sake of the current study the raters were required to evaluate both speaking tasks – a monologue and a role play – separately, although only one grade is usually assigned for the speaking test.

18 raters were randomly selected from Estonian and Russian-based schools to mark 20 video-recordings of oral proficiency interviews produced during the national English language examination of 2012 within the framework of the Estonian Science Foundation Grant No 9037. Most of the teachers participating in the study had an assessment experience in the national examination, the length of service ranging from 2 to 18 years. Out of 18 assessors, only 4 (2 Estonians and 2 Russians) were short of relevant experience. As to their teaching experience, the length of service of the Estonian teachers ranged from less than 5 years to more than 15 years, while the Russian teachers had mostly more than 15 years of service.

Although 7105 students participated in the national English language examination in the spring of 2012 (Kriisa 2012: 3), the choice of the material for the study was limited as the video-recordings were made only with the written consent of all the parties: a candidate, an interviewer and a rater. Thus the permission for recording was obtained only from 20 Estonian 12th-formers studying in one Tallinn and two Harju county secondary schools. In spite of the absence of Russian test-takers, which is a drawback of the present study, the recorded performances represented all levels of oral proficiency in English with the students' scores ranging from the highest to the lowest ones. The database of 20 video-recordings comprises 300 minutes of interview time in total. The recordings as well as the raters were assigned letter-codes and numbers in a random order: Clip1 to Clip20 for the recordings, E1 to E9 for the raters from Estonian-based schools and R1 to R9 for the Russian-based ones. All the participants were female and non-native speakers of English.

In order to elicit more data about the raters' attitudes to and their behaviour during the assessment procedure, a questionnaire study was implemented among the same participants. The data obtained from the Estonian and Russian raters' assessment sheets and questionnaire forms were analysed with the aim of establishing any distinctive differences in the assessment of the aforementioned cultural groups. This was further analysed by statistical methods (ANOVA, the 95% Confidence Interval for Mean, Chi-Square Tests) to ascertain if the collected variations were statistically relevant.

3. Discussion and results

The section presents the findings of the two research instruments obtained from the assessment of the oral performance of 20 test-takers by the two rater groups, and a questionnaire study conducted among the same assessors.

3.1. Assessment of oral proficiency interviews

The results of the assessment of each task of the speaking test are analysed by means of juxtaposition of the ratings of the two cultural groups to discover any differences in the respective assessment procedures.

3.1.1. Monologue and follow-up questions (task 1)

The analysis of the marks assigned for the first task is based upon the four criteria of the marking scale (REKK 2008). Table 1 below reflects the mean scores assigned for each aspect of the monologue, these slightly exceeding 4 points.

Table 1. The mean scores for task 1

Nationality	Communication	Vocabulary	Grammar	Fluency & pronunciation
Estonian	4.36	4.14	4.04	4.36
Russian	4.41	4.31	4.20	4.39

Although no statistically significant differences were revealed in the distribution of the scores afforded for the candidates' communication skills and fluency & pronunciation, since more than 55% of the 20 performances received the highest scores for these aspects, there is slight variation in the raters' opinions as regards the candidates' grammar and vocabulary usage, this demonstrating the respective raters' different perceptions of what constitutes good knowledge of grammar or vocabulary. The Estonian raters were more reluctant to award either the lowest or the highest scores for these aspects. Only 29% of the test-takers received the highest score for their grammar by the Estonian rater group, whereas with the Russian group the number of such candidates rose to 41%. The scores for vocabulary usage were 32% and 47% respectively.

More controversy emerged in connection to three recordings (Clip2, Clip9, and Clip11) where the highest scores assigned by the Russian rater group were in marked contrast to those of the Estonian group, the latter evaluating the same students as hesitant speakers with a limited choice of vocabulary and grammar (mark 3).

Despite having different perceptions of the proficiency level of some of the candidates (these being statistically irrelevant), both rater groups emphasized in their notes the importance of a logical structure and varied vocabulary choice in a monologue.

3.1.2. Role play (task 2)

Similarly to the rating of monologues, the mean scores for the second task do not reveal any statistically significant differences, these being even between the two groups while assessing communication skills and fluency & pronunciation, and slightly higher for lexis and grammar usage with the Russian raters (see table 2).

Table 2. The mean scores for task 2

Nationality	Communication	Vocabulary	Grammar	Fluency & pronunciation
Estonian	4.35	4.21	4.00	4.43
Russian	4.38	4.38	4.11	4.45

The same divergence of opinions exists in the assessment of grammar and lexis here: a smaller number of candidates received the highest scores from the Estonian

raters. In comparison to the Russian assessors who awarded 56% of the students with the highest score for their vocabulary, the Estonian raters considered 41% worthy of this mark. With grammar assessment, the percentages were 36% and 26% respectively. Yet the frequency of awarding the lowest score for the candidates' role playing was higher with the Russian raters, as up to 8 % of the candidates received only 1 point for any of the criteria in the marking scale, whereas with the Estonian assessors it was only 3%.

To sum it up, both rater groups share similar perceptions about the candidates' good role playing performance in their notes, emphasizing the presence of both the introduction and conclusion to the role play, natural turn-taking patterns and question formation skills.

3.1.3. The total scores for oral proficiency interviews

After the points for both tasks had been awarded, the raters were to decide upon the final scores. As can be seen in table 3, there are no obvious differences in the mean scores of the two rater groups. Although no apparent severity or leniency is exposed in their behaviour, the results still show that the Russian assessors tended to give slightly higher points for each category, hence their total mean score (17.27) is slightly higher than that of the Estonian raters (16.96). The 95% Confidence Interval for Mean does not, however, show any notable discrepancies between the groups.

Table 3. The overall mean scores for tasks 1 and 2

Nationality	Communication	Vocabulary	Grammar	Fluency & pronunciation
Estonian	4.37	4.18	4.02	4.39
Russian	4.46	4.32	4.14	4.43

As to the frequency of awarding minimum and maximum points, the Russian raters were more critical of the candidates' communication skills and vocabulary usage by giving the lowest mark for these, whereas the minimum score assigned by the Estonian raters for the same categories was 2 points. In case of fluency & pronunciation, the score distribution was the reverse: 1 point received from the Estonian raters and 2 points from the Russian ones (see table 4).

Table 4. Maximum and minimum points awarded for each aspect of the candidate's performance

Task	Nationality	Minimum	Maximum
Communication	Russian	1	5
	Estonian	2	5
Vocabulary	Russian	1	5
	Estonian	2	5
Grammar	Russian	1	5
	Estonian	1	5
Fluency & pronunciation	Russian	2	5
	Estonian	1	5

The score distribution between the rater groups was also analysed with respect to their work experience as a national examination assessor and the length of service (see table 5).

Table 5. Correlation between the length of service and the number of scores awarded for fluency & pronunciation (N – the total number of the raters being multiplied by the number of the videorecordings assessed)

Length of service	N	Mean	Std. deviation	Std. error	95% confidence interval for mean		Min	Max
					Lower bound	Upper bound		
1–5	80	4.23	.763	.085	4.06	4.39	2	5
6–10	40	4.55	.639	.101	4.35	4.75	3	5
11–15	40	4.55	.677	.107	4.33	4.77	2	5
>15	200	4.44	.741	.052	4.33	4.54	1	5

The analysis of the variance does not reveal any significant differences between the awarded scores and the rating experience. However, there is statistically relevant correlation (according to the 95% Confidence Interval for Mean) between the rater’s teaching experience and the score awarding for fluency & pronunciation. The mean score (4.23) assigned by less experienced raters with less than five years of service was the lowest in comparison to that of the more experienced teachers. The scores assigned by the most experienced raters with more than 15 years of service was higher (4.44) than the previous group’s but lower than that of the participants with 6-10 and 11-15 years of teaching (4.55). Less experienced raters (with 6-10 years of service) tended to give higher scores (3 points) for what the most experienced teachers (with more than 15 years of experience) afforded only the minimum amount of points (1 point), which resulted thus in the confusion, on the part of the former group, while distinguishing between a hesitant speaker and a very laconic speaker, (see REKK 2008) and unfair grading of such learners.

3.2. Questionnaire study results

The aim of the questionnaire study was to elicit data about the assessors’ opinions of the quality of the existing marking scale (i.e., whether it needed any further improvement), of the assessment of various aspects of a candidate’s oral performance (i.e., the effect of a candidate’s accent on the rating, the importance of content, pronunciation, lexis, fluency and grammar use in a candidate’s output) as well as of their own rating behaviour (i.e., the existence of any distracting factors affecting the rating, interference of a rater’s cultural identity with the assessment, effect of a test taker’s self-confidence, willingness for cooperation with the interlocutor, display of interest in the topic, etc.). In addition, the participants were required to provide background information about themselves, as to the length of service as a teacher and national examination assessor, and whether they experienced any need for further improvement of their rating skills.

3.2.1. The raters' impressions of the marking scale

When asked about the raters' perceptions of the existing marking scale for speaking, a notable variance between the two rater groups is exposed in their need for further improvements. There are statistically relevant differences between the groups: 89% of the Estonian raters and only 11% of the Russian ones were in favour of further improvements (see table 6).

Table 6. The necessity to improve the rating scale for speaking

Nationality	Rating scale needs improvement		Total
	0	1	
Russian	88.9%	11.1%	100.0%
Estonian	11.1%	88.9%	100.0%
Total	50.0%	50.0%	100.0%

Unlike their Russian colleagues, the majority of the Estonian raters criticized the existing marking scale for lack of proper criteria in assessing speaking and of appropriate levels for each category. The most controversial aspects in the rating scale referred to the ambiguity in the terms like occasional mistakes, an independent speaker versus a good speaker, and communication versus fluency (see REKK 2008). In other words, the raters felt that some assessment criteria seemed to overlap. The participants also complained of the difficulties while applying the same criteria to assess both the monologue and role playing.

Finally, both groups were also highly critical of the existing marking scale – 67% of all the participants doubted its validity.

3.2.2. Factors affecting the raters' behaviour

The research reveals that even after having passed standardised assessor training 78% of the interviewed raters admitted the presence of a number of factors disturbing the assessment process, in terms of either tasks or procedural matters.

The first objection raised concerns an obligation to record oral proficiency interviews. According to rater R6, this distracts not only a candidate but also an assessor/interviewer who has to worry about the quality of the recording.

In connection with the tasks, the raters opted for more variability in the question and answer section, as sometimes student cue cards contain prompts that are difficult to match with the answers in the interviewer script, this resulting in awkward interaction between the candidate and the interviewer, which is difficult to assess.

The participants were further asked whether their cultural identity may interfere with the assessment process. 94% of all the respondents denied this.

Based on the research by Winke et al. (2012), the raters were also inquired of a possible effect of a candidate's accent on the rating. Strangely enough, neither the Russian (89%) nor the Estonian (67%) rater group had ever experienced such an influence, which may be accounted for by the respondents' insufficient experience with various learner accents. However, both rater groups claimed to lower their

marks in case of serious miscomprehension occasioned by a test taker's strong accent.

Next, it was hypothesized in the study that both rater groups may be affected by factors outside the marking scale, such as their preconceived impressions of what constitutes good speaking proficiency.

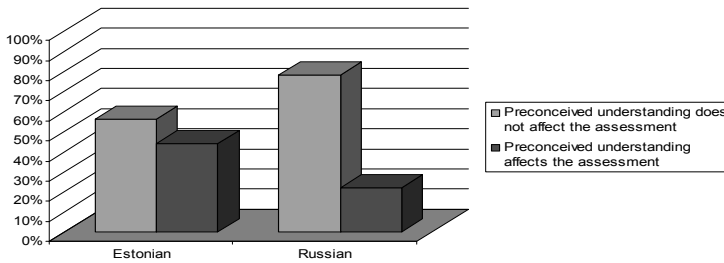


Figure 1. The influence of preconceived understanding of language proficiency levels on the candidates' assessment (x-axis: nationality, y-axis: percentage of raters)

Figure 1 demonstrates slight balance in the opinions of the Estonian rater group where 56% denied and 44% admitted of their rating behaviour being affected by some preconceived understanding of different speaking proficiency levels. The majority of the Russian group denied (78%) such an influence, as only 22% admitted its effect. Some raters were concerned in their notes about their tendency to compare candidates against each other, rather than checking a candidate's skills against the marking scale criteria.

3.2.3. Features of the candidate's language and the rater's behaviour

This part of the questionnaire contained closed-ended questions as to which features of the test takers' ability (content, pronunciation, lexis, fluency, grammar) may affect the scoring process.

Both rater groups (66% of the Estonians and 44% of the Russians) regarded the content to be the most essential aspect in assessing students' oral performances. Next, a wide range of vocabulary was appreciated by 56% of the Estonian raters and 33% of the Russian ones. Strangely enough, the presence of a candidate's proper pronunciation skills was ignored by both group. As to the importance of fluency and grammar, the variance analysis ANOVA demonstrates statistically significant differences between the two groups (see table 7). The mean degrees of the importance of fluency and grammar are 4.000 and 2.56 respectively within the Russian group, thus grammar being emphasized over fluency by them. This means that if a candidate is not able to maintain the flow of speech but nevertheless uses correct grammar structures, this will have no effect on the rater's judgements. In the Estonian group, the figures are 2.833 for fluency and 3.78 for grammar, with fluency being stressed over grammar (see table 7). Furthermore, none of the respondents within this group considered fluency to be the least essential aspect in assessing the speaking skill, whereas 33% thought of grammar as the least important in a candidate's output. Thus the rating of fluency and grammar knowledge in oral proficiency tests is affected by the assessors' cultural background.

Table 7. Comparison of the relevance of fluency and grammar usage in speaking between the two rater groups

Feature	Nationality	N	Mean	Std. deviation	Std. error	95% confidence interval for mean		Min	Max
						Lower bound	Upper bound		
Importance of fluency	Russian	9	4.000	1.000	.333	3.231	4.769	2	5
	Estonian	9	2.833	1.2748	.4249	1.853	3.813	1	4
Importance of grammar use	Russian	9	2.560	1.014	.338	1.780	3.330	1	4
	Estonian	9	3.780	1.302	.434	2.780	4.780	1	5

The raters were further inquired about the most distracting factors in a candidate’s speech, such as mispronunciation, hesitation, self-repetition, limited vocabulary choice and grammatical inaccuracy. Contrary to the similarities in the attitude of both groups towards mispronunciation, hesitation and repetition, statistically relevant differences between the groups emerge only with respect to a candidate’s vocabulary choice. In contrast to the Russian raters (22%), the Estonian ones (66%) are highly distracted by a candidate’s limited wordstock (see table 8). Obviously those candidates whose vocabulary usage is restricted will not obtain high scores from the latter group, notwithstanding a high level of proficiency in the other aspects of their speech (e.g., pronunciation, grammatical accuracy).

Although the data about the last category of the distracting factors, i.e., ungrammatical forms, show no statistically relevant differences between the two cultural groups, the Russian participants admitted this distraction in 56% of all cases. Conversely, the Estonian raters with the same response were represented by 78%. Such a high percentage of the raters admitting the distraction of grammatical inaccuracy seems inconsistent since most of the raters had claimed earlier of the prevalence of content over grammar in assessing the speaking skills.

Table 8. The most distracting factor when rating speaking – vocabulary use

Nationality	Distracting vocabulary use		Total
	0	1	
Russian	77.8%	22.2%	100.0%
Estonian	33.3%	66.7%	100.0%
Total	55.6%	44.4%	100.0%

Finally, the study aimed at analysing whether the raters’ impressions may be affected by the following aspects in a candidate’s performance, such as the candidate’s self-confidence, display of interest in the topic, willingness to interact with the interlocutor, elaboration on one’s ideas, and personal response. There were no statistically relevant differences between the groups, apart from the candidate’s willingness to interact with the interlocutor. Contrary to the Russian raters (56%), all the Estonians (100%) regarded the test taker’s lack of communication strategies as detrimental to his or her positive assessment (see table 9). This may refer to Russian assessors’ willingness to overrate test takers in spite of a lower level of

their communication skills. The statistical analysis confirms the significance of this discrepancy between the two cultural groups.

Table 9. The effect of willingness for interaction on the perception of communication and fluency

Nationality	The effect of willingness for interaction on the perception of communication and fluency		Total
	0	1	
Russian	44.4%	55.6%	100.0%
Estonian	0%	100.0%	100.0%
Total	22.2%	77.8%	100.0%

4. Conclusion

The present research is a first attempt to compare the behaviour of the two cultural groups of non-native raters within the context of rating oral proficiency interviews of the national examination in English. The analysis of the assessment of 20 video recordings by 18 Estonian and Russian raters as well as the questionnaire study confirmed the hypothesis to some extent.

In spite of the absence of statistically significant differences in the scores awarded by both groups, the Russian raters tended to assign the highest as well as the lowest points for the four criteria (communication, vocabulary, grammar, and fluency & pronunciation) of the candidates' performance more frequently than their Estonian colleagues. The research revealed significant correlation between the raters' teaching experience and their severity of rating: the assessors with less than 5 years of experience were the strictest in assessing a candidate's fluency & pronunciation while the raters with 6-10 and 11-15 years of service were the most lenient, awarding the highest points for the same criteria.

The data obtained from the questionnaire study refers to the following statistically relevant differences in the perceptions and behaviour of the two rater groups: a) attitude towards the existing rating scale, since the Estonian group was in favour of some improvements to make the rating process more reliable, whereas the Russian one exposed no need for that; b) perception of the degree of relevance of fluency or grammar in the assessment, as the Russians emphasised grammar knowledge over fluency, and the Estonians attached more importance to fluency rather than to grammatical accuracy; c) perception of the most distracting factors in rating oral proficiency interviews, since most of the Estonians (66%) were disturbed by the candidates' limited vocabulary choice, whereas only 22% of the Russians found this distracting; d) attitude towards the candidate's interaction skills where all the Estonians (100%) admitted that their perception of a test taker's fluency and communication skills is affected by his or her ease of interaction with the interlocutor, and the proportion of the Russian raters sharing the same view comprised only 56%.

In view of a number of statistically insignificant differences in the raters' behaviour that emerged in the course of the research, it is difficult to claim whether the elicited differences between the two rater groups may be attributed to their cultural background, or whether they are the result of a limited number of the participants.

Nevertheless, the findings of the study provide valuable insight into rater behaviour which is imperative for developing valid testing tools, training raters and generating sound test results.

References

- Abedi, Jamal 2011. Assessing English language learners: Critical issues. – Maria del Rosario Basterra, Elise Trumbull, Guillermo Solano-Flores (Eds.). *Cultural Validity in Assessment: Addressing Linguistic and Cultural Diversity*. New York: Routledge, 49–71.
- Alas, Ene 2010. *The English Language National Examination Validity Defined by Its Oral Proficiency Interview Interlocutor Behaviour*. Tallinn University Dissertations on Humanities 22. Tallinn: Tallinn University Press.
- Bachman, Lyle F.; Lynch, Brian K.; Mason, Maureen 1995. Investigating variability in tasks and rater judgements in a performance test of foreign language speaking. – *Language Testing*, 12 (2), 238–257. <http://dx.doi.org/10.1177/026553229501200206>
- Brown, Anne 1995. The effect of rater variables in the development of an occupation-specific language performance test. – *Language Testing*, 12 (1), 1–15. <http://dx.doi.org/10.1177/026553229501200101>
- FitzGerald, Helen Gay 2003. *How Different Are We? Spoken Discourse in Intercultural Communication: The Significance of the Situational Context*. *Languages for Intercultural Communication and Education* 4. Clevedon: Multilingual Matters.
- Fulcher, Glenn 2003. *Testing Second Language Speaking*. London: Longman/Pearson Education.
- Hofstede, Geert 1997. *Culture and Organisations: Software of the Mind*. New York: McGraw-Hill.
- Kriisa, Kristel 2012. 2012. aasta inglise keele riigieksami lühianalüüs. [Analysis of the National Examination in the English Language of 2012.] <http://ekk.edu.ee/valdkonnad/uldharidusvalishindamine/riigieksamite-materjalid-2012/inglise-keel> (10.3.2013).
- Lumley, Tom; McNamara, Tim F. 1995. Rater characteristics and rater bias: Implications for training. – *Language Testing*, 12 (1), 54–71. <http://dx.doi.org/10.1177/026553229501200104>
- Luoma, Sari 2004. *Assessing Speaking*. Cambridge: Cambridge University Press. <http://dx.doi.org/10.1017/CBO9780511733017>
- McNamara, Tim F. 2000. *Language Testing*. Oxford: Oxford University Press.
- Pajupuu, Hille 2001. *Kuidas kohaneda võõras kultuuris: käsiraamat*. [How to Adapt Oneself to a Foreign Culture: A Handbook.] Tallinn: Eesti Keele Sihtasutus.
- Pajupuu, Hille 2007. *Kuidas hinnata suure panusega testide hindajaid*. [How to assess the raters of high-stakes tests.] – *Eesti Rakenduslingvistika Ühingu aastaraamat*. *Estonian Papers in Applied Linguistics*, 3, 221–233. <http://dx.doi.org/10.5128/ERYa3.15>
- Pollitt, Alastair; Murray, Neil L. 1996. What raters really pay attention to. – *Performance Testing, Cognition and Assessment*. *Selected Papers from the 15th Language Testing Research Colloquium*, Cambridge and Arnhem. Cambridge: Cambridge University Press, 74–91.
- REKK 2008 = National Examinations and Qualification Centre of Estonia 2008. *Marking scale for speaking*. www.ekk.edu.ee/valdkonnad/uldharidusvalishindamine/riigieksamite-materjalid-2008/inglise-keel (26.12.2012).
- Tannen, Deborah 1984. *Conversational Style: Analysing Talk Among Friends*. Norwood, NJ: Albex.
- Trompenaars, Fons 1994. *Riding the Waves of Culture: Understanding Cultural Diversity in Business*. London: Nicholas Brealey.

- Wigglesworth, Gillian 1993. Exploring bias analysis as a tool for improving rater consistency in assessing oral interaction. – *Language Testing*, 10 (3), 305–319. <http://dx.doi.org/10.1177/026553229301000306>
- Winke, Paula; Gass, Susan; Myford, Carol 2012. Raters' L2 background as a potential source of bias in rating oral performance. – *Language Testing*, 30 (2), 231–252. <http://dx.doi.org/10.1177/0265532212456968>

Irina Stassenko (Tallinn Polytechnic School) is a graduate of Tallinn University (2013) who has also been trained as a national examination interviewer.
Pärnu Rd 57, 10135 Tallinn, Estonia
irina.stsnk@gmail.com

Liljana Skopinskaja (Tallinn University), her research interests include intercultural communication, foreign language teaching methodology and language teacher education.
Narva Rd 25, 10120 Tallinn, Estonia
liljana.skopinskaja@tlu.ee

Suliko Liiv (Tallinn University), her research interests are contrastive linguistics, language policy, intercultural communication and teacher education.
Narva Rd 25, 10120 Tallinn, Estonia
liiv@tlu.ee

KULTUURILISTE ERINEVUSTE UURING SUULISE KEELEPÄDEVUSTESTI HINDAJA KÄITUMISES

Irina Stassenko, Liljana Skopinskaja, Suliko Liiv

Tallinna Ülikool

Artiklis analüüsitakse kultuurilisi erinevusi keelepädevustesti hindaja käitumises inglise keele riigieksami läbiviimisel. Autoreid huvitab, kas eesti- ja venekeelsete hindajate käitumises esineb statistiliselt olulisi erinevusi, mis võiksid riigieksami tulemuste usaldusväärsust olulisel määral mõjutada. Uurimuse aluseks on kahekümne inglise keele riigieksami videosalvestuse hindamine, mille viisid läbi kaheksa eesti- ja kaheksa venekeelset hindajat, ja sellele järgnev ankeetküsitlus samade hindajate seas. Tulemused näitavad küll kehtestatud hindamisreeglite üldist järgimist mõlemas uurimisrühmas, kuid samas ilmnevad statistiliselt olulised erinevused nii hindajate tegevuses kui ka nende poolt antud hinnangutes kõneoskuse hindamiskaala eri aspektide kohta.

Võtmesõnad: hindamine, hindaja variatiivsus, hindaja usaldusväärsus, kultuuriline valiidsus, suuline keelepädevustest

ILUKIRJANDUSKEELEST UUE PILGUGA

Jekaterina Trainis, Kais Allkivi

Ülevaade. Artiklis kirjeldatakse 20. sajandi teise poole eesti ilukirjanduskeele kasutusmustreid, mis näitavad, kuidas emakeelekõneleja eelistab oma mõtteid kirjalikult väljendada. Uurimisaines on saadud Tartu Ülikooli pindsüntaktiliselt märgendatud ja ühtlustatud ilukirjanduskorpusest. Rakendatud on klasteranalüüsi, mis toob tekstivalimis sageduse alusel esile ühesuguseid morfo- ja süntaksimärgendite lineaarseid järgnevusi. Valitud on trigrammid, mis esinesid korpuses kaks ja enam korda.

Keelekasutuse üldised mustrid on välja toodud trigrammide esikomponendi sõnaliigilise kuuluvuse järgi. Nii tulevad esile morfoloogilised klassid ja alamklassid, mis moodustuvad ühise algus- ja lõpukomponendiga sõnaliigijärgenditest. Iga alamklass hõlmab morfoloogilisi klastreid, kuhu trigrammid koonduvad keskmise komponendi sõnaliigi varieerumise alusel. Morfosüntaktilise trigrammi moodustavad sõnajärgendid, mille komponendid kattuvad nii sõnaliigi ja vormi kui ka süntaktilise funktsiooni poolest. Morfoloogiliste klasside, alamklasside, klastrite ja morfosüntaktiliste trigrammide hierarhia toob esile keelekasutuse üldised tendentsid ja morfosüntaktilise ning leksikaalse varieerumise piirid.

Võtmesõnad: ilukirjanduskeel, klasteranalüüs, morfosüntaktiline varieeruvus, leksikaalne varieeruvus, eesti keel

1. Meetod ja teoreetilised lähted

Uurimuse eesmärk on kirjeldada eesti ilukirjanduskeele morfoloogilist, morfosüntaktilist ning leksikaalset varieerumist, välja tuua regulaarselt/ebaregulaarselt ilmnevad grammatilised ja leksikaalsed valikud. Ühtlasi on eesmärk näidata lingvistilise klasteranalüüsi kasutuspotentsiaali keelekirjelduses.

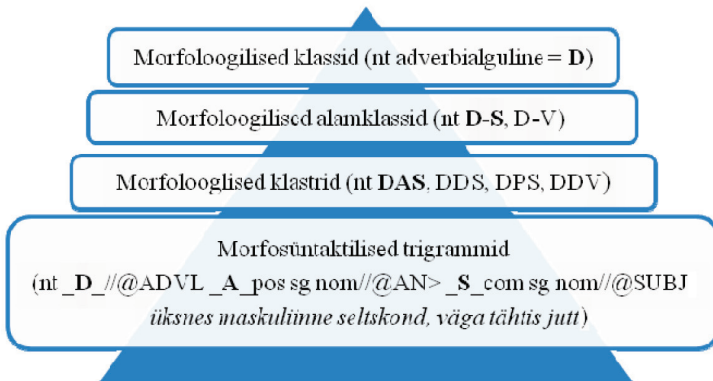
Klasteranalüüs on meetodina objektiivne, kui uurija paneb eelnevalt paika kindlad klasterdamisreeglid: erinevate klasterdamismeetoditega saadud klassifikatsioonid võivad erineda (Remm jt 2012: 75–76). Käesolevas uurimuses on

rakendatud S. Otsa programmeeritud klastrileidjat, mille sisendiks on eesti keele parseri väljund (Ots 2012). Klastrileidja otsib tekstist sarnaseid morfo- ja süntaksimärgendite lineaarseid järjendeid ning fikseerib nende sageduse. Kasutatud on nn libisevat tükeldamist: programm otsib järjendeid põhimõttel sõna1 sõna2 sõna3; sõna2 sõna3 sõna4 jne. Klastrileidja väljundit on töödeldud Excelis.

Varem ei ole eesti ilukirjanduskeelt sellise meetodi abil analüüsitud, ent sama meetodit on kasutatud eesti ilukirjandus- ja õppijakeele morfoloogiliste ja morfosüntaktiliste mustrite võrdlemisel (vt Esilon 2013 ja Esilon 2014 käesolevas kogumikus).

Uurimuse tarvis on klastrileidja otsing piiratud kolmest komponendist koosnevate üksuste ehk trigrammidega. Excelis reastatakse tekstikasutuses kaks ja enam korda esinenud trigrammid tähestikuliselt esikomponendi sõnaliigilise kuuluvuse alusel. Nii moodustuvad morfoloogilised klassid (nt adverbialguline ehk D). Alamklasse moodustavad trigrammid, mis tulevad esile mustritena ühesuguse algus- ja lõpukomponendi alusel, keskmise komponendi sõnaliik varieerub (nt adverb-substantiiv ehk D-S). Morfoloogilised klastrid eristuvad kõigi kolme komponendi morfoloogilise varieerumise alusel alamklassis (nt adverb-adjektiiv-substantiiv ehk DAS). Iga klaster sisaldab morfosüntaktiliselt ja leksikaalselt varieeruvaid trigramme, mis põhinevad samasugustel morfo- ja süntaksimärgendite järgnevustel (nt *_D_//@ADVL _A_pos sg nom//@AN> _S_com sg nom//@SUBJ* ehk adverb adverbialina + adjektiivne eestäiend ainsuse nominatiivis + substantiiv ainsuse nominatiivis subjektina: *pisut piinlik vaikus*).

Morfosüntaktilised trigrammid ja morfoloogilised klastrid, alamklassid ning klassid moodustavad hierarhilise liigituse (vt joonis 1): iga klastri sagedaim trigramm on selle klastri invariant; iga alamklassi sagedaim klaster on selle alamklassi invariant; iga klassi sagedaim alamklass on kindla sõnaliigiga algava klassi invariant. Kõige enam esinenud klass on keelekasutuse invariant.



Joonis 1. Lingvistilise klasteranalüüsi hierarhia

Siinses uurimuses on klasterdamist mõistetud kui lingvistiliste objektide rühmitamist morfo- ja süntaksimärgendite regulaarse koosinemise ja varieerumise põhjal. Morfosüntaktiliste kirjelduste alusel pannakse samm-sammult omavahel kokku lingvistiliselt sarnased objektid. Liitmisprotsess lõpeb, kui kõik ühiste tunnustega objektid on ühes klastris. Analüüsi tulemusel ilmnevad tekstikasutuse lingvistilised

mustrid, mille alusel saab kirjeldada grammatika ja leksika omavahelisi seoseid ning keelendite funktsioone.

Uurimuse keeleaines pärineb Tartu Ülikooli eesti kirjakeele süntaktiliselt analüüsitud ja ühestatud ilukirjanduskorpusest. Uurimuse tarvis on sellest leitud kõik trigrammid, mida valimis on esinenud kaks ja enam korda, k.a (osa)lause piiril.

Kokku oli ilukirjanduskorpuses 49 211 morfosüntaktilist trigrammi, sellest kaks ja enam korda kasutatuid 14 316, mis teeb trigrammide üldmahuks 147 633 sõnet (trigrammide arv $49\,211 \times$ pikkus 3 sõnet). Kaks ja enam korda kasutatud trigrammide üldmaht on 42 948 sõnet ($14\,316 \times 3$). Suhe kaks ja enam korda esinevate trigrammide ($14\,316 \times 100\% : 49\,211 = 29,09\%$) ning ainukordsete ($34\,895 \times 100\% : 49\,211 = 70,9\%$) vahel on $29,09\% : 70,9\%$. Seega kehtib 1929. aastal kirja pandud Zipfi needus (vt ka Tuldava 1977, Tullo, Hurford 2003), mis näitab, et keeles esineb suhteliselt väike rühm väga sageli kasutatud konstruktsioone ning palju keskmise ja väikese sagedusega konstruktsioone (suhe $1/3$ ja $2/3$).

Optimaalse klastriarvu määramist peetakse klasteranalüüsi kitsaskohaks, kuna see etapp on valdavalt subjektiivne (Burns, Burns 2008: 557). Alamklasside kirjeldamisel on arvestatud optimaalse trigrammide hulgaga, mis leitud valemi $k \approx \sqrt{n} : 2$ järgi, kus n on objektide arv (siin: kaks ja enam korda esinud trigrammide arv), mille järgi optimaalne trigrammide hulk k on 85 (vt lisad, märgitud halli värviga). Trigrammide statistilist olulisust hinnates lähtutakse sellest arvnäitajast. Morfosüntaktiliste trigrammide tasandil analüüsitakse klasside kahe sagedasema alamklassi klastreid, ülejäänud alamklasse kirjeldatakse ülevaatlikumalt.

Kuna klasteranalüüsis ei ole ühte ainuõiget lahendit, rõhutavad mitmed autorid, et kvantitatiivse analüüsi tulemust peab tõlgendama teooria seisukohast (Burns, Burns 2008: 552, Remm jt 2012: 76, vt ka Niglas s. a.). Uurimuse aluseks on kasutuspõhine lähenemine, mis võrsunud konstruktivismi (Croft 2003) ja funktsionaalse grammatika (Halliday 1994) seostest.

M. Barlow ja S. Kemmer (2000) ning J. Bybee (2003) on selgitanud, et keelekasutusmudelid (ingl *models of the language use*) on sageduse põhjal esile toodud keeleliste teadmiste mudelid, nt võistlusmudel (*Competition Model*), leksikaalse tähendusvõrgustiku mudel (*Lexical Network Theory*) jm. P. Norvig (1989) on tutvustanud leksikaalse tähendusvõrgustiku mudelit, mis aitab toime tulla üksiksõnade ja sõnadevahelise tähendusega. A. Veismann (2008) on kaassõnade kirjeldamisprobleemide üle arutledes välja toonud radikaalse tähendusvõrgustiku mudeli, kus kesksed liikmed on sõna tähenduste prototüüpsemad esindajad.

Ka kasutuspõhine grammatikakirjeldus (ingl *usage-based grammar description*) on olemuselt standardiseeritud grammatikakirjeldusest erinev ja kasutuspõhisest keelekäsitlusest (*usage-based language description*) kitsam mõiste. Vastupidiselt generatiivsele lähenemisele, mis toetub seisukohale, et keelekasutaja memoreerib vaid ebareeglipäraseid vorme, püüab kasutuspõhine keelekirjeldus näidata, et esinemissageduse tõttu jätab keelekasutaja meelde ja rakendab valdavalt, enamasti või harva vaid teatud liiki konstruktsioone (Barddal 2003). Morfoloogilisi universaale uurivas loomuliku morfoloogia teoorias tuuakse välja, et teatud nähtused on nii vormi- kui ka sõnamoodustuses ootuspärasemad ja sagedasemad, st keelekasutaja silmis loomulikud kui teised (Habicht 2006: 13–15).

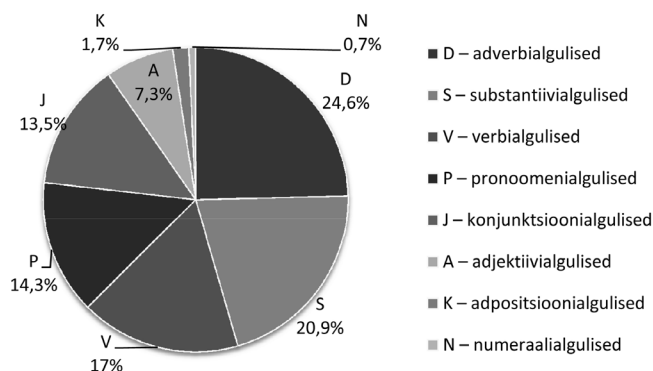
Esinemissageduse olulisust on keelekasutuse uurimisel rõhutanud ka J. Bybee, P. Hopper (2001:1–24) ja M. Haspelmath (2006, 2008). Sagedus dikteerib

keelekasutusmustrit. Seejuures on kõrge esinemissagedusega vormid vähem muutusealtid kui madalama sagedusega vormid, mis on omakorda sõnade ja nende koosinemise suhtelise autonoomia väljendus keeles (Bybee 2010: 50, 75).

Käesolev keelekasutusest tulenev empiiriline uurimus tugineb invariandi, esinemissageduse ja osakaalu mõistetele ning morfosüntaktilise ja leksikaalse varieeruvuse piiridele (vt ka Eslon 2013: 13–38).

2. Adverbiaalguline klass

Trigrammide jagunemine morfoloogilistesse klassidesse esikomponendi sõnaliigi alusel näitab, et ilukirjanduskeeles on suurim osakaal adverbiga algavatel trigrammidel: 24,6% valimi mahust (vt joonis 2, lisad 1–6). Suhteliselt sagedased on ka substantiivi- (20,9%), verbi- (17%), pronoomeni- (14,3%) ja konjunktsioonialguline (13,5%) klass, väiksema kasutussagedusega adjektiivialguline (7,3%). Harva esinenud adpositiooni- (1,7%) ja numeraalialgulist (0,7%) klassi artiklis ei kirjeldata. Järgnevalt antakse ülevaade morfoloogiliste klasside osakaalust valimis.



Joonis 2. Morfoloogiliste klasside osakaal valimis

Adverbiaalgulise klassi invariantne alamklass on adverb-substantiiv (D–S) *nii selgesti meeles* (vt lisa 1), järgnevad adverb-verb (D–V) *ennevanasti on olnud*, adverb-adverb (D–D) *seejärel silmitseb mõtlikult*. Alamklassid adverb-pronoomen (D–P) *veel kaua-kaua tema*, adverb-adjektiiv (D–A) *hirmus ilusad valged*, adverb-konjunktsioon (D–J) *nägupidi omad ja* ning adverb-adpositioon (D–K) *otse silmade ees* on juba oluliselt vähem esindatud. Alamklass adverb-numeraal (D–N) *möödas juba seitsesada* on teiste klassidega võrreldes marginaalne (vt tabel 1).

Tabel 1. Adverbiaalguliste alamklasside osakaal valimis

Alamklass	D–S	D–V	D–D	D–P	D–A	D–J	D–K	D–N
Osakaal klassis	29,4%	25,4%	19,1%	7,5%	7,5%	6,5%	4,4%	0,2%

2.1. Alamklass D–S

Adverbiga algava klassi sagedaim alamklass D–S koosneb üheksast klastrist: DAS (33,7%), DDS (16,5%), DPS (15,8%), DVS (13,9%), DSS (8,2%), DJS (4,5%), DNS (4,5%), DKS (2,6%) ja DGS (0,3%). Adverbidest korduvad alamklassis D–S enim koha-, seisundi-, rõhu- ja ajamäärsõnad. Klatri trigrammid näitavad, et üks komponent seob teisi, tõmmates oma poole magnetina vaid kindlaid elemente.

Alamklassi D–S invariantne klaster on **DAS**. Selles esineb koha- (*üles, seal, kõrvale*), viisi- ja seisundi- (*rahutult, kapriüsselt, kollakalt*), kvantiteedi- (*liiga, palju, üsna*) ja rõhumäärsõnade (*kahjuks, liiga, üha*) leksikaalset varieerumist. Klatri invariantne morfosüntaktiline trigramm on adverb adverbialina + adjektiivne eestäiend ainsuse nominatiivis + substantiiv ainsuse nominatiivis subjektina, nt *mõneti kummaline lugu, tõesti süütu kurat*.

Klatri **DDS** invariantne trigramm on kaks adverbilist adverbialina + substantiiv ainsuse partitiivis objektina, nt *liigagi täpselt aimu, ka mõnikord pahandust*. Adverbide valik ja kombineerumine on väga mitmekesine, tavaliselt esinevad järjestikku aja- ja rõhumäärsõna (*ju ometi juhtnööre, siis ootamatult kaela*).

Klatri **DPS** invariantne trigramm on adverb + possessiivpronoomeni ainsuse genitiivis + substantiiv ainsuse partitiivis objektina, nt *ainult oma väitekirja, siin igal õhtul*. Pronoomeni valik on liigiti piiratud: *oma* on kujunemas kõrge esinemissageduse tõttu stampsõnaks.

Klatri **DVS** invariantne trigramm on lausealguline adverb adverbialina + öeldisverb indikatiivis imperfekti ainsuse 3. pöördes + substantiiv ainsuse nominatiivis subjektina, nt *Siis helises uksekell, Vahel ulatus vesi*. Adverbidest korduvad lause alguses *Siis, Pärast, Ees*. Pea pooltel juhtudel on trigrammi moodustavaks verbivormiks *oli*, mis on klatri DVS kasutusele tüüpiline.

2.2. Alamklass D–V

Adverbiga algava klassi alamklass D–V koosneb seitsmest klastrist: DDV (39,8%), DVV (24,8%), DJV (12,6%), DSV (11,2%), DPV (7,9%), DAV (3,5%), DXV (0,2%). Alamklassi D–V adverb väljendab põhiliselt subjektiivset modaalsust, rõhutab öeldut või viitab öeldule.

Invariantse klatri **DDV** invariantne trigramm on kaks adverbi adverbialina + eitav abiverb, nt *nüüd küll ei, enam sugugi ei*. Adverbi on kasutatud subjektiivse modaalsuse väljendamise ja millegi rõhutamise eesmärgil (*tegelikult, tingimata*), tagasi viite vahendina lause alguses (*Küllap, Alles, Kui*), harva kohamäärusena.

Klatri **DVV** invariantne trigramm on adverb adverbialina + öeldisverb eitava kõne indikatiivi imperfektis, nt *veel ei liikunud, muidugi ei olnud*. Adverbe iseloomustab leksikaalne rikkus ja funktsioonide mitmekesisus, neid on kasutatud sageli osalause alguses sidendina (*siis ei osanud*), subjektiivse modaalsuse väljendamiseks (*nähtavasti ei pidanud*) või millegi rõhutamise eesmärgil (*tõepoolest ei alustanud*).

Klatri **DJV** invariantne trigramm on adverb + rinnastav sidesõna + transitiivne öeldisverb indikatiivis imperfekti ainsuse 3. pöördes, nt *üles ja küsis, pooleks ja virutas*. Adverbide leksikaalne varieeruvus on suur. Enamjaolt esineb adverb ühendverbide osana (<panen> *kinni ja jätan*, <andis> *tagasi ja läks*).

Klastrid **DSV** invariantne trigramm on adverb adverbialina + substantiiv ainsuse partitiivis objektina + partitsiip öeldisverbi osana, nt *ikka puhtust pidanud, ka märki tabanud*. Eitavas kõne puhul rõhutab adverb öeldut (*Isegi Kraaving ei*) või seob (*siis last ei*).

2.3. Alamklass D–D

Alamklass D–D koosneb kuuest klastrist: DDD (40%), DVD (18,9%), DSD (13,7%) ja DJD (13,4%), DPD (13,1%), DAD (0,9%). Alamklassi adverbivalik on rikkalik.

Invariantse klastrid **DDD** lekikaalne varieeruvus on suur: adverbid rõhutavad (*ometi natuke aegsamini*), toovad esile (*pealegi veel nii*) ja seovad (osa)lauseid (*Kas tõesti nii*). Kõige enam korduvad *kuidas, nüüd, mitte, võib-olla*. Klastrid **DVD** on eelistatud *olema*-verbi indikatiivis preesensi ainsuse 3. pöördes. Lausealgulistest adverbidest korduvad *Siin* ja *Nii*. Klastrid **DSD** täidab substantiiv adverbiala, objekti ja subjekti rolli. Klastrid **DJD** esineb rohkesti viisimäärust (*tõiselt ja ka, üli-viisakalt ja korralekutsuvalt*). Adverbide leksikaalne varieeruvus on suur. Klastrid **DPD** on eelistatud pronoomenit ainsuses.

2.4. Alamklassid D–P, D–A ja D–K

Alamklassi D–P moodustavad kuus klastrit: DVP (66,8%), DDP (20,6%), DPP (5%), DJP (3,3%), DSP (3,1%) ja DAP (1,2%). Pea kõikide klastrite trigrammid on tüüpiliselt (osa)lause alguses. Adverbil on enamasti sidususfunktsioon, teistest enam korduvad *Nüüd, samas, siis*.

Alamklass D–A koosneb kaheksast klastrist: domineerib invariantne klaster DDA (58,1%), järgnevad DPA (16,5%), DVA (8,9%), DAA (8,4%), DSA (4,2%), DJA (2,3%), DKA (0,8%), DNA (0,8%). Kõikide alamklasside adverbivalik on rikkalik.

Alamklassis D–K on neli klastrit: DSK (82,4%), DPK (14,4%), DDK (2%) ja DNK (1,3%). Alamklassi iseloomustab samuti adverbide rikkalik leksikaalne varieerumine.

3. Substantiivialguline klass

Osakaalult teine morfoloogiline klass on substantiivialguline, osakaal 20,9% (vt joonis 2, lisa 2). Substantiivialgulised trigrammid liigituvad kaheksasse alamklassi (vt tabel 2). Tüüpilisim on substantiiv-adverb (S–D) *meeste puhul ühtmoodi*, osakaal valimis 33,3%. Järgnevad substantiiv–substantiiv (S–S) *linna karget värskest* ja substantiiv-verb (S–V) *pilgu mujale keeranud*. Oluliselt vähem on esindatud klassid substantiiv-konjunktsioon (S–J) *töö pooleli ja*, substantiiv-adpositsioon (S–K) *ema selja taha*, substantiiv-adjektiiv (S–A) *süli oli pehme*, substantiiv-pronoomen (S–P) *sisetunne ütles mulle*, ja substantiiv-numeraal (S–N) *ema tütar esimesest*.

Alamklasside keskmine komponendi varieerumine on üsna avar. Nii jagunevad S–D ja S–J kuueks ning S–V ja S–A seitsmeks, alamklass S–S kaheksaks klastriks.

Tabel 2. Substantiivialguliste alamklasside osakaal valimis

Alamklass	S-D	S-S	S-V	S-J	S-K	S-A	S-P	S-N
Osakaal klassis	33,3%	24,0%	22,5%	7,1%	5,5%	5,2%	2,4%	0,1%

3.1. Alamklass S-D

Alamklass S-D sisaldab kuut morfoloogilist klastrit: SVD (58,2%), SDD (18,1%), SKD (11%), SSD (7,3%), SJD (3,9%) ja SPD (1,6%).

SVD on dominantne klaster; selle invariantne trigramm koosneb substantiivist ainsuse nominatiivis subjektina + öeldisverbist indikatiivis imperfekti ainsuse 3. pöördes + adverbist adverbialina, nt *helk kumas otsekui, töö edenes vaevaliselt*. Trigrammi leksikaalne varieerumine on avar substantiivil, veidi piiratum adverbil ja kõige piiratum verbil, kus ligi pool näidetest sisaldab *olema*-verbi, nt *isa oli seal, kool oli tookord, meel oli väga*. Korduvalt esinevad ka tuumverbid *hakkama, tulema* ja *liikuma*, mida on samuti eelistatud kasutada imperfektis ja ainsuse 3. pöördes. Adverbidest on enim viisi- ja seisundiadverbe (*häirivalt, kiuslikult, külmakartlikult, mänglevalt, ropult, rõõmsalt*), tuleb ette temporaal- (*ikka, täna, tookord, siis, varsti, möödas*), lokaal- (*kohal, kõrvale, peale, seal, siia, siinsamas*), rõhu- (*ka, küll, lihtsalt, nii*) ning hulgaadverbe (*väga, üleni*).

Klastris **SDD** invariantne trigramm on substantiiv ainsuse nominatiivis subjektina + kaks adverbilist adverbialina, nt *mees lõpuks välja, käsi nii enesestmõistetavalt*. Kõige sagedamini on kasutatud rõhuadverbi *ka* (*hing ka sisse, teadus siis ka*).

Klastris **SKD** invariantne trigramm on substantiiv ainsuse genitiivis laiendina + adpositsioon + adverbiline adverbial, nt *juhtunu üle järele, puu najal kummaliselt*. Adpositsioonidest esineb enim *pärast*, adverbidest korduvad samad sõnad, mis sagedased ka eespool kirjeldatud klastrite puhul (*välja, veel* ja *ka*). Korduvalt esineb adverbialijärjendeid *üle järele* (<*mõtles*> *juhtunu üle järele*) ning *alt välja* (<*paistis*> *teki alt välja*).

3.2. Alamklass S-S

Alamklassis S-S on kaheksa klastrit: SJS (35,4%), SAS (21,9%), SVS (17,8%), SKS (10,8%), SDS (6,9%), SPS (3,6%), SSS (3,0%) ja SNS (0,5%). Osakaalu poolest tõusevad esile klastrid, mille keskmine komponent on konjunktsioon, adjektiiv ja verb.

Invariantse klastris **SJS** invariantse trigrammis seob rinnastav sidesõna kahte subjekti, mõlemad ainsuse nominatiivis, nt *tüli ja pahandus, jutt või luuletus*. Konjunktsioonide varieeruvus on väga väike, ülekaalukalt sagedaim on *ja*, vaid viiel korral esines *või*.

Klastris **SAS** invariantne trigramm on substantiiv ainsuse genitiivis eestäiendina + adjektiiv ainsuse partitiivis eestäiendina + substantiiv ainsuse partitiivis objektina, nt *autori igapäevast elamist, naise südpehmet kõhtu*.

Klastris **SVS** valitseb suur vormivarieeruvus: klastris 135 sõnajärjendit liigituvad 55 morfosüntaktiliselt erineva trigrammi alla. Selget tüüpilist trigrammi esile ei kerki. Sagedaim trigramm on lausealguline substantiiv ainsuse nominatiivis

subjektina + transitiivne öeldisverb indikatiivis imperfekti ainsuse 3. pöördes + substantiiv ainsuse genitiivis objektina, nt *Jumal lõi maailma, Tüdruk surus nina*.

3.3. Alamklassid S–V ja S–K

Alamklassis S–V on seitse klastrit: SVV (29,7%), SDV (25,8%), SKV (16,9%), SSV (13,4%), SJV (13,1%), SAV (0,6%) ja SPV (0,6%). Alamklassi klastreid iseloomustab suur vormivarieeruvus: enamikku trigrammidest on kasutatud vaid kahel-kolmel korral.

Invariantses klastris **SVV** järgneb substantiivile tavaliselt eitusverb (*isa ei lähe, uhkus ei lubanud*), harvem verbi mineviku liitajavorm (*pilk oli otsinud, tähtedega oli kirjutatud*) või finiitverbi ja infinitiivi ühend (*naine võib sattuda, juuksed läikima pani*). Klaster **SDV** sisaldab enamasti ühendverbi, tavaliselt mineviku liitaja (<oli> *šokolaadi ära söönud*) või da-infinitiivi vormis (<võid> *ukse lahti teha*). Klaster **SKV** täidab adpositsioonifraas enamasti adverbiaali rolli, adpositsioonidest korduvad enim *ees, eest, juures, poole, taha* ja *otsa*. Klaster **SSV** esineb sageli lause alguses (*Mehe vaim oli, Riide võilillekiri oli, Vannitoa uks avanes*). Klaster **SJV** hõlmab osalausepiiril esinevaid sõnajärgendeid, mille keskmine komponent on eranditult rinnastav sidesõna, tavaliselt *ja* (*kausi ja tõstis, õhuluugi või paneb*).

Alamklassi S–K nelja klastriga seas on selge dominant **SSK**, mille osakaal on 86,6% (*poeg tooli najal, viulikast käe otsas*). Ülejäänud kolme klastriga osakaal pole kuigi märkimisväärne: SPK (8,7%), SVK (3,5%) ja SNK (1,2%).

4. Verbialguline klass

Verbiga algava klassi osakaal valimis on 24,6% (vt joonis 2, lisa 3). Invariantne alamklass on verb-adverb (V–D) *mõtleb Leopold nukralt*. Teisel kohal on verb-substantiiv (V–S) *kõlas katkendlik kellatirin*. Pea kolm korda harvem esineb alamklass verb-adjektiiv (V–A) *algab üks tüütu*. Ülejäänud alamklasside osakaal valimis pole kuigi märkimisväärne: verb-verb (V–V) *võib pihta hakata*, verb-konjunktsioon (V–J) *paistsid kaugele otsekui*, verb-pronoomen (V–P) *möödudes vahetasid nad*, verb-adpositsioon (V–K) *kadus vahemüüri taha* ja verb-numeraal (V–N) *astuda need kakssada* (vt tabel 3).

Tabel 3. Verbialgulistele alamklassidele osakaal valimis

Alamklass	V–D	V–S	V–A	V–V	V–J	V–P	V–K	V–N
Osakaal klassis	41,7%	26,3%	9,5%	8,2%	5,2%	4,6%	4,5%	0,2%

4.1. Alamklass V–D

Invariantne alamklass V–D sisaldab viit klastrit, neist neli olulist: VDD (36,6%), VSD (22,0%), VPD (21,1%), VVD (17,9%), VJD (2,4%). Verbivalik on küll rikkalik, kuid eelistatakse siiski *olema*-verbi ainsuse 1. ja 3. pöördes.

Invariantses klastris **VDD** varieeruvad verbi ainsuse ja mitmuse pöörded ning *nud*-partitsiip, ka leksikaalne varieeruvus on avar, kuid sagedaim verb on *olema* (*oli üsna kõrgelt, olin juba päris, on ju lõpuks, oleks nagu veel*). Klastris **VDD** invariantne trigramm on öeldisverb indikatiivis imperfekti ainsuse 3. pöördes + kaks adverbilist adverbiaali, nt *oli väga ammu, muutus ajapikku väga*.

Klastris **VSD** invariantne trigramm on transitiivne öeldisverb indikatiivis imperfekti ainsuse 3. pöördes + substantiiv ainsuse nominatiivis objektina + adverb adverbialina, nt *pööras pilgu ära, võttis buketi vastu, pani noa kõrvale*.

Klastris **VPD** invariantne trigramm on *olema*-verb indikatiivis imperfekti ainsuse 3. pöördes abitegusõnana + ainsuse 3. isiku pronoomen nominatiivis + adverb adverbialina, nt *oli ta salaja, oli ta siinkandis, oli ta vahepeal*. Verbivormidest domineerib ainsuse 3. pööre, nagu ka ainsuse 1. pööre (*sain ma liigagi, püüdsin ma küllalt*).

Klastris **VVD** invariantne trigramm on öeldisverb indikatiivi eitava kõne preesensis + adverb adverbialina, nt *ei kao kuhugi, ei suuda kauem, ei lähe sinna, ei ole võlgu, ei ole ainult*. Klastris trigrammides tuleb hästi esile verbide leksikaalne rikkus.

4.2. Alamklass V–S

Alamklass V–S koosneb seitsmest klastrist: VDS (27,2%), VAS (24,0%), VPS (17,9%), VSS (14,4%), VVS (10,4%), VJS (3,4%), VNS (2,7%). Verbide leksikaalne varieerumine on selles alamklassis avar, kuid siiski enim korduvad ainsuse 3. pöördes *olema, leidma ja tegema*.

Klastris **VDS** on verb üle poole juhtumitest imperfekti ainsuse 3. pöördes, kuid see pole seotud *olema*-verbiga. Samas klastris invariantne trigramm sisaldab koopulat (*olema* indikatiivi imperfekti ainsuse 3. pöördes) + adverb adverbialina + substantiiv ainsuse nominatiivis predikatiivina, nt *oli seal keskkoolidirektor, oli rohkem hingearst, oli teadagi kaunitar*.

Klastris **VAS** invariantne trigramm on öeldisverb indikatiivis imperfekti ainsuse 3. pöördes + adjektiiv ainsuse nominatiivis eestäiendina + substantiiv ainsuse nominatiivis subjektina, nt *tekkis ähmane lootuskiir, kasvas kõrge vili*. Klastris verbivalik on leksikaalselt piiratud: ligi poole näidetest moodustavad trigrammid *olema*-verbiga, varieerub vaid aja- ja pöördevorm. Seevastu adjektiivide leksikaalne varieeruvus on avar.

Klastris **VPS** korduvad verbid *olema, leidma ja tegema*. Substantiiv esineb nii subjekti (*näisid kõik mõtted*), objekti (*täita mingit ülesannet*), adverbialina (*tuleb meile külla*), predikatiivi (*on minu nimi*) kui ka tagasõna laiendina (*istus ka klaveri <taha>*). Invariantne trigramm on öeldisverb indikatiivis preesensi ainsuse 3. pöördes + demonstratiivpronoomen ainsuse nominatiivis eestäiendina + substantiiv ainsuse nominatiivis subjektina, nt *tuleb see nüanss, on too paratamatus*.

Klastris **VSS** esineb suur verbi ja substantiivi leksikaalne varieeruvus. Verbidest on korduvalt kasutatud *olema* ainsuse ja mitmuse 3. pöördes. Invariantne trigramm on koopula (*olema* indikatiivi preesensi ainsuse 3. pöördes) + substantiiv ainsuse genitiivis eestäiendina + substantiiv ainsuse nominatiivis predikatiivina, nt *on asjatundja pilk, on inimese tegu*.

4.3. Alamklassid V–A ja V–K

Alamklass V–A koosneb kuuest klastrist: VDA (65,0%), VPA (14,6%), VSA (9,3%), VAA (7,5%), VVA (2,7%), VKA (0,9%). Dominantne klaster **VDA** vastandub osakaalu poolest ülejäänutele. Alamklassile on üldiselt iseloomulik piiratud verbivalik: domineerivad *olema*-verbi erinevad vormid. Adverbidest on korratud lekseeme *vist, nii, pisut, õige ja väga*, mis väljendavad valdavalt kõneleja suhtumist või rõhutavad öeldut.

Alamklass V–K koosneb neljast klastrist: VSK (80,2%), VPK (14,2%), VDK (3,8%), VNK (1,9%). Selge dominant on **VSK**, ülejäänud klastrite osakaal on samuti väike, nagu alamklassis V–A. Erinevus viimasest seisneb verbi tunduvalt avaramas leksikaalses varieerumises, kõige enam korduvad *olema, tulema, ronima* ja *minema* erinevad vormid.

5. Pronoomenialguline klass

Pronoomeniga algava klassi osakaal on valimi üldmahust 14,3% (vt joonis 2, lisa 4). Invariantne alamklass on pronoomen-adverb (P–D) *ma arvan küll*, järgnevad pronoomen-verb (P–V) *ta lõpuks saavutanud*, pronoomen-substantiiv (P–S) *selle võika oleni*, pronoomen-adpositsioon (P–K) *miski nende vahele*, pronoomen-pronoomen (P–P) *ta seda endale*, pronoomen-adjektiiv (P–A) *enda vastas paljast*, pronoomen-konjunksioon (P–J) *selline näolapp ja* ning pronoomen-numeraal (P–N) *tal umbes üks* (vt tabel 4). Alamklasside P–D, P–V ja P–S osakaal on pronoomenialgulises klassis valdav, ülejäänud alamklasside osakaal ei ole kuigi oluline.

Tabel 4. Pronoomenialguliste alamklasside osakaal valimis

Alamklass	P–D	P–V	P–S	P–K	P–P	P–A	P–J	P–N
Osakaal klassis	39,6%	22,8%	20,4%	7,6%	3,7%	3,6%	2,2%	0,1%

5.1. Alamklass P–D

Alamklass P–D koosneb viiest klastrist: PDD (32,0%), PVD (30,0%), PSD (26,3%), PPD (6,9%) ja PKD (4,8%). Alamklassi pronoomenivalik on piiratud, domineerivad personaalpronoomen *ta* ja demonstratiivpronoomen *see*.

Klastris **PDD** invariantne trigramm on ainsuse 3. isiku personaalpronoomen subjektina + kaks adverbilist adverbiaali, nt *ta mõnikord ka, ta ükskord vahele*.

Klastris **PVD** invariantes trigrammis on lausealguline demonstratiivpronoomen + öeldisverb indikatiivis imperfekti ainsuse 3. pöördes + adverb adverbialina, nt *See oli nii See hakkas nüüd*.

Klastris **PSD** korduvad demonstratiivpronoomen *see* ja impersonaalsuse marker *üks*; mõlemat lekseemi on kasutatud erinevates käändevormides. Klastris invariantne trigramm on demonstratiivpronoomen ainsuse nominatiivis substantiivi eestäiendina + substantiiv ainsuse nominatiivis subjektina + adverb adverbialina, nt *see külamees liiga see asi siin*.

5.2. Alamklass P–V

Alamklass P–V sisaldab kuut klastrit: PVV (33,1%), PDV (31,4%), PSV (29,0%), PPV (3,9%), PKV (2,2%) ja PAV (0,4%). Kolme esimese klastris osakaal valimis on oluline, kolme viimase osakaal aga marginaalne. Alamklassis domineerivad demonstratiivpronoomeni ainsuse 1. ja 3. isiku lühivormid ning demonstratiivpronoomen *see*.

Klastris **PVV** keskmise komponendi leksikaalne varieerumine on piiratud, kasutatakse *olema*-verbi jaatava ja eitava kõne vorme, pronoomenitest eelistatakse ainsuse 1. ja 3. isiku lühivormi (*ma ei esitanud, ta ei kiirustanud*). Invariantne trigramm on lausealguline personaalpronoomen ainsuse 3. isikus + pluskvamperfekti liitaja vorm öeldisena, nt *Ta oli sõitnud Ta oli maganud*.

Klastris **PDV** domineerivad personaalpronoomeni 1. ja 3. isiku lühivormid. Invariantne trigramm on ainsuse 1. isiku pronoomen subjektina + adverb adverbialina + verbi eitus, nt *ma lihtsalt ei mina muidugi ei*.

Klastris **PSV** on sage demonstratiivpronoomen *see* (*see töö edenes*), personaalpronoomenite pikk vorm ainsuses ja mitmuses (*meie linnas sündis, minu naisel oli*) ning lühivorm ainsuses (*ta prillid läigatavad, mu grimassid vaheldusid*). Invariantne trigramm on lausealguline demonstratiivpronoomen *see* ainsuse nominatiivis substantiivi eestäiendina + substantiiv ainsuse nominatiivis subjektina + öeldisverb indikatiivis imperfekti ainsuse 3. pöördes, nt *See tunne oli, See mullatükk paindus*.

5.3. Alamklassid P–S ja P–K

Alamklass P–S koosneb seitsmest klastrist: PAS (53,5%), PDS (15,0%), PVS (10,7%), PSS (8,7%), PPS (7,8%), PNS (2,4%) ja PKS (1,9%). Alamklassi P–S pronoomenite valik ei ole nii piiratud, kui eespool kirjeldatud alamklassides. Domineeriva klastris **PAS** (pronoomen-adjektiiv-substantiiv) invariantes trigrammis on kasutatud vaid possessiivpronoomenit *oma* nt *oma vana sõpra oma professionaalset üleolekut*. Esile tuleb substantiivi ning sellega ühilduva adjektiivi käände varieerumine. Kõrvuti grammatiliste käänetega kasutatakse ka adessiivi (*tollel vihmasel detsembripäeval*), elatiivi (*nendest uutest apsudest*) ja inessiivi (*ta koduses kabinetis*). Käändevalik on eelnevas kontekstis kasutatud põhisõna reksioonimallist (nt *nägin keda? millal? teda tollel vihmasel detsembripäeval*).

Alamklassis P–K on kolm klastrit: dominantne **PSK** (94,1%) ning osakaalu poolest väheolulised **PPK** (3,3%) ja **PDK** (2,6%). Pronoomenitest korduvad

peamiselt personaal-, demonstratiiv- ja refleksiivpronoomenid, eriti *see, oma* ja *tema* grammatiliste käänete vormid.

6. Konjunktsioonialguline klass

Konjunktsioonialgulise klassi osakaal valimis on 13,5% (vt joonis 2, lisa 5). Klass koosneb seitsmest alamklassist. Tüüpiline alamklass on konjunktsioon-verb (J–V) *aga võib-olla oli*; enam-vähem võrdse osakaaluga esindatud alamklassid on konjunktsioon-adverb (J–D) *ning veel rohkem* ja konjunktsioon-substantiiv (J–S) *ja soojad kindad*; alamklasside konjunktsioon-pronoomen (J–P) *ja vaevalt ta*, konjunktsioon-adpositsioon (J–K) *kuid aja jooksul*, konjunktsioon-adjektiiv (J–A) *kui ta kõrvalseisva*, ning konjunktsioon-konjunktsioon (J–J) *ja selge nagu* osakaal pole märkimisväärne (vt tabel 5). Kõik konjunktsioonialgulised alamklassid on suure klastrivarieeruvusega (igas 5–7 klastrit).

Tabel 5. Konjunktsioonialguliste alamklasside osakaal valimis

Alamklass	J–V	J–D	J–S	J–P	J–K	J–A	J–J
Osakaal klassis	38,6%	24,9%	22,3%	6,9%	3,7%	3,1%	0,5%

6.1. Alamklass J–V

Suurima osakaaluga alamklass J–V kätkeb kuut klastrit: JPV (34,9%), JSV (33,2%), JDV (26,3%), JVV (4,7%), JJV (0,5%) ja JAV (0,3%).

Invariantnes klastris **JPV** sisaldub 84 erinevat morfosüntaktilist trigrammi, mida on kasutatud enam-vähem ühtlaselt. Seetõttu ei tule esile tüüpilist trigrammi, kuid sagedaim on alistav sidesõna + ainsuse 3. isiku personaalpronoomen nominatiivis subjektina + öeldisverb indikatiivis preesensis ainsuse 3. pöördes, nt *et ta seisab, kui tema hakkab*. Selles trigrammis on kasutatud valdavalt alistavat sidesõna *et*, vaid korra *kui* ja *sest*. Verbidest domineerib *olema* indikatiivi preesensi vormis. Personaalpronoomenite puhul on eelistatud ainsuse 3. isiku lühivormi (*ta*).

Vormilise mitmekesisusega paistab silma ka klaster **JSV**, mille all 71 erinevat trigrammi. Invariantne trigramm on alistav sidesõna + substantiiv ainsuse nominatiivis subjektina + öeldisverb indikatiivis preesensi ainsuse 3. pöördes, nt *et trepp on, kuni tee kaob*. Konjunktsioonidest eelistatakse alistavat sidesõna *et*, *olema*-verbi indikatiivi preesensis, harvem on samas vormis kasutatud verbe *hakkama*, *kaduma*, *muigama*, *painduma*, *pöörduma* ja *surisema*.

Klastris **JDV** sisaldub 47 morfosüntaktilist trigrammi. Invariantne on lausealguline rinnastav sidesõna + adverb adverbialina + öeldisverb indikatiivis imperfektis ainsuse 3. pöördes, nt *Aga siis juhtus, Ja nõnda tuli*. Konjunktsioonidest on ülekaalus *aga*, verbidest *olema*, adverbidest *siis*.

6.2. Alamklass J–D

Alamklass J–D jaguneb viieks klastriks: JVD (35,5%), JDD (29,8%) ja JPD (20,2%), JSD (12,8%) ja JJD (1,7%). Konjunktsioonidest on valdav rinnastav sidesõna *ja*.

Klastri **JVD** invariantne trigramm on rinnastav sidesõna + öeldisverb indikatiivis imperfekti ainsuse 3. pöördes + adverb adverbiaalina, nt *ja tuli uuesti, ja naeratas salapäraselt, vaid lamas kinnisilmi*. Trigramm eristub rikkaliku leksikaalse varieerumise poolest, eriti verbide osas. Nii on tuumverbide kõrval võrdselt esindatud täistähenduslikud verbid, nt *rebenes, moondus, räntsatas, paindus ja heldis*. Adverbidest korduvad sageli *siis, jälle, üles ja ometi*, ent samas esineb ka ilukirjanduskeelele iseloomulikke onomatopoeetilisi adverbe, nt *kärinal* ja *rabinal*.

Klastri **JDD** invariantne trigramm koosneb rinnastavast sidesõnast + kahest adverbilisest adverbiaalist, nt *ning omamoodi ka, ja isegi nagu, aga siit enam*. Esile tuleb kaks mustrit: 1) *ja* trigrammi alguses ja *nagu* lõpus, nt *ja isegi nagu, ja vist nagu, ja pisut nagu, ja kuskilt nagu* ning 2) rõhusõna *ka* trigrammi keskel või lõpus, nt *ja ka natuke, aga ka päris, ja muidugi ka, ning omamoodi ka*.

6.3. Alamklass J–S

Alamklass J–S on konjunktsioonialgulisel klassis kõige suurema klastrivarieeruvusega: JAS (34,8%), JVS (20,9%) ja JPS (19,3%), JSS (11,8%), JDS (8,5%), JJS (2,4%) ja JNS (2,4%). Invariantse klastri **JAS** invariantne trigramm algab rinnastava sidesõnaga (tavaliselt *ja*), järgneb adjektiiv eestäiendina ja substantiiv ainsuse nominatiivis, nt *ja valge karikakar, aga tüütu karvanäss*. Substantiivi leksikaalne varieeruvus on võrdlemisi suur.

7. Adjektiivialguline klass

Adjektiivialgulise klassi osakaal valimis on 7,3% (vt joonis 2, lisa 6). Selle klassi invariantne alamklass on adjektiiv-substantiiv (A–S) *märga läikivat nahkkuube*. Järgnevad adjektiiv-verb (A–V) *noored naised meeldivad* ja adjektiiv-adverb (A–D) *vanad trepiastmed rõõmsalt*. Ilukirjanduskeelele vähem iseloomulikud alamklassid on adjektiiv-adpositsioon (A–K) *suletud ukse ees*, adjektiiv-adjektiiv (A–A) *kauge ja salapärane*, adjektiiv-konjunktsioon (A–J) *hea sõber ning*, adjektiiv-pronoomen (A–P) *hilissuvisel päikesel enda* ja adjektiiv-numeraal (A–N) *hõõguvad peopesad tuhandetest* (vt tabel 6).

Kõige suurema osakaaluga alamklassid A–S, A–V ja A–D on samas ka suurima klastrivarieeruvusega, adjektiivi kasutatakse enamjaolt substantiivi eestäiendina. Alamklassid A–K, A–J ja A–N koosnevad vaid ühest klastrist, mille keskmine komponent on substantiiv.

Tabel 6. Adjektiivialguliste alamklasside osakaal valimis

Alamklass	A–S	A–V	A–D	A–K	A–A	A–J	A–P	A–N
Osakaal klassis	26,7%	18,7%	16,6%	14,5%	11%	10,7%	1,5%	0,2%

7.1. Alamklass A–S

Alamklass A–S on ilukirjanduskeelele tüüpiline. Keskmise komponendi varieerumise alusel eristub neli klastrit: AAS (54,3%), ASS (42,0%), AJS (2,9%) ja APS (0,7%).

Dominantse klastri **AAS** vormide varieerumine on avar (kokku 40 morfosüntaktilist trigrammi). Varieeruvad adjektiivivõrdlusaste ning substantiivi ning adjektiivivõrdlusaste arv- ja käändekategooria. Invariantne trigramm on kaks adjektiivset eestäiendit + substantiivainsuse (harvem mitmuse) nominatiivis subjektina, nt *mahe kurblik hääl, keskealine tüise naisterahvas; ilusad valged lilled, pikad rohekad juuksed*. Trigrammi substantiivide leksikaalne varieerumine on maksimaalne – enamus lekseeme on kasutatud ainukordselt. Adjektiivide hulgas esineb kordusi, nt *pikad, suur ja väike*.

Klaster **ASS** on veelgi suurema vormide varieerumisega: invariantne trigramm on adjektiivne ja substantiivne eestäiend ainsuse genitiivis + substantiivainsuse nominatiivis subjektina, nt *kauge automootori surin, valge pesuriide pind*.

7.2. Alamklass A–V

Alamklass A–V koosneb neljast klastrist: ASV (95,9%) on ülitugev dominant, üksiknäidetega on esindatud klastrid AVV (2,1%), AJV (1,0%) ja ADV (1,0%).

Klaster **ASV** sisaldab 65 morfosüntaktilist trigrammi, grammatiline varieerumine on avar. Vahelduvad verbi aja- ja pöördvormid, finiitsus-infiniitsus, tegevusnime- või kesksõnatunnus. Klastrisse kuuluvad ka eitava kõne trigrammid, nagu *kurnatud inimene ei*, kus eituspartiklile järgneb verbi finiitvorm või moodustub liitpredikaat. Adjektiividel varieerub käänne ja arv, võrdlusastmetest on valdav positiiv, vaid korra esines komparatiiv. Klastri ASV invariantne trigramm on adjektiivainsuse nominatiivis eestäiendina + substantiivainsuse nominatiivis subjektina + öeldisverb indikatiivis imperfekti 3. pöördes, nt *täielik hullumeelsus oli, paha lugu juhtus, vintske vanamees suri. olema-verbi kasutus on tavaline*.

7.3. Alamklassid A–D, A–K, A–J ja A–A

Alamklassi A–D dominantne klaster on **ASD**, osakaal valimis 82,6%. Ülejäänud klastrite osakaal jääb selgelt alla (AJD 14,0%) või on marginaalne (ADD 2,3%) ja AVV (1,2%). Klastris **ASD** täiendab adjektiiv enamasti objekti või adverbiaali rollis olevat substantiivi, harvem on substantiiv subjekti funktsioonis. Iseäranis siis, kui substantiivi on kasutatud objektina, kuulub adverb ühendverbi koosseisu (<tegi> *kohustusliku töö ära*). Ehkki võiks oodata, et adjektiivide leksikaalne varieeruvus on suurem kui adverbidel, leidub korduvaid lekseeme umbes sama palju, nt sagedaimat adjektiivivõrdlusaste (*suur*) ja adverbivõrdlusaste (*juba*) on kasutatud võrdses mahus.

Alamklassis A–K on üks klaster **ASK**, mis hõlmab 23 erisugust trigrammi, nt *kaharate okste all*. Adpositsioonidest kipuvad korduma eest, juures, alla, käes ja pärast.

Ka alamklass A–J koosneb ühest klastrist **ASJ**. Stereotüüpelt on kasutatud rinnastavat sidesõna *ja*, nt *suur südamevalu ja*.

Alamklassis adjektiiv-adjektiiv (A–A) on kolm klastrit: dominantne AJA (83,3%), vähese osakaaluga ASA (14,9%) ja marginaalne AAA (1,8%). Klasteris **AJA** kasutatakse alati rinnastavat sidesõna *ja*, nt *tühi ja vaikne*.

8. Kokkuvõte

Uurimuse põhjal saab väita, et eesti ilukirjanduskeeles esineb kõige sagedamini adverbi-, substantiivi- ja verbialgulisi trigramme (sagedaimad sõnaliigijärjendid on SVD, VDD, DDV ja DAS). Ajavormidest eelistatakse indikatiivi imperfekti, preesens esineb harvem; ainsus on sagedasem kui mitmus; 3. isik domineerib *mina*-vormi üle; sagedasemad pronomeniid on *see* ja *ta*. Rinnastuse puhul eelistatakse sidesõna *ja*, alistuse puhul *et*, pronomeni paarides *tema-ta* ja *mina-ma* lühivormi. Leksikaalselt rikkalikult esindatud sõnaliigid on substantiiv, adjektiiv ja adverb. Verbidest domineerib *olema*, väga sageli kasutatakse ka *hakkama*-verbi, harva onomatopoeetilisi verbe (*kõssitama*, *potsatama*, *räntsatama*) ja kestust tähistavaid verbe (*sugenema*, *valendama*).

Ilukirjanduskeele lingvistiline klasteranalüüs tõi esile seose vormide ja lekseemide varieerumise vahel: mida suurem on vormivarieeruvus, seda väiksem kipub olema leksikaalne, ja vastupidi. Nt alamklassis verb-adpositsioon on verbi leksikaalne varieeruvus väga suur, kuid vormivalik piiratud: VSK (*liikus ukse poole*) on nelja klasteri seas dominant.

Hea näide suurest vormi- ja piiratud leksikaalsest varieerumisest on pronomeni- ja konjunktsioonialgulised trigrammid. Esimene hõlmab kaheksat alamklassi, teine seitset; mõlemal on suur klasterivarieeruvus, kuid pronomeni ja konjunktsiooni leksikaalne varieeruvus on väga piiratud. Klasterites ei kerki esile tüüpilisi ja stereotüüpseid trigramme, nagu adverbi-, substantiivi- ja adjektiivialgulistes klasterites. Selgeimad dominandid joonistuvad välja adjektiivialgulistes klasterites, kusjuures alamklassisisene klasterivarieeruvus on sel juhul äärmiselt väike. Samas ei ole välistatud ka avar vormivarieeruvus koos leksikaalse rikkusega. Ühes klasteris võib samaaegselt olla hulgaliselt trigramme koos vaheldusrikka leksikaga. Niisugused on substantiivi- ja verbialguliste klasside alamklassid, kus klasterivarieeruvus väga suur ning klasterite invariantseid trigrammid ei ole ülejäänud trigrammidest oluliselt sagedamad. Osati käib see ka adverbiaalulise klassi kohta.

Niisiis toob eesti ilukirjanduskeele klasteranalüüs esile kindlaid keelekasutusmustrid. Sageli korduvad sõna(vormi)de järjendid kinnistuvad, teatud sõnade ja vormide vahele tekib justkui magnetväli, mis neid üksteise poole tõmbab. Nt klasteri DVP (*samas leidis ta*, *samas märkas ta*, *samas oli ta*) aluseks on muster *samas* + öeldisverb lihtmineviku 3. pöördes + 3. isiku personaalpronomeni. Adverbi ja verbi leksikaalse varieerumise alusel (*siiski lükkas ta*, *siiski jättis ta*) kinnistub see struktuur keelekasutusmustrina. Niisugusest teabest on kasu eesti keele automaattöötlusvahendite tarvis, aitab seletada kasutus põhise keelekäsitluse üld- ja üksikküsimusi, mh võrrelda eesti keele arengutendentse diakroonilises aspektis.

Ilukirjanduskeele klasteranalüüsi tulemuste võrdlus eesti keele grammatika, käsiraamatu, keeleõpikute ja koolile mõeldud programmide sisuga aitab leida vastuseid küsimusele, kuid võrd need allikad kajastavad autentselt keelekasutusele omaseid tendentse. Nähtustele, mis on ilukirjanduskeeles harvad või marginaalsed,

võiks keeleõppes pöörata vähem tähelepanu, samas kui keelekasutusmuutritele ja suurema esinemissagedusega keelestruktuuride kasutamisele võiks panna rohkem rõhku. Ilukirjanduskeel kui kirjaliku keelekasutuse ideaalvariant sobib malliks nii emakeelekõnelejale kui ka eesti keele kui teise keele omandajale.

Viidatud kirjandus

- Barddal, Johanna 2003. Case in Icelandic. – A Synchronic, Diachronic and Comparative Approach. *Lundastudier i Nordisk Språkvetenskap A 57*. Department of Scandinavian Languages, Lund University. <http://org.uib.no/iecastp/barddal/chapter2.pdf> (24.6.2013).
- Barlow, Michael; Kemmer, Suzanne (Eds.) 2000. Introduction. – Usage Based Models of Language. Stanford: CSLI Publications. <http://www.press.uchicago.edu/ucp/books/book/distributed/U/bo3624529.html> (10.2.2014).
- Burns, Robert; Burns, Richard 2008. Cluster analysis. – Business Research Methods and Statistics Using SPSS. Additional advanced chapters online: <http://www.uk.sagepub.com/burns/chapters.htm> (5.8.2013). SAGE Publications Ltd.
- Bybee, Joan L. 2003. Phonology and language use. Cambridge: Cambridge University Press.
- Bybee, Joan L. 2010. Language, Usage and Cognition. Cambridge University Press. <http://dx.doi.org/10.1017/CBO9780511750526>
- Bybee, Joan L.; Hopper, Paul J. 2001. Frequency and the Emergence of Linguistic Structure. *Typological Studies in Language 45*. John Benjamins Publishing Company. <http://dx.doi.org/10.1075/tsl.45>
- Croft, William 2003. Typology and Universals. Second edition. Cambridge Textbooks in Linguistics. Cambridge University Press.
- Eslon, Pille 2013. Kahe keelekasutusvariandi võrdlus: morfoloogilised klassid ja klastrid. [The comparative study of language use: Morphological classes and clusters.] – *Lähivõrdlusi. Lähivertailuja*, 23, 13–38. <http://dx.doi.org/10.5128/LV23.01>
- Eslon, Pille 2014. Morfosüntaktilise ja leksikaalse varieerumise piiridest. – *Eesti Rakenduslingvistika Ühingu aastaraamat*, 10, 55–71. <http://dx.doi.org/10.5128/ERYa10.18>
- Habicht, Külli 2006. Meie muutuv kirjakeel. [Our Literary Language in Change.] – *Oma Keel*, 1, 12–19.
- Halliday, Michael A. K. 1994. *An Introduction to Functional Grammar*. London: Arnold.
- Haspelmath, Martin 2006. Against markedness (and what to replace it with). – *Journal of Linguistics*, 42 (1), 25–70. <http://dx.doi.org/10.1017/S0022226705003683>
- Haspelmath, Martin 2008. Frequency vs. iconicity in explaining grammatical asymmetries. – *Cognitive Linguistics*, 19 (1), 1–33. <http://dx.doi.org/10.1515/COG.2008.001>
- Niglas, Katrin s. a. Klasteranalüüs. [Cluster Analysis.] <http://www.cs.tlu.ee/~katrin/wp/wp-content/uploads/2013/11/klaster.pdf> (24.6.2013).
- Norvig, Peter 1989. Building a large lexicon with lexical network theory. – Proceedings of the IJCAI Workshop on Lexical Acquisition. USA: University of California at Berkeley. <http://krextown.googlecode.com/svn/trunk/papers/lnt.pdf> (26.9.2013).
- Ots, Sander 2012. Statistikapõhise tarkvara loomine morfoloogiliste kollokatsioonide eraldamiseks eesti keele tekstidest. [Software for Morphosyntactic Cluster Extraction from Estonian Texts.] *Bakalaureusetöö*. Tallinn: Tallinna Ülikooli informaatika instituut.
- Remm, Kalle; Remm, Jaanus; Kaasik, Ants 2012. Ruumiliste loodusandmete statistiline analüüs. [Statistical Analysis of Spatial Data.] *Õpik-käsiraamat*. Tartu Ülikooli ökoloogia ja maateaduste instituut. http://kalleremm.ee/RASA/RASA_teooria.pdf (4.8.2013).

- Tuldava, Juhan 1977. Sagedussõnastik leksikostatistilise uurimise objektina. [The frequency dictionary as an object of lexicostatistical investigation.] – Tõid keelestatistika alalt. Keelestatistika II. Tartu: Tartu Riikliku Ülikool, 141–171.
- Tullo, Catriona; Hurford, James R. 2003. Modelling Zipfian distributions in language. – Proceedings of Language Evolution and Computation Workshop / Course at ESSLLI. University of Edinburgh, 62–75. <http://www.ling.ed.ac.uk/~jim/zipfjrh.pdf> (16.2.2013). Tartu Ülikooli ühtlustatud ilukirjanduskorpus. [Disambiguated Fiction Corpus of Estonian.] http://lepo.it.da.ut.ee/~heli_u/SA/ (24.6.2013).
- Veismann, Ann 2008. Eesti kaassõnade tähenduste kirjeldamise probleemidest. [A cognitive semantic analysis of the Estonian adpositions *üle* 'over' and *peale* 'on, onto': Some questions and queries.] – Keel ja Kirjandus, 5, 335–352. <http://keeljakirjandus.eki.ee/Veismann.pdf> (25.12.2013).

Jekaterina Trainis (Tallinna Ülikool) on käsitlenud Johannes Aaviku Rootsi-perioodi sõnavara, doktoritöös uurib semantilisi käändeid diakroonsest aspektist.
Narva mnt 25, 10120 Tallinn, Eesti
jekaterina.trainis@tlu.ee

Kais Allkivi (Tallinna Ülikool) on uurinud alaleütleva käände kasutust, magistritöös võrdleb eesti kirjakeele ja õppijakeele verbikonstruktsioone.
Narva mnt 25, 10120 Tallinn, Eesti
kais.allkivi@tlu.ee

ON BELLETRISTIC LANGUAGE FROM A NEW PERSPECTIVE

Jekaterina Trainis, Kais Allkivi

Tallinn University

The article describes the usage patterns of the Estonian belletristic language, which indicate how native speakers prefer to express their thoughts in writing. Research material is taken from the syntactically analyzed and disambiguated corpus of Estonian literary language. Cluster analysis that has been applied brings forth identical linear subsequences of morphological and syntactic tags based on frequency. Morphosyntactic trigrams that appeared in the corpus at least twice have been selected for the analysis.

The patterns of language use have been presented on the basis of the lexical category of the first component of constructions. Every morphological class consists of subclasses that have a common beginning and ending component, and every subclass consists of clusters that are distinguished by the variation of the middle components of trigrams. A cluster contains trigrams that have the same subsequence of morphological and syntactic tags. This hierarchy reveals general tendencies of language use as well as the limits of morphosyntactic and lexical variability.

Keywords: belletristic language, cluster analysis, morphosyntactic variability, lexical variability, Estonian

Lisa 1. Adverbialguline klass

Alam-klassid	Klastrid	Klastrisagedus
D-S	DAS	344
	DDS	169
	DPS	161
	DVS	142
	DSS	84
	DJS	46
	DNS	46
	DKS	27
	DGS	3
Kokku		1022
D-V	DDV	352
	DVV	219
	DJV	111
	DSV	99
	DPV	70
	DAV	31
	DXV	2
Kokku		884
D-D	DDD	265
	DVD	125
	DSD	91
	DJD	89
	DPD	87
	DAD	6
Kokku		663
D-P	DVP	175
	DDP	54
	DPP	13
	DJP	9
	DSP	8
	DAP	3
Kokku		262

Alam-klassid	Klastrid	Klastrisagedus
D-A	DDA	151
	DPA	43
	DVA	23
	DAA	22
	DSA	11
	DJA	6
	DKA	2
	DNA	2
	Kokku	
D-J	DDJ	75
	DVJ	59
	DAJ	51
	DSJ	41
Kokku		226
D-K	DSK	126
	DPK	22
	DDK	3
	DNK	2
Kokku		153
D-N	DNN	4
	DDN	2
Kokku		6
Üldarv kokku		3476

Lisa 2. Substantiivialguline klass

Alam-klassid	Klastrid	Klastrisagedus
S-D	SVD	611
	SDD	190
	SKD	116
	SSD	77
	SJD	41
	SPD	17
Kokku		1052
S-S	SJS	268
	SAS	166
	SVS	135
	SKS	82
	SDS	52
	SPS	27
	SSS	23
	SNS	4
Kokku		757
S-V	SVV	211
	SDV	183
	SKV	120
	SSV	95
	SJV	93
	SAV	4
	SPV	4
Kokku		710
S-J	SDJ	79
	SVJ	54
	SKJ	44
	SSJ	41
	SAJ	4
	SJJ	2
Kokku		224

Alam-klassid	Klastrid	Klastrisagedus
S-K	SSK	149
	SPK	15
	SVK	6
	SNK	2
Kokku		172
S-A	SVA	80
	SDA	37
	SKA	24
	SJA	14
	SSA	4
	SAA	2
SPA	2	
Kokku		163
S-P	SVP	44
	SJP	16
	SDP	8
	SKP	7
Kokku		75
S-N	SVN	2
	SVN	2
Kokku		4
Üldarv kokku		3157

Lisa 3. Verbialguline klass

Alam-klassid	Klastrid	Klastrisagedus
V-D	VDD	362
	VSD	218
	VPD	209
	VVD	177
	VJD	24
Kokku		990
V-S	VDS	170
	VAS	150
	VPS	111
	VSS	90
	VVS	65
	VJS	21
	VNS	17
Kokku		624
V-A	VDA	147
	VPA	33
	VSA	21
	VAA	17
	VVA	6
	VKA	2
Kokku		226
V-V	VDV	75
	VJV	41
	VSV	29
	VPV	24
	VVV	24
	VAV	2
Kokku		195

Alam-klassid	Klastrid	Klastrisagedus
V-J	VSJ	51
	VDJ	32
	VAJ	27
	VVJ	13
Kokku		123
V-P	VVP	64
	VDP	31
	VPP	8
	VJP	6
Kokku		109
V-K	VSK	85
	VPK	15
	VDK	4
	VNK	2
Kokku		106
V-N	VDN	2
	VPN	2
Kokku		4
Üldarv kokku		2377

Lisa 4. Pronoomenialguline klass

Alam-klassid	Klastrid	Klastrisagedus
P-D	PDD	256
	PVD	240
	PSD	210
	PPD	55
	PKD	39
Kokku		800
P-V	PVV	153
	PDV	145
	PSV	134
	PPV	18
	PKV	10
	PAV	2
Kokku		462
P-S	PAS	221
	PDS	62
	PVS	44
	PSS	36
	PPS	32
	PNS	10
	PKS	8
	Kokku	
P-K	PSK	145
	PPK	5
	PDK	4
Kokku		154

Alam-klassid	Klastrid	Klastrisagedus
P-P	PVP	34
	PDP	15
	PSP	14
	PPP	6
	PJP	4
	PKP	2
Kokku		75
P-A	PDA	35
	PVA	20
	PAA	6
	PSA	6
	PKA	5
Kokku		72
P-J	PSJ	30
	PDJ	6
	PAJ	4
	PPJ	4
Kokku		44
P-N	PDN	2
Kokku		2
Üldarv kokku		2022

Lisa 5. Konjunksioonialguline klass

Alam-klassid	Klastrid	Klastrisagedus
J-V	JPV	257
	JSV	245
	JDV	194
	JVV	35
	JJV	4
	JAV	2
Kokku		737
J-D	JVD	169
	JDD	142
	JPD	96
	JSD	61
	JJD	8
Kokku		476
J-S	JAS	148
	JVS	89
	JPS	82
	JSS	50
	JDS	36
	JJS	10
	JNS	10
Kokku		425

Alam-klassid	Klastrid	Klastrisagedus
J-P	JDP	58
	JVP	30
	JJP	20
	JPP	15
	JSP	7
	JKP	2
	Kokku	
J-K	JSK	56
	JPK	15
Kokku		71
J-A	JDA	21
	JPA	14
	JSA	12
	JVA	8
	JAA	5
Kokku		60
J-J	JSJ	6
	JAJ	3
Kokku		9
Üldarv kokku		1910

Lisa 6. Adjektiivialguline klass

Alam-klassid	Klastrid	Klastrisagedus
A-S	AAS	150
	ASS	116
	AJS	8
	APS	2
Kokku		276
A-V	ASV	186
	AVV	4
	AJV	2
	ADV	2
Kokku		194
A-D	ASD	142
	AJD	24
	ADD	4
	AVV	2
Kokku		172
A-K	ASK	150
Kokku		150
A-A	AJA	95
	ASA	17
	AAA	2
Kokku		114
A-J	ASJ	111
Kokku		111
A-P	ASP	14
	AVP	2
Kokku		16
A-N	ASN	2
Kokku		2
Üldarv kokku		1035

ADJEKTIIVIDE POLÜSEEMIA KORPUSES JA SÕNARAAMATUS

Maria Tuulik

Ülevaade. Artikli eesmärk on uurida eesti omadussõnade semantikat ja eelkõige polüseemiat koondkorpuse 100 sagedama adjektiivinäitel. Esimeses pooles vaadeldakse valimi jagunemist mitmetähenduslikkuse aspektist ja selleks, et selgitada polüseemia ja struktuuri suhteid, analüüsitakse ka adjektiivide morfoloogilise struktuuri varieerumist. Teises pooles keskendutakse semantilistele tüüpidele, mille abil saaks mitmetähenduslike adjektiivide tähendusniheid paremini kirjeldada. Tutvustatakse semantiliste tüüpide rakendusvõimalusi sõnaraamatute koostamisel ja pakutakse välja eesti adjektiivipeamised semantilised tüübid.

Võtmesõnad: semantika, leksikoloogia, leksikograafia, sõnamoodustus, semantilised tüübid, eesti keel

1. Sissejuhatus

Adjektiive pole traditsioonilises leksikaalsemantikas nii ulatuslikult uuritud kui nimisõnu või verbe, ometi on nad semantiliselt vähemalt sama mitmekesised (I. Peters, W. Peters 2000). Näiteks võib adjektiivide tähendus muutuda vastavalt sellele, millist konkreetset substantiivi tähenduskomponenti ta rõhutab ja samuti vastavalt sellele, mis tüüpi substantiivi täiendina ta esineb.

Artiklis käsitletakse eesti keele adjektiivide polüseemiat korpusandmete ja leksikograafiliselt koostatud kirjelduste põhjal. Sõnaraamatuesituste uurimisel kasutatakse peamise allikana praegu koostatavat ühekõitelist eesti keele seletavat sõnaraamatut (Langemets jt 2010), kus polüseemia ühtlasema esituse huvides märgitakse sõnatähenduste juurde semantilisi tüüpe. Spetsiaalselt adjektiivide jaoks pole eesti keeles semantilisi tüüpe veel välja töötatud, vaid kasutatakse substantiivide analüüsil tuvastatud tüüpe (Langemets 2010). Kuna adjektiivide semantilised tüübid sobituvad substantiivi omadega vaid osaliselt, siis arutletakse, milliseid oleks veel otstarbekas lisada.

Et leksikograafias toetab semantilise tüübi lisamine sõnatähendustele sama tähendusvälja sõnade sarnasemat kirjeldust, siis on adjektiivide semantiliste tüüpide väljatöötamine ka oluline etapp Eesti Keele Instituudi leksikograafi tööriista – EELexi sõnastikusüsteemi – arendamisel. Artikli rakenduslik väljund ongi eesti adjektiivide semantiliste tüüpide loend, mida saab sõnaraamatutöös kasutada.

Lisaks aitab semantiliste tüüpide selgitamine tulevikus tuvastada eesti adjektiivide süstemaatilise polüseemia malle. Margit Langemets (2010: 252) toob välja, et süstemaatilisi malle kodeerides on võimalik ühtlustada semantilise info esitust keelevaras, näidata sõnatähenduste vahelisi loogilisi seoseid ja vähendada seeläbi polüseemia loetlevat esitust, mis on omane traditsioonilisele sõnaraamatule.

2. Materjali valik

Korpuse töötlemise programmi Sketch Engine¹ abil on Tartu Ülikooli koondkorpusest tehtud adjektiivide sagedusloend (100 adjektiivide) ja puhastatud valim automaatse morfoloogilise analüüsi vigadest, et sisse jääksid ainult nii-öelda tõelised omadussõnad. Kuid millised on tõelised adjektiivid ja mida tuleks sõnade adjektiiviks määratlemisel arvesse võtta?

“Eesti keele grammatika” (EKG 1995: 20) defineerib omadussõnu kui tüüpjuhul käändes, arvus ja võrdlusastmetes muutuvaid sõnu, mis väljendavad (asja) omadust ning võivad lauses esineda ühilduva atribuudi või teatud liiki verbide laiendina ka predikatiivi ja seisundimääruse positsioonis. Kuid kõik semantilised adjektiivid pole keeles grammatiliselt adjektiivid. Eesti keeles on keeruline piiri tõmmata näiteks partitsiipide või käändumatute adjektiivide puhul. Käändumatud adjektiivid, mis automaatse analüüsiga esile tulid, jäeti semantika ja süntaksi põhjal valimisse sisse, nt *eri*. Samas partitsiipide iga juhtumit vaadati eraldi, kuna sagedusloendis leidis läbisegi realselt adjektiividena kasutuses partitsiipe ning puhtaid verbivorme, mis olid korpuses samamoodi adjektiiviks märgendatud.

Sagedusloendi 100 esimese adjektiivide seas oli 17 sõna, milles võis kahelda. Need olid korpusetextides ainult verbina kasutusel olevad sõnad (nt *olev*, *pandud*, *hakanud*, *kuuluv*), peamiselt verbina kasutusel olevad sõnad (nt *läinud*) ja sõnad, kus adjektiivne kasutus esines verbivormidega pooleks (nt *tulev*, *kehtiv*). Kuigi sõna *läinud* puhul on võimalik ka adjektiivne kasutus tähenduses ‘eelmine’ (nt *läinud aasta*), siis korpust ja sõnavisandeid uurides ilmnes, et valdav osa *läinud* kasutustest olid siiski verbivormid (*hästi läinud*, *tööle läinud*, *meelest läinud* jne). Tuli ette ka keerulisemaid juhtumeid, kus adjektiivne kasutus paistis korpuses verbivormidega pooleks esinevat, nt sõnad *tulev* ja *kehtiv*. Sõnavisanditest võis näha, et üsna palju on adjektiivset kasutust, nt *tuleval aastal*, *nädalal*, *hooajal*, aga samuti leidis hulgaliselt verbivorme, nt *välismaalt tulev*, *arutusele tulev*.

Sellistel juhtudel hinnati, kas partitsiibi adjektiivse kasutuse tähendus on nihkunud, see tähendab, et ei kattu enam verbi omaga (nt *elav* tähenduses ‘liikuv, elurõõmus’). Siin võis mängida rolli ka partitsiibi semantiline läbipaistmatus. Seejärel vaadati, kas võrdlusastmed on võimalikud, näiteks saab võrrelda *särav*, *säravam*, *säravaim*, ent mitte *kuuluv*, *kuuluvam*, *kuuluvaim*. Ja viimaks prooviti, kas partitsiibile on võimalik adverbi ette panna, näiteks kõlab täiesti loomulikult

¹ Sketch Engine on tekstikorpuste töötlemisprogramm, mille abil saab koostada ja töödelda konkordantse, teha sagedusloendeid, sõnavisandeid jm (Kilgarriff jt 2004). Selle eestikeelse mooduli kohta vt Kallas 2013.

väga erinev, kuid mitte *väga toimuv*. Kokkuvõtvalt võib öelda, et lähtuti leksikaal-semantilisest vaatepunktist ja valimisse jäeti sisse leksikaliseerunud, sisulises või vormilises mõttes oma paradigmast väljunud partitsiibid.

Automaatse analüüsiga sattus adjektiivide sagedusloendisse veel sõna *kogu*, millel EKSS-is adjektiivset tähendust polegi. Kokku tõsteti kõrvale 17 sõna ja võeti sama palju juurde n-ö õigeid adjektiive, mis sagedusloendis järgmiseks tulid. Eesti keele kõigi sõnade 100 sagedama sõna hulka mahtus adjektiividest üksnes kolm: *uus*, *suur* ja *viimane*.

3. Ülevaatlikult valimist

Koondkorpuse põhjal on kõige sagedamad adjektiivid eesti keeles *uus*, *suur*, *viimane*, *hea* ja *võimalik*. Võrdlusena võiks tuua Oxfordi korpuse² põhjal inglise keele sagedamad adjektiivid *good* 'hea', *new* 'uus', *first* 'esimene', *last* 'viimane' ja *long* 'pikk'. Kolm viiest sagedamast adjektiivist (*uus*, *hea* ja *viimane*) kattuvad ning ka sõna *esimene* sobiks oma sageduse poolest loetelusse, kuid on koondkorpuses märgendatud numeraalina ja seetõttu siinses eesti adjektiiviloendis esile ei tule.

Võrreldes ülejäänud eesti ja inglise keele 100 sagedamat omadussõna, leidis kinnitust asjaolu, et valdav osa sagedatest adjektiividest on universaalsed. Samas paistab eesti omadussõnade tagant korpuse iseloom. Kuna suur osa koondkorpuse tekstidest on ajakirjanduslikud, siis on valimis sees sõnad nagu *majanduslik*, *sotsiaalne*, *rahvuslik*, *juriidiline* ja *erapooletu*, kuid esile ei tule näiteks meeleolu märkivad adjektiivid nagu *rõõmus*, *kurb* või *vihane*, mis mõjuvad märksa igapäevasemana.

Värvidest kuuluvad eesti 100 sagedama adjektiivi hulka *must*, *valge*, *punane* ja *roheline*. See langeb kokku Berliini ja Kay (1969) hierarhiaga, mille järgi on *must* ja *valge* põhivärvinimed ehk kõige primaarsemad. Kui keeles on kolmas värvisõna, siis on see *punane*, ja kui neljas, siis on see *roheline* või *kollane*. Valimist nähtub, et samas järjekorras tulevad värvisõnad ka sageduse põhjal esile. Kuid miks tuleb eesti keeles esimesena, peaaegu poole suurema sagedusega välja *roheline*, mitte *kollane*? Korpuse lauseid analüüsides selgub, et *rohelise* sagedusele aitab kaasa keskkonnakaitsjate tähendus ja keskkonnakaitsesega seotud väljendite rohkus (nt *roheline mõtteviis*, *liikumine*, *energia* jne). Ka *kollasel* on sagedaid kollokatsioone teises tähenduses 'põnevust, sensatsiooni taotlev', nt *kollane ajakirjandus*, *ajaleht*, aga üldiselt ei tundu *kollase* teine tähendus nii produktiivne ja semantiliselt avatud kui *rohelise* oma.

4. Morfoloogiline struktuur ja polüseemia

Morfoloogilise struktuuri järgi jaotub valim järgmiselt: juuradjektiive on 34, tuletisi 57, liitsõnu 3 ja *v*-kesksõnu 6 (vt tabel 1). Sõnade juuradjektiivideks, tuletisteks ja liitsõnadeks jagamisel tugineti Silvi Vare sõnapere sõnaraamatule (2012).

² Vt <http://oxforddictionaries.com/words/the-oxford-english-corpus> (15.10.2013).

Tabel 1. Valimi jaotumine morfoloogilise struktuuri järgi

Valim	Kokku	Näited
juuradjektiivid	34	<i>uus, suur, hea, vana, noor, vaba, väike</i>
tuletised	57	<i>tõeline, vajalik, järgmine, edukas, ilus</i>
liitsõnad	3	<i>rahvusvaheline, käesolev, lugupeetud</i>
v-kesksõnad	6	<i>erinev, huvitav, sobiv, vastav, tulev, elav</i>

Reet Kasiku (2004: 140) järgi on oluline osa eesti keele adjektiividest *lik-*, *line-* või *ne-*tuletised ja juuradjektiive on suhteliselt vähe. Vaadeldud omadussõnade loendis moodustasid juuradjektiivid kolmandiku ja sageduse poolest olid need esil, esimese 20 seas oli tüvisõnu isegi üle poole, kinnitades seega Zipfi (1949) vähima pingutuse seadust, mille järgi kõige sagedamad sõnad on ühtlasi ka kõige lühemad. Eesti sagedamad juuradjektiivid on enamasti “kulunud” 3-4-täheliseks.

Tuletised jagunevad järgnevalt: 28 *ne-*tuletist, 13 *lik-*tuletist, 10 *line-*tuletist ja 6 muud, mille hulgas omakorda leidus *tu-*liitelisi, *s-*liitelisi ja *ik-*liiteline tuletis. Tabel 2 esitab tuletusliidete järgi jagatud tuletiste esinemisarvud koos näidetega uuritud valimist.

Tabel 2. Tuletised liidete järgi

Tuletised	Kokku	Näited
<i>ne-</i> tuletised	28	<i>tänane, lihtne, tõsine, endine</i>
<i>lik-</i> tuletised	13	<i>ohtlik, võimalik, kohalik, vajalik</i>
<i>line-</i> tuletised	10	<i>tavaline, poliitiline, roheline, tõeline</i>
muud (<i>tu-</i> , <i>ik-</i> , <i>s-</i> tuletised)	6	<i>edukas, ilus, tähtis, võimatu, üksik, erapooletu</i>

Adjektiivide tähenduste arv pärineb koostamisel oleva ühekõitelise eesti keele sõnaraamatu (Langemets jt 2010) tööbaasist. Võrreldes suure seletava sõnaraamatuga (EKSS 2009), kus tähendusi loetleti paljuski lineaarselt, esitatakse siin polüseemiat prototüüpsete tuumtähenduste kaupa, millest lähtuvad teised, tuletatud tähendused. See tähendab EKSS-i peente tähenduseristuste üldistamist, ümberrühmitamist ja sageli kokkuvõtmist (Langemets jt 2010: 799–780). Praegu on märksõnastiku ligikaudu 9000 adjektiivist koostatud umbes pooled.

Polüseemseid sõnu, st selliseid, kus sõnaraamatus on eristatud kaks või enam põhitähendust, oli 100 sagedama adjektiivi hulgas kokku 80, sealhulgas ka üks homonüüm (*täis*). Monoseemseid adjektiive, millel sõnaraamatus on üks põhitähendus, oli valimis 20. Polüseemsete adjektiivide suur ülekaal on ootuspärane, kinnitades samuti juba Zipfilt pärinevat väidet, et sageli esinevad sõnad kipuvad olema polüseemsemad (Zipf 1945: 251–256).

Uuritud 100 adjektiivil oli põhitähendusi kokku 341. Tabel 3 annab ülevaate sellest, kuidas sagedamad adjektiivid jagunevad vastavalt tähenduste arvule.

Tabel 3. Valimi jagunemine põhitähenduste arvu põhjal

Tähenduste arv	1	2	3	4	5	6	7	8
Sõnu vastava tähenduste arvuga	20	31	8	10	9	9	8	5

Substantiivide analüüsil on Langemets (2010: 88) välja toonud, et kõige lühemad sõnad ehk tüvisõnad on kõige polüseemsemad, seetõttu huvitas uurijat, kas sama tendents kehtib ka adjektiivide juures. Sel eesmärgil arutati eraldi juuradjektiivide, tuletiste, liitsõnade ja partitsiipide polüseemia indeks (tabel 4).

Tabel 4. Adjektiivide polüseemia indeks ehk keskmine põhitähenduste arv sõna kohta

Valim	Kokku	Polüseemia indeks
juuradjektiivid	34	5
tuletised	57	2,6
liitsõnad	3	1,3
v-kesksõnad	6	2,3

Tabelist 4 nähtub, et juuradjektiivide polüseemia indeks on 5, tuletistel aga 2,6 ehk ligikaudu poole väiksem. Järgnevad v-kesksõnad (2,3) ja liitsõnad (1,3). Kõik 6-tähenduslikud ja valdav osa 7-tähenduslikest sõnadest olid juuradjektiivid. Monoseemseid juuradjektiive leidis 100 adjektiivi seas ainult kaks, *eri* ja *kuulus*. Kõige polüseemsemad tüvisõnad ja tuletised valimis olid sellegipoolest tähenduste arvu järgi võrdsed: *hea* ja *vana* 8 tähendusega ning *viimane* ja *positiivne*, samuti 8 tähendusega. Kõige vähem polüseemsed olid liitsõnad, mis taas toetab Zipfi (1945, 1949) väidet, et sagedamad sõnad on lühemad. Seega võib arvata, et liitsõnad (“tänu” oma pikkusele) kalduvad olema pigem monoseemsed.

5. Semantilised tüübid ja nende rakendamisvõimalused

Semantilised tüübid on abstraktsed abivahendid tähendusüksuse juures, mille (kombinatsioonide) abil on võimalik võrrelda omavahel semantiliselt seotud sõnu, kirjeldada sõna sisemisi tähendusmalle. Näiteks Langemetsa järgi (2010: 198–200) võib isikuomaduste sõnade puhul üsna regulaarselt esilduda suhe OMADUS–INIMENE, mis tähendab, et üks ja sama sõna tähistab nii (füüsilist) omadust kui ka isikut, keda selle omaduse järgi tuntakse või iseloomustatakse. Semantilised tüübid on vastavalt OMADUS ja INIMENE_OMADUS (näide 1).

- (1) OMADUS–INIMENE
 Kõlas mahe bass. (OMADUS)
 Esines kuulus bass. (INIMENE_OMADUS)

Semantilised tüübid võivad olla uurijale abiks polüseemsete sõnade analüüsil, eriti süstemaatilise polüseemia selgitamisel. Ühtlasi võivad semantilised tüübid sõnaraamatutöös olla ka praktiliseks abivahendiks sõna semantika kodeerimisel. Kui teatud sõnade tähenduste vahelduses ilmnevad regulaarsused on juba kindlaks tehtud, siis võib vastavaid malle kodeerides sõnaraamatus semantilise info esitust korrastada ning näidata ja selgitada sõnatähenduste vahelisi loogilisi seoseid (Langemets 2010: 252).

Peale selle võimaldab semantiliste tüüpide kasutamine sõnaraamatutöös samasse tähendusvälja kuuluvaid sõnu ühtlasemalt kirjeldada. Näiteks koostati

eesti keele põhisõnavara sõnastiku (PSV 2014) substantiivid Langemetsa tüpoloogialusel mõistegruppide kaupa (linnud, loomad, kehaosad, kuud jne), mitte tähestiku järjekorras, nagu on sõnastik ise. Sellises koostamismudelis on võimalik sama tüüpi sõnade puhul kasutada sarnast leksikograafilist malli (nt sarnase ülesehitusega definitsioone), mille tulemuseks on semantilise info ühtlasem esitus sõnaraamatus.

Teine väga oluline semantiliste tüüpide rakendusvaldkond on korpusete semantiline märgendamine, mida Eestis pole seni veel tehtud. Semantilise märgendamise peamine eesmärk on sõnatähenduste ja sõnaühendite tähenduste ühestamine, mis on vajalik näiteks masintõlkeprogrammide loomisel. Sisuliselt tähendab see polüseemse sõna tähenduste eristamist tekstis. Semantiline tüüp aitab märgendada sõnade tähenduslikku kuuluvust: korpusete esinevale sõnale või sõnaühendile lisatakse märges vastavalt selle semantilisele kategooriale. Sel moel saaksid semantilised tüübid toetada automaatset semantilist keeleanalüüsi ja sõnatähenduste ühestamist.

6. Eesti adjektiivide semantilised tüübid

Semantiliste tüüpide väljatöötamisel analüüsiti ühekõitelise seletava sõnaraamatu põhjal 100 sagedama adjektiivide 341 põhitähendust. Tähtsused jagati semantikast lähtudes rühmadesse ja uuriti definitsioone, et vaadata, kas teatud tüüpi adjektiivid on ka sarnasel viisil defineeritud ja ehk annab see, millist tähenduse elementi või suhet definitsioonis välja tuuakse, vihje semantilise tüübi määramiseks.

Ühekõitelises eesti keele seletavas sõnaraamatus märgitakse adjektiivide juurde substantiivide analüüsil tuvastatud tüüpe (Langemets 2010). Kuna adjektiivide semantilised tüübid substantiivide semantiliste tüüpidega päriselt ei sobitu, siis ongi käesoleva artikli üks eesmärk pakkuda välja tüübistik just spetsiaalselt omadussõnadele.

Adjektiivide semantiliste tüüpide arv võib erinevates käsitlustes olla väga erinev. Uurimuses kasutati eeskujuna ja võrdlusmaterjalina Dixoni (1982) seitset olulisemat adjektiivide semantilist tüüpi, leksikaalse andmebaasi SIMPLE koostamisel väljatöötatud 18 tüüpi (Lenci jt 2000: 30–33) ning Hundsnurcheri ja Spletti (1982) uurimuse põhjal loodud GermaNeti 15 adjektiivide semantilist klassi, millel on omakorda 78 alaliiki (vt ka Orav 2000).

Tuntud adjektiivide uurija Robert Dixon on välja toonud 7 olulisemat adjektiivide semantilist tüüpi: DIMENSIOON (nt *suur, pikk, lai*); VANUS (nt *uus, vana*), HINNANG (nt *hea, halb, veider*), VÄRVUS (nt *valge, must, punane*), FÜÜSILISED OMADUSED (nt *kõva, tugev, puhas*), KIIRUS (nt *kiire, aeglane, äkiline*) ja INIMESEGA SEOTUD OMADUSED (nt *vihane, rõõmus, armukade*). Neli esimest (DIMENSIOON, VANUS, HINNANG ja VÄRVUS) on Dixoni järgi kõige enam seotud adjektiivide kategooriaga ja moodustavad omadussõnade semantilise prototüübi, see tähendab, et keeltes, kus on piiratud hulk omadussõnu, kuuluvad need tõenäoliselt just nendesse tüüpidesse (Dixon 1982: 46).

Semantilise tüübistiku koostamisel katsuti silmas pidada ka rakenduslikku eesmärki, et seda oleks võimalik sõnaraamatutöös päriselt kasutada. See tähendab, et tüüpe ei tohiks olla nii palju, et need muutuksid asjaks iseeneses, samuti ei tohiks tekkida olukorda, kus leksikograafil on iga tähenduse juures kõvasti peamurdmist,

millise tüübi kasuks otsustada. Seetõttu üritati ka tüüpide nimetused hoida võimalikult läbipaistvad ja arusaadavad. Samas oli analüüsi huvides jagada tüübid peeneteralisemaks, kui need hetkel ühekõitelises seletavas sõnaraamatus on. Kuna selleks, et oleks võimalik järgmise sammuna selgitada adjektiivide regulaarseid tähendusvaheldusi ehk avada süstemaatilise polüseemia malle, peavad nende semantilised tüübid olema senistest (vt tabel 5) märksa kitsamad.

Tabel 5. Ühekõitelise seletava sõnaraamatu omadussõnade semantilised tüübid

Semantiline tüüp	Kirjeldus
OMADUS_FÜÜS	füüsikaline omadus (mass, kõrgus, kiirus, heli, värv jne)
OMADUS_FÜÜS_VÄRV	värv, värvus
OMADUS_KVAL	hinnanguline omadus
OMADUS_MAX	intensiivsust, suurust rõhutav omadus
OMADUS_MIN	väiksust, märkamatus rõhutav omadus
OMADUS_PSÜHH	psüühikaga (tunnetega, tajuga) seotud omadus

Praegu on ühekõitelises seletavas sõnaraamatus adjektiivide jaoks kasutusel kuus semantilist tüüpi (lisaks veel üldine tüüp OMADUS). Sellise tüübistikuga ei tule tähendusvaheldused alati esile, näiteks *vana* tähenduses ‘kunagine’, mis praegu on märgitud lihtsalt omaduseks, võiks olla OMADUS_AEG, *vana* tähenduses ‘teatava vanusega’, mis on samuti omaduseks märgitud, võiks olla OMADUS_VANUS jne.

Tabelis olevatest tüüpidest jäid uude tüpoloogiasse samal kujul OMADUS_FÜÜS ja OMADUS_PSÜHH, OMADUS_MAX ja OMADUS_MIN sulatati kokku tüübiks OMADUS_RÕHUTAMINE, kuna materjali uurides torkas silma, et muidu võis tüübi nime ja selle sisu vahel kohati vastuolu tekkida, näiteks *viimase* tähendus ‘muude seas vähima väärtusega, kõige madalamal asuv, kõige viletsam’ (näitelause: *Kas mina olen siis see viimane, kellega tantsidagi ei kõlba?*) on märgitud tüübiks OMADUS_MAX ja on arusaadav, et mõeldakse ‘maksimaalselt vähim, viletsaim’, kuid siiski võiks seda tähendust semantika põhjal tõlgendada ka tüübina OMADUS_MIN ehk ‘kõige vähem väärtuslik’.

OMADUS_FÜÜS_VÄRV ja OMADUS_KVAL jäid samuti uude tüübistikku sisse, kuid need nimetati ümber tüüpideks OMADUS_VÄRV ja OMADUS_HINNANG. Esimene seetõttu, et sobituks vormi poolest paremini ülejäänud tüüpidega ja teine lihtsalt nimetuse läbipaistvuse huvides.

Valimi analüüsist koorus 15 omadussõna semantilist tüüpi, mis on esitatud tabelis 6 koos seletuse ja näidetega. Näited esindavad vastavate adjektiivide EKSS-i esimest põhitähendust (mitte kõiki). Kui silmas peetakse mitteprimaarset tähendust, on näite taha lisatud selgitus. Tärniga (*) on tähistatud näited, mis valimisse ei sattunud.

Ideaalis võiksid kõik semantilised tüübid üksteist välistada, kuid tegelikkuses on sellist tulemust väga raske – kui mitte võimatu – saavutada. Tabelis 6 toodud tüüpidest on mõni laiem kui teised, näiteks OMADUS_FÜÜS hõlmab ka tüüpe OMADUS_MASS, OMADUS_VÄRV, OMADUS_MÕÕDE jms. OMADUS_FÜÜS on mõeldud kasutamiseks teise valikuna ehk siis, kui adjektiivi tähendust ei saa paigutada kitsamasse tüüpi.

Tabel 6. Adjektiivide semantilised tüübid

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

Näites (2) on lühendatud kujul esitatud EKSS-i artikkel *endine*, mille tähendusi on varasematesse semantilistesse tüüpidesse keeruline liigitada. Nende järgi klassifitseeruksid mõlemad tähendused üldsõnalise tüübi OMADUS alla, kuid uusi tüüpe kasutades võiks esimene tähendus olla OMADUS_AEG ja teine tähendus OMADUS_SUHTELISUS.

(2) **endine** <adj> <-se 5 või -se 4>

- 1 [OMADUS_AEG] varem kellekski või millekski olnud, varem kasutusel olnud, kunagine. *Endine õpetaja, direktor.*
 - varasemasse, möödunud aega kuuluv, muistne. *Endised ajad, inimesed.*
- 2 [OMADUS_SUHTELISUS] selline nagu varem, muutumatu. *Elab endist viisi.*

Loodud tüübiga ei ole kindlasti ammendav ja võib edasise analüüsi käigus täiendada. Kuid loodetavasti saab siiski ka praegu asendada sõnastikubaasis lihtsalt omaduseks märgitud tähendused täpsema tüübiga. Ja leksikograafi vaatepunktist võib polüsemse sõna puhul juba semantiliste tüüpide lisamine aidata mõtteid organiseerida ja paremini tähendusjaotust struktureerida.

7. Kokkuvõte ja tulevik

Artiklis kirjeldati ja analüüsiti eesti 100 sagedamat adjektiivii nii polüseemia kui morfoloogilise struktuuri seisukohast. Eraldi vaadeldi juuradjektiivide, tuletiste, liitsõnade ja partitsiipide polüseemia indeksit ehk keskmist tähenduste arvu sõna kohta. Kõige polüseemsemad on eesti keeles juuradjektiivid (polüseemia indeks 5). Tuletiste polüseemia indeks on 2,6 ehk ligikaudu poole väiksem. Kõige monoseemsemad olid 100 sagedama adjektiivii hulgas liitsõnad (polüseemia indeks 1,3).

Artikli teises pooles tutvustati semantiliste tüüpide rakendusvõimalusi, nii teoreetilise abivahendina polüseemsete sõnade analüüsil kui ka praktilise abivahendina sõnasemantika kodeerimisel sõnaraamatutöös. Analüüsides ühekõitelise seletava sõnaraamatu põhjal 100 sagedama adjektiivii põhitähendusi, pakuti välja 15 semantilist tüüpi spetsiaalselt eesti keele (polüseemsete) adjektiivide kirjeldamiseks.

Järgmise sammuna on plaanis tuvastada eesti adjektiivide süstemaatilise polüseemia mallid, mida nüüd, tänu olemasolevatele semantilistele tüüpidele, on võimalik teha. Selle töö eesmärk omakorda oleks uurida eesti adjektiivii polüseemiat ja luua ka teoreetiline alus, mille põhjal saaks ühtlustada adjektiivide polüseemia esitust sõnaraamatus.

Viidatud kirjandus

- Berlin, Brent; Kay, Paul 1969. *Basic Color Terms. Their Universality and Evolution*. Los Angeles.
- Dixon, Robert M. W. 1982. *Where have all the adjectives gone? And other essays in syntax and semantics*. *Janua linguarum. Series Maior*, 107. Berlin: Walter de Gruyter. <http://dx.doi.org/10.1515/9783110822939>
- EKG 1995= Erelt, Mati; Kasik, Reet; Metslang, Helle; Rajandi, Henno; Ross, Kristiina; Saari, Henn; Tael, Kaja; Vare, Silvi 1995. *Eesti keele grammatika I*. Eesti Teaduste Akadeemia, Eesti Keele Instituut. Tallinn.
- EKSS = Eesti keele seletav sõnaraamat 1–6. (“Eesti kirjakeele seletussõnaraamatu” 2., täiendatud ja parandatud trükk.) [The Explanatory Dictionary of Estonian.] Toimetanud Margit Langemets, Mai Tiits, Tiia Valdre, Leidi Veskis, Ülle Viks, Piret Voll. Eesti Keele Instituut. Tallinn: Eesti Keele Sihtasutus, 2009.
- Hundsnurher, Franz; Splett, Jochen 1982. *Semantik der Adjektive im Deutschen: Analyse der semantischen Relationen*. Westdeutscher Verlag. <http://dx.doi.org/10.1007/978-3-322-87624-9>
- 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.
- Kasik, Reet 2004. *Eesti keele sõnatuletus*. [Derivation in Estonian.] Tartu: Tartu Ülikooli Kirjastus.
- Kilgarriff, Adam; Rychly, Pavel; Smrž, Pavel; Tugwell, David 2004. *The Sketch Engine*. – Proceedings of the 11th EURALEX International Congress (Lorient, 6–10 July 2004). Lorient: Université de Bretagne-Sud, 105–117.
- Langemets, Margit 2010. *Nimisõna süstemaatiline polüseemia eesti keeles ja selle esitus keelevaras*. Tallinn: Eesti Keele Sihtasutus.
- Langemets, Margit; Tiits, Mai; Valdre, Tiia; Voll, Piret 2010. *In spe: ühekõiteline eesti keele sõnaraamat*. – *Keel ja Kirjandus*, 11, 793–810.

- Lenci, Alessandro; Bel, Nuria; Busa, Federica; Calzolari, Nicoletta; Gola, Elisabetta; Monachini, Monica; Ogonowski, Antoine; Peters, Yvonne; Peters, Wim; Ruimy, Nilda; Villegas, Marta; Zampolli, Antonio 2000. SIMPLE: A General Framework for the Development of Multilingual Lexicons. – *International Journal of Lexicography*, 13 (4), 249–263. <http://dx.doi.org/10.1093/ijl/13.4.249>
- Orav, Heili 2000. Adjektiivid kui semantiline probleem: wordnet-tüüpi tesaauruste koostamise kogemused. – Tiit Hennoste (Toim.). *Arvutilingvistikalt inimesele*. Tartu Ülikooli üldkeeleteaduse õppetooli toimetised 1. Tartu: Tartu Ülikooli Kirjastus, 153–166.
- Peters, Ivonne; Peters, Wim 2000. The Treatment of Adjectives in Simple: Theoretical Observations. – *Proceedings of LREC 2000: 2nd International Conference on Language Resources and Evaluation*.
- PSV 2014 = Eesti keele põhisõnavara sõnastik. [The basic dictionary of Estonian.] Toimetanud Jelena Kallas, Mai Tiits, Maria Tuulik. Tallinn: Eesti Keele Sihtasutus, 2014.
- Zipf, George Kingsley 1945. The meaning-frequency relationship of words. – *Journal of General Psychology*, 3, 251–256. <http://dx.doi.org/10.1080/00221309.1945.10544509>
- Zipf, George Kingsley 1949. *Human behavior and the principle of least effort*. Addison-Wesley Press.
- The Oxford English Corpus. <http://oxforddictionaries.com/words/the-oxford-english-corpus> (15.10.2013).
- Vainik, Ene 2002. Millest on tehtud eestlaste emotsioonisõnavara? – *Keel ja Kirjandus*, 8, 537–553.
- Vare, Silvi 2012. Eesti keele sõnapered. Tänapäeva eesti keele sõnavara struktuurianalüüs I–II. Tallinn: Eesti Keele Sihtasutus.

Maria Tuulik (Eesti Keele Instituut, Tartu Ülikool) on eesti keele põhisõnavara sõnastiku töörühma liige ja kirjutab doktoritööd eesti adjektiivide polüseemiast. Põhilised uurimisvaldkonnad: leksikaalne semantika, leksikograafia, tekstilingvistika.
 Roosikrantsi 6, 10119 Tallinn, Estonia
maria.tuulik@eki.ee

ADJECTIVAL POLYSEMY IN CORPORA AND DICTIONARIES

Maria Tuulik

Institute of the Estonian Language

The article describes and analyzes the 100 most frequent Estonian adjectives with regard to both polysemy and morphological structure. Also examined is the polysemy index, i.e. the average number of meanings per word, for root adjectives, derivatives, compound words and participles. The most polysemous of these are root adjectives, with a polysemy index of 5. The polysemy index of derivatives is roughly half that, 2.6. Among the 100 most frequent adjectives, the least polysemous are compound words (polysemy index 1.3).

The second half of the article discusses the application of semantic types, both as a theoretical aid in the analysis of polysemous words and as a practical tool in the coding of semantic information in dictionaries. Based on the primary meanings of the 100 most frequent adjectives as given in the single-volume explanatory dictionary of Estonian, 15 semantic types are proposed in order to describe Estonian (polysemous) adjectives.

The next planned step, made possible by the existence of the semantic types, is to introduce the patterns of systematic adjectival polysemy in Estonian. The aim of this would be to study Estonian adjectival polysemy and establish a theoretical framework on the basis of which the presentation of adjectival polysemy in dictionaries could be harmonized.

Keywords: semantics, lexicology, lexicography, word formation, semantic types, Estonian

SÕNASEMANTIKA JA TEKSTI MÕISTMINE PÕHIKOOI ESIMESES JA TEISES KOOLIASTMES

Krista Uibu, Maile Timm

Ülevaade. Sõnasemantika ehk sõna tähenduse tundmine on oluline tervikteksti sisu mõistmisel. Kui sõnade tähendus ei kinnistu, võib teksti mõistmine kujuneda õpilastele probleemiks kõigis kooliastmetes. Käesoleva pikiuurimuse eesmärk oli välja selgitada, mil määral muutub sõnatähenduse tundmine ja teksti mõistmine põhikooli esimeses ja teises kooliastmes. Artiklis analüüsitakse 3.–5. klassi õpilaste sõnatähenduse ning kahe raskusastmega tekstimõistmise ülesannete sooritust. Uuringus osales 508 õpilast (267 tüdrukut ja 241 poissi) 30 Eesti üldhariduskoolist. Selgus, et teksti mõistmine oli parem õpilastel, kes tundsid paremini erinevate sõnade tähendust. Soolises võrdluses edestasid tüdrukud poisse kõigil kolmel aastal sõnatähenduse tundmises ja teksti mõistmise keerukamas ülesandes, mis eeldas head info integreerimise oskust. Tüdrukud tundsid poistest paremini ka nende sõnade tähendust, mida nad olid varem õppinud. Et toetada õpilaste semantilist pädevust ja teksti mõistmist, on oluline arvestada nende arengulist eripära ning pakkuda lugemiseks võimetele ja huvidele vastavaid tekste.*

Võtmesõnad: sõnasemantika, teksti mõistmine, soolised erinevused, pikiuurimus, eesti keel

1. Sissejuhatus

Teksti mõistmine on keerukas kognitiivne protsess (Cain jt 2004), mis on seotud sõnasemantika, sh sõnade tähendussuhete mõistmisega (Häkkinen 2007). See hõlmab kõike alates üksikute sõnade kodeerimisest kuni tervikteksti sisu mõistmiseni (Kabilan jt 2010). Teksti mõistmise ja sõnavara arendamine on olulisimaid ülesandeid põhikoolis (Põhikooli riiklik õppekava 2010). Nii on esimese kooliastme eesmärk lugemistehnikate omandamine ning sõnatähenduse, lause ja tervikteksti

* Uurimust on toetanud Haridus- ja Teadusministeeriumi grant nr 3-2/TA5966 "Eesti põhikooli efektiivsus".

sisu mõistmine. Teises kooliastmes on pöhirõhk eakohastest tekstidest arusaamisel ja tekstist vajaliku teabe leidmisel (Põhikooli riiklik õppekava 2010). Kui laps mõistab sõnade tähendust, siis saab ta tõenäoliselt aru ka lausetest ja tervikteksti mõttest. Samas tekitab sõnasemantika paljudele õpilastele probleeme, sest erinevatel põhjustel (nt ajanappus, tähelepanematus, teksti keerukus) ei pruugi uued sõnad kinnistuda ning nende tähendus jääb arusaamatuks. Seepärast on oluline märgata raskusi põhjustavaid sõnu (Soodla jt 2013) ja pühendada nende kinnistamisele piisavalt aega.

Teksti mõistmise ja sõnavara tundmise uurimustest on selgunud, et õpilastel esineb raskusi eakohastest tekstidest arusaamisega, arendamist vajab sõnasemantika (Henno jt 2007, Pandis 2002, Puksand jt 2010). Samas puuduvad pikiuurimused õpilaste sõnatähenduse tundmisest ja seostest teksti mõistmisega. Käesoleva uurimuse eesmärk oli välja selgitada pikaajalised muutused põhikooli esimese ja teise kooliastme õpilaste sõnatähenduse tundmisest ning seosed sõnatähenduse ja ilukirjandusliku teksti mõistmise vahel.

2. Sõnasemantika ja teksti mõistmise areng

2.1. Keele ja sõnatähenduse omandamine

Noam Chomsky (1965) käsitluse järgi on keele omandamise süsteem inimesel kaasa sündinud ja keel omandatakse kuulmise abil. Seevastu kognitiivpsühholoogias valitseb arvamus, et lapsed ei omanda keelt mitte loomupäraselt, vaid õpivad seda kasutama kogemuslikult (Tomasello jt 1997). Saxtoni (2010) järgi tähendab keele kogemuslik õppimine individuaalse kõne kujunemist. Oluline periood emakeele arengus on nn *kriitiline periood*, mis algab 6. eluaastast ja kestab varase puberteedieani (umbes 12. eluaastani). Maksimaalse tulemuse saavutamiseks tuleks teha keele omandamisega tööd selleks ettenähtud ajal, sest pärast seda ei ole võimalik keelt täielikult omandada (Saxton 2010). Uurimusi keele õppimise kriitilisest perioodist on ka teistelt teadlastelt (Lenneberg 1967, Mayberry 1993, Meisel 2008), kes on leidnud, et esmatähtis on arendada sõnavara, mille kaudu laps õpib nii verbaalset suhtlemist kui ka lugemist. Sõnade õppimise tasandil analüüsib õppija iga sõna täht-tähelt, iga täht on eraldi üksus ning need kokku moodustavad uue tähendusega terviksõna (Schroeder 2011). Uut sõna korrates ja analüüsides kinnistatakse sõna tähendus ning see liigub lühiajalisest mälust edasi pikaajalisse mällu (Tulving, Madigan 1970).

Nii Zhang jt (2008) kui ka Prado ja Plourde (2011) on leidnud, et sõnavara omandatakse etapi viisi. Esimeses etapis toimub tajutav sõnatöötlus ehk sõna seostatakse reaalse eluga. Näiteks, sõna *kullerkupp* puhul soovitatakse kasutada visuaalset materjali (pildikaarte, fotosid, videoid), mille abil tekivad lapsel kiiremini seosed ja mälopildid. Teises, liigendamise etapis kujuneb õppijal reaalne seos sõna ja objekti vahel (vt ka Tulving, Madigan 1970: 437–484), nt *kullerkupp näeb välja selline nagu pildikaardil, ta on kollast värvi* (õpetaja räägib juurde, et lill õitseb kevadel). Lapsel on tekkinud õpitud sõna kohta tajupildid ning need on kandunud üle reaalsesse ellu. Kolmandas etapis oskab laps õpitud sõna juba sobivas kontekstis kasutada, nt vestlusel kevadlilledest teab ta, et *kullerkupp on*

kollane kevadlill. Kui kõik need kolm etappi on läbitud, on suur tõenäosus, et õpitu kinnistub (Zhang jt 2008).

On leitud, et sõnaliigiti on sõnade omandamine erinev. Kuna sõna on abstraktne mõiste, siis enne tähenduse andmist sõnale peab õppija selle enda jaoks kategoriseerima (Bloem, La Heij 2003, Caramazza 1997), nt *kullerkupp on lill*, nimisõna. Kõige lihtsam on omandada konkreetse tähendusega nimisõnu, sest neile saab osutada otse (nt *lill, vihmavari, laud*). Keerulisem ja aeganõudvam on omandada abstraktse tähendusega nimisõnu (nt *kurbus, lein, õnn*), mida pole võimalik otseselt näha ega katsuda (Hirsh-Pasek, Golinkoff 2006). Veelgi keerukam on tegu- ja omadussõnadega (Imai jt 2005), sest sõnale tähenduse andmine tähendab erinevate kasutuskontekstide ühendamist ja laiendamist (Cruse 2004). Lisaks vajab laps tegu- ja omadussõnade puhul mitmekordset kogemust, enne kui ta sõna tähenduse omandab (Saxton 2010).

Ainuüksi sõnade tähenduse mõistmisest aga ei piisa, et tekstist tervikuna aru saada. Sõnade produtseerimises mängivad ühtmoodi olulist rolli nii semantika, süntaks kui ka sõnavara, kusjuures keeleõppes toimivad need tasandid järjestikkuse põhimõttel ehk kõigepealt tehakse keeleõppes tööd sõna tähendusega (semantiline tasand), seejärel lauseõpetusega (süntaks) ja viimasena sõnavara avardamisega (leksikaalne tasand) (Bock 1982, Butterworth 1989, Caramazza 1997). Et tekstist aru saada, on oluline mõista sõna lause tasandil ja tajuda lauset kui tervikut. Samuti peab lugeja mõistma, kuidas laused on omavahel seotud ja kuidas nad moodustavad tervikteksti (Karlep 2003).

2.1.1. Elteadmiste tähtsus teksti mõistmisel ja uue sõnavara omandamisel

Begeny ja Silber (2006) leidsid oma uurimuses, et paljude algklassiõpilaste jaoks tekitab teksti mõistmine raskusi. Ka teised uurijad (Masterson, Apel 2006, Schroeder 2011) on märkinud, et teksti sisu mõistmine on lugeja jaoks keeruline protsess, sest tekstides esinevad sõnad on sageli erineva raskusastmega ning neid võidakse mõista erinevalt. Nii võib lihtsamate sõnade omandamisele kuluda mitu korda vähem aega kui keerukamate sõnade õppimisele. Tekkida võib olukord, kus laps loeb justkui lünkteksti: ta mõistab tuttavaid ja kiiremini õpitavaid sõnu, kuid raskemad sõnad jäävad talle arusaamatuks. Varem omandatud sõnavaral ja elteadmistel on seega tähtis ülesanne (Broek, Espin 2012, Soalt 2005). Elteadmised aitavad luua seoseid uue ja tuttava info vahel, lihtsustades võõra teksti mõistmist (Prado, Plourde 2011, Primor jt 2011, Schroeder 2011).

Eason jt (2012) tõdeavad, et lapsed, kellel on minimaalne kokkupuude uue sõnavaraga, ei mõista täielikult uue teksti tähendust. Tundmatute sõnade mõistmise juures on tähtis ka see, et sõna, mida õppija kuuleb, oleks tema jaoks tuttava kõlaga (fonoloogiline teadlikkus) (Begeny, Silber 2006, Pečjak jt 2011). Samuti aitab tähenduse mõistmisele kaasa see, kui õpitud sõnu asetatakse erinevasse konteksti (Cruse 2004). Nii saab sõnade tähendus kiiremini selgeks võrreldes sellega, kui sõnu kasutatakse vaid ühes kontekstis.

Sama oluline kui tundmatute sõnade märkamine tekstis on nende esitamise viis (Hall, Sabey 2007). Uute sõnade selgitus ja kasutus peavad olema piisavalt

sagedased (8–10 tähenduslikku väljendit uue sõnaga), külluslikud (sõnale leitakse sünonüüme, paronüüme ja antonüüme) ning avarad (sõna kasutatakse erinevas kontekstis). Sõnade täpse vaste teadmine lisab lastele enesekindlust ja huvi lugemispala vastu (Wise jt 2010). See omakorda seostub eduelamusega, mis innustab edasi lugema ning uurima. Bray ja Barron (2003) juhivad aga tähelepanu sellele, et laste jaoks huvitavate tekstide leidmine võib osutada küllaltki keeruliseks.

2.1.2. Teksti mõistmise arendamine põhikoolis

Lugemise peamine eesmärk on arendada teksti mõistmise oskust (Primor jt 2011). Lugeja ülesanne on anda sõnadele ja tekstile tähendus vastavalt kogemustele ja omandatud sõnavara mahule (Puksand jt 2010). Nii saab arendada ka abstraktsemat mõtlemist ja keerukamate seoste mõistmist.

Põhikooli esimeses kooliastmes (1.–3. klassis) on eesti keel integreeritud õppeaine, mis hõlmab nii emakeelt kui ka kirjandust. Peamine eesmärk on lugemistehnikate arendamine ja õpitud tekstide mõistmine. Tekstid esimeses kooliastmes on lihtsad ja selgelt sõnastatud (Uibu, Voltein 2010). 4.–6. klassis on eesti keel ja kirjandus eraldi õppeained, kus olulisel kohal on info integreerimise, analüüsi- ja järeldamisoskuse arendamine, mida aitab saavutada keerukamate tekstide lugemine (Põhikooli riiklik õppekava 2010). Tekstid muutuvad tähendusrikkamaks, laste sõnavara avardub. Kuigi teksti mõistmise eesmärgid esimeses ja teises kooliastmes on mõnevõrra erinevad, on need omavahel hierarhilises seoses (Zhang jt 2008). Seega kui laps ei ole esimeses kooliastmes omandanud elementaarset lugemisoskust, siis ei saa temalt eeldada ka abstraktsema teksti sisu mõtestamist – tal puuduvad selleks vajalikud eeluskused ja ettevalmistus.

2.2. Ilukirjandusliku teksti mõistmine

Tekstiliigil võib olla määrav tähtsus, kui hästi laps mõistab teksti sisu (Eason jt 2012). Kõige üldisemalt võib tekstid, millega õpilased põhikooli esimeses ja teises kooliastmes kokku puutuvad, jaotada teabe- ehk õppekirjanduseks ja ilukirjanduseks. Ilukirjanduslikud tekstid on hea võimalus õpetada lapsele juba varases eas teksti sidusust ja loogilisust, mis aitavad mõista loetu sisu (Chun 2009). Kui tekst on kirjutatud lihtminevikus, jutustavas vormis, kasutatud on igapäevaseid väljendeid ja tegelaskujusid, hõlbustab see teksti mõistmist. Kasutatakse ka ilukirjanduslikke väljamõeldisi, mis on lastele tuttavad. Ilukirjanduslik tekst pakub sageli elamust, milles võib peituda ka kasvatuslik moment. Sidus tekst võib sisaldada uusi sõnu ja väljendeid, mis kuuluvad konteksti ning mida on lastel seetõttu kergem mõista ja omandada.

Põhikooli teises kooliastmes muutuvad ilukirjanduslikud tekstid mahukamaks. Oluline on arendada lapse fantaasiat ja loogilist mõtlemist (Soalt 2005), samuti oskust kirjeldada ja analüüsida sündmusi ja tegelaskujusid, liigendada teksti ning eristada peamõtet. Lisaks seostatakse olemasolevad teadmised uutega ja kujundatakse arusaamu (Block, Pressley 2002).

Üks probleeme ilukirjandusliku tekstiga töötamisel on selle keekekasutus (Müürsepp 1986). Autor peab tabama lugeja sõnavara mahtu ja teksti mõistmise

võimet. Sihtrühmaks olevad lapsed peavad suutma loetut haarata, kuid tekst peab neile ka midagi uut pakkuma. Mõningatel juhtudel on vaja kasutada varasemaid teadmisi, et täita lünki ja sõnastada tervikmõte (Eason jt 2012, Müürsepp 1986).

2.3. Soolised erinevused teksti mõistmises

Poisid ja tüdrukud erinevad teksti mõistmise poolest üksteisest. Põhikoolis on tüdrukute vaimne areng poistest kiirem (Logan, Johnston 2010, Ning jt 2010, Prado, Plourde 2011), see mõjutab muu hulgas nende võimet tekstist aru saada (Bray, Barron 2003). Rätty jt (2004) leidsid, et humanitaarainetes, eelkõige keeleõppes, edestavad tüdrukud algklassides poisse. Kelley ja Deckeri (2009) sõnul peavad tüdrukud alates teisest kooliastmest lugemisoskust vajalikumaks kui poisid ning loevad rohkem koolivälisest kirjandusest. See omakorda avaldab positiivset mõju sellele, kuidas tüdrukud teksti mõistavad (Puksand jt 2010). Ka rahvusvahelise PISA funktsionaalse lugemise testidest on selgunud, et tüdrukud saavad teksti mõistmise ülesannetes paremaid tulemusi kui poisid – vahe on võrdsustatav isegi terve õppeaastaga (OECD 2009, 2013).

Kuigi tüdrukute lugemisoskuse areng on kiirem, on võimalik poisse tüdrukutega samale tasemele viia, kuid see nõuab järjepidevat tööd (Bruemmer 2006, Corthals 2010). Selleks on vaja õpetaja teadlikku suunamist, kannatlikkust ja aega, et loetud pala põhjalikult analüüsida ja ümber jutustada ning moodustada teksti kohta küsimusi. Oluline on suunata poisse süvenema loetusse, mitte anda neile tavapäraseid ülesandeid. Selline õpetamisviis arendab mõtlemist ja analüüsivõimet (Kratwohl 2002).

3. Uurimuse eesmärk ja hüpoteesid

Sõnatähenduse tundmine on tähtis komponent teksti kui terviku mõistmisel. Uurimuse eesmärk oli välja selgitada muutused õpilaste sõnatähenduse tundmises Eesti üldhariduskoolide 3., 4. ja 5. klassi poiste ja tüdrukute sõnatähenduse tundmises ja teksti mõistmises. Uurimuses püstitati viis hüpoteesi:

- 1) õpilased, kes tunnevad sõnade tähendust paremini 3. klassis, teavad nende tähendust paremini ka 4. ja 5. klassis;
- 2) 3.–5. klassis tunnevad tüdrukud sõnatähendust märksa paremini kui poisid (Ning jt 2010, Prado, Plourde 2011);
- 3) 3.–5. klassi sõnavaraülesandes esinenud sõnade tundmine paraneb aastatega ning tüdrukud edestavad siin poisse;
- 4) teksti mõistmise ülesanded, mis eeldavad tekstist info äratundmise oskust, on jõukohased nii poistele kui ka tüdrukutele; seevastu soolisi erinevusi eeldati ülesandes, mis nõuab head info integreerimise oskust (Saxton 2010);
- 5) oletati, et õpilaste sõnatähenduse tundmine seostub tekstist arusaamisega – erinevatel kognitiivsetel tasemetel.

4. Meetod

4.1. Valim ja protseduur

Pikiuurimuses osalesid põhikooli esimese ja teise kooliastme õpilased (3.–5. klass) 30 Eesti üldhariduskoolist. Valimi moodustamisel arvestati kooli asukohta ja piirkonda (maa- ja linnakool), kooli tüüpi (põhikool ja gümnaasium) ning klassi suurust (väikesed ja suured klassid, liitklassid). Kõik õpilased õppisid põhikooli riikliku õppekava (2007) järgi. Lapsi testiti kolme järjestikuse õppeaasta alguses. 3. klassis osales uurimuses 798 õpilast, 4. ja 5. klassis mõlemas 800 õpilast. Emakeeletestid saadeti koolidesse kinnistes ümbrikes, testi täideti emakeeleõpetajate juhendamisel eesti keele tunni ajal. Testi jaoks oli ette nähtud üks ainetund (umbes 45 minutit).

Kuna artikli autorite eesmärk oli välja selgitada pikaajalised muutused õpilaste sõnatähenduse tundmises ja teksti mõistmises, analüüsitakse järgnevalt ainult nende õpilaste vastuseid, kes sooritasid emakeeletesti kõigil kolmel aastal. Selliseid õpilasi oli kokku 508, neist 241 (47,4%) poisid ja 267 (52,6%) tüdrukud. Õpilaste keskmine vanus 3. klassis oli 9,08 (standardhälve 0,37); 4. klassis 10,00 aastat (0,36); ja 5. klassis 10,98 aastat (0,35).

4.2. Mõõtevahendid

Teksti mõistmist ja sõnatähenduse tundmist hinnati kolme ülesandega: 1) teksti mõistmise lihtsam ülesanne, 2) teksti mõistmise keerukam ülesanne, 3) sõnavaraülesanne. Testi väljatöötamisel võeti aluseks Bloomi hierarhilise taksonoomia edasiarendus (Krathwohl 2002), emakeele ainekava nõudmised ja riiklikud emakeele tasemetööd (Uibu jt 2010, Uibu, Tropp 2013). Kolm emakeeletesti ülesannet koostati teksti põhjal: 3. klassis oli selleks Miloš Macoureki “Kaelkirjak” (184 sõna), 4. klassis Ivan Krõlovi valmi “Haug” kohandatud versioon (161 sõna), 5. klassis sama valmi täiendatud versioon (185 sõna). Ülesannete tüübid ja tööjuhendid olid kolmel aastal samad. Õpilaste vastused kodeeriti dihhotoomselt: õige 1-ga, vale või märkimata vastus 0-ga.

1. Teksti mõistmise lihtsamal ülesandes mõõdeti tekstist info leidmise ja ära-tundmise oskust, mis nõudis eelkõige tähelepanu ja meelespidamist. Õpilasel tuli esmalt läbi lugeda lugemispala ja seejärel otsustada, kas etteantud laused on teksti järgi õiged või väärad. Lause, mis ei vastanud teksti sisule, tuli läbi kriipsutada. 3. klassi ülesandes oli kaheksa lauset (nt *Kaelkirjak ei osanud aritmeetikat. Ta nuttis palju* [tekstis: *ohtralt*]), 4. klassis samuti kaheksa lauset (nt *Haugi süüdistati varguses. Rebane oli prokuröriks*). 5. klassis kasutati valmi täiendatud versiooni, mis oli natuke pikem, ja ülesanne sisaldas 11 lauset (nt *Kohut mõistma kutsuti ka kaks isast lammast. Kohtunikud olid targad ja õiglased*). Ülesande sisereleiaablus (Cronbachi alfa) oli 3. klassis 0,84, 4. klassis 0,54 ja 5. klassis 0,87.

2. Teksti mõistmise keerukamas ülesandes mõõdeti oskust integreerida ümbersõnastatud ja kontekstis sisalduvat infot. Ülesande lahendamise edukus oli seotud analüüsi-, järeldamis- ja tõlgendamisoskusega. Õpilane pidi ühendama kaks osalauset, mis kuulusid sisult kokku (nt 4. ja 5. klassis pealause: *Haug anti kohtu alla, sest ...*; kõrvallause: a) *ta tahtis rahus surra*; b) *ta ei lasknud teistel*

elada; c) *haug andis rebasele altkäemaksu* [tekstis: *pistist*]). Seejuures ühildusid kõik kõrvallused pealausega nii morfoloogiliselt kui ka süntaktiliselt. Kokku oli 3. ja 4. klassi ülesandes kuus sisult kokkukuuluvat lauset, 5. klassis seitse lauset. Ülesande sisereliaablus oli 3. klassis 0,93, 4. klassis 0,83 ja 5. klassis 0,87.

3. Sõnavaraülesandega hinnati õpilaste sõnatähenduse tundmise oskust. Ülesandesse valiti sõnad, mis olid teksti mõistmise seisukohalt olulised, samuti arvestati valikul sõnaliigiga (nt konkreetne ja abstraktne nimisõna, omadussõna, tegusõna, määrsõna). Õpilase ülesanne oli leida tekstis esinenud sõnadele etteantud loendist õige tähendusega sünonüüm. 3. klassi ülesanne sisaldas kolme sõna (*aritmeetika, kohmakas, ohtralt*), 4. klassi testis lisati neile veel kolm sõna 4. klassi tekstist (*ränk, nõõrima, prokurör*). 5. klassi ülesandes küsiti peale 3. ja 4. klassi sõnade ka sõnu 5. klassi tekstist (*pistis, noobel, valm, võlts ja kronu*). Sõnu oli kokku kümme. Sõnad, mille tähendust kontrolliti kõigil kolmel aastal, olid *kohmakas* ja *ohtralt* (edaspidi: osaülesanne). Tervikülesande sisereliaablus oli 3. klassis 0,62; 4. klassis 0,90 ja 5. klassis 0,86.

4.3. Andmeanalüüsi meetodid

Andmeid analüüsiti statistikapaketiga SPSS Statistics (versioon 20.0). Analüüsimisel kasutati kahesuguseid skooore. Esiteks arvutati ülesannete keskmised väärtused, kus iga õpilase keskmine väärtus märkis õigete vastuste osa kõigist võimalikest vastustest. Teiseks arvutati iga ülesande summaskoor, mida kasutati erinevuste tuvastamiseks õpilaste soo põhjal 3., 4. ja 5. klassis.

Muutuste väljaselgitamiseks õpilaste sõnatähenduse tundmises 3.–5. klassis kasutati ülesannete keskmisi väärtusi ja tehti kordusmõõtmiste dispersioonanalüüs (ingl *repeated measure ANOVA*). Detailseid erinevusi rühmade vahel mõõdeti *post hoc*-analüüsi Bonferroni testiga. Poiste ja tüdrukute erinevusi kolme aasta sõnavaraülesannetes esinenud sõnade tundmises ja teksti mõistmises hinnati ühefaktorilise dispersioonanalüüsiga (ingl *one-way ANOVA*), kasutades ülesannete summaskooore. Statistilise olulisuse väljendamiseks rühmade vahel kasutati peale *p*-väärtuse ka mõjufaktori suurust η^2 , mille tõlgendamisel toetuti Coheni jt kriteeriumidele: $\eta^2 = 0,01–0,05$ nõrk mõju; $\eta^2 = 0,06–0,13$ keskmine mõju; $\eta^2 \geq 0,14$ suur mõju (Cohen jt 2007: 522).

Analüüsimaks õpilaste sõnatähenduse tundmise ja teksti mõistmise seoseid, kasutati Pearsoni lineaarkorrelatsiooni kordajat. Korrelatsiooni tugevuse interpreteerimisel olulisusnivool $p < 0,01$ arvestati järgmisi kriteeriume: $r > 0,65$ tugev seos; $0,35 < r < 0,65$ mõõdukas seos; $r < 0,35$ nõrk seos (Cohen jt 2007: 536–537).

5. Tulemused

Esmalt tehti kirjeldavad analüüsid, et selgitada muutusi õpilaste sõnatähenduse tundmises. Tervikülesande võrdlemisel kolmel aastal kasutati ülesannete keskmisi skooore. Summaskooride alusel võrreldi kolmel aastal ülesannetes esinenud sõnade tundmist. Tulemused kogu valimi ja poiste-tüdrukute kohta eraldi on esitatud tabelis 1.

Tabel 1. Sõnavaraülesannete keskmised väärtused (M) ja standardhälbed (SD)

Ülesanne	Kogu valim (N = 508)		Poisid (N = 241)		Tüdrukud (N = 267)	
	M	SD	M	SD	M	SD
Tervikülesanne*						
3. klass	0,6	0,33	0,57	0,34	0,62	0,32
4. klass	0,82	0,27	0,76	0,29	0,87	0,24
5. klass	0,56	0,22	0,53	0,23	0,59	0,21
Osaülesanne**						
max = 2						
3. klass	0,92	0,65	0,86	0,65	0,97	0,65
4. klass	1,55	0,68	1,42	0,73	1,68	0,6
5. klass	1,61	0,61	1,47	0,65	1,73	0,54

Märkus. * keskmine skoor, ** summaskoor

Muutust õpilaste sõnatähenduse tundmises analüüsiti 3 (ülesanne) × 2 (sugu) kordusmõõtmiste dispersioonanalüüsiga. Selgus, et ajal oli statistiliselt oluline mõju tervikülesande lahendusele: $F(2, 506) = 192,36, p < 0,001, \eta^2 = 0,28$. *Post hoc*-analüüsist selgus, et õpilaste kolme aasta tulemuste vahel oli statistiliselt oluline erinevus ($p < 0,05$), seejuures oli kogu valimi puhul sõnatähenduse tundmine kõrgeim 4. klassis ja madalaim 5. klassis (vt tabel 1). Hüpotees, et õpilaste sõnatähenduse tundmine paraneb ajaga, leidis osaliselt kinnitust.

Õpilaste sool oli samuti oluline mõju tervikülesande sooritusele: $F(1, 506) = 17,33, p < 0,001, \eta^2 = 0,03$. *Post hoc*-analüüs näitas, et tüdrukute tulemused olid statistiliselt oluliselt kõrgemad kui poistel kõigis klassides ($p < 0,001$). Ometi aja ja soo koosmõju tervikülesande sooritusele ei leitud, mis tähendab, et võrreldes poistega oli tüdrukute sõnatähenduse tundmine märksa parem, kuid erinevus poiste ja tüdrukute vahel ajaga ei suurenenud.

Järgnevalt uuriti poiste ja tüdrukute erinevusi kolme aasta sõnavaraülesandes küsitud sõnade tundmises. Ühefaktoriline dispersioonanalüüs näitas, et 3. klassi tekstis esinenud sõnade (*kohmakas* ja *ohtralt*) tähenduse tundmine oli tüdruku- tel statistiliselt oluliselt parem kui poistel 4. klassis ($F(1, 505) = 19,57, p < 0,001, \eta^2 = 0,037$) ja ka 5. klassis ($F(1, 506) = 12,82, p < 0,001, \eta^2 = 0,045$). Olulist erinevust poiste ja tüdrukute vahel ei ilmnenu 3. klassis ($p = 0,06$).

Et hinnata erineva raskusastmega teksti mõistmise ülesannete sooritust poiste ja tüdrukute võrdluses, kasutati ülesannete summaskoore. Tabelis 2 on esitatud andmed kolme aasta ülesannete kohta.

Ühefaktorilisest dispersioonanalüüsist selgus, et teksti mõistmise lihtsamas ülesandes olid tüdrukute tulemused statistiliselt oluliselt paremad kui poistel 4. klassis ($F(1, 506) = 7,73, p = 0,006, \eta^2 = 0,015$) ja 5. klassis ($F(1, 506) = 12,82, p < 0,001, \eta^2 = 0,025$). Teksti mõistmise keerukamas ülesandes edestasid tüdrukud poisse kõigil kolmel aastal: 3. klassis $F(1, 501) = 6,34, p = 0,012, \eta^2 = 0,013$; 4. klassis $F(1, 506) = 5,02, p = 0,026, \eta^2 = 0,01$ ja 5. klassis $F(1, 506) = 11,40, p = 0,001, \eta^2 = 0,022$. Selline tulemus oli kooskõlas püstitatud hüpoteesiga.

Kasutades Pearsoni lineaarkorrelatsiooni kordajat, leiti seosed õpilaste sõnatähenduse tundmise ja teksti mõistmise ülesannete keskmiste tulemuste vahel (vt tabel 3).

Tabel 2. Teksti mõistmise ülesannete kirjeldavad näitajad: keskmised väärtused (M) ja standardhälbed (SD)

Ülesanne	Poisid (N = 258)		Tüdrukud (N = 286)	
	M	SD	M	SD
Teksti mõistmise lihtsam ülesanne				
3. klass (max = 8)	6,91	1,40	7,12	1,36
4. klass (max = 8)	6,40	1,34	6,74	1,40
5. klass (max = 11)	9,05	2,20	9,64	1,48
Teksti mõistmise keerukam ülesanne				
3. klass (max = 6)	4,85	1,59	5,18	1,32
4. klass (max = 6)	4,07	1,73	4,41	1,63
5. klass (max = 7)	5,54	1,75	6,03	1,53

Tabel 3. Korrelatsioonid sõnatähenduse ja teksti mõistmise vahel

Ülesanne		3. klass			4. klass			5. klass		
		1	2	3	4	5	6	7	8	9
3. klass	1. Sõnatähenduse tundmine	-								
	2. Teksti mõistmise lihtsam ülesanne	0,2	-							
	3. Teksti mõistmise keerukam ülesanne	0,21	0,32	-						
4. klass	4. Sõnatähenduse tundmine	0,31	0,24	0,4	-					
	5. Teksti mõistmise lihtsam ülesanne	0,18	0,24	0,33	0,39	-				
	6. Teksti mõistmise keerukam ülesanne	0,16	0,21	0,29	0,41	0,33	-			
5. klass	7. Sõnatähenduse tundmine	0,29	0,24	0,31	0,48	0,38	0,36	-		
	8. Teksti mõistmise lihtsam ülesanne	0,22	0,17	0,31	0,33	0,38	0,24	0,41	-	
	9. Teksti mõistmise keerukam ülesanne	0,24	0,24	0,44	0,38	0,36	0,37	0,42	0,43	-

Märkus. Tabelis esitatud väärtused on olulisusnivool $p < 0,001$

Statistiliselt olulised seosed ilmnemise kõigi aastate sõnatähenduse tundmise ja teksti mõistmise ülesannete vahel. Järgnevalt esitatakse neist vaid tähtsamad. Tugevaim seos avaldus 4. ja 5. klassi sõnatähenduse tundmise vahel ($r = 0,48$). Mõnevõrra suurem korrelatsioon ($r > 0,40$) leiti 5. klassis sõnatähenduse ja mõlema teksti mõistmise ülesande vahel ning 4. klassis sõnatähenduse ja teksti mõistmise keerukama ülesande vahel. Lisaks seostus 4. klassi sõnatähenduse tundmine mõõdukalt ($r > 0,35$) 3. klassi teksti mõistmise keerukama ja 4. klassi lihtsama ülesandega.

Sõnatähenduse tundmine 5. klassis korreleerus mõõdukalt mõlema 4. klassi teksti mõistmise ülesandega.

Mõõdukad seosed avaldusid ka erinevate aastate teksti mõistmise ülesannete vahel. Nii korreleerus 5. klassi teksti mõistmise keerukam ülesanne sama ülesandetiübiga 3. ja 4. klassis. Lisaks leiti nii 4. klassis kui ka 5. klassis olulised seosed mõlema teksti mõistmise ülesande vahel. Seejuures oli tugevaim seos 5. klassis ($r = 0,48$). Mõõdukalt korreleerusid ka 4. ja 5. klassi teksti mõistmise lihtsamad ülesanded (vt tabel 3).

6. Arutelu ja järeldused

Õpilaste üks peamisi ülesandeid põhikooli esimeses ja teises kooliastmes on omandada hea teksti mõistmise oskus (Lerikkanen 2007, Pandis 2008). Käsitletavas pikiuurimuses analüüsiti muutusi 3.–5. klassi õpilaste sõnatähenduse tundmises ja teksti mõistmises ning hinnati nende seoseid. Nii nagu varasemates uurimustes (nt Rätty jt 2004, Bray, Barron 2003), ilmnisid ka siin kõigis klassides erinevused poiste ja tüdrukute sõnatähenduse tundmises ja teksti mõistmise keerukamas ülesandes. Samuti edestasid 4. ja 5. klassis tüdrukud poisse nende sõnade tundmises, mis esinesid 3. klassi sõnavaraülesandes. Uurimusest selgus, et õpilased, kes tundsid paremini sõnade tähendust, said tekstist paremini aru. Eriti oluliseks osutus sõnatähenduse tundmine teksti mõistmise keerukama ülesande lahendamisel.

Esmalt uuriti õpilaste sõnatähenduse tundmist. Oletus, et õpilaste sõnatähenduse tundmine paraneb aastatega, leidis osaliselt kinnitust. Õpilaste sõnavaraülesande tulemused olid kõrgeimad 4. klassis, kuid langesid 5. klassis. Üks põhjuseid võib olla see, et sõnaloend, kust õpilane pidi 5. klassis valima õige tähendusega vaste, sisaldas kõige rohkem valikuid. On märgitud, et mitme valikuga ülesannetes võib valikute arv oluliselt mõjutada sooritust, sest valikute suurenedes väheneb tõenäosus vastata juhuslikult õigesti (Rodriguez 2005). Teise põhjusena on nimetatud sõnade tundmist sõnaliigiti (Saxton 2010). Kõige lihtsamini omandatakse ja teatakse konkreetse tähendusega nimisõnu, keerulisem on tegu- ja omadussõnade ning abstraktse tähendusega nimisõnade õppimine (Hirsh-Pasek, Golinkoff 2006, Imai jt 2005). Mitmes ülesandes küsitud sõnade (nt *prokurör*, *kohmakas*) tähendus võis olla õpilastele võõram, kuigi kõik sõnad pärinesid tekstist, mille laps oli läbi lugenud. Lisaks oli võimalik konteksti järgi sõnade tähendust lahti mõtestada. Kolmas põhjus võis olla õpetaja tegevus ja õpetamisviis. Praeguses uurimuses kontrolliti laste sõnatähenduse tundmist kolmel sügisel, kuid õpetaja tegevusi mõõtmiskatsete vahel ei uuritud. Sõnade tähenduse kinnistamisel on aga tähtis, et sõna kasutatakse piisavalt sageli ja erinevas kontekstis (Cruse 2004, Wise jt 2010). Seepärast on vaja, et õpetaja kordaks õpilastega sõnu. Varasemate uurimuste järgi avaldab õpetaja tegevus mõju õpilaste emakeele tulemustele (Uibu jt 2010).

Analüüsides erinevusi poiste ja tüdrukute sõnatähenduse tundmises, leiti, et tüdrukute sõnavaralised teadmised olid kõigil kolmel aastal paremad kui poisitel. Sellised tulemused on kooskõlas nii rahvusvahelise PISA testi kui ka riiklike tasemetööde tulemustega (Henno jt 2007, OECD 2009, 2013, Puksand jt 2010). Põhjus võib olla see, et tüdrukud erinevad poistest verbaalsetelt võimetelt, mida on peetud oluliseks teguriks sõnavara omandamisel ja sõnade tähenduse eristamisel

(Saxton 2010). Samas ei leitud praeguses uurimuses aja ja soo koosmõju õpilaste sõnatähenduse tundmisele. Järelikult, tüdrukute ja poiste sõnavarateadmistes ilmnevad erinevused, kuid aastatega need ei suurene.

Varem õpitud sõnade tundmist kontrolliti kahe sõnaga (*kohmakas* ja *ohtralt*), mis esinesid sõnavaraülesandes kolmel aastal. Selgus, et 3. klassis ei olnud poiste ja tüdrukute teadmistes erinevusi, kuid 4. ja 5. klassis oli tüdrukute sooritus oluliselt parem kui poistel. Et mõista sõnade *kohmakas* ja *ohtralt* tähendust, pidi laps aru saama lause sisust. Arvestades, et põhikoolis on tüdrukute vaimne areng poiste omast kiirem (Logan, Johnston 2010, Ning jt 2010), on mõistetav, miks varem tekstis esinenud sõnade tundmine valmistab poistele hiljem raskusi. Lisaks, sõnade abstraktne tähendus võis olla takistuseks nende omandamisel (Hall, Sabey 2007, Saxton 2010).

Toetudes Krathwohli (2002) seisukohtadele, oletati, et teksti mõistmise keerukama ülesandega, mis nõudis oskust integreerida ümbersõnastatud infot, tulevad tüdrukud paremini toime kui poisid. Teises teksti mõistmise ülesandes, mis eeldas info äratundmise oskust, soolisi erinevusi ei oodatud. Ootuspäraselt ilmneski, et kui 3. klassis oli teksti mõistmise lihtsam ülesanne jõukohane nii poistele kui ka tüdrukutele, siis keerukama ülesandega, kus laps pidi ühendama kaks osalauset, arvestades kontekstis sisalduvat infot, tulid tüdrukud poistest paremini toime kõigil kolmel aastal. Varasemates uurimustes on leitud, et teksti mõistmist hõlbustab erinevate süntaktiliste ja grammatiliste konstruktsioonide tundmine (Cain, Oakhill 2007) ning võime infot töödelda ja ekstrapoleerida (Krathwohl 2002). Näib, et esimeses ja teises kooliastmes on tüdrukutel need oskused paremini arenenud.

Viimasena uuriti seoseid õpilaste sõnatähenduse tundmise ja teksti mõistmise vahel. Ilmnes, et õpilased, kes tundsid 4. klassis paremini sõnade tähendust, lahendasid 5. klassis paremini nii sõnatähenduse kui ka teksti mõistmise ülesandeid. Olulised seosed leiti ka 4. klassi sõnatähenduse ja 3. klassi teksti mõistmise keerukama ülesande vahel. Niisiis, lapsed, kes said 3. klassis teksti mõistmise keerukama ülesandega paremini hakkama, näitasid järgmisel aastal paremaid teadmisi sõnatundmises. Lisaks avaldus oluline seos 4. klassi sõnatähenduse ja teksti mõistmise lihtsama ülesande vahel. Samuti tundsid 5. klassis paremini sõnade tähendust õpilased, kellel teksti mõistmise mõlema ülesande sooritus oli 4. klassis parem. Sellest võib järeldada, et sõnavaralised teadmised on määrav tegur teksti mõistmisel. Sõnade tähendust tundmata ei ole võimalik mõista lausete tähendust ning teksti kui terviku sisu (Karlep 2003).

Olulised seosed leiti ka erinevate aastate teksti mõistmise lihtsama ja keerukama ülesande vahel. Õpilased, kes said 3. klassis teksti mõistmise keerukama ülesandega paremini hakkama, lahendasid selle paremini ka 4. ja 5. klassis. Teksti mõistmise hierarhilist iseloomu on märkinud Zhang jt (2008). Teises kooliastmes saavutavad teksti mõistmises paremaid tulemusi õpilased, kellel on need oskused välja arenenud esimeses kooliastmes.

Analüüsides teksti mõistmise lihtsama ja keerukama ülesande seoseid aastate kaupa, selgus, et 5. klassis saavutasid paremad tulemused teksti mõistmise keerukamas ülesandes õpilased, kes said paremini hakkama ka lihtsama ülesandega. Sarnane seos teksti mõistmise ülesannete vahel leiti ka 4. klassis. On selge, et kui õpilane mõistab teksti sisu, siis ei valmista talle raskusi tekstist info äratundmine ega ka ümbersõnastatud ja kontekstist tuleneva info seostamine (Krathwohl 2002).

Praegusel uurimusel olid mõned piirangud, mida tuleks tulemuste üldistamisel arvesse võtta. Esiteks, 3. klassi testis ning 4. ja 5. klassi testis kasutati erinevaid lugemispalasisid. Teiseks, ülesannete tüübid ja tööjuhendid olid kõigil aastatel samad, kuid sõnade arv sõnavaraülesannetes oli erinev. On märgitud, et mitme valikuga ülesannetes võib valikute arv mõjutada ülesande sooritust (Rodriguez 2005). Peale valikute arvu tuleks sõnavaraülesannete koostamisel arvestada ka seda, et sõnatähendus on kontekstitundlik. Kolmandaks, uurimuses ei hinnatud selliseid sõnatähenduse tundmist ja teksti mõistmist mõjutavaid tegureid nagu õpilase lugemise täpsus ja motivatsioon ning õpetaja tegevused jmt. Suurem tunnuste variatiivsus võimaldanuks saada mitmekesisema pildi õpilaste sõnatähenduse tundmist ja teksti mõistmist mõjutavatest teguritest põhikoolis.

Ometi saab uurimuse kestuse (kolm aastat) ja esindusliku valimi põhjal teha järeldusi esimese ja teise kooliastme õpilaste sõnatähenduse tundmise ja teksti mõistmise kohta. Esiteks, teksti mõistmine ja sõnatähenduse tundmine on omavahel seotud ning sõnade tähendus kontekstitundlik. Teiseks, tüdrukud edestavad poisse nii sõnatähenduse tundmises kui ka teksti mõistmises. Kolmandaks, tüdrukud teavad poistest paremini varem õpitud sõnade tähendusi. Järelikult tuleks poiste sõnavara arengu ja tekstimõistmise oskuse toetamiseks arvestada nende keelelist arengut ja pakkuda neile võimete ja huvide kohaseid tekste.

Viidatud kirjandus

- Begeny, John C.; Silber, Jennifer M. 2006. An examination of group-based treatment packages for increasing elementary-aged students' reading fluency. – *Psychology in the Schools*, 43 (2), 183–195. <http://dx.doi.org/10.1002/pits.20138>
- Block, Cathy C.; Pressley, Michael 2002. What comprehension instruction could be. – M. Pressley, C. C. Block (Eds.). *Comprehension Instruction*. New York: Guilford, 383–392.
- Bloem, Ineke; La Heij, Wido 2003. Semantic facilitation and semantic interference in word translation: Implications for models of lexical access in language production. – *Journal of Memory and Language*, 48 (3), 468–488. [http://dx.doi.org/10.1016/S0749-596X\(02\)00503-X](http://dx.doi.org/10.1016/S0749-596X(02)00503-X)
- Bock, Kathryn J. 1982. Toward a cognitive psychology of syntax: Information processing contributions to sentence formulations. – *Psychological Review*, 89 (1), 1–47. <http://dx.doi.org/10.1037/0033-295X.89.1.1>
- Bray, Gayle Babbitt; Barron, Sheila 2003. Assessing reading comprehension: The effects of text-based interest, gender, and ability. – *Educational Assessment*, 9 (3–4), 107–128.
- Broek, Paul van den; Espin, Christine A. 2012. Connecting cognitive theory and assessment: Measuring individual differences in reading comprehension. – *School Psychology Review*, 41 (3), 315–325.
- Bruemmer, Katherine Hlavach 2006. Teaching to biological gender preferences in an all-boys Catholic school. – *English Journal*, 95 (6), 31–41. <http://dx.doi.org/10.2307/30046625>
- Butterworth, Brian 1989. Lexical access in speech production. – W. Marslen (Ed.). *Wilson Lexical Representation and Process*. Cambridge, MA: MIT Press.
- Cain, Kate; Oakhill, Jane; Bryant, Peter 2004. Children's reading comprehension ability: Concurrent prediction by working memory, verbal ability, and component skills. – *Journal of Educational Psychology*, 96 (1), 31–42. <http://dx.doi.org/10.1037/0022-0663.96.1.31>
- Cain, Kate; Oakhill, Jane 2007. Reading comprehension difficulties: Correlates, causes, and consequences. – Kate Cain, Jane Oakhill (Eds.). *Children's Comprehension Problems*

- in Oral and Written Language: A Cognitive Perspective. New York, London: The Guilford Press, 41–76.
- Caramazza, Alfonso 1997. How many levels of processing are there in lexical access? – *Cognitive Neuropsychology*, 14 (1), 177–208. <http://dx.doi.org/10.1080/026432997381664>
- Chomsky, Noam 1965. *Aspects of the Theory of Syntax*. MIT Press.
- Chun, Christian W. 2009. Critical literacies and graphic novels for English-language learners: Teaching maus. – *Journal of Adolescent & Adult Literacy*, 53 (2), 144–153. <http://dx.doi.org/10.1598/JAAL.53.2.5>
- Cohen, Louis; Manion, Lawrence; Morrison, Keith 2007. *Research Methods in Education*. 6th ed. London, New York: Routledge.
- Corthals, Paul 2010. Nine- to twelve-year olds' metalinguistic awareness of homonymy. – *International Journal of Language & Communication Disorders*, 45 (1), 121–128. <http://dx.doi.org/10.3109/13682820902745446>
- Cruse, Alan D. 2004. *Meaning in Language: An Introduction to Semantics and Pragmatics*. Oxford: Oxford University Press.
- Eason, Sarah H.; Goldberg, Lindsay F.; Young, Katherine M.; Geist, Megan C.; Cutting, Laurie E. 2012. Reader–text interactions: How differential text and question types influence cognitive skills needed for reading comprehension. – *Journal of Educational Psychology*, 104 (3), 515–528. <http://dx.doi.org/10.1037/a0027182>
- Hall, Kendra M.; Sabey, Brenda L. 2007. Focus on the facts: Using informational texts effectively in early elementary classrooms. – *Early Childhood Education Journal*, 35 (3), 261–268. <http://dx.doi.org/10.1007/s10643-007-0187-2>
- Henno, Imbi; Tire, Gunda; Lepmann, Tiit; Reiska, Priit; Ehala, Martin 2007. Ülevaade rahvusvahelise õpilaste õpitulemuslikkuse hindamise programmi PISA 2006 tulemustest. [The Official Review of Students Results of PISA 2006.] http://uuringud.ekk.edu.ee/fileadmin/user_upload/documents/PISA_16pparuanne_041207.pdf (5.5.2013).
- Hirsh-Pasek, Kathy; Golinkoff, Roberta M. (Eds.) 2006. *Action Meets Words: How Children Learn Verbs*. Oxford: Oxford University Press. <http://dx.doi.org/10.1093/acprof:oso/9780195170009.001.0001>
- Häkkinen, Kaisa 2007. Keeleteaduse alused. [Basics of Linguistics.] Tallinn: AS Pakett.
- Imai, Mutsumi; Okada, Hiroyuki; Haryu, Etsuko 2005. Mapping novel nouns and verbs onto dynamic action events: Are verb meanings easier to learn than noun meanings for Japanese children? – *Child Development*, 76 (2), 340–355. http://dx.doi.org/10.1111/j.1467-8624.2005.00849_a.x
- Kabilan, Muhammad Kamarul; Seng, Mak Kem; Kee, Ooi Ae 2010. Reader-text transaction in text comprehension. – *Journal of Language Studies*, 10 (3), 127–142.
- Karlep, Karl 2003. *Kõnearendus. Emakeele abiõpe II*. [Tutoring Native Language II.] Tartu: Tartu Ülikooli Kirjastus.
- Kelley, Michelle J.; Decker, Emmeline O. 2009. The current state of motivation to read among middle school students. – *Reading Psychology*, 30 (5), 466–485. <http://dx.doi.org/10.1080/02702710902733535>
- Krathwohl, David R. 2002. A Revision of Bloom's Taxonomy: An Overview. – *Theory into Practice*, 41 (4), 212–218. http://dx.doi.org/10.1207/s15430421tip4104_2
- Lenneberg, Eric H. 1967. *Biological Foundations of Language*. Oxford: Wiley.
- Lerkkanen, Marja-Kristiina 2007. *Lugema õppimine ja õpetamine alus- ja algõpetuses*. [Teaching and Learning How to Read in Pre- and Primary-Education.] Tartu: Tartu Ülikooli Kirjastus.
- Logan, Sarah; Johnston, Rhona 2010. Investigating gender differences in reading. – *Educational Review*, 62 (2), 175–187. <http://dx.doi.org/10.1080/00131911003637006>

- Masterson, Julie J.; Apel, Kenn 2006. Effect of modality on spelling words varying in linguistic demands. – *Developmental Neuropsychology*, 29 (1), 261–277. http://dx.doi.org/10.1207/s15326942dn2901_13
- Mayberry, Rachel I. 1993. First-language acquisition after childhood differs from second-language acquisition: The case of American sign language. – *Journal of Speech and Hearing Research*, 36 (6), 1258–1270. <http://dx.doi.org/10.1044/jshr.3606.1258>
- Meisel, Jurgen M. 2008. Child second language acquisition or successive first language acquisition? – B. Haznedar, E. Gavruseva (Eds.). *Current Trends in Child Second Language Acquisition. A Generative Perspective*. Amsterdam: Benjamins, 55–80.
- Müürsepp, Mare 1986. Lasteraamatu keel kasvatusmõjurina. [The Language of Children's Books in the Impact of Education.] – L. Talts (toim.). *Algõpetuse aktuaalseid probleeme II. [Actual Problems in Pre-Education II.]* Tallinn: TPÜ Kirjastus, 99–111
- Ning, Hailiang; Dai, Xue; Zhang, Fachun 2010. On gender difference in English language and its causes. – *Asian Social Science*, 6 (2), 126–130.
- OECD 2009. *Creating effective teaching and learning environments: First results from TALIS*. Paris: OECD. <http://www.oecd.org/education/school/43023606.pdf> (14.4.2013).
- OECD 2013. *PISA 2012 Results: Ready to Learn – Students' Engagement, Drive and Self-Beliefs (Volume III)*. Pisa: OECD Publishing. <http://dx.doi.org/10.1787/9789264201170-en>
- Pandis, Meeli 2002. *Lugemisoskus ja sellega seotud tegurid. [Reading Ability and Related Factors.]* Magistritöö. Tallinn: TPÜ kasvatuseduste teaduskond.
- Pandis, Meeli (koost.) 2008. 3. klassi eesti keele ja vene keele tasemetöö analüüs. http://www.ekk.edu.ee/vvfiles/o/3_kl_eesti_ja_vene_keeles_tasemetoo_analyys_2008.pdf (30.4.2013).
- Pečjak, Sonja; Podlesek, Anja; Pirc, Tina 2011. Model of reading comprehension for 5th grade students. – *Studia Psychologica*, 53 (1), 53–67.
- PISA = The Programme for International Student Assessment. [Rahvusvaheline õpilaste õpitulemuslikkuse hindamisprogramm.] <http://www.oecd.org/pisa/> (14.2.2014).
- Prado, Ludivina; Plourde, Lee A. 2011. Increasing reading comprehension through the explicit teaching of reading strategies: Is there a difference among the genders? – *Reading Improvement*, 48 (1), 32–43.
- Primor, Liron; Pierce, Margaret E.; Katzir, Tami 2011. Predicting reading comprehension of narrative and expository texts among Hebrew-speaking readers with and without a reading disability. – *Annals of Dyslexia*, 61, 242–268. <http://dx.doi.org/10.1007/s11881-011-0059-8>
- Puksand, Helin; Henno, Imbi; Lepmann, Tiit 2010. *Pisa 2009 – Eesti tulemused. [Pisa 2009 – Estonian results.]* G. Tire (toim.). http://www.ekk.edu.ee/vvfiles/o/PISA_2009_Eesti.pdf (28.4.2013).
- Põhikooli ja gümnaasiumi riiklik õppekava 2007. Vabariigi Valitsuse 25. jaanuari 2002. a määrus nr 56. <https://www.riigiteataja.ee/akt/12888846> (20.3.2013).
- Põhikooli riiklik õppekava 2010. Vabariigi Valitsuse 28. jaanuari 2010. a määrus nr 14. <https://www.riigiteataja.ee/akt/13273133> (15.4.2013).
- Rodriguez, Michael C. 2005. Three options are optimal for multiple-choice items: A meta-analysis of 80 years of research. – *Educational Measurement: Issues and Practice*, 24 (2), 3–13. <http://dx.doi.org/10.1111/j.1745-3992.2005.00006.x>
- Räty, Hannu; Kasanen, Kati; Kiiskinen, Johanna; Nykky, Merja 2004. Learning intelligence – children's choices of the best pupils in the mother tongue and mathematics. – *Social Behavior and Personality*, 32 (3), 303–312. <http://dx.doi.org/10.2224/sbp.2004.32.3.303>
- Saxton, Matthew 2010. *Child Language. Acquisition and Development*. SAGE Publications.
- Schroeder, Sascha 2011. What readers have and do: Effects of students' verbal ability and reading time components on comprehension with and without text availability. – *Journal of Educational Psychology*, 103 (4), 877–896. <http://dx.doi.org/10.1037/a0023731>

- Soalt, Jennifer 2005. Bringing together fictional and informational texts to improve comprehension. – *The Reading Teacher*, 58 (7), 680–683. <http://dx.doi.org/10.1598/RT.58.7.8>
- Soodla, Piret; Vija, Maigi; Pajusalu, Renate 2013. Eesti ja soome sõnalugemistestide võrdlus. [Comparison of Estonian and Finnish word reading tests.] – *Eesti Rakenduslingvistika Ühingu aastaraamat*, 9, 279–296. <http://dx.doi.org/10.5128/ERYa9.18>
- Zhang, Lawrence Jun; Gu, Peter Yongqi; Hu, Guangwei 2008. A cognitive perspective on Singaporean primary school pupils' use of reading strategies in learning to read in English. – *British Journal of Educational Psychology*, 78 (2), 245–271. <http://dx.doi.org/10.1348/000709907X218179>
- Tomasello, Michael; Akhtar, Nameera; Dodson, Kelly; Rekau, Laura 1997. Differential productivity in young children's use of nouns and verbs. – *Journal of Child Language*, 24 (2), 373–387. <http://dx.doi.org/10.1017/S0305000997003085>
- Tulving, Endel; Madigan, Stephen A. 1970. Memory and verbal learning. – *Annual review of Psychology*, 21, 437–484. <http://dx.doi.org/10.1146/annurev.ps.21.020170.002253>
- Uibu, Krista; Kikas, Eve; Tropp, Kristiina 2010. Teaching practices, their dynamics, associations with self-reported knowledge and students' language achievement. – A. Toomela (Ed.). *Systemic Person-Oriented Study of Child Development in Early Primary School*. Frankfurt am Main: Peter Lang Verlag, 47–71.
- Uibu, Krista; Voltein, Elve 2010. Eesti keel. [Estonian language.] – E. Kikas (toim.). *Õppimine ja õpetamine esimeses ja teises kooliastmes*. [Learning and Teaching in the First and Second Stage of Primary School.] Tallinn: Eduko, 215–242
- Uibu, Krista; Tropp, Kristiina 2013. Longitudinal study of the development of language competence among the Estonian 4th and 5th grade students. – J. Mikk, P. Luik, M. Veisson (Eds.). *Change in Teaching and Learning*. Frankfurt am Main: Peter Lang Verlag, 45–63.
- Wise, Justin C.; Sevcik, Rose A.; Morris, Robin D.; Lovett, Maureen W.; Wolf, Maryanne; Kuhn, Melanie; Meisinger, Beth; Schwanenflugel, Paula 2010. The relationship between different measures of oral reading fluency and reading comprehension in second-grade students who evidence different oral reading fluency difficulties. – *Language, Speech, and Hearing Services in Schools*, 41, 340–348. [http://dx.doi.org/10.1044/0161-1461\(2009/08-0093\)](http://dx.doi.org/10.1044/0161-1461(2009/08-0093))

Krista Uibu (Tartu Ülikool) peamine uurimisvaldkond on laste keeleline areng ja õpetajate õpetamistegevuste mõju õpilaste keelepädevustele.
Salme 1a, 51013 Tartu, Estonia
krista.uibu@ut.ee

Maile Timm (Kõrveküla lasteaed), uurimisvaldkond on õpilaste sõnasemantika areng põhikoolis.
Lasteaia 4, 60512 Tartu vald, Estonia
[timmmaile@gmail.com](mailto:timmermaile@gmail.com)

WORD SEMANTICS AND TEXT COMPREHENSION IN THE FIRST AND SECOND STAGE OF THE PRIMARY SCHOOL

Krista Uibu, Maile Timm

University of Tartu, Preschool of Kõrveküla

Word recognising is important to understand the text. When the meaning of the word does not fix in the memory then text comprehension might become a problem for a student in all school stages. The aim of the present longitudinal study was to find out to what extent student word recognition and text comprehension changes in primary school's first and second stage. The article analyses student word recognition and two different level text comprehension tasks in Grades 3 to 5. 508 students (267 girls and 241 boys) from 30 schools participated in this study. The research revealed that students who knew the meaning of different words better also had better text comprehension skills. In gender comparison, girls outperformed boys in word recognition and higher-level text comprehension task, which requires the ability to integrate information in all three consecutive years. Girls also outperformed boys in previously learned word recognition. To support students' word semantic and text comprehension skills it is necessary to take into consideration students' developmental differences and offer them age-appropriate texts.

Keywords: word recognition, text comprehension, gender differences, longitudinal study, Estonian

KELLE EMOTSIOONILE ANDA TEKSTI ETTE LUGEDES HÄÄL? VAATLUS MITMEST PERSPEKTIIVIST

Ene Vainik

Ülevaade. Artikli eesmärgiks on esile tuua, milliste inimloomusest tulenevate tõsiasjadega peaks arvestama, kui on plaanis hakata sünteeskõnele emotsioone matkivat prosoodiat lisama. Sel eesmärgil on kokku koondatud lähenemised erinevatest inimest puudutavatest teadusharudest nagu neuroloogia, kirjandusteadus, sotsiaalpsühholoogia, filosoofia jne. Eri perspektiividena käsitletakse kuulaja, teksti peategelase, autori ja lugeja emotsioonide kujunemist. Seejärel kirjeldatakse, kuidas lisab vokaalseid emotsioone inimesest ettelugeja ning kuidas lähenetakse teksti ja emotsioonide vahekorrale afektirehkenduse valdkonnas. Artikli lõpuosas esitatakse visioon keskmise lugeja emotsionaalset reageeringut jälgendavast virtuaalsest *Emonculus*'est kui võimalikust tulevikuperspektiivist emotsionaalse prosoodia lisamisel sünteeskõnele.*

Võtmesõnad: kirjalik tekst, sünteeskõne, emotsioonid, afektirehkendus, perspektiivide mitmekesisus, eesti keel

Ajal, mil paljud tehnoloogilised seadmed kannavad austavat täiendit *nuti* (-telefon, -TV), on õigustatud ootus, et ka kirjalikku teksti ettelugev kõnesüntesaator oleks mingil määral nutikas. Tahame, et ta oskaks aru saada, kas tegu on formaalsema või isiklikuma tekstiga, ning loeks teksti ette inimesega võimalikult sarnasel viisil – seal, kus kohane, mõningase emotsionaalse kaasaelamisega (vt nt Mihkla jt 2012). Selle üle, kas ja kuidas on arvutil võimalik kirjalike tekstilausete seast emotsionaalselt laetuid ära tunda, on arutletud mujal ja uurimistööd selles suunas pole kaugeltki lõpetatud (vt nt Vainik 2010, 2012a ja 2012b). Siinses käsitluses pakub huvi eelkõige küsimus, kelle emotsioonile anda teksti ette lugedes hääl? Kas inimliku emotsionaalsuse jälgendamiseks sünteeskõnes tuleks luua mulje, et kõneleb teksti autor, vahendada teksti keskse tegelase emotsionaalset seisundit või anda hääl hoopis kellelegi kolmandale? Vastuse otsimisel lähtutakse seisukohast, et enne, kui hakata

* Uurimust on toetanud sihtrahastus "Eesti keele alusuuringud keeletehnoloogiliste rakenduste teenistuses" (SF0050023s09).

mingit inimlikku omadust arvutile õpetama, on kasulik võimalikult põhjalikult välja uurida, mida on selle nähtuse kohta juba teada saadud inimest uurides.

Selle kohta, kuidas inimesed lugemise käigus mõistavad teksti ja selle sees või taga paiknevate isikute emotsioone, ning mida nad seejuures ise emotsionaalselt läbi elavad, on kirjutatud palju ja mõistetavalt eri distsipliinide raames (kirjandusteadus, psühholoogia, filosoofia). Kuna tegemist on kahtlemata kompleksse ja mitmetahulise probleemistikuga, on käesolevas käsitluses mindud selle eri perspektiividena eritlemise teed. Esimese perspektiivina vaadeldakse küsimust, kas psühholoogiliselt realistlik prosoodia on üldse tarvilik lõpptarbija, st ette loetava teksti kuulaja, seisukohast ja milliste inimese loomusest tulenevate asjaoludega seda põhjendatakse. Teise perspektiivina tuuakse esile tekstist endast tulenevad võimalused keskendada tähelepanu kas teksti peategelase või siis autori emotsioonidele. Kolmas perspektiiv heidab valgust sellele, mida on arvatud teksti lugeja emotsioonide kujunemise kohta. Neljanda perspektiivina vaadeldakse ettelugemise situatsiooni tavainimesest ettelugeja positsiooni silmas pidades ning viiendana esitatakse n-ö arvuti perspektiiv. Käsitluse käigus võrreldakse teoreetilistes allikates kirjapandut mõningal määral ka Eesti emotsionaalse kõne korpuse¹ koostamisel saadud empiiriliste tulemustega ja lähenemisega teksti emotsionaalsuse määramisele (nt Altrov 2007, 2008, Pajupuu jt 2012). Lõpuosas arutletakse tulevikuperspektiivide üle ja esitatakse kokkuvõtte ning järeldus.

Kuulaja perspektiiv

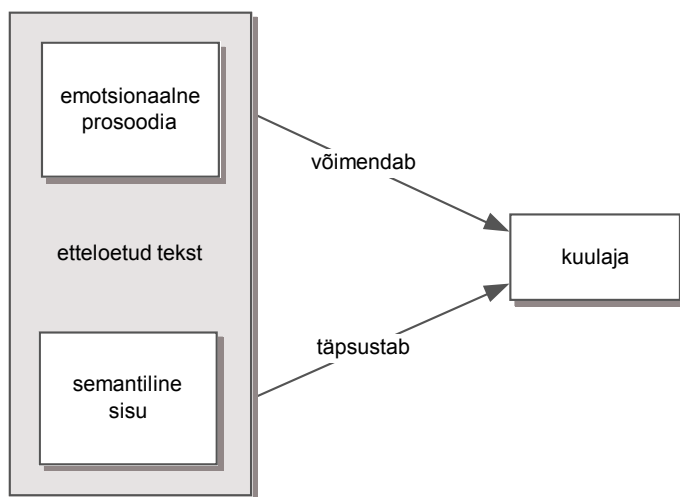
Kas ja miks peaks kõnesüntesaator teksti ette lugedes üldse püüdma inimlikku emotsionaalsust jäljendada? Üks vastuseid tuleneb sellest, et meie kuulmismeel on kohanenud jälgima ja tõlgendama eelkõige spontaanseid lausungeid. Loomulikus kõnesituatsioonis kaasneb verbaalse sõnumiga detailidest rikas situatiivne kontekst, mis pakub tuge kuuldu interpreteerimiseks. Kirjaliku teksti ettelugemine erineb loomulikust suhtlusest mitte üksnes lausete varieeruva pikkuse ja kompleksuse poolest, vaid ka selle poolest, et loomulikust multimodaalsest suhtlussituatsioonist jääb toimima ainult auditiivne ülekandekanal, seda eriti olukordades, kus ettelugemine toimub mitte näost näkku, vaid mingi seadme vahendusel (nt raadio, audioraamat). Seetõttu tundub otstarbekas, et auditiivsel kanalil esitatav sõnum kompenseeriks mingil määral vahetus suhtlusolukorras esinevaid visuaalseid infoallikaid, näiteks kõneleja näoilmet või kehaasendit, mis annavad tihti märku just kõneleja emotsionaalsest seisundist.

Lisaks intuiitivsele inimsarnasuse eelistusele räägivad emotsioonide jäljendamise kasuks ka uurimused, mis on näidanud, et lausetest aru saamine toimub kõige paremini (täpsemini, kiiremini) siis, kui edastatava sõnumi semantilise sisuga sobib kokku hääletoon e prosoodilise väljenduskanali informatsioon (vt ülevaadet Pell jt 2011). Uurijad ei ole päris ühte meelt selles osas, kumb informatsioon on ülimalik. Ühest küljest on leitud, et inimehäälele prosoodiliste vihjete põhjal emotsiooni omistamine on protsess, mis toimub tahtmatult ja kiiresti (Pell jt 2011). Juhtumil, kui semantiline sisu on emotsionaalse tooni suhtes kas neutraalne või mitmeti tõlgendatav, on suurem roll täita auditiivsel kanalil. Näiteks oleneb lause *Riina*

sai palgakõrgendust interpretatsioon kõneleja hääletoonist. Üldjuhul saavad kuulajad üsna hästi aru, kas kõneleja on sündmuse üle siiralt rõõmus või kõlab tema hääles kadeduse noot (Paulman jt 2011). Teisalt on leitud ka, et reaajas teksti kuulates võib semantilise kanali informatsioon mõnel juhul osutada prosoodiast olulisemaks (Pell jt 2011). Näiteks on lastel ja teise keele õppijatel raske aru saada lausungi iroonilisest toonist. Seetõttu tõlgendavad nad öeldut otsesõnu, lähtudes semantilise tasandi infost. Kolmandaks on kirjanduses väljendatud seisukohta, et kuna vastuvõtja töötleb signaali niikuinii omal viisil, milles on palju objektiivselt signaalist sõltumatut, ei määra emotsiooni atribuutsiooni lausungile mitte sõnumi sisu ega selle objektiivselt mõõdetavad prosoodilised parameetrid, vaid situatiivne kontekst. Näiteks võib isegi täpselt samal viisil intoneeritud lausung *ei või olla* väljendada sõltuvalt olukorrast nii suurt imestust, üllatust, aga olla öeldud ka suure hirmu või kohkumisega (Reilley, Seibert 2003).

On leitud, et kuigi neuroloogilisel tasandil võib tegemist olla osaliselt iseseisvate protsessidega, mis leiavad aset eri ajupiirkondades, toimub semantilise sisu ja akustilise väljenduse analüüs samaaegselt ning need protsessid on omavahel seotud. Emotsionaalse prosoodia toimimise mehhanismina on välja toodud selgitus, mille kohaselt on prosoodia roll aktiveerida või võimendada emotsioonidega seostuvat ainet mälus, mis omakorda soodustab sõnaliste stiimulite kognitiivset töötlust (Paulmann jt 2011). Siinkirjutaja julgeb teha oletuse, et prosoodia annab (vastuvõttu võimendades) emotsiooni edasi kiiresti ja suhteliselt robustselt, sellal kui semantiline kanal lubab teha täpsemaid eristusi ning järeldusi.

Kokkuvõttes kujutab emotsionaalse kõne kuulamine endast arvatavasti mitmeastmelist protsessi. Esimene aste on emotsioonivihjete (iseloomulikud prosoodilised vihjemustrid, valents ja aktivatsioon) kokkurehendamine, teise astmena võib järgneda lause konnotatsiooni seostamine mälus leiduva emotsionaalse informatsiooniga ning kolmanda astmena lause lõplik interpreteerimine. Kuigi loomult mitmeastmeline, on suulise kõne emotsionaalsusest arusaamine normaalse täiskasvanud inimese puhul millisekundite küsimus (Paulmann jt 2011).



Joonis 1. Emotsionaalse kaasaelamisega etteloetud teksti vastuvõtt

Skemaatiliselt esitab teksti ettelugemisel esinevat olukorda joonis 1, milles hall kast kujutab kuulajale esitatavat kaheplaanilist helisõnumit. Osa sellest moodustab tekst, st sõnumi semantiline sisu, ja osa ettelugeja kõnetoon, mis järgib etteloeitava tekstiga kaasnevaid emotsionaalseid tõuse ja langusi. On oluline, et prosoodiline väljendus oleks kooskõlas, või vähemasti mitte vastuolus semantilise sisuga, kuna on üldine seaduspärasus, et vastuolulise informatsiooniga toimetulek nõuab kognitiivset lisapingutust, mis väljendub töötlusaja pikenemises (vt Kahneman 2011). Semantilise sisuga kooskõlas olev prosoodiline väljendus seevastu hõlbustab kuuldava teksti vastuvõttu reaajas, kui kõnet ei ole võimalik tagasi kerida ega kahtluse korral uuesti kuulata.

Tekstisised perspektiivid: peategelane ja autor

Kuidas toimub teksti peategelase e protagonistide emotsioonide mõistmine? Psühholoogia valdkonnas tehtud katsed on näidanud, et lugejad konstrueerivad oma meeles (kuigi mitte teadlikult) mingisuguse ettekujutuse teksti keskse tegelase emotsionaalsest seisundist ning on võimelised jälgima selle muutumist tegevuse arenedes. Seejuures on leitud, et lugejad tulevad toime representatsioonide varal, mis on abstraktsemad konkreetsetest emotsiooninimetustest (ei eristata nt *kahetsust*, *piinlikkust* ja *häbi*), kuid on samas spetsiifilisemad valentsimäärangutest nagu *positiivne* vs. *negatiivne* (Gygax jt 2003, de Vega jt 1996). Opereeritakse mingisuguste vahepealsel spetsiifilisuse astmel olevate klastritaoliste või grupiviisiliste representatsioonidega, mille olemuse kohta ei ole aga veel täit selgust (Molinari jt 2009, Gygax jt 2003). Järeldusi protagonistide emotsioonide kohta tehakse jooksvalt ning automaatselt ning seda isegi juhul, kui tekstis otsesõnu ühelegi emotsioonile viidatud ei ole. See on võimalik, kuna lugejad mõistavad tekstis kirjeldatud sündmusi – seal esinevate isikute tegusid, eesmärke ja suhteid teiste tegelastega (Molinari jt 2009, Gernsbacher, Robertson 1992, viidatud de Vega jt 1996 kaudu).

Kui inimene loeb teksti, siis ta ei analüüsi igas üksikus lauses peituda võivat emotsiooni, vaid talitab ökonoomsemalt: mingi ettekujutus teksti (lõigu) peategelase emotsioonist konstrueeritakse teksti algusosas ja hiljem tulevatest tekstiosadest korjatakse üles vihjed, mis esialgset ettekujutust kas kinnitavad või kummutavad. Kui sündmuste edenedes ilmneb uusi asjaolusid, tehakse peategelase emotsionaalses representatsioonis muudatus. Peategelase emotsioonide viimast, “kehtivat” representatsiooni hoitakse aktiivsena alal ka neutraalsete lausete vältel, mida võib emotsionaalset infot sisaldavate vahel olla kuni neli-viis. Kui edasises tekstiosas lisandub informatsiooni, mis on emotsionaalses mõttes “ebaloogiline”, st ei klapi mudeliga, mille lugeja on oma peas peategelase seisundist jooksvalt konstrueerinud, siis selliste vastuoludega toimetulek nõuab lugejalt lisapingutust, mis väljendub töötlusaja pikenemises (de Vega jt 1996). Üldiselt on lugejatel ootus teksti peategelase e protagonistide emotsioonide loogilise arengu suhtes. Viimast on peetud koguni “liimiks”, mis narratiivi erinevaid osi sidusaks ja loogiliseks tervikuks seob (de Vega jt 1996 ja seal viidatud kirjandus).

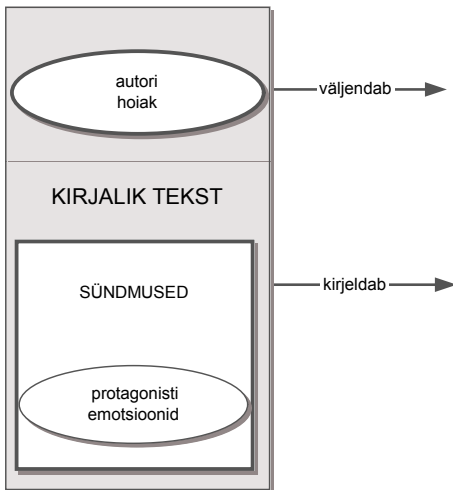
Kuidas avaldub autori emotsionaalne suhtumine? Selle põhjal, kas ja kuidas teksti autor oma peategelase tegemisi kas vähem või rohkem emotsionaalselt ning kaasaelamisega või distantseeritult kirjeldab, saab lugeja midagi järeldada teksti autori hoiaku ja emotsioonide või siis nende puudumise kohta. Kirjandusteaduses

on alternatiivseid võimalusi käsitletud eri fokaliseerimistüüpidega, millega kaasneb kas subjektiivne (kaasaelav, ingl *involved*) või objektiivne (mittekaasaelav, *uninvolved*) emotsionaalne vaatepunkt (Rimmon-Kenan 2002: 82, viidatud Bednarek 2008: 185 vahendusel). Teiseks on teksti- ja žanrianalüüsis tegeldud hoiaku (*stance*) käsitlemisega, mis kätkeb endas autori isiklike tundeid ja hinnanguid. Eristatakse episteemilist hoiakut (kahtlus, kindlus) ja suhtumishoiakut (positiivne vs. negatiivne suhtumine, tunded) (Conrad, Biber 2000: 57, viidatud Bednarek 2008: 16 järgi).

Autori emotsionaalne kohalolek vs. distantseeritus on üks olulisi aspekte, mille poolest tekstižanrid erinevad üksteisest. Ilukirjanduse autorilt eeldatakse, et ta loob sellised karakterid, kelle tegemiste ja tunnetega nad tunnevad empaatilist sidet (Keen 2006: 221). Ajakirjanduses on žanriti ja väljaandeti erinevusi. Liigne emotsionaalne kaasaelamine võib ajakirjanduses olla isegi taunitav. Näiteks esitas riigikogusaadik Igor Gräzin käesoleva artikli kirjutamise ajal protesti Eesti Rahvusringhäälingu suhtes, kuna leidis, et keelekasutus riigiteleviisioonis ei ole piisavalt neutraalne.² Päevik, blogipostitus ja isiklik kiri kujutavad seevastu žanreid, mille sees autoril on vabadus väljendada ja/või üsnagi üksikasjalikult kirjeldada iseenda tundeid. Tal on ka vabadus rääkida täie kaasaelamisega kolmandatest isikutest iroonilise alatooniga või naljatledes, neid andunult ülistades, hukka mõistes, halvustades jne.

Ka autori emotsioonide puhul ei saa üldjuhul eeldada, et need oleks väljendatud ja määratletavad kategoriaalselt ja otsesõnu (*rõõm, viha, hirm, üllatus, vastikus, kadetus* jne). Pigem kujuneb lugejal teksti autori emotsionaalsest seisundist samuti mingit laadi abstraktsem representatsioon (nt psühholoogilise distantsi ja/või hinnangu näol).

Juhtumil, kui tekst on kirjutatud *mina*-vormis, st esimese isiku positsioonilt, on autor ja protagonist üks ja seesama isik. Kuivõrd taolises intiimsemas žanris varustatakse lugejat informatsiooniga autor-protagonisti plaanide, tegevuse, edu ja ebaedu ning suhete-suhtlemise kohta teiste inimestega, on lugejal märksa rohkem pidepunkte, mille najal konstrueerida endale ettekujutus autor-protagonisti emotsioonidest, kui neid ka otsesõnu ei mainita.



Joonis 2. Autori ja protagonistide emotsioonide vahendamine tekstiga

² <http://www.epl.ee/news/eesti/igor-grazin-peab-err-i-uudiste-tooni-lubamatuks.d?id=65327458> (22.9.2013).

Skemaatiliselt esitavad tekstis endas sisalduda võivaid emotsioone ja/või hoiakuid joonisel 2 leiduvad ovaalid. Kui on tegemist minajutustusega, langevad teksti kirjeldav ning väljenduslik plaan ning kaks võimalikku eri perspektiivi kokku üheks. Objektiivsusust taotlevates žanrides seevastu on emotsionaalne distants autori ja peategelase vahel suurem või puudub autori hoiakute väljendusplaani sootuks ning piirduakse kirjeldusega.

Lugeja perspektiiv

Selle kohta, kuidas inimesed tekstis leiduvat emotsionaalset informatsiooni kasutavad, on samuti mitmeid eri teooriaid. Eespool oli juttu teksti peategelase emotsioonidest dünaamilise mudeli konstrueerimisest (de Vega jt 1996). On leitud aga, et selle kõrval, et lugeja n-ö kognitiivne aparaat protagonistide ja autori emotsioonidest mingil tasemel ettekujutuse loob, elab ta ka ise midagi emotsionaalsel tasandil läbi või kaasa.

Lugeja emotsioonide vastu on huvi tuntud nii erinevates humanitaarse suunitlusega teadusharudes (nt lingvistiline pragmaatika, kirjandusteadus, filosoofia, sotsiaalpsühholoogia, kriitiline tekstianalüüs) kui ka inimese ja masina suhtlusele keskenduvatel aladel, nt afektirehkenduses (ingl *affective computing*) või ka turunduses ja reklaaminduses. Emotsioonide esilekutsumise kaudu on inimesed teatavasti hõlpsasti mõjutatavad ja ka manipuleeritavad. See oli teada juba antiikses retoorikaõpetuses (Aristotelese "Retoorika", viidatud Caffi, Janney 1994: 334 vahendusel).

Filosoofias ja kirjandusteaduses on arutletud selle üle, kas teksti lugemisel tekkivad tunded on tõelised emotsioonid või mingisugused libaemotsioonid (ingl *quasi emotions, make-believe emotions*, vt Walton 1990: 195–204, viidatud Sklar 2009 vahendusel). Emotsionaalses reageerimises just eriti ilukirjandustekstile on nähtud filosoofilises mõttes paradoksi, kuna emotsiooni esilekutsujana esineb nähtus, mille kohta kogejatel on kindel teadmine, et seda – nt romaani peategelast ja tema kannatusi – pole üldse reaalsuses olemaski (Sutrop 1999). Seepärast on oletatud, et inimesed reageerivad lugedes mitte reaalse elu nähtustele, vaid teatud sorti kujutlustele või representatsioonidele. Välja on pakutud nn simulatsiooniteooria, mis tähendab kirjanduses esitatavate narratiivide n-ö mahamängimist meie siseses multimodaalsete representatsioonide teatris, koos vastavate emotsionaalsete reaktsioonidega (nt Currie 1998). On ka väidetud, et lugejad süüvivad (*immerse*) teksti sellisel määral, et seisund on võrreldav unenägude ajal kogetavaga, viimast seisukohta näivad toetavat uurimused nn tekstiga *kaasa-minemise (transportation)* fenomenist (Gerrig 1993, viidatud Sklar 2009 vahendusel). Sotsiaal-psühholoogias läbi viidud uurimustes on leitud aga, et tekstist indutseeritud emotsioonid ei pruugi kvalitatiivselt siiski erineda vahetult elusituatsioonides kogetavatest emotsioonidest, võtmeteguriks on, kui võrd lähedaselt seotuna tunneb lugeja end tekstis kirjeldatud tegelaste, sündmuste ja asjaoludega (Myyry, Helkama 2007: 251, viidatud Sklar 2009 vahendusel).

Arvatakse ka, et ilukirjanduse lugemisel tekkiva emotsionaalse kaasaelamise võlgneme empaatiale³ kui võimele mõista ja mingil määral jagada teiste inimeste tundeid. Empaatiat aluseks on samastumisvõime e oskus tajuda maailma teise olendi

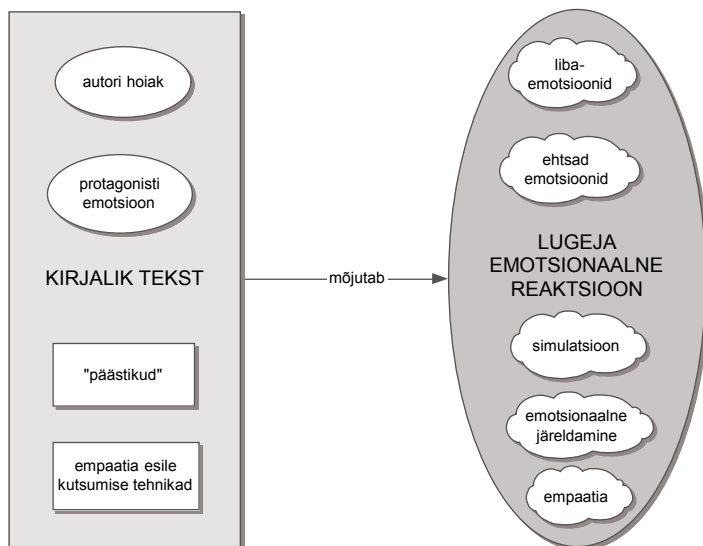
³ Juba sada aastat on tunnete ülekandumise fenomeni kirjeldatud empaatia mõiste abil. Termin *empaatiat* on kasutusele võtnud Edward Titchener tõlkevastena saksakeelsele terminile *Einfühlung* 'sissetundmine', mille kasutuselevõtja au kuulub Theodor Lippsile, ehkki loodud on see kontseptsioon veel märksa varem (1873) Robert Vischeri poolt (Montag jt 2008).

perspektiivist (de Waal 2009). Nagu paljude evolutsiooniliselt omandatud asjadega, ei ole meil sisuliselt valikut, kas olla empaatiline või mitte ja seetõttu rakendub see võime ka siis, kui loeme ilukirjandust, vaatame filmi või loome ise erksaid kujutluspilte, mille keskmes on teiste inimeste edu või ebaedu. Empaatia määr varieerub mingil määral isikuti ja on keskestlābi naistele omasem kui meestele (vt nt Baron-Cohen 2003). Empaatilisi inimesi on peetud “paremateks lugejateks” (Keen 2006) ja arvatavasti on seepärast just naistele suunatud nii suur hulk spetsiifilist kirjandust kui ka ajakirjandust (nn “naistekad”).

Teisest küljest pole saladus, et kirjanikud teevad lugejate emotsionaalseks kaasahaaramiseks ka teadlikke pingutusi. Kirjandusteaduses on kirjeldatud järgnevaid empaatia esilekutsumise tehnikaid: jutustamine esimese isiku vaatepunktist, karakteriloomise vahendid nagu tegelasele isikupärase nime andmine, välimuse kirjeldus, iseloomuomadustele osutamine, kirjeldatavate tegevuste laad, liikumine, kõnemaneeer jne. Arvatakse, et vahetu kohaloleku ja situatsiooniga ühendatuse tunnet suurendab oleviku aja kasutamine mineviku asemel. (Keen 2006: 215–216)

Ka ajakirjanduse uurimise kontekstis on tegeldud lugeja emotsioonide kujunemisega. Ungerer (1997) on osutanud, et artiklis sündmuse tasandil kirjeldatud (protagonisti) emotsioonid ei pruugi kokku langeda nendega, mida on kavas äratada lugejates. Ajakirjaniku või toimetuse poolt kavatsatud mõjutusemotsioon võib olla siiski kirjeldatuga mingisuguses järelduslikus seoses: nt kirjeldades mõrtsuka rõõmu, äratatakse inimestes viha, kirjeldades mõrtsuka pisaraid, omistatakse talle kahetsus ja tekitatakse lugejates kaastunnet. Ajakirjanduse lugejas reaalselt tekkivaid emotsioone on raske täpselt määratleda sellepärast, et lugeja emotsioon tekib *emotsionaalsete järelduste* tulemina, millega ta reageerib tekstis leiduvatele päästikutele/vihjetele (*trigger/cue*) (Ungerer 1997: 309). Need vihjed või päästikud panustavad kas ühele või mitmele nn emotsionaalse ruumi dimensioonile, milleks on pakutud psühholoogias ammu tuntud semantilise ruumi kolme dimensiooni (Ungerer 1997, Caffi, Janney 1994, dimensioonid pärit Osgood jt 1957). Dimensioonid on: a) hinnang (meeldiv *vs.* ebameeldiv); b) aktiivsus (erutatus, intensiivsus); c) võim, kontroll (ingl *potency*). Hinnang ja intensiivsus on tekstis kergemini ära tuntavad, kuid võim, kontroll raskemini. Selle n-ö potentsiidimensiooni teenistuses on need elemendid tekstis, mis viitavad lähedusele, spetsiifilisusele, evidentsiaalsusele, tahtelisusele (Caffi, Janney 1994). Toda raskestitabatavat dimensiooni on lingvistikas nimetatud ka pühendumiseks ja seotuseks (*commitment, involvement*) (Ungerer 1997: 312).

Ka lugejas tekkiv emotsioon võib olla rohkem või vähem spetsiifiline. Näiteks, kui esitatakse fakte võimurite valelikkusest, siis reaktsioonina tekib lugejates tõenäoliselt nõrdimus ja viha. Vähem spetsiifiline on emotsionaalne kaasaelamine, kui tekstiga hinnanguid ei kaasne, või kui see sisaldab vastuolulisi hinnanguid. Näiteks, kui kirjeldatakse sportlase ponnistusi teel olümpiale. Lugejad elavad kaasa edule ja tagasilöökidele, kuid raske oleks väita, et nad ühe lause juures tunnevad täiemahulist rõõmuemotsiooni ja järgmise lause juures kurbust – seda eriti tingimustes, kui lõplik otsus sportlase saatuse kohta polegi veel tehtud. Lisaks hinnangule mõjutab lugeja emotsiooni tugevust ja selgepiirilisust ka see, kas tekstis on juttu lähedastest või kaugetest, elulistest või abstraktsetest, võimuga seotud või mitteseotud sündmustest jpm. Emotsionaalne järeldamine on kahtlemata mitmetasandiline ja kompleksne protsess, mille mehhanismid vajavad edasist uurimist.



Joonis 3. Lugeja emotsionaalse reaktsiooni kujunemine

Kokkuvõtlikult võib öelda, et lugeja võimes tekstile emotsionaalselt reageerida ei kahtle keegi, küll on aga erinevaid teooriaid selle kohta, mida need lugeja emotsioonid endast kujutavad ja kuidas tekivad. Skemaatiliselt esitab olukorda joonis 3. Joonise vasakut poolt, mis kujutab teksti, on võrreldes joonisega 2 lihtsustatud ja lisatud tekstis sisalduda võivad lugeja emotsioone esile kutsuvad elemendid. Pilvekujulised väljad, mis on paigutatud lugeja emotsionaalset reaktsiooni kujutava ovaali sisse, esitavad oletusi, mida on kirjanduses lugeja emotsionaalse reaktsiooni olemuse ja selle kujunemise mehhanismide kohta tehtud.

Ettelugeja perspektiiv

Mida teeb tekstiga inimesest ettelugeja? Teksti ettelugemine nii, et see mõjuks kuulajale loomulikuna, ei ole ka inimese jaoks ülesanne, millega igäüks hõlpsasti ja ilma eriettevalmistuseta hakkama saaks. Kuigi nn ilmekas lugemine kuulub meil kohustusliku alghariduse hulka,⁴ kalduvad tavainimesedki teksti esitama mehhaaniliselt ja kaasaelamiseta. Näiteks eesti keele emotsionaalse kõne korpuse⁵ koostamise käigus kogeti, et soovitava ilmekuse kohta juhiseid saamata kaldusid mitteprofessionaalidest ettelugejad teksti esitama tuimalt. Rahuldav esituse loomulikkus saavutati instruksiooniga “tuua esile teksti meeleolu”. On mõistetav, et ettelugejad soovisid ülesandega toimetulekuks tekstiga eelnevalt tutvuda ning ettelugemist koguni harjutada (Altrov 2007: 24).

Nagu eelnevas osas välja toodud, on ettelugejal tekstiga tutvumise järel võimalik esiletoomiseks valida kahest erinevast allikast tuleneva emotsionaalse värvingu vahel: tekstis tegutseva tegelaskuju (protagonisti) emotsioonide või siis teksti autori emotsionaalse hoiaku vahel (vt joonis 2). On ka kolmas võimalus: ettelugeja ei püüa

⁴ Vt http://www.teatoimeta.ee/Lea_Nurkse_1904-1960_288.htm (13.9.2013).

⁵ Eesti emotsionaalse kõne korpus <http://peeter.eki.ee:5000/> (13.9.2013).

mitte niivõrd empaatiliselt sisse elada teksti autori või tekstis kujutatud inimeste rolli, vaid laseb hääletoonis kõlada iseenda emotsioonidel (joonis 3). Just viimane oligi nähtavasti eesti keele emotsionaalse kõne korpuse loojate eesmärgiks, kui otsustati kasutada mitte näideldud kõnet vaid n-õ tavainimese poolt sisse loetud teksti. Eeldatakse, et ettelugejal kujuneb konteksti najal mingi oma hinnang tekstis kujutatud sündmustest, ja et sellele hinnangule tugineva emotsiooni kannab ta ette loetavale tekstile häälega üle (Altrov, Pajupuu 2012: 50). Ettlugejatele antud juhend “tuua esile teksti meeleolu” tähendab seega eelkõige juhendit tuua välja ettelugejas endas tekkinud emotsionaalne seisund: kui tekst teda ärritab, kandub ärritus ette loetavale tekstile, kui tekst masendab, siis loeb ta teksti ette rusutud hääletoonil jne. Seega on nn “teksti meeleolu” ja selle tulemusena etteloeavatele tekstilõikudele (ja lausetele) lugeja poolt rakendatav emotsioon loomult projektiivset laadi, mis võib, aga ei pruugi olla vastavuses teksti protagonistiga ja/või autori emotsiooniga.

Selle tõestuseks, et “teksti peal” paiknev projektiivne ettelugeja emotsioon ja tekstis sõnadega väljendatud (kas autori või protagonistiga) emotsioon ei ole üks ja seesama, on ka tulemused, mis on saadud eesti keele emotsioonikorpuse etteloeatud lausete emotsioonide määramisel (vt nt Altrov, Pajupuu 2012). Kuivõrd eraldi katses ja erinevatel katseisikutel lasti määrata samade lausete emotsioon vastavalt kas ainult kirjalikul kujul (lugemistest) või etteloeatuna (kuulamistest), siis tuli välja, et keskmiselt 45% lausetest on ettelugeja lugenud mingi muu emotsiooniga, kui see, mis äratuntavalt esineb kirjalikus tekstilauses (vt tabel 1). Osutub ka, et ettelugeja kaldub teksti esitama pigem neutraalsemalt ja kurvemalt, kui lausete sisu eeldaks – tekstile neutraalsuse ja kurbuse n-õ “peale lugemine” on protsentuaalselt kõrgem. Teksti lugemist rõõmsamana, kui sisu eeldaks, tuleb teiste tonaalsustega võrreldes ette vähem, kuid siiski ligi 30%-l lausetest. Järelikult on ettelugeja projektiivsed emotsioonid (nii nagu nad testkuulajate hinnangute põhjal identifitseeriti) suhteliselt lõdvalt seotud tekstilausete sisuplaanis väljendatud emotsioonidega (nii nagu need ilma helita lugemiskatsetes identifitseeriti).

Ennatlik oleks ettelugeja projektiivsete emotsioonide ja tekstis sisuplaanis identifitseeritavate emotsioonide erinevust käsitleda lugeja vähese kompetentsusena, pigem seletab nähtust asjaolu, et lugeja oli orienteeritud terviktekstile (lõigule) ja mitte üksikute lausete esitamisele, nagu tulemusi kuulamis- ja lugemistestides hinnata lasti. Teine põhjus tõlgenduste erinevuseks tuleneb ilmselt sellest, et nii tekstis kui hääletoonis väljenduvad emotsioonid varieeruvad oma spetsiifilisuse (s.o kategoriaalse määratletavuse) astmelt ning see loob ruumi vastuste hajuvusele.

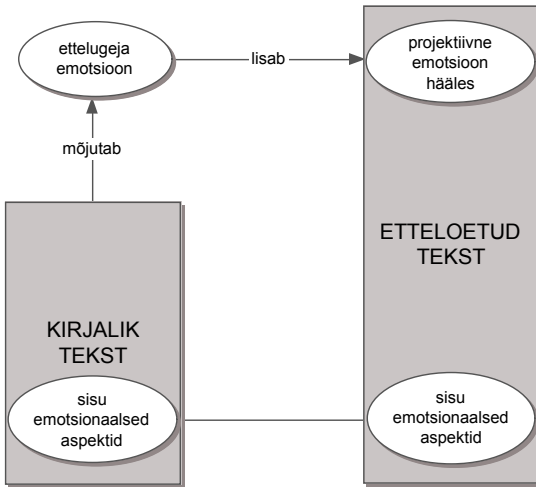
Tabel 1. Lausete distributsioon emotsioonimäärangute vahel võrrelduna lugemiskatse ja kuulamiskatse lõikes

Emotsioon	Lauseid	“Sisu mõjutab”	“Sisu ei mõjuta”	Projektiivsete omistuste %
rõõm	232	163	69	29.74
viha	277	177	100	36.10
kurbus	191	88	103	53.93
neutraalne	208	87	121	58.17

Märkus. Andmed pärinevad Altrovi ja Pajupuu (2012) artiklist, hallid alad on autori poolt juurde arvatud.

Nagu käesolevast jaotisest välja tuli, võib tavainimesest ettelugeja lisada ja keskmiselt iga teise lause puhul lisaski kirjutatud tekstile omalt poolt teistsuguse emotsionaalse interpretatsiooni, kui oli antud lause sisu tasandil. Tulemuseks on see, et ettleotud teksti kuulaja saab rohkem ja osaliselt teistsugust informatsiooni, kui seda oli kirjalikuna esitatud lausetes, kuna ettelugeja on teksti “rikastanud” enda emotsioonidega.

Kokkuvõtlikult kujutab olukorda joonis 4. Joonise vasakul poolel kujutatud sisu emotsionaalsed aspektid on samad, millest eelpool on olnud juttu (joonis 3): autori hoiak ja protagonistide võimalik emotsioon, tekstis kasutatud empaatia tekitamise tehnikad, “päästikud”. Teksti sisu emotsionaalsed aspektid kanduvad üle ettleotud lausesse, ettelugeja poolt väljendatav projektiivne emotsioon võib kas olla teksti sisust tuleneva emotsiooniga kooskõlas või mitte.



Joonis 4. Ettlugeja emotsiooni lisandumine tekstile

Arvuti perspektiiv

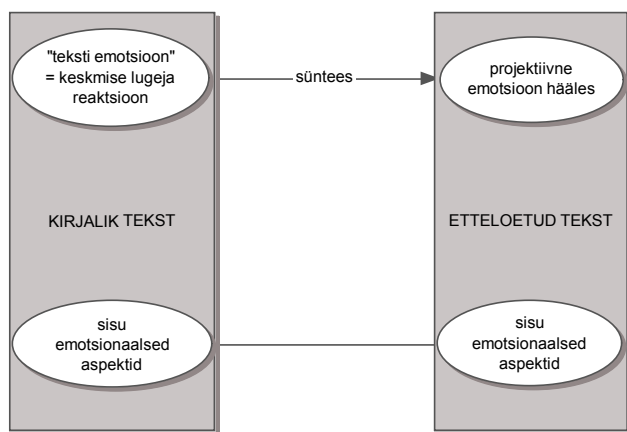
Kuidas otsustab arvuti (kõnesüntesaator), millise emotsionaalse registriga tuleks loetavat teksti esitada? Eelnevast inimlike perspektiivide ülevaatest nähtus, et teksti esitamisel relevantsemad emotsioonid võivad kuuluda erinevatele isikutele (protagonist, autor, ettelugeja) ning varieeruda spetsiifilisuse astmelt. Kelle perspektiivi peaks süntesaator esmajoones võtma ja millisel täpsusastmel emotsioone väljendada?

Nn afektirehkenduse (ingl *affective computing*) tehnoloogiavaldkonnas ning selle vajadustest tõukuva uurimisparadigma raames püütakse leida võimalikult hõlpsasti teostatavaid lahendusi emotsionaalsuse realiseerimiseks arvutil. Kuivõrd tegeldakse rakendustega ning mitte eeskätt teooriate ja avastustega inimeste kohta, on kasutusele võetud erinevate isikute perspektiive tasalülitav termin “teksti emotsionaalsus”, mis käsitleb emotsioone justkui teksti ja selle osade (nt lõikude, lausete) omadustena (vt nt Pajupuu jt 2012 ja seal esitatud viited). Käsitus, mis omistab emotsioone tekstile (lõikudele, lausetele) ja mitte eeskätt isikuile, saab olla üksnes

metafoorne.⁶ Terminoloogiline piltlikkus ei sega loomulikult tekstis sisalduvate võtmesõnade valentside põhjal “teksti emotsiooni” välja arvatamast – ka eesti keele jaoks on juba loodud emotsioonidetektor,⁷ mis kalkuleerib tekstilõikude valentsi selles sisalduvate emotsionaalset värvingut omavate sõnade (nn “võtmesõnade”) väärtuste (positiivne vs. negatiivne) summana. Taolise tekstuaalse üldafekti kalkuleerimise juhul oleks siinkirjutaja arvates psühholoogilise realistlikkuse huvides siiski korrektsem rääkida mitte teksti või lõigu emotsioonist ja selle tuvastamisest, vaid emotsiooni atribuutsioonist tekstile (lõigule, lausele). See, kas tehnoloogilisest vajadusest ellu kutsutud lähenemised peaksid üldse olema ka psühholoogiliselt realistlikud või mitte, on muidugi omaette arutlusteema.

Kui süntesaator hakkaks teksti ette lugedes juhinduma taolisest tekstile omistatud üldafektist, siis kelle emotsioone ta õigupoolest jälgendaks? Osutub, et kui kirjanduses räägitakse “teksti emotsionaalsusest”, siis vaikimisi toimib siiski selle nähtuse samastamine tekstivälise isiku – lugeja – reageeringuga (vt joonis 3). Seda käsitatakse teatud mõttes ülimuslikuna võrreldes teksti autori või tekstis esinevate tegelaste emotsioonidega. Pajupuu ja kolleegid (2012: 184) ei jäta selles suhtes mingit kahtlust: “[lugejahinnang] on kõige usaldusväärsem ehk õige, sest just lugeja annab tekstile tähenduse”. Paraku, nagu tuli välja ettelugeja perspektiivi käsitlusest, oli ühe üksiku lugeja poolt tekstilausele projitseeritud emotsioon üksnes lõdvalt seotud emotsioonidega, mida samade stiimulite puhul “tundsid ära” teised lugejad, kes osalesid lugemiskatses. Niisamuti olid ka lugemistestides osalenute endi hinnangud keskeltläbi üsna hajuvad (vt Vainik 2010).

Kirjanduses on seatud erinevaid konsensusnõudeid, kuid emotsioonide nelikotsustuste puhul on juba määrangute 50%-st üksmeelt peetud piisavaks (Vogt jt 2008). Oletatavasti peaks süntesaator teksti analüüsima ja ette lugema nii, nagu teeks seda lugejate 51%-line enamik. Seega on tekstist võtmesõnade põhjal kokku rehkendatava üldafekti e “teksti emotsiooni” näol tegemist hüpoteetilise keskmise lugeja potentsiaalse emotsionaalse reageeringuga.



Joonis 5. Emotsionaalse prosoodia sünteesimine

⁶ Niiviisi mõtlema ja emotsioone elututele asjadele omistama suunab meid tavakeel. Näiteks omadussõna *kurb* kasutatakse nii tähenduses ‘alanenud meelolu’, nt *kurb tüdruk*, kui ka ‘alanenud meelolu põhjustav’, nt *kurb film*. Me saame siiski aru, et tüdruk ja film on kurvad hoopis iseiviisi: tüdrukul on meeleolu alanenud, kuid filmi kohta seda väita ei saa.

⁷ <http://peeter.eki.ee:5000/valence> (22.9.2013).

Üldistatult võtab afektirehkenduses omaks võetud asjade käsituse kokku joonis 5. Nn teksti emotsioon e keskmise lugeja oletatav reaktsioon tuleb sünteesida kui projektiivne emotsioon, mida väljendatakse häälega “teksti peal”. Kuna teksti sisu jääb samaks, siis paratamatult kanduvad etteloeetavasse üle ka selle emotsionaalsed aspektid (autori hoiak ja protagonisti võimalik emotsioon, tekstis kasutatud empaatia tekitamise tehnikad, “päästikud”).

Kas teksti põhjal kalkuleeritav hüpoteetilise keskmise lugeja emotsionaalne reaktsioon tuleks määratleda kategoorianimetustena (nt *hirm, rõõm, kurbus, viha*) või iseloomustustena dimensioonidel (nt valents, aktivatsioon)? Afektirehkenduse raames on kritiseeritud tekstide emotsionaalsuse klassifitseerimist pelgalt valentsi alusel, kuivõrd selline jaotus ei ole vastavuses keeletehnoloogilistes rakendustes vajatava kategooriate süsteemiga (Karlgrén jt 2012). Ka eesti keele puhul on emotsionaalse prosoodia süntees planeeritud kategooriaalse mudeli põhjal – st kavas on esialgu luua vihale, rõõmule, kurbusele ja neutraalsusele omased intonatsioonid (vt nt Tamuri 2013). Siiski kujutavad emotsioonikorpuse lausete testimisel rakendatud kategooriad teatud mõttes üldisema tasandi nähtusi, kui seda on kitsalt vastavate emotsiooninimetuste (rõõm, viha, kurbus) tähendused. Kategooriad on ekstensiooni abil määratletud emotsioonide tavanimetustest laiemalt: *rõõm* sisaldab *tänulikkust, õnnelikkust, meeldivust* ja *elevust*; *kurbus* sisaldab *üksindust, muret, lohutamatus* ja *lootusetust* ning *viha* *kibestumist, irooniat, tõrksust, põlgust, pahatahtlikkust, raevu*. Sellised klastritaolised kategooriad sobivad kokku uurimistulemustega, mis osutavad, et inim-meel eelistabki emotsioonidest luua ning mõttes jälgida mitte täiesti spetsiifilisi kategoorianimesid ega ka pelgalt valentsihinnanguid vaid pigem mingil vahepealsel spetsiifilisuse astmel olevaid representatsioone (Gygax jt 2003, de Vega jt 1996, Molinari jt 2009).

Seega tundub praegu emotsioonidetektori poolt tekstile antav määratlus *positiivne, negatiivne, neutraalne* või *vastuoluline* jäävat lõppeesmärki silmas pidades alamääratletuks ja seda tuleks täiendada vähemasti mingit sorti üleminekumooduliga, mis muudaks dimensionaalse kirjelduse (nt määratluse *negatiivne*) kategooriaalseks (nt *kurbus*). Kuna kategooriaalsed määratlused on loomult detailsemad ja kitsamad, siis vajab taoline moodul võrreldes valentsimääranguga lisainformatsiooni.

Tulevikuperspektiiv

Kui kirjalikul kujul esinevale tekstile hakata emotsionaalsete rõhuasetustega prosoodiat sünteesima, siis peaks tegema esmalt valiku, millise isiku häälele ja emotsioonidel lasta kõlada. On võimalik valida, kas süntesaator peaks teksti ette lugema empaatiliselt – tuues välja teksti autori ja/või teksti keskse tegelase e protagonisti emotsioonid – või pigem nagu tavainimene, kes etteloeetavale tekstile oma emotsionaalse tõlgenduse lisab. Praeguse seisuga on tehtud valik tavainimese jäljendamise kasuks (Mihkla jt 2012). Projektiivset laadi emotsionaalse prosoodia puhul võib kooskõla teksti autori või protagonisti emotsioonidega olemas olla, kuid võib ka puududa.

Näib, et emotsionaalseks kõnesünteesiks läheb vaja mingit laadi *Emonculus*'t, virtuaalset olendit, kes tekstis esinevatele stiimulitele “keskmise lugeja” sarnaselt

reageerib ning oma virtuaalsed emotsioonid tekstile üle kannab, viimast prosoodiliselt rikastades. Seetõttu on eriti oluline saada suuremat selgust selle kohta, kuidas kujunevad ning mida endast kujutavad teksti lugemisel tekkivad emotsioonid (vt jaotist “Lugeja perspektiiv”). Otstarbekaks võib osutuda ka virtuaalsele *Emonculus*’ele erinevate isiksusetüüpide modelleerimine. Näiteks võiks olla nii, et neurootiline *Emonculus* rakendab enam ärevat või masendunud hääletooni, samas kui ekstravertne kalduks rohkem ekspluateerima reibast ja joviaalset kõnemaneeeri.

Emonculus’te loomine lahendaks ka probleemi, kuidas minna üle tekstilausele võtmesõnade põhjal kalkuleeritud dimensionaalsetest määrangutest spetsiifilistemade emotsioonide juurde. Olukorras, kus informatsiooni tekstis (lauses) napib (nt pole võimalik identifitseerida, kas negatiivsus peaks lugejat tegema pigem vihaseks või kurvaks) lähtub *Emonculus* enda sisseprogrammeeritud loomupärasest kalduvusest – neurootiline rakendab kahtluse korral rusutud ja ekstravertne ärritunud hääletooni. Sama teevad ka tavainimestest ettelugejad.

Õppetunnid eri perspektiividest

Artiklis refereeritud kirjandus, nende põhjal loodud skeemid ning arutelu aitasid lugejal loodetavasti aru saada, et tekstiga võivad olla seotud erinevad inimlikud vaatepunktid ja sellest tulenevalt on mingil määral ka erinevad lähenemised ning arusaamised emotsionaalsetest seisunditest. Ülevaatlikkuse mõttes võetakse järgnevas kokku olulisemad õppetunnid.

Kulaja perspektiivile keskendumine õpetas, et sisuga kooskõlas oleva emotsionaalse prosoodia ülesanne on aktiveerida emotsionaalset ainet vastuvõtja mälus ning seega hõlbustada kuuldava teksti vastuvõttu reaaliajas. Emotsionaalsel prosoodial on suurem roll nende lausete tõlgendamisel, mille sisu on kas vastuoluline või neutraalne.

Peategelase perspektiivi kohta selgus, et lugejad konstrueerivad tema emotsionaalsest seisundist oma peas dünaamilise mudeli, ning et lugejal on ootus tema emotsioonide loogiliseks arenemiseks, vastuoluline info nõuab töötlusel lisa-aega.

Autori perspektiivi arvessevõtt juhatas mõistma, et ta võib kirjeldatava suhtes olla kas emotsionaalselt osavõtlik või distantseeritud. Kui tekst on *mina*-vormis, langevad autori ja peategelase perspektiiv kokku.

Teksti lugeja perspektiivile keskendumine tõi esile, et selles osas valitseb kirjanduses oletuste ja seletuste paljusus. Levinud seletus lugeja emotsioonidele on empaatiavõime, kuid selle kõrval on välja pakutud ka simulatsiooni ning teksti põhjal emotsionaalsete järelduste tegemist. Viimase kirjeldamisel on peetud otstarbekaks nn semantilise ruumi kolme dimensiooni: hinnangut, aktiivsust ja potentsi.

Ettelugeja perspektiivist vaadates osutus, et tavalugeja projitseerib etteeloetavale tekstile enda emotsionaalse seisundi ega samastu ilmingimata alati lause peategelase ja/või autori emotsiooniga: ligikaudu poolte lausete puhul väljendas ta oma hääles äratuntavalt mingisugust muud emotsiooni kui see, mis oli teksti põhjal lausele konsensuslikult teiste lugejate poolt omistatud.

Iga tekstiga seotud isiku emotsioonide puhul võis tõdeda, et need saavad olla väljendunud ja äratuntavad kas spetsiifilisena või mittespetsiifilisena. Kui

konkreetsel emotsioonikategoriale otsesõnu tekstis viidatud ei ole (puuduvad emotsiooninimetused, vt nt Vainik 2002), siis saadakse hakkama mingit laadi abstraktsema, vahepealsel spetsiifilisuse astmel oleva representatsiooniga, mis on siiski komplekssem pelgast valentsimäärangust (positiivne vs. negatiivne).

Arvuti perspektiivi tundmaõppimine tõi esile, et rakenduslikust eesmärgist lähtudes saab ja võib nähtustest luua tegelikkusega võrreldes lihtsustatud mudelid. Hoiduda tuleks siiski nende mudelite pidamisest tegelikkuseks.

Kõike arvesse võttes tundub ühest küljest, et kui kõnesüntesaator on mõeldud järgima suhtlejate kooperatiivsuse printsiipi, siis peaks ta püüdma kuulajale vastu tulla ning varustama ette loetava teksti sisuga kooskõlas oleva prosoodiaga. Teisest küljest ei näi teatud kõrvalekalded sisu ja prosoodilise väljenduse vahel teksti mõistmisel siiski liiga suureks takistuseks olevat, kuni ei esine kuulajalt liiga suurt lisapingutust nõudvat ilmselget vastuolu. Kuniks ettelugeja intonatsioon on ligilähedane sellele, mida kuulaja eelneva tekstiosa ja peategelase emotsioonide arengu loogika põhjal eeldab, jääb vastuvõtja peas konstrueeritud emotsioonimudel kehtima vähemasti mõne järgneva lause vältel, tekst säilitab emotsionaalse sidususe ja hinnalist tööt-lusaega ei kaotata. Teksti prosoodiline esitus ei pea seega olema ideaalne, kuid ei tohiks vastuvõttu ja arusaamist emotsionaalse ebalooilisusega ka segada.

Teksti ettelugemisega seotud isikute, nende perspektiivide ja emotsioonide võimalike lähte kohtade ülevaatus näitas, et nii lugeja kui ka kuulaja on mitte lihtsalt tekstis sisalduva emotsionaalse info vastuvõtjad ja edastajad, vaid selle aktiivsed tõlgendajad, tähenduste ning seoste otsijad.

Viidatud kirjandus

- Altrov, Rene 2007. Emotsionaalse kõne korpuse loomine eesti keele tekst-kõne sünteesi jaoks. Tekstimaterjali evalvatsioon viha näitel. Magistritöö. Tartu Ülikool.
- Altrov, Rene 2008. Eesti emotsionaalse kõne korpus: teoreetilised toetuspunktid. – Keel ja Kirjandus, 4, 261–272.
- Altrov, Rene; Pajupuu, Hille 2012. Estonian Emotional Speech Corpus: Theoretical base and implementation. – The 4th International Workshop on Corpora for Research on Emotion Sentiment & Social Signals (ES3), Istanbul. L. Devillers, B. Schuller, A. Batliner, P. Rosso, E. Douglas-Cowie, R. Cowie, C. Pelachaud (Eds.). Istanbul, 50–53.
- Altrov, Rene; Pajupuu, Hille; Pajupuu, Jaan; Tamuri, Kairi 2012. Kõne ja teksti emotsionaalsuse statistilised mudelid. – Ettekanne riikliku programmi “Eesti keeletehnoloogia 2011–2017” konverentsil 2.–3. oktoobril Tartus, AHHA keskuses. <http://www.keeletehnoloogia.ee/konverentsid/ekt-esimene-konverents/EKT1.pdf> (3.9.2013).
- Baron-Cohen, Simon 2003. The Essential Difference: Men, Women and the Extreme Male Brain. Penguin/Basic Books.
- Bednarek, Monika 2008. Emotion Talk Across Corpora. Houndmills, New York: Palgrave Macmillan. <http://dx.doi.org/10.1057/9780230285712>
- Caffi, Claudia; Janney, Richard W. 1994. Towards a pragmatics of emotive communication. – Journal of Pragmatics, 22 (3–4), 325–373. [http://dx.doi.org/10.1016/0378-2166\(94\)90115-5](http://dx.doi.org/10.1016/0378-2166(94)90115-5)
- Conrad, Susan; Biber, Douglas 2000. Adverbial marking of stance in speech and writing. – Susan Hunston, Geoffrey Thompson (Eds.). Evaluation in Text. Oxford: Oxford University Press, 56–73.
- Currie, Gregory 1998. Realism of character and the value of fiction. – Jerrold Levinson (Ed.). Aesthetics And Ethics: Essays at the Intersection. Cambridge: Cambridge University Press, 161–181.

- de Waal, Frans 2009. *The Age of Empathy*. Harmony Books.
- de Vega, Manuel; Inmaculada, Leon; Diaz, Jose M. 1996. The representation of changing emotions in reading comprehension. – *Cognition and Emotion*, 10 (3), 303–321. <http://dx.doi.org/10.1080/026999396380268>
- Gernsbacher, Morton A.; Robertson, Rachel W. 1992. Knowledge activation versus sentence mapping when representing fictional characters' emotional states. – *Language and Cognitive Processes*, 7 (3–4), 353–371. <http://dx.doi.org/10.1080/01690969208409391>
- Gerrig, Richard J. 1993. *Experiencing Narrative Worlds: On The Psychological Activities Of Reading*. New Haven, CT, London: Yale University Press.
- Gygax, Pascal; Oakhill, Jane; Garnham, Alan 2003. The representation of characters' emotional responses: Do readers infer specific emotions? – *Cognition and Emotion*, 17 (3), 413–428. <http://dx.doi.org/10.1080/02699930244000048>
- Kahneman, Daniel 2011. *Thinking, Fast and Slow*. London: Allen Lane.
- Karlgren, Jussi; Sahlgren, Magnus; Olsson, Fredrik; Espinoza, Fredrik; Hamfors, Ola 2012. Usefulness of sentiment analysis. – Ricardo Baeza-Yates (Ed.). *ECIR 2012, LNCS 7224*. Berlin, Heidelberg: Springer-Verlag, 426–435.
- Keen, Suzanne 2006. A theory of narrative empathy. – *Narrative*, 14 (3), 207–236. <http://dx.doi.org/10.1353/nar.2006.0015>
- Mihkla, Meelis; Hein, Indrek; Kalvik, Mari-Liis; Kiissel, Indrek; Sirts, Risto; Tamuri, Kairi 2012. Estonian speech synthesis: Applications and challenges. – Andrej Kibrik (Ed.). *Computational Linguistics and Intellectual Technologies. Papers from the Annual International Conference "Dialogue"*. Moskva: PITY, 443–453.
- Molinari, Carlos; Burin, Débora; Saux, Gastón; Barreyro, Juan P.; Irrazabal, Natalia; Bechis, María S.; Aníbal, Duarte; Verónica, Ramenzoni 2009. Fictional characters' emotional state representation: What is its degree of specificity? – *Psicothema*, 21 (1), 9–14.
- Montag, Christiane; Gallinat, Jürgen; Heinz, Andreas 2008. Theodor Lipps and the Concept of Empathy: 1851–1914. – *American Journal of Psychiatry*, 165 (10), 1261. <http://dx.doi.org/10.1176/appi.ajp.2008.07081283>
- Myrsky, Liisa; Helkama, Klaus 2007. Socio-cognitive conflict, emotions and complexity of thought in real-life morality. – *Scandinavian Journal of Psychology*, 48 (3), 247–259. <http://dx.doi.org/10.1111/j.1467-9450.2007.00579.x>
- Osgood, Charles E.; Suci, George J.; Tannenbaum, Percy H. 1975 [1957]. *The Measurement of Meaning*. Urbana, Chicago: University of Illinois Press.
- Pajupuu, Hille; Kerge, Krista; Altrov, Rene 2012. Lexicon-based detection of emotion in different types of texts: preliminary remarks. – *Eesti Rakenduslingvistika Ühingu aastaraamat*, 8, 171–184. <http://dx.doi.org/10.5128/ERYa8.11>
- Paulmann Silke; Ott, Derek; Kotz, Sonja A. 2011. Emotional speech perception unfolding in time: The role of the basal ganglia. – *PLoS ONE*, 6 (3), e17694. <http://dx.doi.org/10.1371/journal.pone.0017694>
- Pell, Mark D.; Jaywant, Abishek; Monetta, Laura; Kotz, Sonja A. 2011. Emotional speech processing: Disentangling the effects of prosody and semantic cues. – *Cognition and Emotion*, 25 (5), 834–853. <http://dx.doi.org/10.1080/02699931.2010.516915>
- Reilly, Judy; Seibert, Laura 2003. *Language and emotion*. – Richard J. Davidson, Klaus R. Scherer, Hill H. Goldsmith (Eds.). *Handbook of Affective Sciences*. Oxford University Press, 535–559.
- Rimmon-Kenan, Shlomith 2002. *Narrative Fiction: Contemporary Poetics*. London, New York: Routledge.
- Sklar, Howard 2009. Believable fictions: On the nature of emotional responses to fictional characters. – *Helsinki English Studies* 5. *Electronic Journal*. <http://blogs.helsinki.fi/hes-eng/volumes/volume-5/> (4.9.2013).

- Sutrop, Margit 1999. Sympathy, imagination, and the readers' emotional response to fiction. – Jürgen Schlaeger (Ed.). *Representations of Emotions*. Tübingen: Gunter Narr Verlag, 29–42.
- Tamuri, Kairi 2013. Emotsioonide akustika ja nende mudeldamine. – Ettekanne Keeleteaduse, Filosoofia ja Semiootika Doktorikooli (KFSDK) sügisseminaril “Kõnetehnoloogias filoloogile”, 13.9.2013. https://dl.dropboxusercontent.com/u/33219645/Sangaste%202013/Kairi%20Tamuri_2013.pdf (6.1.2014).
- Ungerer, Friedrich 1997. Emotions and emotional language in English and German news stories. – S. Niemeier, R. Dirven (Eds.). *The Language of Emotions. Conceptualization, Expression, and Theoretical Foundation*. Amsterdam, Philadelphia: John Benjamins, 307–329.
- Vainik, Ene 2002. Emotions, emotion terms and emotion concepts in an Estonian folk model. – *Trames: Journal of the Humanities and Social Sciences*, 6 (4), 322–341.
- Vainik, Ene 2010. Kuidas õpetada kõnesüntesaatorile empaatiat? Emotsiooni automaatse tuvastuse võimalustest eestikeelses kirjalikus lauses sisalduva info põhjal. – *Eesti Rakenduslingvistika Ühingu aastaraamat*, 6, 327–347. <http://dx.doi.org/10.5128/ERYa6.20>
- Vainik, Ene 2012a. Eesti keele sõnavara afektikalduvus ja mis sellega peale hakata. – *Keel ja Kirjandus*, 8–9, 644–657.
- Vainik, Ene 2012b. Kuidas määrata eesti keele sõnavara tundetoone? – *Eesti Rakenduslingvistika Ühingu aastaraamat*, 8, 257–274. <http://dx.doi.org/10.5128/ERYa8.17>
- Vogt, Thurid; Andre, Elisabeth; Wagner, Johannes 2008. Automatic recognition of emotions from speech: A review of the literature and recommendations for practical realisation. – Christian Peter, Russell Beale (Eds.). *Affect and Emotion in Human-Computer Interaction*, LNCS 4868. Heidelberg: Springer, 75–91.
- Walton, Kendall 1990. *Mimesis As Make-Believe: On the Foundations of the Representational Arts*. Cambridge, MA: Harvard University Press.

Võrguaadressid

- Eesti emotsionaalse kõne korpus. <http://peeter.eki.ee:5000/> (22.9.2013).
- Eesti Keele Instituudi emotsioonidetektor. <http://peeter.eki.ee:5000/valence> (22.9.2013).
- Eesti keeletehnoloogia konverentside veebiväljaanne. <http://www.keeletehnoloogia.ee/konverentsid/ekt-esimene-konverents/EKT1.pdf> (3.9.2013).
- Eesti Päevalehe veebiväljaanne. <http://www.epl.ee/news/eesti/igor-grazin-peab-err-i-uudiste-tooni-lubamatuks.d?id=65327458> (22.9.2013).
- Keeleteaduse, Filosoofia ja Semiootika Doktorikooli (KFSDK) materjalid. <http://portaal.eki.ee/tegevusvaldkonnad/246.html> (12.12.2012).
- Rahvusvaheline Psühhiaatria kuukiri. <http://ajp.psychiatryonline.org/data/Journals/AJP/3872/08aj1261.PDF> (12.12.2012).
- Tallinna Ülikooli pedagoogilise seminari abimaterjalid. http://www.teatoimeta.ee/Lea_Nurkse_1904-1960_288.htm (13.9.2013).

Ene Vainiku (Eesti Keele Instituut) uurimisalaks on kognitiivne lingvistika, semantika ning keele ja emotsioonide vastasseosed. Avaldanud uurimusi emotsiooninimetustest, piltlikest ja dünaamilistest emotsioonikirjeldustest ning emotsiooni eestikeelses tekstis tuvastamise problemaatikast.

Roosikrantsi 6, 10119 Tallinn, Estonia

Ene.Vainik@eki.ee

WHO'S EMOTIONS TO EXPRESS WHILE READING OUT LOUD? A MULTI-PERSPECTIVE APPROACH

Ene Vainik

Institute of the Estonian Language

Human emotions is one of the aspects that accounts for the desired goal of human-likeness of text-to-speech synthesis. Present article tackles the question to whose emotions give the opportunity to be expressed in voice? Should it be the Author of the text, the Protagonist or someone else? Literature from different disciplines (psychology, literary science, philosophy, affective computing etc.) is used to shed light on the issue.

As regards to the perspective of the potential listener – the ultimate consumer of the text-to-speech service – the literature confirms the intuition that emotional prosody will facilitate understanding, by activating the emotional aspects in the listener's memory, supposedly. The role of emotional prosody is found to be bigger in the case of neutrality or ambivalence of the semantic content of the sentences.

As regards to the emotions of the persons potentially represented in the text – the Protagonist and the Author – it revealed that the readers use to construct a dynamic representation of the Protagonist's emotions in their minds. The emotional attitude of the Author may be recognized from the involved (subjective) vs. uninvolved (objective) type of focalizing. In the case the text is written in the 1st person's perspective, the Author and the Protagonist (and their emotions) coincide, supposedly. The emotional information is processed in a sequential order as it unfolds in the narrative. Generally, the readers have an expectation for the emotional model to be "logical". Processing "illogical" emotional information, while presented, takes extra processing cost.

There appears to be a lot of theorizing in the literature in respect of the emotions of the reader of the text. Besides empathy as an explanation for this phenomenon, also multimodal simulation and emotional inferences from some textual cues have been offered, among others. The latter approach relies on the three dimensions of the semantic/emotional space (well known from psycholinguistics): evaluation, activation, and potency.

While a human being – a layman – has the task to read out loud a text unfamiliar to him/her with an appropriate emotional prosody, it takes some effort, too. Analysis of the empirical data from the Estonian Emotional Speech Corpus suggests, that in half of the instances the laymen tended to express their own emotions towards the semantic content of the text rather than to empathize with the main character or the author of the text.

The different persons and perspectives represented in the text are not accounted for in the field of affective computing, yet. Instead, the readers' expected reaction is attributed to the text as it's measure of "emotionality" and is supposed to be applicable as an instruction for the synthesizer while mimicking human emotional prosody.

Keywords: written text, synthesized speech, emotions, affective computing, multiplicity of perspectives, Estonian language

Trükitud 150 eksemplari