

**Eesti Rakenduslingvistika Ühing**

**Eesti Rakenduslingvistika Ühingu  
aastaraamat 9**

*Estonian Papers in Applied Linguistics 9*

Toimetanud  
Helle Metslang, Margit Langemets  
ja Maria-Maren Sepper

Tallinn 2013

## **EESTI RAKENDUSLINGVISTIKA ÜHINGU AASTARAAMAT 9** **Estonian Papers in Applied Linguistics 9**

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 Öim (Tartu)

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

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

Ühingu aadress / Contact information:

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

[ery@eki.ee](mailto:ery@eki.ee)

[www.rakenduslingvistika.ee/ajakirjad](http://www.rakenduslingvistika.ee/ajakirjad)

Eesti Rakenduslingvistika Ühingu aastaraamatu 9. numbril väljaandmist on toetanud Haridus- ja Teadusministeeriumi riiklik programm "Eesti keel ja kultuurimälu" ja Eesti Keele Instituut.

Autoriõigus / Copyright: autorid ja Eesti Rakenduslingvistika Ühing

Kujundaja / Design: Piia Ruber

ISSN 1736-2563 (trükkis)

ISSN 2228-0677 (võrguväljaanne)

doi:10.5128/ERYa.1736-2563

## SISUKORD

Comparing speaking situations in three different language tests <i>Kolme erineva keele testi rääkimisosade võrdlustulemused</i> Sari Ahola, Tiina Lammervo, Reeta Neittaanmäki, Sari Ohranen, Henna Tossavainen	7
Acquisition of compounds in Estonian and Russian: Frequency, productivity, transparency and simplicity effect <i>Liitsõnade omandamine eesti ja vene keeles: sageduse, produktiivsuse, lihtsuse ja läbipaistvuse mõju</i> Reili Argus, Victoria Kazakovskaya	23
Quantitative and qualitative aspects of switching between Estonian Sign Language and spoken Estonian <i>Eesti viipekeele ja suulise eesti keele vahelise koodimuutuse kvantitatiivsed ja kvalitatiivsed aspektid</i> Kadri Hein	43
Sõjanduskeele ja kujundkeele ristteel ehk Kas kujund (oskus)keelt pigistab? <i>At the crossroads of military and figurative language</i> Reet Hendrikson	61
Eesti VerbNet'i loomise võimalikkusest <i>The feasibility of Estonian VerbNet</i> Indrek Jentson	75
Katseline semantika: planeerimine ja teostus <i>Experimental semantics: Design and implementation</i> Anni Jürine, Jane Klavan, Ann Veismann	85
Kuidas tuvastada dialoogi struktuuri: korpuse analüüs <i>How to recognise the dialogue structure: Corpus analysis</i> Mare Koit	101

Developing interviewer proficiency: A self-perception survey <i>Intervjueerija kutsemeisterlikkust arendades: ankeetküsitluse tulemuste analüüs</i> Riin Kont-Kontson, Ene Alas, Suliko Liiv	113
Keele kestlikkuse ja kao konstrueerimine poolavalikes diskursustes <i>The construction of language maintenance and loss in semi-public discourses</i> Kadri Koreinik, Kristiina Praakli	131
How to create order in large closed subsets of WordNet-type dictionaries <i>Kuidas luua korda Wordnet'i tüüpi sõnaraamatute suurtes kinnistes alamhulkades</i> Ahti Lohk, Ottokar Tilk, Leo Võhandu	149
Üleminek riigikeelsele gümnaasiumiõppele: vilistlaste seisukohtade Q-metodoloogiline diskursusanalüüs <i>Transition to state-language instruction in upper secondary schools according to alumni: Q-methodological discourse analysis of individual positions</i> Anu Masso, Katrin Kello, Valeria Jakobson	161
Laensõnade tähendussuhetest <i>Semantic relations of loanwords</i> Vilja Oja, Iris Metsmägi	181
Speed in cognitive tasks as an indicator of second/foreign language reading and writing skills <i>Kognitiivsete ülesannete lahendamise kiirus kui võõrkeele / teise keele lugemis- ja kirjutamisoskuse näitaja</i> Sanna Olkkonen	195
Interviewer behaviour during oral proficiency interviews: A gender perspective <i>Intervjueerija käitumise iseärasused keelepädevustesti läbiviimisel: sooline variatiivsus</i> Edith Reemann, Ene Alas, Suliko Liiv	209
Sisestatud tarindid lepingute tekstimoodustusvõttena <i>Inserted constructions as the text formation method of contracts</i> Riina Reinsalu	227
Täiendusi Heinrich Gösekeni saksa-ladina-eesti sõnastiku (1660) ja selle uusväljaande (2010) koostamise kohta <i>Supplement to the compilation of the new edition (2010) of Heinrich Göseken's German-Latin-Estonian Glossary (1660)</i> Kristel Ress	245

Meetodeid tekstide leksikaalsete ja grammatiliste erinevuste tuvastamiseks meditsiiniliste tarbetekstide näitel <i>Methods for identifying lexical and grammatical differences          in medical applied texts</i> Raul Sirel	265
Eesti ja soome sõnalugemistestide võrdlus <i>Comparison of Estonian and Finnish word reading tests</i> Piret Soodla, Maigi Vija, Renate Pajusalu	279
Multilingualism of a Southern Estonian – comparison of Estonian, Võru and Finnish narratives <i>Lõuna-eestlase mitmekeelsus – eesti, võru ja soome narratiivide          võrdlus</i> Liina Tammekänd	297
Õiguslikud väljakutsed ja võimalikud lahendused keeleressursside arendamisel <i>Legal challenges and possible solutions in developing          language resources</i> Arvi Tavast, Heiki Pisuke, Aleksei Kelli	317
Keeleressursside litsentsimise õiguslikke küsimusi <i>Legal considerations in licensing language resources</i> Kadri Vider, Heiki Pisuke, Peeter P. Mõtsküla, Triin Tuulik, Aleksei Kelli	333



## COMPARING SPEAKING SITUATIONS IN THREE DIFFERENT LANGUAGE TESTS

Sari Ahola, Tiina Lammervo, Reeta Neittaanmäki,  
Sari Ohranen, Henna Tossavainen

**Abstract.** The purpose of this study was to compare learner outcome when five similar speaking situations were offered in the speaking subtest in Finnish, Swedish and English intermediate level tests in the Finnish National Certificates language testing system. The overall aim was to investigate the tasks by comparing learner outcomes across the three tests and to seek explanations for the outcome in the learners' self-reported demographic, language, educational and professional background.

The speaking situations were selected from the NC item bank meaning that they have undergone the Item Response Theory based analyses which indicate that the tasks function well in all tests. More information was needed to discover possible connections between performance and background information. Performance data and background factors were analysed using descriptive frequency and percent distribution. Cross-tabulation was used to analyse connections between variables. The tasks are discussed in terms of domains and language functions and their connection with the test takers' background information. Results indicate that though there is some variation in the learner outcomes across languages, situational tasks can be used for different languages.

**Keywords:** second and foreign language testing, spoken language, validity, Finnish, Swedish, English tests

### 1. Introduction

This article discusses the results of a study which compared learner<sup>1</sup> outcomes in a language test situation when the same test tasks were used in three different test languages. In the administration of the autumn 2011 test for the National Certificates (NC) in Finland, five situational tasks were offered in the speaking subtests of the

---

<sup>1</sup> Learner is used as a general term for language learners; test taker and candidate refer to people who take the NC examination.

English, Finnish and Swedish intermediate level tests. Additional data for Swedish was collected in spring 2012. The overall aim was to investigate differences and similarities in the performance of test takers across the three language tests and, furthermore, to seek explanations for the outcome in the candidates' demographic, language, educational and professional background as reported by them. This all relates to the general idea of validity in the use of situational test tasks in testing speaking skills in three different languages.

The NC is a second- and foreign-language proficiency test for adults in which the levels of language proficiency are linked with the Common European Framework of Reference for Languages (CEFR) 2001. The intermediate level tests measure language proficiency at independent language user levels B1 and B2. The test system as a whole is based on Finnish legislation (Act 964/2004; Decrees 1163/2004 and 1109/2011) and is independent of any syllabus or curriculum.

There are nine languages in the test system; the three languages in this study can be regarded as high-stakes. The test with the largest number of candidates, the Finnish test, is often taken by immigrants to demonstrate their language skills for immigration purposes. Finland has two official languages, Finnish (91% of population) and Swedish (5.4%), so the Swedish test can also be taken for the same immigration purpose. Finland has a long tradition of foreign language learning and currently English is the most popular foreign language studied in schools. The English test is commonly taken by Finns who wish to work for the Finnish Defence Forces in international military cooperation roles. Furthermore, the results of all three tests are used as proof of language skills for professional purposes. The Swedish test is taken, for instance, by nurses and other health care professionals working in the Swedish-speaking areas in Finland.

The Finnish tests are taken by candidates with diverse language backgrounds (approximately 170 first languages overall) while the English tests are typically taken by Finns with Finnish or Swedish as their first language (L1). The Swedish test is taken by candidates from both diverse and Finnish language backgrounds.

The Framework of the Finnish National Certificates (2002, 2011) and the NC test specifications, which apply to all tests in the nine languages, provide the guidelines for item writing, test compilation and test development, with information about what constitutes communicative language proficiency and hence also the communicative language ability to speak in the foreign language. The test framework in the NC is based on the models of communicative competence offered in Canale and Swain (1980), and further in Bachman (1990) and Bachman and Palmer (1996), and in the framework offered in the CEFR 2001. Thus the test is designed to evaluate everyday language used for communication purposes rather than focusing on grammatical structures and vocabulary. The test specifications offer a defined set of topic areas and language functions for the purposes of item writing.

The test measures language ability in the subtests of writing, speaking, listening and reading. The tasks discussed in this article belong to the speaking subtest and comprise one of a total of four tasks in the subtest which are carried out in the language laboratory in all three tests. It is stated in the test specifications that the test tasks should be communicative and functional. It is also stated in the specifications that the tasks should take into consideration the interactional view of authenticity, which means that the tasks should consider the interaction between the test taker's



language ability – level of proficiency and how to engage the test taker in the performance – and such situational task characteristics as the features of context and the test task (see ALTE Manual 2011: 12, Douglas 2000: 18).

The overall aim of the current study is to find out if the same test tasks are equally valid in all three subtests of speaking. Validity here refers to context validity since it accounts better for the social dimension of language use than the earlier concept of content validity (Weir 2005: 19). Weir (2005: 20) also states that though there are problems in ensuring that we actually follow the specifications and domains we promise to follow, and that there are problems in the operationalisation of every-day language use in a test situation, we should make attempts to ensure context validity.

Comparing test performance with background information (biodata) functions as a method of construct validation. The intention is to detect bias in the test for or against groups of test takers defined by their particular data (age, gender, education etc.) (Alderson et al. 1995: 185). The current study looks at selected background information to seek possible explanations for the differences in test taker performance.

## **2. Data and methods**

The subtest of speaking consisted of four sections in which the candidates were asked, first, to describe an event, then take part in simulated conversations, react to situations and give a short speech. This study focuses on the candidates' reactions to situations which here are regarded as tasks and in the statistical analyses as items. The five situations represented different levels of formality and domains (personal–work), and also served different language purposes (as per NC item writing specifications). The following functions were required from the five test tasks:

1. The informative function of describing one's hobby and the interactional function of tempting a friend to join in.
2. The informative function of describing how the test takers learn new words and the possible methods that help them remember these words.
3. The emotive function of expressing feelings to a colleague in an upsetting situation (the test takers were asked to come up with the situation)
4. The interactional function of congratulating a friend on becoming a parent.
5. The informative function of describing the qualities of a good boss.

The speaking subtest was taken in a language laboratory where candidates' production was recorded. At the start of each section candidates had time to familiarise themselves with the instructions. A preparation time of 15–20 seconds was reserved for the candidates to prepare before each task/item (read the prompt text) and they had 30 seconds to complete their speaking. Candidates used the test booklet with instructions and task descriptions. Prompts advising when to start speaking came from the master recording.

The language of the test task instructions was not the same in all three tests. In the English test the instructions were in Finnish or Swedish, in effect, the test takers' chosen language of administration. In the Swedish and Finnish tests the

instructions were in the target language. However, in all three languages the pre-recorded prompt heard from the master tape giving the instructions of what to do (e.g. 'start speaking now') were in the target language.

The candidates' recorded speech, their test performance, was assessed using the NC criteria for speaking, calibrated and empirically linked with the CEFR. All NC tests are assessed by trained, registered raters<sup>2</sup>. All tests are assessed in centralised assessment sessions organised at the University of Jyväskylä. Every assessment meeting starts with a two-hour training session discussing assessment criteria, benchmarks and possible task expectations.

In the final score for speaking all four tasks are taken into account. Each situational task/item is given a separate score and these together with the scores for other tasks make up the final score for speaking. This study uses the data from the five situational tasks which were the same in all three languages. Performance data of situational tasks in each language was analysed with the Facets package which is based on the Many-Facet Rasch Measurement model (Eckes 2011). This model allows us to consider difference in item difficulty, rater leniency and the functioning of items. Test items are from the NC item bank and have undergone Item Response Theory based analyses which indicate that the tasks function well in all tests.

The data were gathered from three different language tests administered during the autumn of 2011. As the sample for the Swedish test was fairly small in number compared with the English ( $N = 215$ ) and Finnish ( $N = 1084$ ) tests, additional data were gathered during the 2012 spring test bringing the total sample for the Swedish test to 270. The data-collection sample for all languages includes only the test takers who completed the background information sheet.

A background information questionnaire is distributed with every test. It is not compulsory to fill in the information but typically the response rate is around 95%. In addition to the common biographical and demographic data (gender, first language), the questionnaire covers socio-economic factors (education, occupational status, occupational field), purpose of taking the examination, extent of language studies and self-evaluation of target language skills and language use (with family, friends and acquaintances, reading, writing messages, following the media, transactions, work, study; daily, weekly, monthly, not at all). It should be emphasised that the information is self-reported and in the Finnish and Swedish tests it is given in the candidate's second language.

Background and task performance data were analysed per test (language) using descriptive frequency distribution and percent distribution. Cross-tabulation was used to investigate connections between variables. The chi-square test was used to measure the independence of two categorical variables with significance level of 0.05. Adjusted residuals which are based on the comparison of observed and expected frequencies were used to investigate where possible connections might be found. An adjusted residual that is more than 2.0 indicates that the number of cases with the particular variables is significantly larger than would be expected if the null hypothesis were true, with a significance level of 0.05. An adjusted residual that is less than -2.0 indicates that the number of cases is significantly smaller than would be expected if the null hypothesis were true.

### 3. Results and discussion

#### 3.1. Task performance

In the NC speaking tests the situational tasks appear to be the most challenging overall. Luoma (2004: 158) calls situational tasks mini-simulations of reacting in situations. In practice, in situational tasks, what the candidates produce are one- or two-line samples from a dialogue, though in the test situation they are carried out as monologues. This presumes a clear definition of the context from the task, because the test taker has to jump into the ‘conversation’ for a line or two. The tasks only simulate a dialogue, which raises the question of how interactional these tasks are in terms of authenticity.

Task difficulty varies between test languages and, for instance, the task that was the most difficult for the English test takers was the easiest for the Finnish test takers. Explanations for this variation between the three tests and performance per task within tests can be many. As acknowledged in sociolinguistic-based second language acquisition (SLA) research, second language data do not represent a static phenomenon even at a single point in time. Many external variables, such as the specific task required of a learner, the social status of the interlocutor, gender difference and so forth, affect learner production (Gass, Selinker 2001: 222). This discussion aims to shed light on some issues that emerge from the current data.

**Table 1.** Distribution of scores in the five situational tasks (%)

	<B1			B1			B2		
	En	Swe	Fi	En	Swe	Fi	En	Swe	Fi
Situation 1	7.9	20.4	40.4	43.7	49.8	39.1	48.4	29.8	20.5
Situation 2	7.0	27.1	24.8	40.0	45.5	54.8	53.0	27.4	20.4
Situation 3	7.4	30.0	43.2	47.0	46.4	38.2	45.6	23.6	18.6
Situation 4	13.0	16.5	30.5	34.0	53.0	47.6	53.0	30.5	21.9
Situation 5	15.3	29.1	18.3	42.3	47.1	61.4	42.3	23.8	20.3

According to test results in the five speaking tasks in the three languages, the English test candidates performed the best (highest level of B2 and lowest of below B1) regardless of the task. The results for the Finnish and Swedish tests are more similar in distribution of grades, i.e. the typical score is B1, but Swedish test candidates performed slightly better overall with more scores of B2. One likely factor behind this result is the language of the written instructions and prompts, which in the Swedish and Finnish tests is the target language but in the English test is the L1 of the test takers. The recommendation is that the tasks should be easy to read. After all, preparation times are short and the tasks are not meant to directly measure the test takers’ reading skills. The choice of words can affect the level of difficulty in the task (for instance, ‘persuade’ in task 1). In general, the instructions and prompts should be simpler than the expected performance of the examinee (Luoma 2004: 169). As the general aim is to have equal interpretation of proficiency levels in all three languages, it is important to notice that understanding the written instructions and prompts may be easier for those taking the English test than for

those taking the Swedish and Finnish tests. On the other hand, those taking the English test have to use translation in production while those taking the Swedish and Finnish tests may be able to use at least some of the words directly from the written prompt. The instructions heard from the pre-recorded master tape in the target language only refer to the beginning and ending times of the task.

Situational tasks with brief prompts give a limited definition of the context and test takers have to use their imagination. In test situations, the speakers usually notice such task features that are important and meaningful to them and interpret the tasks in their own way. Thus the products are different and it cannot be expected that all candidates react to a given situation in the same way (Douglas 1998). Situational tasks also require quick reaction and response and may involve adopting a role of some kind. The test taker may or may not be familiar with the role. As the five tasks are different and are presented in quick succession they also require a quick change of roles, which may be both stressful and demanding for candidates.

Knowledge of vocabulary, phrases and idiomatic expressions makes it easier for candidates to succeed in the situation, particularly on level B1. Real life experience in similar speaking situations may not always be sufficient, as the skills of improvising and negotiating meaning are also required (e.g. Bygate 1987: 29). It must be pointed out that all five situational tasks discussed in this study require both routine expressions and improvisation and cannot be responded to by using routine expressions alone. Congratulating in task 4 is well suited for those candidates who know the common expression for the situation, and in task 5 some common adjectives such as *friendly*, *fair* are sufficient. Also, task 2 may be easy to pass at level B1 with fairly easy language: *I write words many times, I read words* etc. In task 1 test takers talked about their own hobby which can be regarded as a basic level skill but the second part of the task ('tempting') required more intermediate level ability. Situation 3 is not a routine one and seemed to require improvisation more than the others.

It is also important to bear in mind that situational tasks may not be interpreted as authentic and relevant everyday tasks by all candidates. For instance, the tasks may be more unfamiliar to immigrants than they are to those with a Finnish background (e.g. having hobbies). The tasks which simulate a work situation (tasks 3 and 5) may be more difficult for those who do not have much experience of working life in Finland.

## **3.2. Background factors and test performance**

### **3.2.1. Age**

While age is an important factor in the field of language learning and acquisition, for the current study it was found to be of little significance. The NC is a testing system targeted at adults as a way of having their language skills assessed regardless of how the skill has been acquired (through formal study or practical experience). While the age of test takers varies in a similar fashion in each of the test populations (Finnish 18–68, English 20–63, Swedish 15–72) there is some difference in which age group is the most represented. For Swedish and Finnish the largest test

taker age group is 31–40 years, to which 31% and 33% (respectively) of our sample belonged. For English, on the other hand, the largest age group is 21–30 years (45%). The younger age of the English test takers relates to the test being used as proof of English for deployment in international military cooperation. A typical English test candidate is a young Finnish male.

Our cross-tabulations show that age appears to be connected to performance in three test tasks but for different language tests. Younger age<sup>3</sup> is associated with better performance in talking about hobbies (task 1) in Finnish ( $\chi^2 = 25.7$ ,  $df = 10$ ,  $p = 0.004$ ) and talking about learning vocabulary (task 2) in Swedish ( $\chi^2 = 19.9$ ,  $df = 10$ ,  $p = 0.03$ ). On the other hand, the older age groups managed better in the congratulating of friends (task 4) ( $\chi^2 = 30.0$ ,  $df = 10$ ,  $p = 0.001$ ). While age was found to correlate with performance in these few instances the connection does not warrant making strong conclusions based on it. In some tasks some age groups performed slightly better than the statistical model expects but this cannot be shown to be at the expense of other age groups, i.e., that a task would be more suited to the younger than the older speaker.

### 3.2.2. Gender

Literature on issues of language and gender is vast, but there is very little that researchers actually agree on. However, Holmes (1998) has formulated a list of ‘sociolinguistic universal tendencies’ and suggests that women tend to focus on the affective functions of an interaction, use linguistic devices that stress solidarity, interact in ways that maintain or increase solidarity more often than men, and are stylistically more flexible than men. The situational tasks in this study did not aim at investigating gender difference, but some indication of difference in performance could be seen to link with these tendencies.

In the current data some correlations were found between gender and performance. In the Finnish test women performed better than men in task 1 ( $\chi^2 = 21.5$ ,  $df = 2$ ,  $p = 0.0001$ ), task 2 ( $\chi^2 = 25.7$ ,  $df = 2$ ,  $p = 0.0001$ ) and task 3 ( $\chi^2 = 13.6$ ,  $df = 2$ ,  $p = 0.001$ ). For English women performed better in task 2 ( $\chi^2 = 6.5$ ,  $df = 2$ ,  $p = 0.04$ ) and task 3 ( $\chi^2 = 9.4$ ,  $df = 2$ ,  $p = 0.009$ ). It must be noticed, however, that the English test candidates are a relatively homogenous group (young, male, with relatively high education) with only 14% ( $N = 31$ ) of test takers being female. This can create issues with interpretations of the statistical model, as quantitatively speaking, 31 female test takers cannot be considered representative of the larger population. However, in this data, in tasks 2 and 3 women performed better than men. In the Swedish and Finnish tests gender distribution is more even. In Swedish no gender effect was found.

On one hand we could see the interactional and emotive functions in tasks 1, 2 and 3 to relate to the solidarity and affection tendencies accredited to women but, on the other hand, the tasks also have an informative function. The result then supports the universal tendencies of gender only in part.

---

<sup>3</sup> E.g. In Finnish, task 1, fewer 21–30 year olds scored below B1 and more of them scored B1 than predicted by the statistical model.

### 3.2.3. Education

Level of education has been associated with language skills and performance, but again the connection is not simple. Education level is inevitably related to time (age and time spent studying) not to mention the sociocultural, attitudinal and motivational factors that are involved. In this data the cross-tabulation results of education level and situational task performance are inconclusive.

Among the Finnish test candidates a correlation was found between compulsory education (9 years of schooling) as the highest level of education and poorer performance in task 1 ( $\chi^2 = 61.6$ ,  $df = 10$ ,  $p = 0.001$ ), task 2 ( $\chi^2 = 61.2$ ,  $df = 10$ ,  $p = 0.0001$ ) and task 4 ( $\chi^2 = 56.6$ ,  $df = 10$ ,  $p = 0.0001$ ), while university education is associated with scores of B1 in the same tasks and polytechnic<sup>4</sup> education with scores of B2. The education levels of the test takers in the Finnish test cover the range from compulsory education to university education and this connection with low education level and poor performance is in that context logical. On the other hand, among the English test candidates, a slight dependence was found between vocational school education (high school level) and poorer performances (below B1) in most tasks. It should be noted though that this was indicated by the cross-tabulation residuals and not the chi-test which would have been affected by the homogeneity of the population. In this sample only 2% reported having the lowest education level i.e. compulsory education only. In years of study in the Finnish education system vocational school and high school are equal (12 years of schooling), so the difference is not in time of study but something else, which in this data appears in the education level cell. In the Swedish test, no dependence was found between education and performance.

### 3.2.4. Employment status

The background information sheet asked test takers to indicate their employment status, i.e., whether they are employed, self-employed, unemployed, student, student in labour market training, pensioner, stay at home parent or something else, which they could specify under the category 'other'. The assumption is that in the Finnish context this would have an effect on target language use opportunities and perhaps performance.

Among candidates for the Finnish test there was a correlation between being employed and performing better in the tasks. There was a clear finding of employed people gaining more scores of B2 in all tasks<sup>5</sup>. This is logical and relates to data discussed in the next section on language use. In the Swedish test, the effect was not as clear, but for tasks 2 ( $\chi^2 = 30.8$ ,  $df = 14$ ,  $p = 0.006$ ) and task 4 ( $\chi^2 = 32.4$ ,  $df = 14$ ,  $p = 0.004$ ) a correlation was found between being employed and better performance was found. There was also a connection between being unemployed and scoring below B1 for task 3 based on adjusted residual results. For the English test no dependence was found, but since the vast majority of candidates for this test are 'employed', with only a small percentage reporting any different status, the question does not appear relevant.

---

<sup>4</sup> In the Finnish system, polytechnic refers to tertiary level education: institutes of technology and universities of applied sciences.

<sup>5</sup> Task 1  $\chi^2 = 49.3$ ,  $df = 14$ ,  $p = 0.0001$ , task 2  $\chi^2 = 28.3$ ,  $df = 14$ ,  $p = 0.01$ , task 3  $\chi^2 = 41.8$ ,  $df = 14$ ,  $p = 0.0001$ , task 4  $\chi^2 = 55.7$ ,  $df = 14$ ,  $p = 0.0001$ , task 5  $\chi^2 = 47.6$ ,  $df = 14$ ,  $p = 0.0001$ .

### 3.2.5. Language use

Language use as reported by our sample reflects the Finnish context and the role of the target language as foreign or second language. The language use profile of candidates for the Finnish test indicates that for the majority the target language is used *almost daily* in the public domain: media 70%, reading 58%, transactional activities 66%, work 60% and education 58%. Target language use with family is divided more evenly on the scale, so that 42% report using Finnish almost daily and 30% not at all. Home is the most important domain for first language maintenance in the immigrant context which is also apparent in these language use figures. Finnish use with friends, on the other hand, is frequent; 57% reported almost daily use of Finnish in this domain.

Candidates for the English test present a completely different language use profile. The domain with most frequent use of English is following the media (54% almost daily) while family and transactional activities typically have no use of English (72% and 62% respectively). English use in education is spread more evenly with 30% claiming almost daily use and 34% no use. Language use with friends and acquaintances is the private domain where English is used to some extent, 53% reporting monthly use.

The language use profile of candidates for the Swedish test should be interpreted with caution. Since the population consists of both second language learners and learners of the second national language, the language use profile is an average of two potentially very different profiles. In the private domain, 24% report to using Swedish with the family almost daily and 50% not at all. In the friendship domain, 33% report almost daily use of Swedish and 17% no use. Daily Swedish use to follow the media is not as high as in the other languages: 42%, but on the other hand, only 5% do not follow Swedish media at all. Transactional activities (29% almost daily, 30% no use) and education (33% almost daily, 42% no use) as Swedish use domains are more diverse. Swedish use in the work domain could have been more frequent to reflect the cohorts as expected. After all, the typical motivation for taking the test in Swedish as a second national language is for work purposes and, on the other hand, a Swedish-as-a-second-language learner could be expected to use Swedish also as their work language, if they had chosen Swedish as the official language to test for immigration purposes. Nevertheless, in the current data 44% reported using Swedish at work almost daily and 22% not at all.

Cross-tabulations of language use data and task scores revealed various correlations. Among the Finnish test takers only four language use contexts out of the possible eight<sup>6</sup>, showed a correlation with the tasks. Daily use of Finnish with friends linked with higher scores for task 1 ( $\chi^2 = 17.0$ ,  $df = 6$ ,  $p = 0.009$ ), task 4 ( $\chi^2 = 15.9$ ,  $df = 6$ ,  $p = 0.02$ ) and task 5 ( $\chi^2 = 19.4$ ,  $df = 6$ ,  $p = 0.004$ ). Daily use of Finnish in the work domain was connected with better performance in task 3 ( $\chi^2 = 22.6$ ,  $df = 6$ ,  $p = 0.001$ ), task 4 ( $\chi^2 = 20.0$ ,  $df = 6$ ,  $p = 0.003$ ) and task 5 ( $\chi^2 = 30.1$ ,  $df = 6$ ,  $p = 0.0001$ ). There was also dependence between task 5 and daily use of Finnish with the family ( $\chi^2 = 16.0$ ,  $df = 6$ ,  $p = 0.01$ ) and for study ( $\chi^2 = 13.7$ ,  $df = 6$ ,  $p = 0.03$ ). The finding for language use with study is unexpected, since it indicates that test takers who use the target language less for this purpose perform better in the task.

---

<sup>6</sup> Language use contexts: family, friends and acquaintances, reading, writing messages, following the media, transactions, work, study.

It must be noted though that the way test takers interpreted the category 'study' can vary. It is likely that many who are in fact unemployed have stated that they study because they take part in labour-market training. Thus the study they refer to when reporting language use is in fact Finnish language studies.

For the English test data, connections with reported language use are much fewer. It does not appear that the overall strong performance of the English test takers, when compared with results from Finnish and Swedish tests, is influenced by language use. Only daily English use in the contexts of study and work was correlated with performance in the tasks. Good performance in task 2 correlated with daily use of English for study. There was a correlation between weekly English use at work and scores of B1 and not using English at work with scores below B1. Interestingly good performance in the description task (task 5) was only connected with daily English use for study. Very limited or no use of English in many, and particularly the private domains, does not appear to be connected with performance in these tasks. As above with the English data, these dependences are based on adjusted residual results.

For the Swedish test data, connections with candidates' reported language use are almost nonexistent. We suspect that the combined statistics of two different cohorts flatten the data into being less representative of either separate group.

Overall then language use in the family and friendship domains and, on the other hand in the work and education domains, has a connection with performance in these speaking situation tasks and in these language tests. This makes sense considering that the topics are from these domains. Although target language use, which in the case of second language learners is at a different level to foreign language learners, helps and prepares for the communicative test tasks, it is possible to acquire the relevant skills also through formal learning as demonstrated by the candidates for the English test.

## **4. Conclusions (and recommendations)**

The variation in results of rated performances in the three test languages in the five situational tasks was not great. Analysis of performance data with the Facets package indicates that the test items function adequately. Dependencies were found between rated performance in individual test situations and candidates' background factors, but these do not appear systematic across languages and do not, in the light of test success data, represent cases of bias. However, some differences warrant further discussion.

In light of the differences between foreign and second language learner performance it was surprising that English test takers performed so well in these situational tasks. Finland has a long tradition of classroom foreign language teaching/learning. Traditionally the formal setting has not supported everyday language use and the learners are not immersed in the target language community. However, in contemporary Finland, English is encountered on a daily basis through audio-visual mass media and various forms of popular culture. Still everyday language use situations may not be familiar unless they are practised as part of formal learning. Learning a second language while immersed in the language community implies also



acquisition and ability to use the language in a natural environment. Considering that for candidates for the Finnish and (some candidates for the) Swedish test the target language is a second language, the situational tasks should have favoured them (e.g. Sajavaara 1999).

Though some correlations were found between background information given by the test takers and situational task performance, they were not so significant as to have been crucial in passing the test. It is more likely that success and failure are in fact influenced by many background variables simultaneously rather than one single piece of information. The test taker's education, cultural background and frequency of language use may be important factors in determining success. For instance, it is no surprise that Estonian candidates did well in the Finnish test as they have the advantage of speaking a language which belongs to the same language family as Finnish. It is typical that Finnish learners with Estonian as their L1 learn very fast at the beginning stages of language learning and easily gain a level of proficiency which is not gained by someone with a more distant L1 background, though with the same duration of learning the language (Jantunen 2011, also Suni 1996). Spaan (2000: 35) has noticed that the results of a language test are influenced by the distance between the target language and the test taker's own language, the familiarity of the test form and test taker's education.

In general, situational tasks are best suited for testing the speaker's pragmatic, sociolinguistic and sociocultural knowledge and skills (register, politeness, social relations, idiomaticity etc.). Often expressions of emotions (regret, gratitude, negative and positive feelings) presume knowledge of certain phrases and the pragmatic features of language. These are also affected by the test takers' own social and cultural context. For instance, in the Finnish test, Swedish and Estonian candidates perform better than other candidates perhaps because their social context is closer to the Finnish one. To minimise these factors in a language test, construct and content validity, reliability and test usefulness should be focused on.

Passing the test is also always influenced by many other factors including the topic, the purpose of communication and the test taker's language learning history, as well as the physical situation, the channel and the test taker's own expectations (see e.g. Huhta 2010: 56). As the majority of Finnish and a part of the Swedish test takers took the test for immigration purposes, the stress in the situation is not without its effect in test behaviour. For instance, being in a language lab may be unfamiliar to some of these candidates. High-stakes tests – important to the test takers' future life – usually represent the value system and practices of the target culture and are often designed for a very homogenous target group. Though the NC is a proficiency test which can be taken by any person with no specified language learning background, and the test is not designed for any type of candidate in particular, the test still carries Finnish language testing conventions and traditions (e.g. use of the language lab), which may in part explain why the candidates for the English test performed so well. This raises the issue of fairness in the language test (Messick 1996, Shohamy 2000, Tarnanen, Mäntylä 2006: 116, 120).

The language of task instructions and prompts is significant to the speaking test. Second language learners are expected to have good reading and also listening comprehension skills in addition to speaking. How important understanding the task rubric is for passing the situation is difficult to ascertain on the basis of test

results only, but we may assume that it has some significance particularly for those whose comprehension skills are not very good.

The idea of authenticity is regarded as important for the construct validity of any language test. The aim is to write as authentic situational tasks as possible, but one must bear in mind that external similarity between a test task and a communicative situation does not necessarily fully reflect the nature of the situation. To be authentic, test tasks should be meaningful to the test takers, challenging enough, and engage them in activities that mirror their real-life contexts (e.g. O'Malley, Valdez Pierce 1996: 5, Douglas 2000: 18). The important issue here is perceived relevance. If we wish to have the test takers perform at their very best in the test situation, we should relate the test tasks in type and topical content to the target language use tasks outside the test situation. To be able to make interpretations on the test task and its correspondence with the 'real world' we also need information on how the test taker processes language and how he/she approaches the task (Bachman 1990, Huhta, Takala 1999).

Since the language test is aimed at any adult regardless of their language learning background, the test results indicate that the situational tasks discussed in this study are well suited for measuring test takers' ability to use spoken language in various language use situations.

## References

- Alderson, Charles J.; Clapham, Caroline; Wall, Dianne 1995. *Language Test Construction and Validation*. Cambridge: Cambridge University Press.
- ALTE Manual 2011 = *Manual for Language Test Development and Examining*. For use with the CEFR. Produced by ALTE [Association of Language Testers in Europe] on behalf of the Language Policy Division, Council of Europe, April 2011.
- Bachman, Lyle F. 1990. *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.
- Bachman, Lyle F.; Palmer, Adrian S. 1996. *Language Testing in Practice: Designing and Developing Useful Language Tests*. Oxford: Oxford University Press.
- Bygate, Martin 1987. *Speaking*. Oxford: Oxford University Press.
- Canale, Michael; Swain, Merrill 1980. Theoretical Bases of Communicative Approaches to Second Language Teaching and Testing. *Applied Linguistics* 1 (1). Oxford: Oxford University Press.
- CEFR 2001 = *Common European Framework of Reference: Learning, Teaching, Assessment 2001*. Council of Europe. Cambridge: Cambridge University Press.
- Douglas, Dan 1998. Testing methods in context-based second language research. – Lyle F. Bachman, Andrew D. Cohen (Eds.). *Interfaces Between Second Language Acquisition and Language Testing Research*. Cambridge: Cambridge University Press, 141–155.
- Douglas, Dan 2000. *Assessing Languages for Specific Purposes*. Cambridge: Cambridge University Press.
- Eckes, Thomas 2011. *Introduction to Many-Facet Rasch Measurement: Analyzing and Evaluating Rater-Mediated Assessments*. Frankfurt am Main: Peter Lang.
- Gass, Susan M.; Selinker, Larry 2001. *Second Language Acquisition: An Introductory Course*. 2nd edition. Mahwah, NJ: Lawrence & Erlbaum Associates Inc.
- Holmes, Janet 1998. Women's talk: The question of sociolinguistic universals. – Jennifer Coates (Ed.). *Language and Gender: A Reader*. Oxford/Malden, MA: Blackwell, 461–483.

- Huhta, Ari 2010. Suullisen kielitaidon arviointi: Mitä, miten ja miksi – ja voiko Euroop-  
palainen viitekehys auttaa siinä? – Sabine Grasz, Joachim Schlabach, Edeltraud  
Sormunen, Ari Huhta (Hrsg.). QualiDaF – Qualitätssicherung, Lernziele und Beru-  
teilungskriterien für den fachbezogenen Deutschunterricht. Jyväskylä: Jyväskylän  
yliopisto, Soveltavan kielentutkimuksen keskus, 31–56.
- Huhta, Ari; Takala, Sauli 1999. Kielitaidon arviointi. – Kari Sajavaara, Arja Piirainen-Marsh  
(Toim.). Kielenoppimisen kysymyksiä. Soveltavan kielentutkimuksen teoriaa ja  
käytäntöä 7. Jyväskylä: Jyväskylän yliopisto, Soveltavan kielentutkimuksen keskus,  
179–228.
- Jantunen, Jarmo H. 2011. Kansainvälinen oppijansuomen korpus (ICLFI): typologia, taust-  
tamuuttajat ja annotointi. – Lähivördlusi. Lähivertailuja, 21, 86–105. <http://dx.doi.org/10.5128/LV21.04>
- Luoma, Sari 2004. Assessing Speaking. Cambridge: Cambridge University Press. <http://dx.doi.org/10.1017/CBO9780511733017>
- Messick, Samuel 1996. Validity and washback in language testing. – Language Testing, 13  
(3), 241–256. <http://dx.doi.org/10.1177/026553229601300302>
- O'Malley, Michael J.; Valdez Pierce, Lorraine 1996. Authentic Assessment For English  
Language Learners. Practical Approaches For Teachers. Addison-Wesley Longman.
- Sajavaara, Kari 1999. Toisen kielen oppiminen. – Kari Sajavaara, Arja Piirainen-Marsh  
(Toim.). Kielenoppimisen kysymyksiä. Soveltavan kielentutkimuksen teoriaa ja  
käytäntöä 7. Jyväskylä: Jyväskylän yliopisto, Soveltavan kielentutkimuksen keskus,  
73–102.
- Shohamy, Elana 2000. Fairness in language testing. – Antony J. Kunnan (Ed.). Fairness and  
Validation on Language Assessment. Selected papers from the 19th Language Testing  
Research Colloquium, Orlando, Florida. Studies in Language Testing 9. Cambridge:  
Cambridge: Cambridge University Press, 15–19.
- Spaan, Mary 2000. Enhancing fairness through a social contract. – Antony J. Kunnan (Ed.).  
Fairness and Validation on Language Assessment: Selected papers from the 19th Lan-  
guage Testing Research Colloquium, Orlando, Florida. Studies in Language Testing 9.  
Cambridge: Cambridge: Cambridge University Press, 35–38.
- Suni, Minna 1996. Maahanmuuttajaoppilaiden suomen kielen taito peruskoulun päättövai-  
heessa. – Opetushallituksen moniste 11/1996.
- Tarnanen, Mirja; Mäntylä, Katja 2006. Toisen ja vieraan kielen oppijat yleisissä kielitutkin-  
noissa. – Päivi Pietilä, Pekka Lintunen, Heini-Marja Järvinen (Toim.). Kielenoppija  
tänään – Language Learners of Today. AFinLa vuosikirja 2006. Suomen soveltavan  
kielitieteen yhdistyksen julkaisuja 64. Jyväskylä: Jyväskylän yliopistopaino, 105–123.
- The Finnish National Board of Education and the University of Jyväskylä 2002. The Frame-  
work of the Finnish National Certificates. Helsinki: Edita Prima Oy.
- The Finnish National Board of Education and the University of Jyväskylä 2011. The Frame-  
work of the Finnish National Certificates. Tampere: Tampereen yliopistopaino.
- Weir, Cyril J. 2005. Language Testing and Validation: An Evidence-Based Approach. Bas-  
ingstoke: Palgrave Macmillan.

**Sari Ahola** (Centre for Applied Language Studies, University of Jyväskylä, Finland) is involved in language assessment and activities for the development of the Finnish National Certificates of Language Proficiency.  
sari.ahola@jyu.fi

**Tiina Lammervo** (Centre for Applied Language Studies, University of Jyväskylä, Finland) is involved in language assessment and activities for the development of the Finnish National Certificates of Language Proficiency.  
tiina.lammervo@jyu.fi

**Reeta Neittaanmäki** (Centre for Applied Language Studies, University of Jyväskylä, Finland) is involved in language assessment and activities for the development of the Finnish National Certificates of Language Proficiency.  
reeta.neittaanmaki@jyu.fi

**Sari Ohranen** (Centre for Applied Language Studies, University of Jyväskylä, Finland) is involved in language assessment and activities for the development of the Finnish National Certificates of Language Proficiency.  
sari.ohranen@jyu.fi

**Henna Tossavainen** (Centre for Applied Language Studies, University of Jyväskylä, Finland) is involved in language assessment and activities for the development of the Finnish National Certificates of Language Proficiency.  
henna.tossavainen@jyu.fi

# KOLME ERINEVA KEELETESTI RÄÄKIMISOSADE VÕRDLUSTULEMUSED

Sari Ahola, Tiina Lammervo, Reeta Neittaanmäki,  
Sari Ohranen, Henna Tossavainen

Jyväskylä Ülikool

Artikkel tutvustab uuringut, milles võrreldi õppijate tulemusi kolme erineva keeleksami rääkimise allosas. Tulemusi püütakse seletada eksaminandide poolt antud info kaudu nende demograafilise, haridusliku, professionaalse ning keeleõppe tausta kohta. Uuringu eesmärk oli panustada eksamiülesannete konstrukti (mõõdetava omaduse) valiidsusse ja selgitada välja, kas ülesanded olid ühtmoodi valiidsed kõigi kolme keele puhul.

Uuringus viidi 2011. aasta sügisel läbi katse, mille käigus võrreldi rääkimise osa Soome riiklikel kesktaseme eksamitel (*The National Certificate of Language Proficiency*): soome, rootsi ja inglise keeles. Eksaminandidele anti selleks viis sarnast suhtlussituatsiooni. Rootsi keele testi puhul koguti andmeid veel ka 2012. aasta kevadel. Tulemusi ja taustafaktoreid analüüsiti sagedus- ja protsentjaotusandmete põhjal, muutujatevahelisi seoseid risttabelite abil.

Üksikvastuste teooria (*Item Response Theory*) põhine analüüs näitas, et eksamikorraldaja küsimustepangast valitud ülesanded toimivad kõigi kolme keele puhul hästi nii küsimuste raskuse kui ka eristusvõime osas. Uuringu käigus selgus aga, et selleks, et piisava adekvaatsusega kirjeldada eksaminandide tulemuste seost nende taustainfoga, on vaja enam andmeid.

Uuringu tulemused näitavad, et eksami sooritamise edukus sõltub tõenäoliselt paljudest taustamuutujatest ning muudest testi situatsiooniga seotud muutujatest. Ilmnes ka sõltuvusi eri testsituatsioonides saadud tulemuste ja kandidaatide taustatunnuste vahel, kuid need ei esinenud eri keelte testide vahel süstemaatiliselt ning ei ole seetõttu selgelteristuvaks eduka testisoorituse mõjufaktoriks. Siiski võivad eksaminandi haridus, kultuuritaust ja keelekasutuse sagedus olla olulisteks soorituse edukust mõjutavateks teguriteks.

**Võtmesõnad:** teise keele ja võõrkeele testimine, suuline keel, valiidsus, soome keele eksam, rootsi keele eksam, inglise keele eksam



## ACQUISITION OF COMPOUNDS IN ESTONIAN AND RUSSIAN: FREQUENCY, PRODUCTIVITY, TRANSPARENCY AND SIMPLICITY EFFECT

Reili Argus, Victoria Kazakovskaya

**Abstract.** This article discusses the early phases of acquisition of noun compounds in two typologically different languages – Estonian and Russian. Longitudinal data on three Russian and two Estonian typically-developing monolingual children and their caregivers was analysed and compared in relation to first emergence, the development of different compound types, productivity, frequency, simplicity and transparency of compounds, and also to the impact of morphological wealth in the input on the output.

Although the frequency of compounds is somewhat different in the two languages being observed, the first compounds emerge in the speech of Estonian and Russian children almost at the same age and quite early, at the end of the second year of life.

In both Estonian and Russian, children first acquire productive patterns in both languages, Estonian and Russian, respectively. Factors such as simplicity and frequency are intertwined in the acquisition process. Simple structures occur at the same time as frequent ones but it cannot just be assumed that simplicity is a stronger factor influencing the choice of first noun compounds than frequency. Frequency effects are not straightforward: there are differences in the individual usage of compounds by the caregivers, and the frequency of compounds in child-directed speech does not always seem to be reflected in the speech of the children.\*

**Keywords:** language acquisition, compounds, child-directed speech, frequency, productivity, transparency, simplicity, Estonian, Russian

---

\* The research was carried out with the financial support of the RF President's Foundation (grant No 1348.2012.6 'The School of Functional Grammar in St. Petersburg').

# 1. Introduction

Little or no research has been carried out on the acquisition of compounding in Finno-Ugric languages, such as Estonian, nor in Indo-European languages with rich morphology, such as Russian. In general, the acquisition of compounding has been much less investigated than the acquisition of inflection in many languages (cf. Savickienė, Dressler 2007, Stephany, Voejkova 2009). This is probably mainly due to the fact that the study of compounds is quite a recent development.<sup>1</sup> In Balto-Slavic languages, compounds have not been studied since it was assumed that there are few. However, there are works that analyse occasional compounds in Russian children's speech (Kharchenko, Ozerova 1999) and a study of children's compounds in Lithuanian has been started (Dabašinskienė 2010). The acquisition of skills of compounding in Estonian children with specific language impairment is described by Marika Padrik (2010).

This article is an attempt to add data to the areas as yet not covered in first language acquisition and discusses the early phases of acquisition of compounds in two typologically different languages – Estonian and Russian. Investigation into the acquisition of compounds has so far been concentrated on languages that possess many compounds but not a very rich morphological system (Nicoladis 2006, Berman 2009). Both Estonian and Russian are languages with a rich morphological system, although the system of compounding is, to some extent, different and a comparison with the available data of other languages could offer evidence of how the system of compounding with which children are faced can influence the process of acquisition. Thus, we would like to focus on how the differences in two systems reflect in acquisition and try to answer the question whether children acquire productive patterns first, or whether it is merely frequency exerting an influence.

Compounding is a productive word-formation process in many languages and especially in the Germanic ones (Dressler et al. 2010). According to Natural Morphology, a child acquires the more productive parts of grammar first (cf. Dressler 1999, Kilani-Schoch, Dressler 2005). It can therefore be assumed that the **productivity** of compounding in different languages and the productivity of certain compounding patterns can affect the age of emergence of compounds, the choice of first patterns, and the speed of acquisition of compounding.

The age and speed of the acquisition of morphology has been argued to be dependent on the richness of the system children encounter (Laaha, Gillis 2007). **Frequency** of certain structures, like inflections, derivation suffixes, etc., has been noted to have a great importance on the order of acquisition of different grammatical elements (cf. Argus 2008, 2012). As compounding has a different degree of productivity in different languages and as a result different compounds have different frequency in the input (child-directed speech, CDS), it could be assumed to have an effect on the acquisition of compounding.

Eve Clark and her colleagues have argued that the principle of **simplicity**, amongst others, is important in child word formation (Clark, Berman 1984). Simplicity is measured by the degree of change in a form: the less the word changes, the simpler it is (ibid.: 548). For example, it has been argued that English simple noun+noun (N+N) compounds are easier to acquire than compounds with the

<sup>1</sup> See, in particular, a short description of the acquisition of nominal compounds in Dutch, English, French, German, Greek, Hungarian, Italian and Spanish according to different models compounds (Rivista di Linguistica 1992), as well as English (notably Clark et al. 1985, 1986, Nicoladis 2002), Swedish (Mellenius 1996, 1997), Hebrew (Berman, Clark 1989, Clark, Berman 1984, 1987, Berman 2009), Greek (Stephany 2010) and French (Nicoladis 2002, 2007).



suffix *-er* (Nicoladis 2006: 119). Compounds with differing degrees of simplicity can be found in both Estonian and Russian and an early acquisition of simpler patterns could be supposed, based on work previously conducted.

Simplicity can be considered to be connected with **transparency**. Morphosemantically transparent compounds will be more relevant and may be acquired earlier than opaque ones. According to Libben's basic scale of transparency (1998), the most transparent compounds are those with a structure where both head and non-head (modifier) are transparent. In other words: compounds with the meaning of each constituent morpheme contributing directly to the meaning of a compound (*door+bell* in English, *radio+peredacha* 'radio+transmission=broadcast' in Russian, *seene+supp* 'mushroom+soup' in Estonian) are easier to acquire than others. It also has to be taken into account that interfixed compounds are also morphosemantically less transparent than interfixless ones, because the interfix does not contribute to the meaning of the compound (Dressler et al. 2010: 15). Finally, compounds are semantically more descriptive than derivations, therefore morphosemantically more transparent.

Some prosodic constraints have also been considered to influence the acquisition of morphology (Laalo 1994, Argus 2008). In the early stage of acquisition children use shorter words and prefer disyllabic patterns. For example, in Finnic languages the period of disyllabic structures (among them compounds consisting of two disyllabic words) is often to be considered to last for even five or six months (Laalo 1994: 340, Argus 2008: 32). Since compounds are longer than root-words, it could be that the children do not acquire compounds very early and start to use them only after the disyllabic period, or prefer disyllabic patterns also in compounds. Moreover, it is known that compounds are also harder to read than, for example, phrases (Böer et al. 2012) and on this basis it could be supposed that they could also be harder to comprehend. At the same time there can be some factors which facilitate the acquisition of compounds: the formation of compounds is different from ordinary inflection – it is often merely a juxtaposition of two or several stems. Therefore it could be assumed that it could be easier for the child to acquire than, for example, a derivation.

In this paper we discuss the early phases of first language acquisition of noun compounds. The restriction to nominal compounds, i.e. compounds which function as Nouns (Ns), is justified by the rarity or very different structure of verb compounds in many languages. Compounds are analysed comparatively in Estonian and Russian, concentrating on different factors, like frequency, productivity, simplicity, and transparency, which could influence the age and the speed of their acquisition.

The longitudinal data of three typically-developing, monolingual Russian and two Estonian children and their caregivers has been analysed and compared according to first emergence, development of different compound types, productivity, frequency, simplicity and transparency of compounds, and also the impact of morphological wealth in the input on the output (child's speech, CS).

## 2. Compounds in Estonian and Russian

Compounding has a different degree of productivity in the languages under observation and it is a productive strategy of word formation in Estonian. Although compounding includes the formation of compound verbs, nominals (Ns, ADJs, PRONs) and adverbs, it is very productive only in nouns and adjectives. Noun compounds are quite frequent in colloquial speech and also in child-directed speech (henceforth CDS) (e.g. compounds like *lutipudel* ‘bottle with pacifier’, *riisipuder* ‘rice porridge’). But they are even more frequently used in terminology, for example, in scientific and technical disciplines like biology (most parts of names of plants are compounds) and engineering.

Although compounding is quite often used also in Russian, other means of word formation are used more frequently, namely, suffixation, prefixation, a combination of suffixation and prefixation, substantivation etc. (Russian Grammar 1980). Therefore, compounds used to be less frequent, but are now becoming more frequent in CDS and, consequently, in CS.

Compounding is not always used when building new words in Russian, for example, it is uncommon in verb building. Such rare examples are ‘calques’ from Greek: *blagodarit* ‘to thank’ or *zloslovit* ‘to say spiteful things’. However, compounding is a very useful way of building nouns: *knigoljub* ‘book+to love=book lover’, *inoplanetjanin* ‘different+a planet=alien’, and adjectives: *ostrumnyj* ‘sharp+mind=witty’, *svetlo-goluboj* ‘light+blue=light blue’, similar to Estonian.

The Corpus of Spoken Russian data has shown that compounds are not frequent in colloquial speech, but with some exceptions like *kofemolka* ‘coffee+to mill=machine for milling of coffee’, *nervotrepka* ‘nerves+to spoil=hassle’. However, they are favoured in scientific and technical disciplines as in Estonian: *mirovozzrenie* ‘world+view=world-view’, *jazykoved* ‘language+to know=linguist’.

Table 1 presents the classification and comparison of compound nouns in Estonian and Russian according to the word class and inflectional form of the head and modifier.

Compounds in both languages can concatenate components with different structures<sup>2</sup>. Table 2 provides a classification and comparison of compounds in Estonian and Russian according to the degree of complexity of constituents forming the compound.

---

<sup>2</sup> A compound has been defined here according to structural characteristics – a compound is a word consisting of at least two root morphemes (see also Kerge 1990: 4).

**Table 1.** Estonian and Russian compound nouns according to the word class, form of constituents and productivity of formation pattern (Vasilevskaja 1962, Russian Grammar 1980, Zemskaja 1992, Erelt et al. 1995, Gudilova 2005)

Word class	Estonian	Russian
<b>Head: Verb</b>	Compound nouns consisting of head functioning as a verb are missing in Estonian.	The dominating structure of compounds <b>X+V</b> : the modifier can be a noun (or ADJ, PRON, NUM, V) followed by a verb as the second member, a head of the compound.  The compound might be with an interfix (o, e etc.), and with suffixation – zero (∅): <i>vod-o+pad</i> ‘water+fall+∅=waterfall’ or material one: <i>mor-e+plava-tel</i> ‘sea+sail+suffix <b>tel</b> =seaman’. A complex suffixation is the most productive pattern.  Compounds where a verb is the first member, and a head ( <b>V+X</b> ), are relatively rare: <i>sorv-i+golova</i> ‘tear away+head= madcap’.  The verb part with an interfix coincides with IMP.SINGL. The sub-pattern is unproductive.
<b>Head: Noun</b>	All Estonian noun compounds belong to this type.	Compounds might be with an interfix: <i>les-o+step</i> ‘forest+steppe=forest-steppe’ or without one: <i>plachsh+palatka</i> ‘raincoat+tent=duckbill’. This pattern is less productive.
<b>Modifier: Substantive</b>	Most productive and common pattern in Estonian: <i>riisi+puder</i> ‘rice+porridge=rice porridge’.	This pattern is productive: <i>nos-o+rog</i> (horn-nose) ‘nose+horn=rhinoceros’; <i>jug-o+vostok</i> ‘south+east=southeast’.
<b>Modifier: Adjective (or its part in Russian)</b>	The pattern is not very productive: <i>noor+tööline</i> ‘young+worker=young worker’.	The pattern is very productive: <i>sukh-o+phrukt</i> ‘dry+fruit=dried fruit’. Quite often some part of ADJ and N are concatenated (a kind of abbreviation): <i>med(icinskaja)+sestra</i> ‘med(ical)+sister=nurse’.
<b>Modifier: Pronoun</b>	The pattern is unproductive: <i>enese+kaitse</i> ‘self+defence=self defence’.	The pattern is not very productive: <i>sam-o+proverka</i> ‘oneself+check=self-test(ing)’.
<b>Modifier: Numeral</b>	The pattern is unproductive: <i>kahe+kõne</i> ‘two:GEN+speech= dialogue’.	The pattern is not very productive: <i>perv-o+istochnik</i> ‘first+source=original source’.
<b>Modifier: Verb</b>	The pattern is productive: <i>istu-ma+jäämine</i> ‘sit-INF+staying=remain seated’.	The pattern is unproductive: <i>vert-o+let</i> ‘turn+fly=helicopter’.
<b>Modifier: Particle</b>	The pattern is not very productive: <i>all+kiri</i> ‘under+letter/writing=signature’.	The pattern is unproductive: <i>pol-u+mrak</i> ‘semi+darkness=shade’.

**Table 2.** Constituents of compounds in Estonian and Russian and productivity of formation pattern

Constituent	Estonian	Russian
<b>Root</b>	Productive and frequent pattern: <i>paber+raha</i> 'paper:NOM+money=paper money. The most productive pattern is with singular genitive case form which is a base form without any suffixes occurring as a modifier: <i>seene+supp</i> 'mushroom:GEN+soup=mushroom soup'.	May be used as an interfix: <i>sam-o+let</i> (aeroplane) 'oneself+fly= aeroplane' or without one: <i>divan+krovat'</i> 'sofa+bed=sofa bed'. The pattern is productive.
<b>Stem</b>	The pattern is productive: <i>kauba+ndus+töötaja</i> 'trade+derivational suffix+worker=salesperson'.	The pattern is very productive: <i>kreslo+kachalka</i> 'armchair+rocking chair=rocking chair'.
<b>Inflected form</b>	The pattern is productive: all noun case forms can occur as a modifier: <i>kala-l+käik</i> 'fish-ALL'+going:NOM =fishing'; <i>küla-s+käik</i> 'village-ILL+visit=visit'.	The pattern is productive: <i>zav+kafedroj</i> 'head (shortened from <i>zavedujuchshij</i> )+sub-department:INSTR SINGL=head of sub-department'. Typical case form of inflected modifier is INSTR.
<b>Interfixed form</b> (containing a meaningless linking element)	The pattern is not very productive: <i>töötaja-s+pere</i> 'worker-s-family=people who work together'.	The pattern is productive: <i>gr'az-e+vod-o+lechenie</i> 'mud+water+cure =water-and-mud cure'.
<b>Part of a word</b>	The pattern is productive: <i>lamamis+tool</i> ( <i>lama+mise+tool</i> ) 'lay+derivational suffix+chair=reclining chair'; <i>kiir+tee</i> ( <i>kiire+tee</i> ) 'quick+road=highway' – the last phoneme of a stem disappears.	The pattern is very productive: <i>sten(naja)+gazeta</i> 'wall+newspaper=newspaper on the wall'.
<b>Phrase</b>	The pattern is not very productive: <i>kahe+inimese+voodi</i> ( <i>kaks inimest+voodi</i> ) 'two:GEN person:GEN+bed=bed for two persons'.	The pattern is unproductive: <i>vremj-a+provozhdenie</i> 'time+pass= pastime'.

The systems of Estonian and Russian compounds differ (see Table 2).

1. The type of head-word. In Russian it can be a noun or a verb, which mainly follows a noun, while in Estonian it is always a noun, if the noun is a compound itself. However, the similarity is that both languages have compounds with a head that is a noun.
2. The inflected forms. In Estonian compounds the inflected form is first in a compound, whereas in Russian, it is second.
3. The interfixed form. In Russian the interfix (only vowels) is widely used, while in Estonian (where the interfix could be a consonant) it is quite rare. Moreover, in both languages, compounds could concatenate not only the roots of words, but also the stems.
4. The degree of productivity. In Russian language where the modifier is an adjective are productive, while the most productive pattern of compounding in Estonian is one consisting of noun in genitive as modifier following

the patterns where the modifier is an adjective or a verb. However, the compounds in which the modifier is a pronoun, numeral or particle are equally rare in both languages.

### 3. The data

Table 3 gives a general overview of the data used for the analysis. The longitudinal corpora of three Russian (St. Petersburg) and two Estonian (Tallinn and Tartu) children (~135 hours of recorded speech (119.5 hours for the Russian children, 17 hours for the Estonian ones) were transcribed using CHILDES<sup>3</sup>. Although the quantity of Estonian data was significantly smaller, the number of compounds is almost the same as in the Russian corpora.

**Table 3.** Number of compound types and tokens in the children's speech (Vanja, Liza, Philja, Hendrik, Andreas) and the CDS. MLU – Mean Length of Utterance (here counted in words)

Russian	Compound Types	Compound Tokens	Types/Tokens
Vanja 1;9 (MLU 2.008) – 3;6 (MLU 3.210)	22	119	5.36
Vanja's CDS	34	251	7.38
Liza 1;7 (MLU 1.008) – 3;6 (MLU 4.170)	18	73	4.05
Liza's CDS	39	95	2.43
Philja 1;5 (MLU 1.009) – 2;8 (MLU 2.900)	11	68	6.18
Philja's CDS	16	144	9.00
<b>Estonian</b>			
Hendrik 1;8 (MLU 1.651) – 2;5 (MLU 2.811)	28	70	2.50
Hendrik's CDS	93	200	2.15
Andreas 1;7 (MLU 1.093) – 3;0 (MLU 3.044)	167	329	1.97
Andreas's CDS	184	511	2.78

The total number of compound types and tokens differs in sub-corpora. Among the Russian children, Vanja has 22 types and 119 tokens, Liza has 18 and 73, and Philja has 11 and 68 within their material, while one of the Estonian children, Hendrik, has 28 types and 70 tokens<sup>4</sup> but the other, Andreas, has 167 types and 329 tokens in his data.

<sup>3</sup> Child Language Exchange Data System, <http://childes.psy.cmu.edu> (accessed 23 October 2012).

<sup>4</sup> It must be mentioned here that the corpora of Hendrik is somewhat restricted, the quantity of speech samples is much smaller in the early period (only 20 minutes per month).

## 4. Acquisition of compounds

### 4.1. Emergence of first compounds

First compounds emerged in the speech of the Estonian children early: *lume+pall* ‘snow+ball’, *notsu+sig*a ‘pig:DEM+pig=cavy’ (Andreas 1;9), *kaka+nunnu* ‘shit:DEM+poop:DEM’ (Hendrik 1;10). In the beginning both children had very few compounds: there were four compound types and four tokens in Andreas’s speech and only one compound type with two tokens occurring in the speech of Hendrik.

In the speech of Andreas the N+N compositions emerge at the beginning of the protomorphological period<sup>5</sup>, in the case of Hendrik it emerges three months before this period or, in other words, before noun inflection starts to develop productively. Compounds are used only in one case form at first by both children, the first true mini-paradigm appeared only at 2;7 in Andreas’s speech. Compounds as longer words do not cause any difficulties for children. Andreas starts to use first compounds at the end of his disyllabic period. First compounds emerge in the speech of Hendrik at the period when he has almost only disyllabic words. His first compound clearly demonstrates the preference for using trochaic phonological pattern – he uses the diminutive form from one-syllable form *nunn* (junn) > *nunnu* to adapt the whole compound to the pattern consisting of two disyllabic words *kaka+nunnu*.

Comparing compounding with other word formation possibilities, it must be mentioned that compounding emerges in the speech of Andreas at the same time as other word formation strategy starts to develop – he used first compounds at the same time as first derivated verbs. Hendrik’s compounding emerges much (6 months) earlier than derivation (cf. Argus 2012: 10).

According to Nicoladis (2006: 110), children seem to have little trouble ordering N+N compounds. The same can be stated on the basis of the Estonian data. The ordering of parts in compounds is not absolutely clear for both children, but only at the start of using compounds. Hendrik used both *abi+iuiu* and *iuiu+abi* ‘emergency car’ at 1;11, even later, at 2;4 he used a compound *onu+kass* (should be *kassi+onu*) ‘uncle+cat=male cat’ and from the context we can see that he intended to use the correct version *kassi+onu*. Andreas had only one such ordering error *pika+päka* (should be *päkapikk*) ‘dwarf’ at 2;1.

First compounds emerged in the speech of the two Russian children at the same age 1;9: e.g. *sam-o+let*:DIM (‘oneself+to fly=aeroplane’) (2 types/7 tokens for Philja); *photo+apparat* (‘photo+apparatus=camera’)<sup>6</sup>, *sy-r-o+ezh-ka* (‘raw+to eat=russula’) (2 types/4 tokens for Liza). Vanja’s earlier compounds emerge at 2;2: e.g. *metall-o+lom+ø* ‘metal+to break=scrap’, *vert-o+let+ø* ‘to turn+to fly=helicopter’, and also a compound which is typical for colloquial speech *chern-o+plod-ka* ‘black+foetus=black chokeberry’ (4 types/12 tokens).

In the speech of all the Russian children the first compounds emerge at the stage of protomorphology, because the number oppositions of nouns (SING vs. PL) are already observed, and the contrast verb forms (the mini-paradigms) are used. In contrast to early Estonian compounds, the erroneous order of parts is not present in the Russian data.

<sup>5</sup> The protomorphological period is considered to start from the emergence of first mini-paradigms (for criteria, see Bittner et al. 2003: xvi) and signifies the beginning of active acquisition of morphology.

<sup>6</sup> This concept *photo+apparat* ‘camera’ is named by a specific children’s word (*pack*), but at the same time in the girl’s speech the word *apparat* ‘apparatus’ does exist, which she uses for the ‘camera’.

Although the frequency of compounds is somewhat different in the two languages under observation, the first compounds emerge in the speech of the Estonian and Russian children at almost the same age, at the end of second year of life, which is quite early.

#### 4.2. The structure and productive patterns of early compounds

A productive compound pattern is one that is applied freely and unconsciously to new loanwords (Dressler et al. 2005). In the case of Estonian, the most productive pattern among N+N compounds is a pattern with modifier in the genitive like *rüüsi+puder* 'rice:GEN+porridge'. Estonian early compounds are of a similar structure. Almost all compounds in the speech of both children were N+N, where the modifier is a singular genitive stem (see for patterns the tables 1 and 2): *une+laulu* 'sleep:GEN+song=lullaby' (Andreas 1;8), *tati+nina* 'snot:GEN+nose' (Hendrik 2;0).

Only some compounds with different structure (usually the first part as an onomatopoeic word) were found from the early compounds: *iüüü+abi* 'sound of the emergency car+help=emergency car' (Hendrik 1;11). First compounds with different structure start to emerge in Andreas's speech at 2;0: *kustu+kumm* 'erase+gum=rubber'. These compounds belong to different formation patterns, the modifier could be a noun in nominative, like *mootor+ratas* 'engine+bicycle=motorcycle', as well as an adjective *vana+isa* 'old+father=grandfather' (2;4), or a shortened genitive form like *hingamis+rohi* 'breathing+medicine=inhalator', or a verb stem like *röst+sai* 'roast+bread=toast' (2;1). The number of compounds with modifier in singular nominative remained small during the whole observation period, and there were only five such compounds in the data. First compounds with a modifier in plural genitive occur at Andreas's age 2;8: *isa-de+päev* 'father-PL.GEN+day'.

The distribution of different structures of compounds remains quite similar from the age of 2;0 until the end of the observation period: Approximately two-thirds of all compounds in child speech belong to the most productive pattern (N+N, modifier in singular genitive) and one-third of compounds consist of compounds with different structure. The pattern where the modifier is an adjective (e.g. *must+sõstar* 'black+currant' (Andreas 3;0)) remains the most frequent pattern among infrequent patterns.

The frequency of compounds formed according to the most productive pattern in CDS changes when the child gets older: it can vary from 90% at the beginning of the observation period down to 67% at 2;8. Hence it can be argued that productivity and frequency are factors which are intertwined in the process of acquisition of compounds because the frequency of compounds formed according to productive pattern can change in CDS. It can be seen as a carer's fine-tuning.

The structure of children's first compounds is similar in modern Russian. The boys' data shows that the predominance of X+V compounds (like *tepl-o+khod* 'heat+to go=motor boat') is very marked, with X+N type (*zoo+park* 'zoological+garden=zoo', *mul't+phil'm* 'animation+film=animated cartoon') ranking behind them, but the predominance of this pattern was found in the girls' corpus.

Both the patterns X+V and X+N are the most numerous types of early compounds in Russian Child Speech. Moreover, in all cases these first compounds are subordinate endocentric right-headed ones, according to the classification of compounds (Scalise, Bisetto 2005).

Then, within the X+V type of compound, such structures as PRON+V (*sam-o+kat* 'oneself+to roll=scooter'), ADJ+V (*syr-o+ezhka* 'raw+to eat=russula (an edible mushroom)') and V+V (*vert-o+let* 'to turn+to fly=helicopter') were found.

In the frames of the X+N type of compound, in second place in the boys' corpora was the structure N+N: *nos-o+rog* 'nose+horn=rhinoceros' (V. 2;7). Nevertheless, there are less numerous compounds in comparison with the prototypical presenters of these models, namely N+V and ADJ+N.

The members of compound are connected by one of the prototypical interfix *o*: e.g. *metall-o+lom+ø* 'scrap', *sam-o+let+ø*=airplane'.<sup>7</sup> In most cases, the formation of compounds is followed by the addition of suffixation (usually zero one, see examples above), which also corresponds to the leading trends in Russian compounding. Zero suffixation occurs in the verb part: *-let* (*samolet* 'airplane') > *let(et)* 'to fly'+ $\emptyset$ ', etc.

In Vanja's and Liza's data X+N compounds where two words (stem+ending, which is material or zero one) are concatenated by a hyphen prevailed: e.g. *sorok-a+belobok-a*<sup>8</sup> 'magpie+white-sided=magpie' (Vanja 2;10).

As to preferences and more detailed characteristics of Russian children's compounds, it can be observed that Vanja's data displayed the most variation in compounds in both patterns and their types, although earlier compounds emerge quite late (1;9 is the time of appearance of the first compounds in his CDS). This corpus has the middle 'types/tokens ratio' (TTR) of compounds: 5.36 for CS, 7.38 for CDS.

Liza's data shows, as mentioned above, a lot of so-called hyphen (N+N) models of compounds. The fact that there is an abundance of these words, and most of them are heroes of Russian folk tales, can be explained by the significant role of book reading in the life of this family. The girl's mother (who is the main caregiver) works as a teacher. TTR of compounds in this corpus is the lowest: 4.0 for CS, 2.43 for CDS. In Philja's data the list of compounds varied less but they are very frequent. This corpus has the highest TTR of compounds: 6.18 for CS and 9 for CDS.

Thus, the data of Estonian and Russian children's speech reveal a similar tendency in the emergence of first compounds: children start with the most productive patterns of compounding in both languages regardless of their perceived complexity.

### 4.3. Simplicity and morphosemantic transparency

Formation of compounds is a simple process in almost all cases in Estonian, only in some cases will a meaningless element be interfixed. The modifier can, in some patterns, be an inflected form like allative or illative case form, but these patterns are not frequent. The most frequent and productive pattern is with modifier as a root in singular genitive and it cannot be considered to be more complicated than the pattern with singular nominative. The main reason is that the genitive case form serves like a base form for other (semantic) cases, it has no suffix and

<sup>7</sup> It does not correspond with some comments on compound acquisition (Dressler et al. 2010: 15), at least on these Russian data.

<sup>8</sup> Interestingly, this part (*beloboka*) of the compound is itself a compound, formed by concatenated *bel-o+bok-a* 'white+side' with a special rhyming ending.



is often the same form as the nominative. Hence, when Estonian children start with compounds consisting of modifier in genitive case, we cannot claim that they start from more complicated patterns because the compounds with genitive noun as modifier can be as simple for them as those with nominative modifier. These compounds are also transparent, as both members of the compound, head and the modifier are contributing directly to the meaning of a compound (e.g. in the compound *une+laul* ‘sleep+song=lullaby’ both parts are directly contributing to the meaning of the compound).

The first compound with opaque meaning occurs in Andreas’ speech at the age of only 2;0, and it was *meri+siga* ‘sea+pig=cavy’. Hendrik (2;2) used a compound *mesi+mumm* ‘honey+spot=bee’ in which the head is opaque. There were very few compounds in CS where the head or modifier of the compound was opaque: *rohu+tirts* ‘grass+bit=grasshopper’ (Andreas 2;6), *musi+mumm* ‘kiss+spot=dear’. This kind of compounds also seems infrequent in CDS.

In Russian the most frequent compounds are formed according to an X+V pattern which is accompanied by zero suffixation. It seems that the process of suffixation complicates compounding: *molok-o+voz(i-t)* ‘milk+carry=milk tanker’. However, a zero suffix, which is added to the verb stem *voz(i-t)+ø* ‘to carry’, opens this stem, truncating both a thematic vowel (*a, e, i*) and a form-building suffix of an infinitive (*t*). Simultaneously a bare root is ‘uncovering’: *voz*; see also *pyl-e+sos* ‘dust+suck=vacuum cleaner’: *sos > sos(a-t)+ø*. Modifiers are a root: *molok-o* ‘milk’ or *pyl* ‘dust’.

The use of interfixes, in our opinion, is needed for phonetic reasons. Interfixes open syllables (according to ‘the law of open syllable’) and avoid clusters, which are acquired by Russian children with some difficulty (Gvozdev 1995). In other words, we tend to consider the use of interfixes as a bootstrapping element. Consequently, the forming of such compounds is quite a simple process, and as our results have shown, all children start with this pattern. Until 2;5 in Vanja’s and Liza’s data this is the main type of compounds<sup>9</sup>: Liza has 12 such compounds out of 18, Vanja has 12 out of 22, Philja 8 out of 11.

According to an analysis, modifiers might be Ns (6 out of 11 for Vanja, 3 out of 8 for Liza, 4 out of 8 for Philja): *beton-o+meshalka* ‘concrete+mixer=concrete mixer’; PRONs (4 out of 11 for Vanja, 2 out of 8 for Liza and Philja): *sam-o+kat* ‘oneself+to roll=scooter’; Vs (1 out of 11 for Vanja, 2 out of 8 for Philja): *vert-o+let* ‘to turn+to fly=helicopter’; ADJs (3 out of 8 for Liza): *syr-o+ezhka* ‘raw+to eat=russula’.

It should be noted that both parts of X+V compounds are morphosemantically transparent and they are joined by an interfix which is usually a vowel *-o-*: *sam-o+var* ‘oneself+to boil=samovar’ (13 out of 18 for Liza, 17 out of 22 for Vanja, 8 out of 11 for Philja). An interfix *-e-* (*pyl-e+sos* ‘dust+to suck=vacuum cleaner’) was marked only in two compounds (Liza 2;2, Vanja 2;5).

After 2;5 use of the second pattern X+N becomes more active. It is also a typical model of compounds for Russian, but to a lesser extent. The modifiers might be Ns (*nos-o+rog* ‘nose+horn=rhinoceros’ (Vanja 2;7, Liza 2;9)), ADJs (*bos-o+nozhka* ‘bare+foot:DIM=open-toe sandal’ (Liza 2;11)) or just their parts (*plast(icheskaia)+massa* ‘plastic+mass=plastics’ (Vanja 3;0)), and also NUMs (*sorok-o+nozhka* ‘forty+leg:DIM=centipede’ (Vanja 2;11)).

<sup>9</sup> In the corpus for Philja, no compounds were found after 2;3.

In this model an interfix is used rarely, especially in the situation where the stem is cut and, moreover, concatenation of two words (stems with endings) without an interfix but by means of a hyphen: *ptichka+nevelichka* 'bird:DIM+tiny creature-a birdy-wirdy' (Liza 2;10), *zajchik+pobegajchik* 'hare:DIM+to run:N-a running hare':DIM (Liza 3;10). Both compounds name creatures from Russian folk tales.

Thus, all child compounds are morphosemantically transparent ones, except for a few hyphen compounds such as *akula-+karakula* 'shark+rhyiming meaning-less word=a huge shark' or *kaljaka-+maljaka* (two informal rhyiming nouns from 'to daub')=scribble'. Such compounds are a juxtaposition of two stems with their endings. These are left-headed compounds with opacity of the modifier which usually does not get used separately. As was mentioned above, the most use of such compounds is a characteristic of Liza's data, especially her CDS.

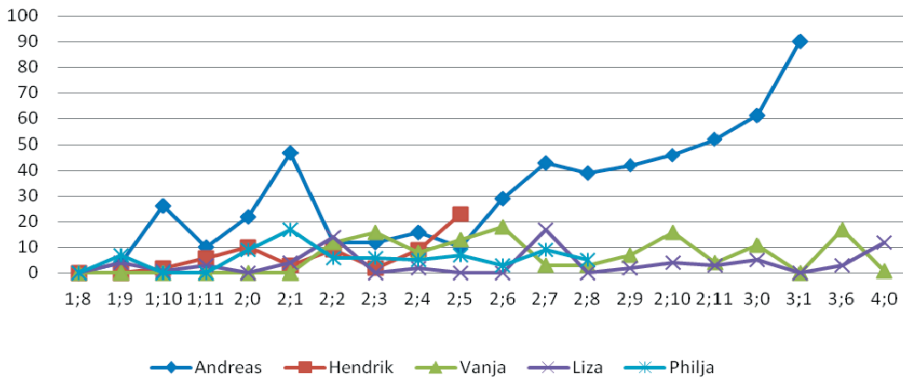
#### 4.4. Frequency effects (input vs. output)

The frequency of noun compounds can vary (from 4.8% to 19.8% among all nouns) in different recordings of both Estonian children. The number of compounds could depend very much on the topics of the conversation: for example, in the recording of Andreas (2;1) the mother and the child were looking at pictures in a book and the compounds *rongi+sõidu+raamat* 'train+ride+book =a book about travelling in a train' and *lamba+tall* 'lamb+son' appeared several times. The percentage of compound Ns still increases from the age 2;5 and does not decrease again during the observation period.

The same tendency can be observed in CDS of Hendrik: At 1;8 there were 5.1% of compounds among all noun tokens and 14.5% of compounds at 2;5. In Andreas's data the percentage of compounds increases from 4.4 (1;8) to 5.8% (2;5) but remains almost stable (aproximately 5% of the nouns were compounds) until the end of observation period.

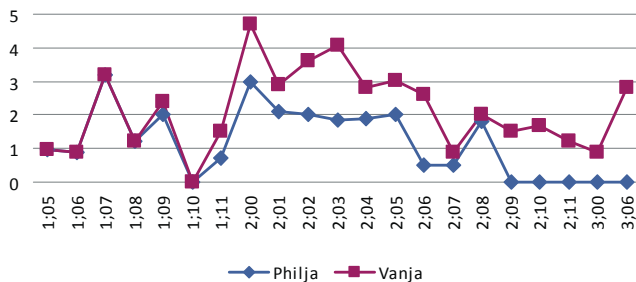
In the data of Estonian children the frequencies of compound tokens are quite similar to those found in CDS. Usually there are about 30% more compound tokens in CDS than in CS, except in Andreas's recordings at the end of the observation period, starting from 2;8 where he has even more compound tokens (48) than his mother (28). The difference in the number of compound types is similar, although slightly less significant, there are usually 10–20% more compound types in the CDS than in CS.

Similarly to the Estonian results, the frequency of Russian compounds depends on the topics of conversation and the activity, in particular, games with a particular toy. The data also demonstrates that there is a direct correlation between the number of tokens in CDS and the number of tokens in CS. However, the surprising thing is that CDS does not have many more token and types than CS but more tokens of the same type. For example, there are 22 children types and 34 adult ones in Vanja's material (that is, half as much again), while Vanja's tokens are 119, and the adult ones are 251 (that is a 110% increase). Almost the same results were revealed in Philja's data (1.45 and 2.1). The indexes in Liza's corpus are 1.2 and 1.3. Figure 1 gives an overview of the volume and increase of compounds in CS.



**Figure 1.** Noun compounds in the speech of Estonian and Russian children

It should be observed that the frequency of compounds in Russian CDS is not as high as in the Estonian data. Compounds consist of approximately 2% or less of all noun tokens: in Liza’s CDS it is 0.7% (noun tokens are 12,420), in Vanja’s CDS it is slightly higher at 1.1% (noun tokens are 22,786). The highest percentage of compounds was found in Philja’s input at 1.63% (noun tokens are 9452). Interestingly, the comparison of both boys’ data where the percentage of noun compounds is more than 1% has shown a quite similar dynamic of their development during the same period (1;8–2;9). Figure 2 gives an overview of the amount and increase of compounds in the caregiver’s speech.



**Figure 2.** Noun compounds in Russian CDS (% of noun tokens)

Some additional comments are required on the pattern of adult compounds used. A definite correlation between the frequency and order of emergence of some models in CDS and CS was observed in Philja’s data, which characterised the biggest percentage of compounds amongst all the Russian corpora. Philja’s mother used only 16 prototypical models of noun compounds where the X+V patterns dominated, and this occurred from the beginning of observation period.

In Vanja’s CDS the amount of X+V and X+N is almost equal. However, nominal models are slightly more prevalent (19, cf. 15). At the same time some of X+N patterns are represented by hyphen N+N compounds concatenated by two words which are not frequent and consequently they do not occur in the CDS with any stable frequency. Finally, a significant feature of Liza’s CDS is an abundance of

such hyphen N+N compounds, which are proper names of fairy-tale folk heroes. These compounds are not frequent (e.g. 1 type / 1 token) in the CDS so they do not enter Liza's speech, which started, similar to the two boys, with the X+V system-prototypical pattern.

#### 4.5. Productive usage of compounds

The emergence of oppositions between compounds and their members demonstrates the child's ability to detect the morphological principle of (de)composing internal form and meaning of words (Dressler et al. 2010: 4). The use of self-invented compounds can also demonstrate the child's ability to use compounding productively.

One of the Estonian children, Andreas, uses compounds productively almost from the beginning of emergence of compounds from age 1;10 when the parts of the compound *kiisu+raamat* 'cat:DEM+book=book about cats' also emerge separately in the recording. Hendrik starts to use compounds productively somewhat later (2;3) when parts of the compound *vanni+tuppa* 'bath+room' (*vann* 'bath' and *tuba* 'room') have also occurred in his speech samples.

First neologisms appeared in the speech of the Estonian children in a different order: Hendrik first starts to use self-invented compounds: *iuiiu+abi* 'sound+help=emergency car' (1;10), and thereafter Andreas *kepi+mana* 'stick+old=grandmother with stick' (2;1). Both compounds are formed according to the most productive pattern of N+N compounding in Estonian. It must be mentioned that Estonian children start to use neologisms before their usage of compounds starts to be productive. Similar results were obtained from the materials of Russian diary observations (Kharchenko, Ozerova 1999) but such examples were not revealed in our Russian corpora.

Little evidence was found to support the claim about the productive use of Russian compounds. However, several frequent verbs were revealed in the speech of our children which can be used as members of compounds and separately (Gagarina 2008) at this time: e.g. *sloMAT* 'to break' (see *metal-o+lom*), *vezTI* 'to carry' (see *par-o+voz*) (Vanja 2;3), *s'jest* 'to eat' (see *syr-o+ezhka*) (L. 1;10).

As for neologisms, only two examples were noted in Liza's data: the type N+N *nos-o+rog-in-a* 'huge rhinoceros' (2;11), and type ADJ+V *syr-o+ezh-ichsh-a* 'huge russula' (4;00). Both words are incidentally-produced ones and are formed from compounds (which were earlier *syr-o+ezhka* (1;9), *nos-o+rog* (2;9)), with the help of augmenting suffixes *-in-* and *-ichsh-* expressing an extraordinarily big size (in contrast to diminutives). So the novelty of these words comes from derivation and this is very different from new compounds found in Estonian, however, both compounds belong to neologisms (Tikhonov 2003).

Moreover, children can form diminutives from compounds, but the diminutives usually appear later than normal compounds: cf. *par-o+voz* (1;10) – *par-o+voz+ik* (Liza 1;11); *sam-o+let* (2;3) – *sam-o+let+ik* (Vanja 3;6). However, in the data of the third Russian child, Philja, his earlier compounds (1;9) were diminutives *parovoz+ik*, *samolet+ik*. The conventional pattern, indicating some productivity of compounds, starts to work in his speech later: *vertolet* (2;1) – *vertolet+ik* (2;3).

## 5. Results and discussion

The main findings of our cross-linguistic study can be summarised as follows.

1. It has been found that compounds in the analysed data of Russian child language appear infrequently, although compounding is quite a productive pattern of modern noun formation. The very first compounds emerge early (1;9), but probably as lexicalised items because little evidence was found to support the claim of productive use of compounds.

First compounds in Estonian appear at almost the same time, at the end of the child's second year of life. In contrast to Russian child acquisition, Estonians also start to use novel compounds and compounding productively early, one child just a month after the emergence of first compounds at 1;10.

2. The coefficients of compound use by each Russian boy (both in types and tokens) are the same when compared with those that appeared in their CDS: To be more precise, in type, the index is 1.5 in Vanja's corpus and 1.45 in Philja's. Regarding tokens, the index is 2.1 for both boys, and 1.3 for the girl.

Individual differences can be found in the Estonian data: Hendrik's mother produced three times more compounds than Hendrik (index 3.3), while the number of compound types in the speech of Andreas and his mother were almost the same (index 1.1). Regarding tokens, a similar difference can be found: Hendrik produced three times fewer compounds than his mother (2.8) and Andreas approximately as many as his mother (1.6).

3. It is worth noting here that the three Russian caregivers used varying numbers of compounds: Vanja's grandmother and Liza's mother produced more than twice (their indexes are 2.1 and 2.4) as many compounds as Philja's mother. It may be stated, then, that frequency of compounds depends not only on language typology, but, to a great degree, on individual preferences and a dialogue topic or activity type.

The number of compounds in the Estonian CDS is also different due the different amount of recordings but at the same time the percentage of compounds among all nouns is similar in the two sub-corpora.

4. The effect of simplicity of a compound is not an explicit one. We can assume that in Russian compounds where the first part (N, ADJ, PRON or even V) is followed by a verb, it might be semantically more transparent for a child and may therefore be preferred for productive usage. But at the same time, it is a frequent pattern within Russian compounds. They have a vowel interfix, which makes the stem simpler in a phonetic sense, and quite often a zero suffix, which renders it more transparent in morpheme aspect. After 2;5 the Russian children start to use the second pattern X+N more actively and within that there are compounds formed by concatenation of two words, one of which could be opaque and the other rhyming. Thus, prototypical Russian noun compounds (which are additionally endocentric, subordinate and interfixed) emerge first and are used more frequently than non-interfixed ones. The insertion of an interfix into a cluster of consonants probably makes articulation easier, which may be a reason for a child to prefer the interfixed forms. In addition, practically all the compounds used are morphosemantically transparent ones; and we can make comment on the transparency of both members of the compound.

There are only a few compounds emerging in the speech of the Estonian children with opaque parts and they emerge later than compounds which are

morphosemantically transparent. Compounds consisting of an opaque constituent or constituents are also rare in the input and that may be the other reason why the Estonian children start to use them later and do not produce them very frequently. The Estonian children did not use compounds containing a meaningless linking element and preferred the most transparent pattern with modifier in genitive.

5. The order of emergence of the different types of Russian compounds reveals a quite clear pattern. According to the data, both X+V and X+N types emerge simultaneously and first in two corpora (Vanja's, Liza's), and subsequently in Philja's data within one month, but model X+V is more frequent. Initially this pattern is represented by four structures of compounds: N+V, PRON+V, ADJ+V, V+V. Conversely, X+N type is not so widely represented, and only two types were found: ADJ+N, N+N in the data of Philja and Liza. In Vanja's corpus NUM+N structure was added.

At the same time, the order of emergence of different types of compounds in Estonian shows a very clear pattern. The first pattern emerging in both CS was N+N type with modifier in genitive case form. This pattern was followed by ADJ+N, after which a pattern with a shortened noun (N+N) emerged. The first acquired pattern was also widely used at the end of the observation period.

6. The main types of compounds and their frequency differ in the three Russian corpora; therefore we cannot assume that the composition of compounds is related to the frequency of occurrence in CS. It seems that compounding serves as a means of enriching the children's lexicon. In the Estonian corpora the frequency has a much stronger effect on the choice of first patterns of compounding.

7. These results may supplement previous cross-linguistic findings about the acquisition of compounds. It can be supposed that Russian is not a compound-poor language, as has been suggested earlier, and it should operate in a similar way to the Hebrew, Baltic (namely Lithuanian) and Romance languages, because compounds in these languages emerge around 2;0. However, children do not use them productively until a later time, in contrast to compound-rich Estonian, where it does not take them very long to start to use them productively. It is also obvious that Russian (an east Slavic language) favours compounding more than west Slavic languages (e.g. Czech (Pačesová 1968)). Thus Russian, like Lithuanian, seems to be placed somewhere between compound-rich languages (Estonian, Germanic, etc.) and compound-poor ones.

In addition to this, a brief comparative analysis has shown that the Russian data is quite similar to that of Lithuanian (Dabašinskienė 2010). The number of Lithuanian child's types is slightly higher than Vanja's, but his tokens are much less than all the Russian children. The same main results were revealed in CDS. The number of types of compounds used by Lithuanian caregivers is higher, whereas they use fewer tokens than in all the Russian data. We can therefore assume that Russian CDS differs from Lithuanian CDS in respect of tokens.

## 6. Conclusion

The results of our comparative investigation led to a conclusion that the system of compounds reflects the process of their acquisition in the following way. The fact that compounding is just one of the productive word-building methods in the Russian derivation system sets a condition for their lower frequency in CDS and consequently in CS, as compared with compound-rich Estonian. Moreover, in Estonian compounds start to function productively earlier than in the speech of Russian children and Estonian children use more innovative compounds than Russian children.

At the same time, in both languages children start to acquire the compounds with those models which are transparent morphosemantically and are productive. However, the factor of frequency of the caregiver's compounds depends not only on the productivity of a definite pattern in language system and some adult preferences, but also to a large extent on the topic of conversation.

## References

- Argus, Reili 2008. Eesti keele muutemorfoloogia omandamine. Tallinna Ülikooli humanitaarteaduste dissertatsioonid 19. Tallinn: TLÜ Kirjastus.
- Argus, Reili 2012. Kausatiivsuse omandamisest eesti keeles. – Eesti Rakenduslingvistika Ühingu aastaraamat, 8, 5–20. <http://dx.doi.org/10.5128/ERYa8.01>
- Berman, Ruth A. 1987. A developmental route: Learning about the form and use of complex nominals in Hebrew. – *Linguistics*, 25 (6), 1057–1086. <http://dx.doi.org/10.1515/ling.1987.25.6.1057>
- Berman, Ruth; Clark, Eve V. 1989. Learning to use compounds for contrast: Data from Hebrew. – *First Language*, 9, 247–270. <http://dx.doi.org/10.1177/014272378900902701>
- Berman, Ruth 2009. Children's acquisition of compounds. – Rochell Lieber, Pavol Štekauer (Eds.). *The Handbook of Compounds*. Oxford: Oxford University Press.
- Bittner, Dagmar; Dressler, Wolfgang U.; Kilani-Schoch, Marianne 2003. Introduction. – Dagmar Bittner, Wolfgang U. Dressler, Marianne Kilani-Schoch (Eds.). *Development of Verb Inflection in First Language Acquisition: A Cross-Linguistic Perspective*. *Studies on Language Acquisition* 21. Berlin/New York: Mouton de Gruyter, vii–xxxvii.
- Böer, Katja; Kotowski, Sven; Härtl, Holden 2012. Compounds vs. phrases: The cognitive status of morphological products. – Presentation on 15th International Morphology Meeting in Vienna. <http://www.wu.ac.at/inst/roman/imm15/abstract-boer-kotowski-haertl.pdf> (13.02.2013).
- Clark, Eve V.; Gelman, Susan A.; Lane, Nancy M. 1985. Noun compounds and category structure in young children. – *Child Development*, 56 (1), 84–94.
- Clark, Eve V.; Hecht, Barbara F.; Mulford, Randa C. 1986. Coining complex compounds in English: Affixes and word order in acquisition. – *Linguistics*, 24 (1), 7–29. <http://dx.doi.org/10.1515/ling.1986.24.1.7>
- Clark, Eve V.; Berman, Ruth A. 1984. Structure and use in acquisition of word-formation. – *Language*, 60, 542–590.
- Dabašinskienė, Ineta 2010. Acquisition of compounds Lithuanian. – *Baltic Linguistics*, 1, 51–86.
- Dressler, Wolfgang U. 1999. What is natural in natural Morphology? – *Travaux du cercle Linguistique de Prague*, 3, 135–144.
- Dressler, Wolfgang U.; Ladányi, Mária 2000. Productivity in Word Formation: a morphological approach. – *Acta Linguistica Hungarica*, 47, 103–144.

- Dressler, Wolfgang U.; Lettner, Laura E.; Korecky-Kröll, Katharina 2010. First language acquisition of compounds – with special emphasis on early German child language. – S. Scalise, I. Vogel (Eds.). *Cross-Disciplinary Issues in Compounding*. Amsterdam: Benjamins, 323–344.
- Erelt, Mati; Kasik, Reet; Metslang, Helle; Rajandi, Henno; Ross, Kristiina; Saari, Henn; Tael, Kaja; Vare, Silvi 1995. *Eesti keele grammatika I*. Tallinn: Eesti Teaduste Akadeemia Eesti Keele Instituut.
- Gagarina, Natalia 2008. Становление грамматических категорий русского глагола в детской речи [Stanovlenie grammaticheskikh kategorij russkogo glagola v detskoj rechi / First language acquisition of verb categories in Russian]. St. Petersburg: Nauka.
- Gudilova, Svetlana 2005. Продуктивные типы образования сложных слов в современном русском языке: На материале неологизмов второй половины XX века [Produktivnye tipy obrazovaniya slozhnykh slov v sovremennom russkomazyke: Na materiale neologizmov vtoroj poloviny XX veka / Productive ways of compounding in modern Russian: On the material of neologisms (the second half of the XX century)]. Moskva: Moskovskij Gosudarstvennyj Universitet.
- Gvozdev, Alexandr N. 1995. Усвоение детьми звуковой стороны русского языка [Usvoenie det'mi zvukovoj storony russkogo jazyka / The Acquisition of Russian phonetics by a child]. St. Petersburg: AKCIDENT.
- Kerge, Krista 1990. Liitsõna. Mõisteid ja seoseid. ENSV Teaduste Akadeemia. Preprint KK1-51.
- Kharchenko, Vera K.; Ozerova, Elena 1999. Сложные слова в детской речи [Slozhnye slova v detskoj rechi / Compounds in child language]. Belgorod: Belgorodskij Universitetskij Press.
- Kilani-Schoch, Marianne; Wolfgang U. Dressler 2005. *Morphologie naturelle et flexion du verbe français*. Tübingen: Gunter Narr.
- Laaha, Sabine; Gillis, Steven 2007. Typological Perspectives on the Acquisition of Noun and Verb Morphology. *Antwerp Papers in Linguistics* 112. Antwerp: University of Antwerp.
- Laalo, Klaus 1994. Kaksitavuvaihe lapsen kielen kehityksessa. – *Virittäjä*, 3, 430–448.
- Libben, Gary 1998. Semantic transparency in the processing of compounds: Consequences for representation, processing, and impairment. – *Brain and Language*, 61, 30–44.
- Mellenius, Ingmarie 1996. Children's comprehension of Swedish nominal compounds. – Carolyn E. Johnson, John H.V. Gilbert (Eds.). *Children's Language*. Mahwah, N.J.: Lawrence Erlbaum. Vol. 9, 167–182.
- Mellenius, Ingmarie 1997. *The Acquisition of Nominal Compounding in Swedish*. Lund: Lund University Press.
- Nicoladis, Elena 2002. What's the difference between 'toilet paper' and 'paper toilet'? French-English bilingual children's crosslinguistic transfer in compound nouns. – *Journal of Child Language*, 29 (4), 843–863. <http://dx.doi.org/10.1017/S0305000902005366>
- Nicoladis, Elena 2006. Pre-school Children's Acquisition of Compounds. – Gary Libben, Gonia Jarema (Eds.). *The Representation and Processing of Compound Words*. Oxford: Oxford University Press.
- Nicoladis, Elena 2007. Acquisition of deverbal compounds by French-speaking preschoolers. – *The Mental Lexicon*, 2 (1), 79–102. <http://dx.doi.org/10.1075/ml.2.1.06nic>
- Pačesová, Jaroslava 1968. *The Development of Vocabulary in the Child*. Brno: Universita Purkyně.
- Padrik, Marika 2010. *Word-Formation Skills in Estonian Children with Specific Language Impairment*. *Dissertationes pedagogicae Universitatis Tartuensis* 8. Tartu: Tartu Ülikooli Kirjastus.
- Rivista di Linguistica*, 4 (1), 1992. Special Issue: The Morphology of Compounding. Sergio Scalise (Guest-ed.).
- Русская грамматика* 1980 [Russkaja grammatika 1980 / Russian Grammar 1980]. Shvedova N. (Ed.). Vol. 1: Фонетика, фонология, ударение, интонация,



словообразование, морфология [Phonetika, fonologija, udarenie, intonacija, slovoobrazovanie, morfologija / Phonetics, phonology, stress, intonation, derivation, morphology]. Moskva: Nauka.

- Savickienė, Ineta; Dressler, Wolfgang U. (Eds.) 2007. The Acquisition of Diminutives. A Cross-Linguistic Perspective. *Language Acquisition and Language Disorders* 43. Amsterdam: John Benjamins.
- Scalise, Sergio; Bisetto, Antonietta 2005. The classification of compounds. – *Lingue e Linguaggio*, IV, 2, 319–332.
- Stephany, Ursula 2010. Preliminary classification of Greek compounds. – Paper, presented at the International Pre- and Protomorphology meeting, 12–14 February 2010.
- Stephany, Ursula; Voejkova, Maria D. (Eds.) 2009. Development of Nominal Inflection in First Language Acquisition: A Cross-Linguistic Perspective. *Studies on Language Acquisition* 30. Berlin: Mouton de Gruyter. <http://dx.doi.org/10.1515/9783110217117>
- Tikhonov, Alexandr N. 2003. Словообразовательный словарь русского языка [Slovoobrazovatel'nyj slovar' russkogo jazuka / Word-formation dictionary of the Russian language]. Vol. 1-2. Moskva: АСТ.
- Vasilevskaja, Evgenia A. 1962. Словообразование в русском языке [Slovoslozhenie v russkom jazyke / Compounding in Russian]. Moskva: Vysshaja Shkola.
- Zemskaja, Elena A. 1992. Словообразование как деятельность [Slovoobrazovanie kak dejatel'nost' / Word-formation as an activity]. Moskva: Nauka.

**Reili Argus** (Tallinn University). Her academic research interests include language acquisition (acquisition of morphology and lexicogrammatical categories).

[reili.argus@tlu.ee](mailto:reili.argus@tlu.ee)

**Victoria Kazakovskaya** (Institute for Linguistic Studies at Russian Academy of Sciences). Her academic research interests include psycholinguistics, functional grammar, conversation analysis, language acquisition (development of communicative competence, acquisition of semantic categories, child-directed speech, negative evidence, fine-tuning).

[victory805@mail.ru](mailto:victory805@mail.ru)

# LIITSÕNADE OMANDAMINE EESTI JA VENE KEELES: SAGEDUSE, PRODUKTIIVSUSE, LIHTSUSE JA LÄBIPAISTVUSE MÕJU

**Reili Argus, Victoria Kazakovskaya**

Tallinna Ülikool, Vene Teaduste Akadeemia

Artikkel keskendub varasele liitsõnamoodustuse arengule kahes tüpoloogiliselt erinevas keeles – eesti ja vene keeles. Kui eesti keeles on liitsõnamoodustus produktiivne ja liitsõnu esineb keeles palju, siis vene keele puhul on liitsõnad enamasti kinnistunud leksikaalsed üksused ning neid leidub igapäevases keelekasutuses vähem kui eesti keeles.

Uurimuse aluseks on pikiuuringuga saadud keeleandmed kahelt eesti ja kolmelt vene keelt omandavalt lapselt. Analüüsitud on nii laste enda kui ka neile suunatud kõnet ning analüüsi keskmes on liitsõnade ilmumine laste keelekasutusse, eri liitmismallide areng ning sellised omandamist mõjutavad tegurid nagu liitmismalli produktiivsus, liitsõnade või mallide sagedus, moodustusmalli lihtsus ja liitsõna morfosemantiline läbipaistvus.

Olgugi et liitsõnade sagedus kahes vaadeldud keeles on erinev, ilmuvad esimesed liitsõnad laste keelde samas vanuses, teise eluaasta lõpus. Mõlemas keeles omandavad lapsed produktiivsed liitsõna moodustamise mallid varakult. Kui eesti lapsed hakkavad liitsõnu kasutama produktiivselt juba mõni kuu pärast esimeste liitsõnade ilmumist, ei saa sama väita vene keelt omandavate laste kohta. Samuti ei kasuta vene keelt omandavad lapsed erinevalt eesti lastest omaloomingulisi liitsõnu. Seega paistab kahe vaadeldud keele liitsõnasüsteemi eripära väga selgelt ka laste keelekasutusest.

Sellised tegurid nagu lihtsus ja sagedus on omandamisprotsessis põimunud: lihtsama ehitusega liitsõnad ilmuvad samal ajal kui sagedased liitsõnad. Sageduse mõju ei ole aga alati ühesuunaline: lapsele suunatud kõne puhul võis täheldada, et mõned vanemad kasutavad lapsega kõneldes enam liitsõnu kui teised ning mitte alati ei pruugi lapsele suunatud kõne liitsõnarikkus peegelduda lapse kõnes.

**Võtmesõnad:** keeleomandamine, liitsõnad, lapsele suunatud kõne, sagedus, produktiivsus, läbipaistvus, lihtsus, eesti keel, vene keel

# QUANTITATIVE AND QUALITATIVE ASPECTS OF SWITCHING BETWEEN ESTONIAN SIGN LANGUAGE AND SPOKEN ESTONIAN

Kadri Hein

**Abstract.** The central research question of the present paper is ‘What variables influence switching (code-switching and code-blending) between Estonian Sign Language and spoken Estonian among bimodal bilinguals in a school environment?’ The data from guided discussions involving deaf and hearing students indicated that bimodal bilinguals predominantly code-blend, i.e. simultaneously produce speech and signs. The main function of switching among the students was to emphasise a constituent in a clause, but they also switched to express themselves in an original way. In some utterances, it proved difficult to apply a single function to a switch, thus a sequential turn-by-turn analysis was employed. A deaf student’s level of switching was influenced most by the hearing status of the interlocutor, and the amount of switching by the interlocutor.\*

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

## 1. Introduction

The bulk of the research on language contact in Estonia considers two spoken/written languages, e.g. Estonian-Russian (Zabrodskaja 2005, Verschik 2008) contacts. The main aim of this paper is to investigate 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 the Tartu Hiie School. The majority of studies on bimodal bilingualism have been conducted on hearing bilinguals (Bishop, Hicks 2005, Emmorey et al. 2008). This study differs from them in the sense that both hearing and deaf bimodal bilinguals participate in the guided discussions. The language contact between Estonian and Estonian Sign Language is unique in that

---

\* The present paper is based on a PhD thesis submitted Birkbeck College, 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.

it is *cross-modal*, or *intermodal* – it involves one language in the visual-spatial and the other in the auditory-vocal modality. Also, elements can be combined both sequentially and simultaneously.

The central research question of the study is “What variables influence code-switching and code-blending between Estonian and Estonian Sign Language (EVK) in a school environment?” Researchers (e.g. Van den Bogaerde 2000, Kuntze 2000) have discussed the effect of different variables on the amount of code-switching. This study focused on the relationship between hearing status, the level of hearing loss, parental hearing status, hearing status of the interlocutor, Estonian language proficiency, interlocutor’s switching and the number of switches.

It is expected that deaf and hearing students predominantly code-blend, i.e. they do not stop signing and start speaking or vice versa, but produce the signs at the time of speaking, and that the Matrix Language of the majority of code-blends is Estonian (Emmorey et al. 2008). Also, it is expected that the word classes used most often in code-blends and code-switches are nouns (Muysken 2004: 153, Van den Bogaerde, Baker 2008).

The functions of switching among the deaf and hearing bimodal bilinguals were determined (McClure 1981, Zabrodskaia 2005). In the cases where it is problematic to determine a single function of a switch, the conversations were analysed turn by turn according to Auer’s (1995) conversational model. Speech act analysis of the excerpts was also carried out (Austin 1962, Searle 1976).

## 2. Estonian Sign Language

A common feature shared by the members of the Estonian deaf community is Estonian Sign Language, or *eesti viipekeel* (EVK). EVK<sup>1</sup> is a minority language with only 1500–2000 signers. In the database of the languages of the world, it is represented with the SIL code ESO (Sutrop 2005). Users of EVK are concentrated around bigger towns in Estonia, such as Pärnu, Tartu, and Tallinn. EVK has been influenced by the neighbouring sign languages, such as Russian and Finnish Sign Language. In Taniroo’s (2007) study, out of 200 signs, 123 were identical in Estonian and Russian Sign Language. EVK was officially recognised on 1 March 2007, when several amendments were made in the Language Act (State Gazette 2007).

It has been found that sign order is relatively flexible in EVK, and usually the topic is signed first, followed by a comment (Laiapea 1992: 2101). Similarly to American and British Sign Language, the interrogative signs are signed at the end of the sentence in EVK. There are three main types of questions in EVK: yes/no questions, alternative questions and *wh*-questions. The signs in EVK are formed in signing space, which can be described as a three-dimensional area extending from the hips to the head, and from one extended elbow to the other. Signing space plays a vital role in expressing time relations. Trükmann (2006) found that the same time relations can be expressed in Estonian and EVK and the ways of expressing time in EVK are similar to those used in other sign languages. However, only one out of the five timelines in EVK, the vertical timeline, has the same location and function as in other sign languages. Signing space is also important in using classifiers.

---

<sup>1</sup> A thorough overview of EVK is given by Paabo (2010).

Classifiers, or classifier constructions, are morphemes used for indicating spatial relations, movement, and the shape and size of the objects (Sandler and Lillo-Martin 2006). Laiapea (2007: 51–53) divides classifiers in EVK as follows: state predicates, process predicates, and action predicates. He also offers four types of classifiers for indicating a human. Some classifiers have lexicalised in EVK and started their life as new signs (Laiapea 2007: 53).

In EVK, 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. As in many other sign languages, it is possible to express the singular, dual, trial, quadral, and plural in EVK (Miljan 2003: 219).

There are nine basic colour terms in EVK: MUST 'black', VALGE 1<sup>2</sup> 'white 1', PUNANE 1<sup>3</sup> 'red 1', ROHELINE 'green', KOLLANE 'yellow', SININE 'blue', PRUUN 'brown', HALL 'grey' and ROOSA/LILLA 'pink/purple' (Hollman 2010).

### 3. Code-switching in bimodal bilingualism

Researchers studying deaf people have discussed the effect of different variables on code-switching, for example, they have found that deaf people code-switch depending on their own hearing status<sup>4</sup>, the hearing status of the parents, the hearing ability of the interlocutor. Language proficiency may also influence code-switching: the higher the deaf person's level of English, the greater the chances of switching (Kuntze 2000: 292). 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.

Emmorey et al. (2008) researched several aspects of Coda<sup>5</sup> communication. The result of the study showed that Coda's prefer code-blending to code-switching when conversing with other bimodal bilinguals, i.e. they rarely stop talking and switch to ASL (ibid.). English was mostly chosen as the Matrix Language for conversations involving ASL-English bimodal bilinguals (Emmorey et al. 2008: 51). The results showed that for bimodal bilinguals, the most often used classes in code-blends and code-switches were verbs, as opposed to unimodal bilinguals, who show more code-blending in nouns (Muysken 2004: 153). Note that Van den Bogaerde and Baker (2008) found that in bimodal bilingualism, nouns dominated over verbs in the code-blends. Besides word classes, other factors play a role in switching, such as semantics and frequency of use (Backus, Verschik 2012), which have not been investigated in this paper.

The written language of Coda's has been studied by Bishop and Hicks (2005) who found that Coda-talk is characterised by systematic code-blending, which differs from SimCom<sup>6</sup> and other forms of bimodal code-blending. Some features of Coda-talk are in line with the characteristics of ASL, such as leaving out the

---

<sup>2</sup> There may be several signs to refer to a particular colour. VALGE 1 is formed by stroking the left hand with the right hand.

<sup>3</sup> This sign is formed by making a round movement with the A handshape on the cheek.

<sup>4</sup> Grosjean (1996: 31–32) suggests that when communicating with a hearing monolingual person, deaf bilinguals deactivate the sign language and function in the monolingual mode. In interaction with other deaf bilinguals, they can mix sign and speech. However, Van den Bogaerde (2000) found that there were almost no instances of code-mixing in deaf children's utterances.

<sup>5</sup> The term *Coda* is used to refer to a hearing child of deaf adults.

<sup>6</sup> SimCom is short for *simultaneous communication*, which is used interchangeably with the term *code-blending* in the present paper.

subjects, copula, determiners, objects, auxiliaries or modals, prepositions and determiners, while others are only characteristic of Coda-talk: verb manipulation, unique glossing and lexicon.

#### **4. Code-switching in unimodal bilingualism**

In unimodal bilingualism, the grammatical (Myers-Scotton 1997) and sociolinguistic views (Auer 1984, 1995, Li Wei 2002) have been the most influential. Hopefully, combining these two approaches in the present paper gives a more objective overview of the contact between EVK and Estonian than only one would. Myers-Scotton (1997) provides a Matrix Language Frame (MLF) model for code-switching, which is based on the assumption that one language dominates in switching (Matrix Language, or ML) over the non-dominant language (Embedded Language, or EL) by supplying more words and morphemes to the sentence. According to the MLF model, there are three types of constituents: ML islands composed of ML grammatical and lexical elements, EL islands consisting of EL grammatical and lexical elements, and mixed constituents that have elements from both languages. The model claims that the ML determines the surface morpheme order in the mixed constituents (the Morpheme Order Principle), and all relevant system morphemes in these constituents come only from the ML (the System Morpheme Principle).

The proponents of sociolinguistic or pragmatic approaches to code-switching have attempted to determine the reasons for switching and describe how code-switching emerges in conversation. Zabrodska (2005) mentioned the following reasons for switches: reported speech; language play; expressive function; side-comment; reiteration, or quasi-translation; the 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), participant make-up, activity type, topic shift, and topic-comment structure (Auer 1995: 120). Each instance of code-switching is unique, and may have multiple functions, a general taxonomy of the functions cannot possibly account for all cases of switching. The entire conversation has to be taken into account in order to be able to determine the reasons for a particular switch.

One of the most influential pragmatic theories on code-switching is the Conversation Analysis (CA) approach to bilingual communication. CA focuses on the individual's actions to achieve interactive goals, which distance the approach from those dominant in macro-sociolinguistic studies. More specifically, CA looks at the way meaning emerges, or is brought about in interaction through code-switching (Li Wei 2002: 167). The advantage of the CA approach is that the language choice of one turn influences the choice for the next turn – the process is sequential; and this limits the researcher's influence, as the analysis focuses on the participants' reflections of their utterances, as seen in their language choices (Auer 1984: 5–6). Thus, one sentence is not enough for analysing code-switching. The previous and the following turns have to be taken into account, as well as the context in which the interaction takes place. Auer (1995) distinguishes four patterns of code choice in conversation (Table 1).

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

Name	Pattern*
Ia	A1 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

\* The letters refer to the languages A and B; the numbers indicate the speakers 1 and 2

Auer (1995) refers to type I switches as discourse-related switches, as they are connected with the organisation of the conversation; type II switching is seen as preference-related, as it relies on the personal preferences of the interlocutor or is done for political reasons; type III switching may be both discourse and participant-related, whereas type IV is termed *transfer*.

## 5. Data

The material was videotaped in two deaf schools in Estonia, one of them an oral school (Tartu Hiie School) and the other bilingual (Tallinn Helen's School). The Tartu Hiie School, which was established in 1941 employs mainly the oral method in educating the deaf students – speaking, lipreading, fingerspelling and writing are used. 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 EVK is taught as the first language and Estonian as the second language. Both hearing and deaf teachers work at the school, instructing mainly in EVK and (written) Estonian<sup>7</sup>. Sign language interpreters help the teachers who do not have the necessary knowledge of EVK. English is taught as a foreign language. Hearing children interested in social work and EVK were admitted to the upper secondary school from 2003 to 2008<sup>8</sup>.

The guided discussions lasted approximately 290 minutes and consisted of 4104 utterances<sup>9</sup>. A total of 47 students, of whom 12 were hearing and 35 were deaf, took part in the guided discussion. In Tallinn Helen's School, 12 deaf students aged 14–20 and 12 hearing students aged 16–19 participated in the guided discussion. As the hearing students in the Tartu Hiie School do not sign, two groups of deaf students were formed by an expert from the school for the purposes of this study: deaf persons with profound hearing loss and deaf persons with slight hearing loss. Thirteen deaf students with profound hearing loss aged 11–16 and ten deaf students with slight hearing loss aged 14–18 participated in the guided discussion. The analysed material from Tallinn Helen's School lasts approximately 141 minutes and from the Tartu Hiie School 149 minutes.

In Tallinn Helen's School, each deaf student participated in two conversations: one with another deaf person and one with a hearing person while the order of

<sup>7</sup> The situation is changing, as more hearing impaired students enter Tallinn Helen's School – their speaking skills need to be developed.

<sup>8</sup> The last class of hearing students finished school in spring 2011.

<sup>9</sup> Utterance was determined by the use of pauses – an utterance was usually surrounded by longer pauses.

signing was randomised. In the Tartu Hiie School, the same design was applied to the two groups formed there. The subjects were told that they did not have to sign all the time; they could use both Estonian and EVK. The students were provided with five topics that were written on a large piece of paper and placed where the participants could see them: introducing yourself to each other (name, sign name, age, form); family; hobbies; spring holidays; free topic. The topics were also explained to the subjects in EVK. The researcher was present when the subjects were conversing about the first four topics and left the room when the students were discussing the free topic.

## 6. Transcription and analysis

Three points were analysed in depth using the ELAN software (Language Archiving Technology 2008): the signs along with relevant gestures (EVK), the spoken components (Estonian), and word classes. The utterances in EVK were transcribed according to the guidelines designed by Toom et al. (2006): the signs were written in upper-case and the letters in fingerspelled words were written in lower-case, separated by dashes. In the first line there are signs in EVK, the second line contains spoken Estonian, the English glosses are in the third line, and the fourth line translates the utterance (see example 1). As pauses and latching form an important part of the analysis of switching, they were separately marked (Table 2).

**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 randomly replaced by codes, including S, number, and D or H, indicating the hearing status of the informant's parents. In the excerpts the students' names were further replaced with names common in Estonia. The analysis of each guided discussion started quantitatively. EL and ML in code-blends of Estonian and Estonian Sign Language cannot be determined by counting the number of morphemes, as the modalities and morphologies differ significantly. Thus, ML and EL in code-blends were determined in considering features of EVK and Estonian. EVK was considered to be the ML if role-shift<sup>10</sup> and classifiers were used. Estonian was the ML if a word was fingerspelled<sup>11</sup> instead of signing. The following aspects were considered in more detail: word order,<sup>12</sup> expressing time, question formation, the category of number, adjectival modification, and agent nouns, as discrepancies in these areas may trigger code-switching.

<sup>10</sup> Role-shift is a common feature in sign languages and is also used in EVK. It refers to the possibility of shifting between the characters in a narration (Laipea 2007: 59).

<sup>11</sup> Although some researchers, e.g. Kuntze (2000: 296) view fingerspelling as a code-switch if it is used instead of a sign, fingerspelled words were not considered as code-switches in this study.

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



Pearson's product-moment correlation coefficient or Pearson's  $r$  was used to study the relationship between the number of switches and the following variables: hearing status of the subject, 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 and the main reasons for switching among deaf and hearing bimodal bilinguals outlined by using the taxonomies compiled by Zabrodskaia (2005) and McClure (1981). Speech acts were analysed according to Searle's (1976) typology. The following speech acts were considered: 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 9 students, which is used for those who study Estonian as a second language.<sup>13</sup> 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 hearing students from Tallinn Helen's School received the best results ( $M = 92.3$ ,  $SD = 5.9$ )<sup>14</sup> in the Estonian language proficiency test, followed by the deaf students from the Tartu Hiie School with slight hearing loss ( $M = 57$ ,  $SD = 22.7$ ) and profound hearing loss ( $M = 44.8$ ,  $SD = 25.0$ ), and deaf students from Tallinn Helen's School ( $M = 44.5$ ,  $SD = 24.8$ ). The deaf students from Tallinn Helen's School had better results (B1-B2) in EVK than the ones from the Tartu Hiie School (A1-A2) and the hearing students from Tallinn Helen's School (A1-A2).

## 7. Quantitative aspects of switching

The guided discussion in Tallinn Helen's School consisted of an elicited conversation under two conditions: deaf-deaf and deaf-hearing interaction. The total number of utterances under both conditions was 2062, of which 1374 were produced by the deaf and 688 by the hearing students. The subjects produced significantly more utterances containing code-blends ( $M = 17.7$ ) than code-switches ( $M = 9.0$ ). Overall, 34.7% of utterances contained some type of switching.

The guided discussion in the Tartu Hiie School consisted of a conversation among deaf students with profound hearing loss and a conversation between deaf students with profound hearing loss and deaf students with slight hearing loss. The total number of utterances in the guided discussion was 2042, of which 1397 were produced by the profoundly deaf students and 645 by the deaf students with slight hearing loss. The informants produced more utterances containing code-blends ( $M = 5.3$ ) than code-switches ( $M = 3.4$ ). Overall, only 9.6% of the utterances contained some type of switching.

The hearing informants in Tallinn Helen's School produced significantly more switches (31.6%) than the deaf informants (14.0%) and the deaf informants with slight hearing loss in the Tartu Hiie School produced significantly more switches

---

<sup>13</sup> The questionnaire and the language test can be found in Hein (2012: 329–350).

<sup>14</sup>  $M$  = mean,  $SD$  = standard deviation. Note that the maximum number of points on the test was 100.

(14.8%) than the ones with profound hearing loss (7.4%). The percentages of code-switches, code-blends and switches in Table 3 show that the students mostly code-blended.

**Table 3.** The percentage of utterances containing code-blends, code-switches and switches produced by deaf and hearing subjects of Tallinn Helen's School and the Tartu Hiie School

Informants	Code-switches	Code-blends	Switches
Deaf – Helen's School	12.4	22.1	24.5
Hearing – Helen's School	22.5	48.5	55.1
Profound hearing loss – Hiie School	3.9	7	7.4
Slight hearing loss – Hiie School	8.7	13	14.8

In order to study the connection between different variables and the number of switches, Pearson's product-moment correlation coefficient was used (Table 4). The interlocutor's hearing status influenced the results most. In Tallinn Helen's School, there was a strong positive correlation between the hearing status of the interlocutor and the number of switches,  $r = 0.54$ ,  $p \leq 0.01$  (one-tailed). In the Tartu Hiie School, a medium positive correlation occurred between the hearing status of the interlocutor and the number of switches,  $r = 0.48$ ,  $p < 0.05$  (one-tailed).

The amount of switching may also be influenced by the amount of the interlocutor's switching. In Tallinn Helen's School, Pearson's  $r$  showed an inexplicable strong negative correlation between the number of switches produced by deaf students and their hearing interlocutors,  $r = -0.66$ ,  $p \leq 0.01$  (one-tailed). In the Tartu Hiie School, Pearson's  $r$  showed a strong positive correlation between the number of switches produced by the deaf students with profound hearing loss and their interlocutors, namely the deaf students with slight hearing loss,  $r = 0.93$ ,  $p \leq 0.01$  (one-tailed).

Besides the variables presented in Table 4, the relationship between EVK proficiency and the number of switches was also studied. The results illustrate the opposite tendency with regard to EVK proficiency and the number of switches: the hearing subjects from the two schools most proficient in EVK switched less than those with the lower EVK abilities.

**Table 4.** Correlation between variables and the level of switching

Variables	Tallinn Helen's School guided discussion	Tartu Hiie School guided discussion
Hearing status	$r = 0.37^*$	$r = 0.26$
Level of hearing loss	$r = -0.20$	–
Parental hearing status	$r = 0.39^*$	$r = 0.02$
Hearing status of the interlocutor	$r = 0.54^*$	$r = 0.48^*$
Estonian language proficiency	$r = 0.33^*$	$r = -0.16$
Interlocutor's switches	$r = -0.66^*$	$r = 0.93^*$

\* The correlation is significant at 0.05 level.

In Tallinn Helen's School, the most often used word classes in code-switches were nouns (26.5%) and verbs (25.2%). In code-blends, nouns (36.1%) were used more often than verbs (23.0%), which may be caused by the fact that verbs are often produced with a particular mouth pattern which does not allow them to be present in code-blends so often. In the Tartu Hiie School, the most often used word classes in code-blends were also nouns (54.9%) and verbs (15.8%). In code-switches, nouns (31.0%) were also most often used, followed by pronouns (23.4%) and verbs (23.0%).

In Tallinn Helen's School, Estonian appeared to be the ML in most of the code-blends (R 20.1%, E1 13.7%, E2 28.9%)<sup>15</sup> rather than Estonian Sign Language (R 11.5%, E1 40.3%, E2 24.3%). In the Tartu Hiie School, most of the code-blended utterances also appeared to have Estonian as the ML (R 30.1%, E1 39.9%, E2 36.6%) rather than Estonian Sign Language (R 24.3%, E1 30.5%, E2 30.7%).

The ML could not always be determined, as some utterances were too short, consisting only of one word. Also, some subjects produced semantically non-equivalent code-blends<sup>16</sup>. Some examples of these types of blends consisted of lexical items produced with differing word order in each language, making the utterance syntactically correct both in Estonian and EVK. In example (1), the utterance in the sign language follows the rules of EVK, as the question word occurs at the end (AGE HOW MANY). At the same time, an Estonian-like utterance is produced: the question word is at the beginning and the pronoun *sa* 'you' has been added.

- (1) S1H: VANUS                      MITU  
           kui            vana    sa                      ole-d  
           how        old    you:2SgNom    be-Pr2Sg  
           'How old are you?'

Thus, the MLF model may not be the best tool for describing the sign language-spoken language contact.<sup>17</sup>

## 8. Functions of switching

The main reason for switching both in Tallinn Helen's School and the Tartu Hiie School was to emphasise a constituent. Sometimes the students also used switching to express themselves in an original way, which took the form of humour and language play. The subjects in Tallinn Helen's School also often switched to side-comment. Other more common functions of switching in Tallinn Helen's School were to attract attention, change of participant make-up (addressee specification; including, excluding, and marginalising participants)<sup>18</sup>, change in topic, and reiteration.

<sup>15</sup> R = researcher's opinion; E1 = first expert; E2 = second expert.

<sup>16</sup> See Emmorey et al. (2005, 2008) for the categories of semantically non-equivalent code-blends in ASL-English hearing bimodal bilinguals.

<sup>17</sup> ML is not always clearly definable in spoken languages either, as found in Russian-Estonian contact, because of the EL word order in the bilingual clause, and problems in determining morpheme type (Zabrodskaja 2009: 60).

<sup>18</sup> The functions of attracting attention and participant make-up were sometimes used in Tallinn Helen's School in connection with the hearing researcher.

## 8.1. Emphasising a constituent

First, the switching function used most often, to provide emphasis on a particular constituent, will be discussed. In some utterances, it was necessary to distinguish between particular signs. In (2), the deaf student (S20D) from Tallinn Helen's School produces a single word during a signed question in order to emphasise the sign VEND 'brother'. First, the sign POISS^VEND<sup>19</sup> is uttered, and to make sure that the hearing interlocutor understands the sign, fingerspelling and speech are used. S46H answers the question in a code-blended mode, as usual. S20D then continues with a signed utterance, which is accompanied by a single word. Here, the word *õde* 'sister' is uttered at the same time as the sign TÜDRUK^ÕDE, which is a combination of the signs TÜDRUK 'girl' and ÕDE 'sister', and is often used to mean 'sister'. The word ÕDE 'sister' is accompanied by speech to distinguish between the terms *sister* and *brother* which share the same sign. The excerpt ends with S46H's code-blended utterance. The last three utterances are statements.

(2)

S20D: SINA OLEMA POISS^VEND v-e-n-d OLEMA POLE  
 – – – vend – –  
 you be boy sibling brother:NomSg be NEG + be  
 'Do you have a brother?'

S46H: POLE  
 Pole  
 NEG + be  
 'No, I don't.'

S20D: MINA OLEMA TÜDRUK^ÕDE  
 – – õde  
 I: 1SgNom be girl sibling:NomSg  
 'I have a sister.'

(2.0)

S46H: – POLE  
 mu-l pole  
 I-1SgAde NEG + be  
 'I don't.'

## 8.2. Side-comment

One of the most prevalent reasons for switching in Tallinn Helen's School was to create a side-comment, which was mainly used by the hearing informants. The hearing subject (S40H) has problems understanding the deaf interlocutor in (3). S38H asks a question about S40H's travel plans for the summer, and the latter does not understand. He/she expresses this in Estonian only; this statement functions as a side-comment.

- (3) S38H: SUVI                      SÕITMA      r-e-i-s              KUHU  
 suve-l                      sõit-ma      reis                  kuhu  
 summer-AdeSg   ride-inf      trip:NomSg      where  
 ‘Where are you going to travel in the summer?’  
 (2.2)
- S40H: ma                      ei              saa      aru  
 I:1SgNom      NEG      get      sense  
 ‘I don’t understand.’

### 8.3. Topic shift

Topic shift also caused some code-switches and code-blends in Tallinn Helen’s School. In the following interaction, the deaf student (S1H) frequently code-blended, but the hearing student (S43H) always signed without switching the code. Example (4) shows that a change in the topic affects S43’s behaviour and he/she starts code-blending as well. S1H asks about the bowling tournament in a code-blended mode. As there is a pause, he/she offers an explanation, also by code-blending. S43H answers in the form of a code-blend that he/she does not know either, which is a statement. S1H and S43H then continue conversing by using both sign and speech. S1H comments that the tournament was supposed to be the day before, which is an explanation. After a pause, S43H gives a reason why the tournament was cancelled, which can also be seen as an explanation. After another pause S43H recalls that the tournament takes place in the same week as open house week, which is a statement.

- (4) S1H: MILLAL TULEMA      BOWLING      VÕISTLEMA  
 millal      tule-b              bõuling              –  
 when      come-Pr3Sg      bowling:NomSg      competition  
 ‘When do we have the bowling tournament?’  
 (1.7)
- S1H: BOWLING  
 bõuling  
 bowling:NomSg  
 ‘Bowling.’  
 (2.0)
- S43H: MITTE\_TEA DMA      APRILL =  
 –                              aprilli-s =  
 not know                      April-IneSg  
 ‘I don’t know; in April.’
- S1H:      = EILE                      PIDAMA                      MUUTMA  
 = eile                      pid-i                              –  
 yesterday:NomSg      have to-Impf3Sg      change  
 ‘It was supposed to take place yesterday but they changed it.’  
 (1.5)

S43H: VÄHE INIMENE INIMENE KOOL  
vähe inimes-i – kooli-s  
little humans-PartPl person school-IneSg  
‘Few people at school.’  
(4.0)

S43H: SAMA NÄDAL – LAHTI\_UKSED –  
sama nädal kui lahtis-te us-te nädal  
same week:NomSg as open-GenPl door-GenPl week:NomSg  
‘The same week as open house week.’

#### 8.4. Reiteration

Reiteration was another function of code-switching utilised in the guided discussion in the two schools. This means that the utterances were repeated in another mode, e.g. simultaneous communication and fingerspelling. In (5), which originates from the Tartu Hiie School, S52H asks a question in Estonian, and as there is no answer within three seconds, he/she utters the question again, this time in the code-blended mode. To make sure that the interlocutor understands, fingerspelling is used. S51H then produces the answer by code-blending and then switches to EVK, which is a statement. S52H then continues in Estonian in his/her statement, and S51H produces the question in EVK. The final statement is given by S52H also in Estonian.

- (5) S52H: kas sa käi-d ekskursioon  
if you:2SgNom go-Pr2Sg school trip:NomSg  
‘Do you take part in the school trips?’  
(3.0)
- S52H: k-a-s s-a k-ä-i-d e-k-s-k-u-r-s-i-oo-n  
kas sa käi-d ekskursioon  
if you:2SgNom go-2Sg school trip:NomSg  
‘Do you take part in the school trips?’
- S51H: KÄIMA LÄTI  
käi-sin –  
go-Impf1Sg Latvia  
‘I have been to Latvia.’
- S52H: Läti (.) mina ka Läti  
Latvia: NomSg I:1SgNom too Latvia:NomSg  
‘Latvia? Me too.’
- S51H: KÄIMA LÄTI VALMIS  
go Latvia:NomSg ready  
‘You have been to Latvia?’
- S52H: Nod  
Jaa  
‘Yes.’

## 8.5. Expressive function

Expressive function was occasionally employed in both schools: the informants changed the code because they wanted to be original, to emphasise something, to make a joke, or to play with the language. In example (6) from the Tartu Hiie School, S68H gives the reasons for giving up basketball training in EVK, which is a statement. S67H comments on it, also in Estonian, which is also a statement. S68H then goes on explaining, and to emphasise the dullness of the daily routine utters the word SAMA ‘the same’ three times simultaneously in EVK and Estonian.

- (6) S68H: KORVPALL LÕPP PRAEGU TRENN LÕPP (.)  
basketball end now practice finish  
‘I finished practising basketball.’
- S68H: LÕPP MITTE-TAHTMA  
finish not want  
‘Finished. Don’t want.’
- S67H: LÕPP  
finish  
‘Finished.’
- S68H: LÕPP IGA\_PÄEV PALJU (.)  
finish every day a lot  
‘Finished. Every day a lot.’
- S68H: ARVUTI KORVPALL SAMA SAMA SAMA IGAV  
– – sama sama sama –  
computer basketball same same same boring  
‘Every day the same: computer and basketball. It is boring.’

In some conversations at Tallinn Helen’s School, the interlocutors signed only. This often happened in deaf-only interactions where code-blending was rarely employed. Also, some hearing subjects chose to sign during the whole conversation with the deaf subject. Some conversations, on the other hand, contained only code-blending, that is, the deaf and hearing participants simultaneously signed and spoke throughout the whole conversation. Sometimes the hearing subject tended to code-blend, whereas the deaf subject signed, and vice versa.

## 8.6. Sequential analysis

In some conversations, a deaf or hearing subject code-switched or code-blended for a particular reason. In others, the function of the switch remained unclear; as the switches were the result of the previous turn. Thus, a sequential analysis of the turns (Auer 1995) was needed. The excerpt below (7) is an example of one such interaction that takes place in the second stage of the guided discussion where the subjects could discuss any topic. The deaf student (S38H) tries to find topics to discuss by code-blending. As S40H produces only a nod for an answer, S38H asks a question in EVK after a 1.6-second pause. Instead of an answer there is another pause, and S38H has to continue signing. As the previous turn is signed, S40H signs as well. The pattern of turns (AB<sub>1</sub> A<sub>2</sub> A<sub>1</sub> A<sub>2</sub>) displays the form of number IIIb in Auer’s (1995) classification, as one of the interlocutors abandons the AB

code and continues talking only in one language. As for speech acts, all utterances are statements except for one question. The overall analysis of speech acts in the data showed that most of the speech acts in excerpts were statements, although questions were also produced.

- (7) S38H: ÕPPIDA IGAV  
 õppi-da igav  
 study-Inf boring  
 ‘It is boring to study.’
- S40H: nod  
 (1.6)
- S38H: SINA ÕPPIMA HEA  
 you study good  
 ‘Do you study well?’  
 (1.5)
- S38H: MAGAMA  
 sleep  
 ‘You are sleeping.’
- S40H: MAGAMA  
 ‘Yes, I am sleeping.’

## 9. Conclusions

The results of the guided discussion show that the students in Tallinn Helen’s School and the Tartu Hiie School mostly code-blend rather than code-switch, as also found by Emmorey et al. (2008). This implies that as opposed to unimodal bilingualism, in sign language-spoken language switching producing two languages simultaneously is common.

It was found that the deaf subjects predominantly signed with their deaf interlocutors, which contrasts with Grosjean’s (1996: 32) suggestion that deaf bilinguals can communicate in the bilingual mode by mixing sign and speech while interacting with other deaf subjects. The results of the present paper are more similar to the study conducted by Van den Bogaerde (2000) who found that there were almost no instances of code-mixing in the deaf children’s utterances.

The previous research on bimodal bilingualism has covered several factors that influence code-switching and code-blending among deaf people. In the present study, the strongest and most significant correlations occurred between the hearing status of the interlocutor, interlocutor’s level of switching and the number of switches. This indicates that the deaf students switched more with hearing interlocutors, whereas the interaction between two deaf persons was mainly signed. The correlation between the interlocutor’s level of switching and the number of switches yielded opposing results, which cannot be easily explained.

Research on bimodal bilinguals has shown mixed results on the use of nouns and verbs in code-blends and code-switches. Emmorey et al. (2008) found that the most often used word classes in code-blends and code-switches are verbs. Van den Bogaerde and Baker (2008) reported the dominance of nouns over verbs in code-blends. In the present study, nouns were used more often in code-blends and



code-switches, which may be caused by the fact that there were more nouns (2886) than verbs (2131) in the data. Besides word classes, other aspects have to be taken into consideration, such as frequency and meaning, which are linked to different types of switching (Backus, Verschik 2012).

The analysis of the Matrix Language and Embedded Language of the code-blended utterances shows that most of the utterances were produced with Estonian as the ML. The results along with the previous findings in bimodal bilingualism (Emmorey et al. 2008) indicate that in sign language-spoken language contact, the spoken language tends to be the ML. Some utterances possessed both Estonian and EVK features, so it was difficult to determine the ML and EL. The results show that using the MLF model (Myers-Scotton 1997) may pose difficulties in sign language-spoken language code-blending, as has also been noted in some spoken language pairs (Zabrodskaia 2009).

In conclusion, the results of the guided discussion show that deaf informants switch for different reasons and use various patterns of code choice. Sometimes it was problematic to apply a single cause for a switch, because it was caused merely by the code choice of the previous turn. In these instances, the sequential model of switching compiled by Auer (1995) was utilised.

## References

- Auer, Peter 1984. *Bilingual Conversation*. Amsterdam: John Benjamins.
- 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.
- 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>
- 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.
- Hollman, Liivi 2010. *Basic Color Terms in Estonian Sign Language*. *Dissertationes linguisticae Universitatis Tartuensis* 14. Tartu: Tartu University Press.
- 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.] *Tööd antropoloogilise ja etnolingvistika vallast* 1. Tallinn: Eesti Keele Sihtasutus.
- Language Archiving Technology. ELAN. Max-Planck-Institute for Psycholinguistics. <http://www.lat-mpi.eu/tools/elan> (01.01.2012).

- 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>
- 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. – Mati Erelt (Ed.). *Estonian Typological Studies V*. Publications of the Department of Estonian of the University of Tartu 18. Tartu: University of Tartu, 169–188.
- Miljan, Merilin 2003. Number in Estonian Sign Language. – *Trames*, 7 (3), 293–223.
- 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]. *Töid antropoloogilise ja etnolingvistika vallast* 3. Tallinn: Eesti Keele Sihtasutus.
- Sandler, Wendy; Diane Lillo-Martin 2006. *Sign Language and Linguistic Universals*. Cambridge: Cambridge University Press.
- Searle, John Rogers 1976. A classification of illocutionary acts. – *Language in Society*, 5 (1), 1–23. <http://dx.doi.org/10.1017/S0047404500006837>
- State Gazette 2007. Language Act. <https://www.riigiteataja.ee/ert/act.jsp?id=12796913> (11.02.2013).
- Sutrop, Urmas 2005. *Estonian Sign Language*. – Raymond Jr. Gordon (Ed.). *Ethnologue: Languages of the World*. Fifteenth edition. Dallas, Texas: SIL International.
- Taniroo, Liisi 2007. Eesti viibete võrdlus vene viibetega Swadesh 200 nimekirja alusel. Bakalaureusetöö. [The Comparison of Estonian Signs with Russian Signs Based on Swadesh 200-word List. BA thesis.] Tartu Ülikooli eripedagoogika osakond.
- Trükmann, Monika 2006. Ajasuhete väljendamine eesti viipekeeles. Magistritöö. [Expressing Time in Estonian Sign Language. MA thesis]. Tartu: Tartu Ülikool.
- 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.] Tallinna Ülikooli eesti filoloogia osakonna toimetised 6. Tallinn: Tallinna Ülikooli Kirjastus.
- Zabrodskaja, Anastassia 2009. *Russian-Estonian Language Contacts: Grammatical Aspects of Language Use and Change*. Tallinn University dissertations on humanities. Tallinn: Tallinn University Press.

**Kadri Hein** (Tallinn University) is a teacher of deaf children at Tallinn Helen's School and a lecturer at Tallinn University. Research interests: sociolinguistics, bilingualism, Estonian Sign Language, second language acquisition.  
[kadri.hein@helen.edu.ee](mailto:kadri.hein@helen.edu.ee)

# EESTI VIIPEKEELE JA SUULISE EESTI KEELE VAHELISE KOODIMUUTUSE KVANTITATIIVSED JA KVALITATIIVSED ASPEKTID

**Kadri Hein**

Tallinna Ülikool

Artiklis uuriti, mis näitajad mõjutavad koodimuutust (koodivahetust ja koodisegunemist) koolikeskkonnas bimodaalsete kakskeelsete hulgas. Suunatud vestlustest kogutud andmed näitasid, et bimodaalsed kakskeelsed eelistavad koodisegunemist, s.t nad viiplevad ja räägivad samaaegselt. Kuigi koodisegunemise maatrikskeel oli enamasti eesti keel, jäi see mõnikord määramata, kuna koodisegunemine võis esineda erineva sõnajärgiga eesti keeles ja eesti viipekeeles.

Kuigi peamine koodimuutuse funktsioon õpilaste seas oli lauseosa rõhutamine, muudeti koodi ka huumori ja keelemängu eesmärgil. Mõnedes lausungites osutus koodimuutuse funktsiooni määramine keeruliseks, seega analüüsiti vestlust vooru järel (Auer 1995). Kurdi õpilase koodimuutuse tase sõltus kõige enam sellest, kas vestluspartner oli kurt või kuulja ning mil määral vestluspartner koodi muutis.

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



# SÕJANDUSKEELE JA KUJUNDKEELE RISTTEEL EHK KAS KUJUND (OSKUS)KEELT PIGISTAB?

Reet Hendrikson

**Ülevaade.** Eesti terminoloogiatöö praktikas on seni piisava tähelepanuta jäänud mõistetasandil avalduv kujundlikkus. Ka on suhtumine erialakeele kujundlikkusesse ambivalentne: kuigi terminiteoreetilistes käsitlustes kohtab üha enam kujundi kui tunnetusvahendi tähtsustamist, on praktikas suhtumine jätkuvalt pigem kõhklev-tõrjuv. Samas on kujundliku taustaga termin sageli erialainimeste loomulik valik.

Artikkel käsitlebki kujundlikkuse rolli oskuskeeles. Uurimisaineks on magistrikursuse ohvitseride terminivalikud ja -põhjendused. Vaatlen, millal kalduvad ohvitserid eelistama kujundlikku terminit ja kuidas valikut põhjendatakse. Lähteks on tees, et funktsionaalse ja juurdumisvõimelise terminoloogia väljatöötamine eeldab teadlikkust sihtgrupi vajadustest, terminieelistustest ja mõistestusviisist.

Sõjanduses, nagu teistelgi erialadel, on käibel kujundliku taustaga termineid, mida on traditsioonilises (oskus)keelekorralduses tavaks taunida kui toorlaene. Selgub, et taolised terminikujud omavad tihti erialainimeste eelistust ja on juurdunud. Ohvitseride terminivalikuid ja -põhjendusi jälgides tundub olevat vähemalt kaalumist väärt hüpotees, et lisaks puhtkeelelisele matkimisele võib taoliste keelendite eelistatus tuleneda inimtunnetuse kalduvusest mõ(is)testada teatavaid nähtusi kujundi abil. See omakorda seab teise valgusse nii mõnegi eesti (oskus)keelekorralduses seni kehtinud põhimõtte.

**Võtmesõnad:** sõjandus, terminoloogia, sotsiokognitiivne terminoloogiateooria, oskuskeelekorraldus, terminikorraldus, kognitiivne metafoorteooria, erialakeele metafoorsus, kujundlikkus, eesti keel

## 1. Taust ja eesmärk

Eesti oskuskeelekorralduses näib praegu olevat põnev otsingute aeg. Ühest küljest – teooriatasandil – on terminoloogiaski saamas üha valdavamaks kognitivism ja mõiste kui tunnetusvahendi tähtsustamine. Teisalt tundub praktilises terminitöös

jätkuvalt valitsevat traditsiooniline paradigma, mis kaldub lähtuma objektivismist (mõiste kui midagi inimtunnetusest sõltumatut), mõiste-termini üksühesest vastavusest ja kui mitte kujundlikkuse eitamisest, siis vähemalt skepsisest kujundlikkuse suhtes<sup>1</sup> ning kalduvusest “tõlkida” kujundlikke termineid mittekujundlikeks. Viimast tendentsi on siinkirjutaja täheldanud esmajoones just – kaitseväelaste sõnakasutust järgides – keeleinimeste, mitte niivõrd kaitseväelaste seas. Kujundlikkuse ja ekspressiivsuse kadumist sihtkeeles on täheldanud ka Temmerman, kes on kõrvutanud ingliskeelset erialateksti selle prantsus- ja hollandikeelse tõlkega (vt Temmerman 2011a: 56–57). Eestikeelses erialatekstis on kujundlike väljendite asendumist otsesõnalistega täheldanud Rohult (2008: 163). Juurdunud tavaarusaama järgi on kujundkeel aga midagi, mis ei käi kokku teaduslikkuse ja täpsustaotlusega, mistõttu tekib küsimus, miks kalduvad erialainimesed nii mõnigi kord eelistama kujundlikke termineid mittekujundlikele<sup>2</sup>. Lisaks on sõjandus teatavasti ala, kus täpsus ja üheselt mõistetavus on iseäranis olulised (seda rõhutavad ohvitserid oma esseedes ja magistritöödes; vt Hendrikson 2012).

Senises keelekesksusele kalduvas terminoloogiapraktikas on piisava tähelepanuta jäänud mõistetasandil avalduva kujundlikkuse osatähtsus. Kaitseväelased, nagu ka paljude teiste erialade esindajad, kasutavad ohtralt võõrmõjulisi, eeskätt inglise keelele taandatavaid keelendeid, millest nii mõnelgi on kujundlik taust. Inglise keele mõju liider- ja malliandva keelena on sõjanduses vaieldamatu, kuid ohvitseride terminivalikuid ja -põhjendusi jälgides tundub olevat vähemalt kaalumist väärt hüpotees, et lisaks puhtkeelisele matkimisele võib kujundlikkus tuleneda ka mõistetasandist, sh inimtunnetuse kalduvusest mõ(is)testada teatavaid nähtusi kujundi abil<sup>3</sup>.

Siinse kirjutise eesmärk ongi käsitleda kujundlikkuse rolli sõjandusterminoloogias. Kaks peamist uurimisküsimust on järgmised: mida lubab aines öelda kujundlikkuse funktsionaalsuse kohta sõjandusterminoloogias ning kuidas võib vastus sellele küsimusele mõjutada sõjandusterminoloogia arendamist? Selleks et neile küsimustele vastata, on oluline vaadelda ka seda, millal kalduvad kaitseväelased eelistama kujundlikku terminit ja kuidas siis oma terminivalikuid põhjendatakse.

Artikkel lähtub eeldusest, et funktsionaalse ja kasutajaskonnas juurdumisevõimelise terminoloogia väljatöötamine eeldab muuhulgas ka teadlikkust sihtgrupi vajadustest ja terminieelistustest. Uurimistulemused peaksid seega olema otseselt rakendatavad sõjandusterminoloogia komisjoni töös ning eeldatavasti aitama lähendada komisjoni tööd sõjandusterminoloogia peamisele kasutajaskonnale – kaitseväelastele.

Esmalt annab käsitus lühiülevaate nüüdisaegsest, sotsiokognitiivsest lähenemisest terminoloogiale, suhestades seda klassikalise terminoloogiateooriaga. Kirjutise teine, rakenduslik osa tugineb peaaesjalikult kahele terminianalüüsi näitele, mis on pärit kahe ohvitseri sõjandusterminoloogiat korrastavast magistritööst.

<sup>1</sup> Üldlevinud suhtumist illustreerib Egle Pulleritsu (2012) tsitaat, mis on küll kontekstist välja rebitud, kuid seda ilmekam: “Loomulikult on keeles olemas konventsionaalsed ehk nn surnud metafoorid, mille puhul ülekantud tähendust enam ei tajugi, nt *käsi* väljendis *kellegi käest küsima* või *kõrv* väljendis *kellegi kõrvale istuma*. Nende tarvitust pole mõtet piirata, sest **olguigi et need on mõnikord välditavad, oleks üsna keeruline neist täielikult loobuda.**” (Minu esiletõst – RH.) Viimane lause ilmestab hästi intuiitivist üldsuhetumist, mis – nagu näha – pole võõras ka keeleteadlikumatele inimestele: kui võimalik, siis vältida!

<sup>2</sup> Samale tulemusele jõuab ka Temmerman (vt nt 2000: 69).

<sup>3</sup> Ka Erelt (2007: 80–81, 197–198) viitab, et sageli on raske kindlaks teha, kas tegu on keelest keelde rändava tõlkelaenuga või on terminiloojad eri keeltes tulnud samale mõttele.

## 2. Klassikaline ja sotsiokognitiivne terminoloogiateooria

Klassikalise terminoloogiateooria järgi saab kindlustada termini objektiivsust, täpsust ja suhtlustõhusust, vältides kujundlikkust kui hägusat väljendusviisi. Kujundlikku keelt käsitleb klassikaline teooria igasuguse teadusliku ja loogilise mõtlemise komistuskivina, mis peale kõige muu ähmastab mõistepiire (vt nt Lotte 1961, Wüster 1979, Felber 1984; vrd ka Ots 1980: 583). Kognitivismi esiletõusuga on hakatud sellist lähenemist üha enam vastustama<sup>4</sup>.

McLaughlini (1995: 86) järgi mängib kujundlikkus väga olulist rolli aladel, mida on traditsiooniliselt peetud kujundkõnet välistavaks, sh teaduses. Sageli ollakse kujundiga lihtsalt nii harjunud, et seda ei märgata. Samas täheldavad teised, et nüüdisaegsed käsitlused küll tähtsustavad üha enam metafoori rolli, ent seavad siiski jäigad piirid metafoori kasutamisele teaduses ja erialakeelis (vt nt Gentner, Jeziorski 1998: 448, Faber 2011: 10). Sotsiokognitiivse terminoloogiasuuna rajajaks peetud Temmermani (nt 2011a: 48) uurimusest selgub aga, et erialasuhtluses on kujundlikkus sageli intuiitiivne ja loomulik valik (vt nt ka Temmerman 2009, Tercedor Sánchez jt 2012). Faber, kelle freimiteooria lähtub muuseas ka neurouuringute tulemustest, toonitab märksõnade *dünaamika* ja *kontekst* olulisust. Nendest järeldub otseselt mittehierarhiliate (*resp.* assotsiatiiv- ehk kujundlike) seoste osatähtsus nii terminoloogias kui tunnetuses (Sager jt 1980: 253, Faber 2011: 14–15, Tercedor Sánchez jt 2012: 1). Kujundlik keelekasutus võimaldab seega tunnetuslikku ligipääsu mõistevaldkondadele, mida oleks muidu raske mõtestada ja keelestada (vrd Black 1962: 37, Boyd 1998: 486, Gentner, Jeziorski 1998: 447, Temmerman 2011a: 48, Temmerman 2011b: 167). Veel enam: Boyd (1998: 482, 483, 486) rõhutab, et metafoor võib tänu oma tähenduslikule avatusele aidata kaasa teadusliku mõtte arengule ja meie maailmamõistmise täiustumisele. Ometi tuleb tõdeda, et kuigi teoorias on üha enam levinud klassikalise koolkonna kriitika ja kognitiivne kanda kinnitamas, oleks terminoloogiatöö praktikas veel vara kõnelda paradigmuuutusest.

Alljärgnev tabel 1 võtab – küll lihtsustatult ja kohati äärmustesse viiduna – kokku klassikalise ja sotsiokognitiivse terminoloogiateooria põhierinevused. Välja on toodud vaid seigad, mis on siinsel juhul olulisimad. Tabel kujutab endast erinevate käsitluste sünteesi, millest johtub niisiis ka tõlgendav alge.

**Tabel 1.** Klassikaline ja sotsiokognitiivne terminoloogiateooria<sup>5</sup>

Klassikaline terminoloogiateooria	Sotsiokognitiivne terminoloogiateooria
<ul style="list-style-type: none"> <li>• Esiplaanil keele sünkroonne mõõde (terminoloogia kui staatiline süsteem);</li> <li>• mõiste kui vajalike ja piisavate tunnuste staatiline kogum;</li> <li>• mõistete vahel on selged piirid;</li> <li>• tähenduse otsimine;</li> <li>• kujundlikkus kui hägus, erialateksti sobimatu väljendusviis;</li> <li>• terminikasutuse kontekstisõltumatus</li> <li>• "El!" sünonüümiale, polüseemiale ja homonüümiale;</li> <li>• terminoloogia ülesanne on eeskätt ühtlustamine ja standardimine.</li> </ul>	<ul style="list-style-type: none"> <li>• Nii reaalsus, keel kui erialane mõte on pidevas muutumises (oluline on termini diakrooniline külg);</li> <li>• mõiste (ehk teadmusüksus) kui tähenduse loomise <i>protsess</i>;</li> <li>• põhiosa mõisteid iseloomustab prototüübistruktuur;</li> <li>• tähenduse konstrueerimine;</li> <li>• kujundlikkus kui mõtlemise hõlbustaja ja tähenduse looja;</li> <li>• nii termini valik kui tähendus sõltuvad kontekstist;</li> <li>• sünonüümia, polüseemia ja homonüümia kui keelele olemuslikud ja funktsionaalsed nähtused;</li> <li>• terminoloogia ülesanne on käsitleda variatiivsust, mõiste kujunemislugu ja kujundlikkust.</li> </ul>

<sup>4</sup> Õigupoolest pole kognitivistika seisukohtades midagi pörutavalt uut, kuigi kognitivistide reetoorikast võiks seda järeldada: nii osutavad juba mitmed romantikud (nt Shelley) keele metafoorsele olemusele ja kognitiivsele funktsioonile. Rousseau on veendunud, et kogu keel põhineb metafoorsusel (sama seisukohta deklareerib hiljem ka Nietzsche), sõnasõnaline keel aga kujutab endast meie olemuselt kujundliku mõtlemise "ärandumist". (Kittay 1987: 4–6)

<sup>5</sup> Põhiallikad: Temmerman 2000 ja 2002, Tavast 2002, Faber 2011.

### 3. Uuritav aines

Edasises vaaten, kuidas sobitub eeslitatuga nii praktiline eriala, nagu seda on sõjandus, kus “täpsus ja üheselt mõistetavus on iseäranis olulised” (Tameri 2009: 22). Ohvitseride arvamusalaldusi ja terminivalikuid jälgides näib küsimus igati õigustatud.<sup>6</sup>

Käsitlus põhineb peaaesjalikult kahel analüüsinäitel, mis on pärit kahest sõjandusterminoloogiat korrastavast magistritööst (vt Paimla 2007, Tameri 2009). Nende koostajad on Kaitseväe Ühendatud Õppeasutuste keskastmeksuse ohvitserid, niisiis oma erialal juba kogenud spetsialistid – nagu ka kõik need, kes tööde autorite korraldatud küsitlustes osalesid. Mõlema töö põhiosa moodustab ühe kitsama sõjandusvaldkonna probleemsete mõistete ja nende võimalike omakeelsete vastete analüüs. Analüüsi tulemusel pakuvad tööde autorid iga käsitletud mõiste tähistamiseks välja terminivariandi, mis osutus nii mõisteliselt kui keeleliselt sobilikemaks. Kõigis uurimustes on analüüsi üks olulisi lähtekohti – kuid valikute puhul siiski mitte määrav tegur – valdkonna asjatundjate seas korraldatud küsitlus<sup>7</sup>. Magistritööde autorid lähtusid oma terminivalikutes niisiis 1) mõistetasan-dist (täenduslik kül), 2) vormitasandist (keeleline kül) ning 3) eksperdirühma terminieelistustest ja/või terminite juurdumusest.

Ohvitseride oskuskeelekäsitust (vt ka Hendrikson 2012) uurides olid huvitavamad kohad, kus kirjutajal tuli valida kujundliku ja mittekujuandliku terminivariandi/-variantide vahel, ning see, kuidas oma valikut põhjendati. Järgnevad kaks näidet on esiteks üsna tüüpilised, teiseks mitmete vormi- ja mõistestunnuste poolest sarnased, ent kummatigi on tehtud valikud erinevad. Ohvitseride terminivalikutega kaasnevate mõistestamismudelite avamiseks peatun põgusalt ka sellel, millist metafoorsust need kannavad. Mõlemas tekstinäites on siinse käsitluse seisukohast olulisimad kohad esile tõstetud.

Näites (1) otsib jalaväetaktika termineid käsitleva magistritöö autor Paimla teiste seas vastet inglise terminile *canalize*<sup>8</sup> ja leiab, et vaste ei saa olla *suunama*, kuna olemas on ka naabermõiste *turn* (Paimla 2007: 60). Seega soovib Paimla kasutada juurdunud terminit *kanaliseerima*, kuigi ÕS-is sõnale täpselt sellist tähendust ei esitata<sup>9</sup>. Töö autor põhjendab oma eelistust järgmiselt.

- (1) “Inglisekeelsele terminile *canalize* vaste määratlemisel tuleb arvestada terminit *turn*, kuna need **asuvad mõistesüsteemis lähestikku**, kuid vastavatel tegevustel on selgelt erinev eesmärk. VL (2006: 480) annab sõnale *kanaliseerima* üheks võimalikuks vasteks *suunama*, kuid sõjanduskeeles on otstarbekas siduda need sõnad erinevate mõistetega. Termin *kanaliseerima* annab analüüsitava mõiste tähenduse paremini edasi, st vastast määratud kohta kokku [esiletõst originaalis] *suunama* (näiteks hävitusalasse), et seal teda hävitada. Neist teise (*turn*) eesmärk ei ole vastast kuhugi kokku suunata ---.” (Paimla 2007: 60)

<sup>6</sup> Ohvitseride oskuskeelevaadetest ilmnis kohati üsna äärmuslikki konservatiivsus (nt ühene seisukoht, et sünonüümial ei tohiks olla sõjanduskeeles kohta, samuti arvamus, et standarditud sõjandusterminoloogia tuleb muuta kogu kaitseväes kohustuslikuks). Samas viitavad nii mõnedki seigid nende terminieelistustes ja -põhjendustes samameelsusele sotsiokognitiivse teooriaga. (Vt Hendrikson 2012: 27–28, 30)

<sup>7</sup> Küsitluste eesmärk oli selgitada välja erialainimeste terminieelistused. Jalaväetaktika termineid käsitleva töö eksperdirühm koosnes 20 ohvitserist ja omakeelseid vasteid otsiti 35 mõistele. Logistikatermineid uurinud töö eksperdirühma suurus oli 51 ohvitseri ja vasteid otsiti 30 mõistele.

<sup>8</sup> Üks lahingutoimingutest (ingl *tactical tasks*): ‘suunama vastast kitsasse liikumiskoridori’ (vrd *turn* ‘sundima vastast liikumissuunda muutma’).

<sup>9</sup> ÕS-is esineb sõna vaid tähenduses *kanalistsiooniga varustama*. Kanaliseeritud asula (ÕS 2006). Tüve pooldas 20 eksperdist 16 (Paimla 2007: 60).



Sarnase, kuid põhimõtteliselt vastupidise lahendusega näite leiab logistikatermineid käsitlevast tööst (Tameri 2009). Tameri magistr töö üks suuremaid probleemkohti on vaste leidmine inglise terminile *cannibalize*, kusjuures ka siin on ohvitseridest koosneva eksperdirühma selge eelistus võõrsõnaline variant, praegusel juhul *kannibaliseerima*<sup>10</sup>. Näide (2) on hea ka seepärast, et peale mõistetasandi ühisjoonte ja vajaduse eristada lähimõisteid (siin lähimõisteks *salvage*, mille eesti vasteks pakutakse samas töös *lammutama*<sup>11</sup>) on näitega (1) sarnane ka vormiline külg. Nii *kannibaliseerima* kui *kanaliseerima* on võõrsõnad, mille puhul on tähistaja ja tähistatava vahel ISO terminikasutusest lähtudes assotsiatiivne seos (vt EVS-ISO 2002: 10, resp. kujundlik suhe<sup>12</sup>). Tameri eelistab aga võõrsõnalisele kujundlikule variandile *kannibaliseerima* mittekujundliku omasõna (*osandama*), põhjendades oma valikuid alljärgnevalt.

- (2) “ÕS annab sõnale *kannibal* tähenduse *inim(ese)sööja*; üle kantud tähenduses *toores, metsik inimene, zooloogias liigikaaslasti sööv loom*. **ÕS-ist lähtudes ei sobi** tüvest *kannibal* tuletatud termin siinse mõiste sisuga. **Kaudne loogika on siiski olemas** – sama tüüpi sõidukite lõhkumine teiste töökorras hoidmise eesmärgil. Tüvest *kannibal* tuletatud termini variandid on ekspertgrupi põhieelistus, kogudes kahe variandi peale 50% vastustest. U. Mereste liigituse järgi on tegemist **kolmandase tuletusega, mis terminiloomeks ei sobi**. Kolmandase tuletamise puhul tõlgitakse võõrkeelne termin otse omakeelde ning seos mõistega on praktiliselt olematu ---. Sõnaartikli analüüsi käigus ei selgunud ühtki ideaalselt sobivat varianti. Ekspertgrupi eelistust [*kannibaliseerima*] töö autor toetada ei saa, sellepärast et kolmandase tuletamise teel ei saada terminit, vaid pigem **võõrkeelse termini tõlge**. --- ÕS (2006) on *osandama* sünonüümidega märkinud *lahti võtma* ja *demonteerima*, mis kinnitab, et *osandama* sobib mõistet tähistama.” (Tameri 2009: 61–62)

## 4. Analüüs

### 4.1. Mida neist näidetest leida on?

Ootuspäraselt ei kasuta kumbagi näite autor põhjenduses sõnu *piltlik, kujundlik, metafoorne* vms. Näites (1) tõdeb Paimla, et *kanaliseerima* “annab tähenduse paremini edasi”<sup>13</sup>. Sellele üldsõnalisele konstateeringule järgneb täpsustus, millega kirjutaja püüab selgitada tekkivat tähenduseseost, võttes abiks tekstigraafika (“vastast määratud kohta **kokku** suunama”). Nimetatud seigad lubavad oletada,

<sup>10</sup> *kannibaliseerima*-tüve pooltas 51 eksperdirühma liikmest 21, *osandama*-tüve 2 ohvitseri, kõiki ülejäänuid mainiti vaid üks kord (Tameri 2009: 61). Üldse pakuti vasteks tervelt 20 (*sic!*) terminivarianti, millest mõni on küll taandatud ühele variandile (*mine-* või *ja-*liite varieerumine ehk sõnaligi erinevus ning õigekirja- või trükkiviga sisaldavad keelekujud; vt Tameri 2009: 61).

<sup>11</sup> Mõistete erinevus seisneb lühidalt selles, kas seade läheb pärast demonteerimist veel kasutusse (ingl *cannibalize*) või mitte (ingl *salvage*; vrd Tameri 2009: 60–63).

<sup>12</sup> Nimelt ei esine viidatud ISO standardis sõnu *kujundlik(kus)*, *metafoor* ega *metonüümia*, mis näib juba iseenesest kõnekas. Nende asemel kõneldakse – põgusalt ja just kui möödaminnes – assotsiatiivseostest, mida defineeritakse kui “kahe sellise mõiste suhet, mida **kogemuse tõttu** (!) ühendab mingi mittehierarhiline seos” (EVS-ISO 1087-1:2002). Seevastu nt Faber (2011: 15) tõdeb kognitiivse psühholoogia ja neuroteaduste uurimistulemustele tuginedes, et kujundlikkusel põhinevad mõistesuhted on tunnetuslikust seisukohast vähemalt sama olulised kui hierarhiaseosed.

<sup>13</sup> Konteksti arvestades on paslik mainida, et kahe maailmasõja vahelisel perioodil oli kaitseväs samas tähenduses kasutusel termin *lehterdama*, millel on samuti kujundlik taust (vrd *lehttrise ajama*, niisiis kujundiskeemilt sarnane terminivariant). Viimase puhul poleks tegu võõrsõnaga ja see ei peaks tekitama ka häirivaid kõrvalseoseid, ent *kanaliseerima* on kõnealuse mõiste tähistamisel ilmselgelt juurdunud. Kujundlik oli ka ainus eksperdirühma väljapakutud vaste (*sängitama*). Niisiis näib, et teatud mõistete tähistamisel kaldutaksegi eelistama kujundlikku terminit.

et näite (1) autor on adunud küll mingit liiki assotsiatiivseoste (selle kohta vt ka 12. joonealust märkust) olemasolu, kuid nende kirjeldamiseks napib metakeelevahendeid. Taolist oletust toetab Temmermani (2002: 211) tähelepanek, et keelendi aluseks olevad kujundiskeemid ei avaldu leksikaalsel tasandil kunagi täielikult ja üksüheselt (vrd ka Black 1962: 37), mistõttu on ka nende kirjeldamine keerukam.

Kui Paimla rõhub oma terminivaliku põhjenduses mõisteseoste ja sõnastikus sisalduvat teavet kasutab vaid ühe argumendina, siis tsitaadis (2) saab määravaks teguriks ÕS-is esitatud. Kuna ÕS-i järgi on sõnal *kannibal* teistsugune tähendus, siis ei sobi Tameri järgi ka sellest johtuv verbituletis. Tõdedes, et ideaalselt sobivat varianti analüüsi käigus ei selgunud, eelistab Tameri mittekujundlikku varianti *osandama*, mis on “keeleliselt ökonoomne, suupärane ja eestikeelne, kuid võiks olla läbipaistvam” (Tameri 2009: 62). Huvitav on tõik, et kuigi *kannibal*-tüve peab Tameri tähenduslikult sobimatuks, tuuakse tsitaadis ometi esile kaudse loogika (!) olemasolu. Sellele järgnevast selgitusest (“sama tüüpi sõidukite lõhkumine teiste töökorras hoidmise eesmärgil”) aimub, et kirjutaja on siiski tajunud mingit laadi sarnasusseost. Sõnastades sel moel metafoori ülekandaluse, läheb Tameri oma analüüsis sammu võrra kaugemale esimese omast.

#### 4.2. Kaks näidet – kaks erinevat oskuskeelekäsitust

Näidetel (1–2) on olulisi ühisjooni: lisaks sellele, et mõlemad terminid on lähtekees kujundlikud, eelistas kujundlikku terminit ka põhiosa ohvitseri-eksperte. Terminivariantide sobivuse hindamisel kasutavad mõlema töö autorid sõnastikus esitatud teavet (vastavalt VL ja ÕS). Samuti viidatakse otsesemalt või kaudsemalt sõnast johtuvatele (kõrval)seostele (“määratud kohta **kokku** suunama”; “kaudne loogika on siiski olemas”).

Ent kui vaadata tehtud valikuid, on eeltoodud analüüsinäited (küll mõnevõrra tinglikult) taandatavad kahele põhimõtteliselt erinevale oskuskeelekäsitusele. Esimene (st valik *kanaliseerima* kasuks) tõukub eeldusest, et kujundlik terminikasutus kätkeb võimalust anda edasi nüansse, mis oleks muul moel raskendatud. Näites (2) on olukord komplitseeritum. Selle autor Tameri täheldab, et *kannibal*-tüve kasutamisel on olemas küll kaudne loogika, kuid otsustamisel saab määravaks sõna-, mitte mõistetasand. Kirjutaja viitab analüüsis Merestele (2000: 85), kelle liigituse järgi on tegu kolmandase terminimoodustuse ilminguga, mis termini-loomeks ei sobi. Tameri väidab, et “seos mõistega on praktiliselt olematu”, teisel aga, et “kaudne loogika on siiski olemas”. Kui lähtuda sotsiokognitiivsest läheneemisest, on terminivariandil *kannibaliseerima* (nagu ka selle inglise vastel) selgelt kujundlik taust, mistõttu see võiks keeleülese kontseptuaalse metafoori ilminguna justkui funktsioneerida ja anda tähendust paremini edasi kui paronüümipesa tekitav *osandama*. Kujundlikkust on ilmselt siiski tähele pannud ka Tameri, viidates kaudse loogika olemasolule. Seega on viimane analüüs tinglikult taandatav klassikalisele terminoloogiateooriale. Kõnekas on, et kujundliku taustaga varianti *kannibaliseerima* pooldas ligi pool eksperdirühmast, teisi variante mainiti (ühe erandiga) vaid üks kord (vt Tameri 2009: 61).

Mõistetasandil peegeldab *kannibaliseerima* (nagu ka selle inglise keelekuju) MASINA mõtestamist ELUSOLENDINA. *Kanaliseerima* sisaldab SUUNA-skeemi ja MAHUTI-skeemi (sõnastatav kontseptuaalse vormelina VASTANE ON VEDELIK, siit

edasi üldisema tasandi metafoorina<sup>14</sup> VASTANE ON OBJEKT). Sel moel kajastavad need mõlemad kaitsevääslaslikku – või kaitsevääslaslikuks peetavat – mõtteviisi: masina mõtestamist inimesena ja inimese käsitamist objektina.<sup>15</sup>

### 4.3. Millal kaldutakse eelistama kujundlikku terminit ja kuidas oma eelistusi põhjendatakse?

Kui vaadelda nende kahe näite kõrval teisi terminivalikuid ja nende põhjendusi (lisaks käsitletud töödele ka Afanasjev 2011), on võimalik täheldada teatavaid üldsuundi või tingimusi, millal kaldutakse eelistama kujundlikku varianti. Ohvitseride eelistus kaldub langema kujundlikule variandile siis, kui 1) on vaja eristada lähimõisteid (mõistekeskne<sup>16</sup> argument) või kui 2) termin on juba lähtekeeles kujundlik (esmapilgul sõnakeskne argument). Sageli esinevad need kaks koos nagu ka mõlemas siinses näites. Otsuse langetamisel on tihti peale määrava tähtsusega see, et 3) tekkivad kõrvalseosed ei häiriks.

Põhjendused jagunevad laias laastus neljaks. Viidatakse 1) lähimõistete eristamise vajadusele, 2) terminikuju juurdumusele, samuti 3) alg- ja sihtkeele terminikujude vormilisele sarnasusele (tekkivale “ühesele seosele” inglise ja eesti variandi vahel) ning 4) ilmnevatele (kõrval)seostele. Põhjendustest on adutav, et tekkivatel assotsiatsioonidel on sageli otsustav roll. Nagu eespool mainitud, osutatakse subjektiivsetele kõrvalseostele ka analüüsinäidetes (1–2). Subjektiivsete seoste olulisust tõendab seegi, et neid nimetatakse sageli ka kombineerituna teiste argumentidega<sup>17</sup>.

Ohvitserid ei mõtesta niisiis oma terminivalikuid skaalal *kujundlik-mittekujundlik*, vaid kõneldakse valdavalt subjektiivselt keeletajust ja sellest johtuvatest assotsiatsioonidest (iseloomulikud sõnad-fraasid: *assotsieerub, peegeldab, seonduv, seostub; väärassotsiatsioonid, negatiivne varjund; tunduvad kummalsed ja võõrad* jt). Sageli nimetatakse, et sõna annab/ei anna tähendust hästi/täpselt edasi nagu ka alljärgnevas näites (3).

- (3) --- vaste *kokkupõrkeala* [ingl *engagement area*] võib tekitada **väärassotsiatsioone** (näiteks seostub alaga, kus toimus autoavarii) ning **ei anna selgelt mõiste sisu edasi**. (Paimla 2007: 51)

Taolisest põhjendusest peegeldub küll eeldus, et terminikuju valikul peab lähtuma mõistest, ent üldjuhul ei selgitata, milles mõistele mittevastavus täpsemalt seisneb.

Kokkuvõttes ilmneb, et ohvitserid kalduvad eelistama võõrsõnalist või tõlkelaenuist terminivarianti, millel on nii mõnigi kord kujundlik taust. Neil juhtudel on tegu

<sup>14</sup> Metafoorid erinevad üldistusastmelt: mida üldisema tasandi metafoor, seda üldisemad on allik- ja sihtala ning seda abstraktsem ja ühtlasi visandlikum metafoor (Kövecses 2002: 38–39, 249–250). Madalama tasandi metafooride üldistamisel toimib metaprojektsioon ÜLDINE ON ÜKSIK (*GENERIC IS SPECIFIC*; vt Lakoff, Turner 1989: 162). Nimetatud üldmetafoori ongi käsitatud tegurina, mis teeb võimalikuks metafooride loovuse ja mitmetähenduslikkuse (täpsemalt vt Krikmann 2003, Hendrikson 2002, Hendrikson 2003: 14–15).

<sup>15</sup> Sama mõtestusviisi peegeldavad sellisedki sõjandusterminid nagu *inimmaterjal, kannatanute käitlemine, masina elutsükl, kokkupõrge* 'lahingukontakt', *surnud ala, üksuste manustamine* 'juurdeandmine' jt. Mõistagi ei saa nimetatud viisi pidada universaalseks ja ainukehtivaks: nagu metafooride puhul ikka, on võimalik tuvastada erinevaid käsitusviise.

<sup>16</sup> Nimetan *mõistekeskseks* põhimõtet, mille puhul lähtutakse terminikuju sobivuse hindamisel esmajoones mõistest ja selle taga olevast mõistesüsteemist, *sõnakeskseks* kalduvust keskenduda vaid keeletasandile (sõna kui konkreetse tähejärjendi sobivus). See vastandus on mõistagi tinglik. Lähedased mõisted on leksikograafias kasutatavad onomasioloogiline ja semasioloogiline lähenemine (pikemalt vt nt Cabré 1999: 33–34, Tavast 2002).

<sup>17</sup> Näiteks lahingutoimingute *siduma* ja *fikseerima* puhul tuuakse esile vajadust eristada lähimõisteid ning sellega seoses nenditakse, et kõnealusel sõnal “on tähenduselt väga lähedased, kuid teine variant [*fikseerima*] viitab rohkem kinnistatusele” (Paimla 2007: 61).

(oskus)keelekorralduses traditsiooniliselt toorlaenudeks peetavate keelekujudega, mis annavad mõistet spetsialistide hinnangul täpsemalt edasi ja on kasutajaskonnas juurdunud. Põhjendustes ollakse sageli üldsõnalised ja hägusad, toetatakse subjektiivsele keeletajule (keelend “annab/ei anna tähendust täpselt edasi”). Sellise tendentsi üks võimalikke seletusi on, et kujundlikkust küll tajutakse, aga oma keeletaju ei osata täpsustada (keelendis, *resp.* mõistetasandil toimuva kirjeldamiseks puudub vajalik metakeel).

## 5. Järeldusi (*resp.* küsimusi) sõjanduskeele arendamise seisukohalt

Niisiis seisavad (sõjandus)terminoloogia korrastajad-arendajad kahe põhimõttelise küsimuse ees.

### 5.1. Subjektiivsus või isomorfsus ehk kuidas kirjeldamatut (?) kirjeldada?

Võiks arvata, et ohvitseride terminipõhjendustes kajastuva üldsõnalisuse tagamaid on muuhulgas seik, et kirjutajateks on sõjandus-, mitte keeletespetsialistid. Kumatigi ei saa üldsõnalisust ja isiklikust keeletajust tekkivate seoste nimetamist argumentidena pidada vaid nende magistritööde jooneks: taolised põhjendused on tavalised – ja argumentidena aktsepteeritavad – ka filoloogide ja terminikomisjonide terminivaidlustes. Kognitivismist lähtudes saab põhjenduste üldsõnalisust ja hägusust seletada nii, et metafoorsed mudelid ja nende aluseks olevad kujundiskeemid ei avaldu keeletasandil kunagi üheselt ja täielikult (vrd nt Temmerman 2002: 211), mistõttu on loogiline, et selliste aspektide kirjeldamine ongi keerukam. Pealegi pole allik- ja sihtala vastavus kunagi täielik – vastasel korral poleks ülekandel mõtet ja taoline olukord pole õigupoolest mõeldavgi.

Omasõnast johtuvate kõrvalseostega on võõrsõna eelistamist põhjendatud teisteski valdkondades (vt nt Erelt 2007: 129). Iseenesest ei ole subjektiivsuses midagi taunitavat, pigem vastupidi: keele tunnetuslikku funktsiooni ja kogemuspõhisust arvestades on see eelduspärane. Kuid terminiarutelude puhul on hea teadvustada, et sõnadega kaasnevad tähenduseseosed on inimestel erinevad. Pealegi võib üks ja sama seos ühte inimest häirida, teist mitte (nt *kokkupõrkeala* seostumine autoavariiga). Sel moel kipuvad arutelud taanduma jah-ei-ehk sobib-ei sobi-vaidlustele. Samale tõdemusele jõuab Tavast (2008: 178), nentides, et absoluutseid tõdesid keelendite sisemiste omaduste kohta ei saa kasutada terminivaidlustes argumentidena ning et rõhk nihkub sel juhul pigem suhtluses osalevatele inimestele, nende hoiakutele ja pädevustele. Loomulikult kaotavad taolised vaidlused mõtte juhul, kui mõtestatakse ümber (eesti) oskuskeelekorralduse üks seniseid põhimõtteid, milleks on muuhulgas ka ühtlustamispuue ja kindlasti küll mitte range preskriptiivsus, kuid siiski suunav-soovitav hoiak.

Standardimise mõttekuse seab küsimärgi alla Temmerman, kes peab seda klassikalise terminoloogiateooria igandiks (vt nt Temmerman 2000: 68). Sõjanduses teeb olukorra siinkirjutaja jaoks huvitavaks tõik, et kaitseväelased ise – sõjandusterminoloogia peamine kasutajaskond – peavad sõjandusterminoloogia arendamise

põhieesmärgiks ja sihiks just standardi loomist ning igasuguse variatiivsuse kõrvaldamist (nt üldlevinud idee, et sõjandusterminoloogia andmebaas Militerm tuleb muuta kohustuslikuks; vt Hendrikson 2012: 22). See, kas erinev mõistekäsitus ja terminikasutus tegelikke mõistmisraskusi tekitab, on eraldi uurimisteema, kuid ohvitseride ainuvaldavale seisukohale tuginedes võib erinevate terminivariantide kasutamine pöördida lahinguväljal kaose ja kaotusega (täpsemalt vt samas). Kui lähtuda sotsiokognitiivse terminoloogiateooria variatiivsusest ja kirjanduses viimasel ajal üha jõulisemalt kõlanud seisukohtadest, siis peaks see võimalus olema vähemalt teoreetiliselt välistatud.

## 5.2. Toorlaen või keeleülene kujund?

Kognitiivne teooria võimaldab näha teises valguses termineid, mida nii mõnigi keeleteimetaja soovitaks vältida kui toorlaene<sup>18</sup>. Kognitiivse metafoorteooria järgi viitab taoline keelekasutus tunnetuse ja keele kogemuspõhisusele, analoogia-tajule ja mõistete (*resp.* mõtlemisüksuste) prototüübistruktuurile (pikemalt vt nt Temmerman 2000: 61–71, samuti Rosch, Mervis 1975: 573–574). Piltlik termin on sageli ilmekam ja selle kasutamine tähendab tihti peale väljendusökonomiat, mis on terminoloogia jaoks iseäranis oluline (samale seisukohale jõuab ka Erelt 2007: 89–90). Fuertes-Olivera ja Nielsen (2011: 172) arvates peaks igasugune kakskeelne terminoloogiatöö lähtuma eeldusest, et lähte- ja sihtkeeles kasutatakse kontseptuaalselt sarnaseid termineid: kui termin on lähtekeeles metafoorne, siis peaks see olema metafoorne ka sihtkeeles, ja kui vastavat metafoori sihtkeeles veel pole, tuleks see keele- ja kultuuriolusid arvestades luua. Kõik nimetatud argumentid osutavad tõigale, et kujundliku termini eelistamine selle kujundlikkuse tõttu lähtekeeles ei viita mitte sõna-, vaid mõistekesksusele.

Kui kahe eelkäsitletud analüüsinäite juurde tagasi tulla, siis võiks küsida, kas mõlemal juhul tuleks eelistada võõrsõnalist, kujundliku taustaga terminit. Siinkirjutaja ei söanda võtta nii resolutsiooni seisukohta: inglise *cannibalization* vastena välja pakutud *osandama* sobib nii tähenduselt kui annab võrdluses naaberterminiga *lammutama* (ingl *salvage*) mõistet täitsa hästi edasi (kuna mõlemad on omasõnad, siis on täidetud ka süsteemsusnõue). Keelendiga *osandama* seoses tuleb siiski mainida, et oskuskeelekorralduses on soovitatud vältida paronüümpesade teket ehk liigset tuletiste moodustamist (Erelt 2007: 164). Teisalt on selge, et *kannibaliseerima* on peale võõrsõnalisuse pikem-kohmakam kui *osandama*. Ka Erelt (2007: 137) leiab, et *kannibaliseerima* kasutamine (tähdenduses 'varuosadeks lammutama') ei sobi. Siiski viidatakse teisel kujundlike terminite elujõule<sup>19</sup> ja vastustatakse nii äärmuslikke oma- kui võõrtermini kaitsjaid (vt nt Erelt 2007: 127).

Ka termin *kanaliseerima* on suunamistähenduses põhjustanud keeleinimeste pahameelt (sest "ÕS-is sõna *selles tähenduses* ei esine!"). Kummatigi ei ole mõistetasandil välja tuua midagi, mis räägiks sõna *kanaliseerima* sobimatusest kõnealust mõistet tähistama. Pealegi on mõlemale võõrsõnalisele terminivariandile kuulunud kasutajaskonna ülekaalukas poolehoid (vt Paimla 2007: 60, Tameri 2009: 61). Kas

<sup>18</sup> Termin (?) *toorlaen*, mida kasutatakse terminivaidlustes sageli argumentina (ja mida on siinkirjutaja ka ise teinud), on siiani defineerimata (seda ei esine ei "Eesti keele käsiraamatus" ega raamatus "Terminiõpetus"). Toorlaenuks peetava termini üks lihtsamaid näiteid sõjanduses on *võtmeala* (ingl *key terrain*), mida analüüsitakse ka jalaväetaktika terminite töös (vt Paimla 2007: 49). Kõnekas on, et kõik eksperdid eelistasid tõlkelaenuist täiendosa *võtme-*, samas on keeleteimetajad selle kasutamist tauninud (vt nt Mäekivi 2010: 31).

<sup>19</sup> "Kindlasti läheb käibebe kujundlik, osavalt piltlik omatermin ----" (Erelt 2007: 147). Erelt ja Kull (1997) nendivad ka, et metafoorsuse suhtes võiks eesti terminoloogias inglise keelest mõistuse piires eeskju võtta.

lähtuda niisiis põhimõttest, et võõrsõna sobib *vaid juhul*, kui sobivat omasõna pole, või peaks esiplaanil olema termini võime anda mõistet ilmekalt edasi ja juurdumus kasutajaskonnas? Juhul kui aktsepteerida laenulise taustaga kujundlike terminite eluõigust, tuleb olla valmis ka eesti (oskus)keelekorralduse senises praktikas valitsenud põhimõtete revideerimiseks.

## 6. Kokkuvõte ja arutelu

Kaitseväelaste terminieelistused annavad tunnistust sellest, et traditsiooniliselt toorlaenudeks peetavad, kujundliku taustaga terminikujud omavad kaitseväelaste selget eelistust ja on juurdunud. Põhjenduste üldsõnalisus (mis ilmses ka ohvitseride oskuskeelevaadete käsitlesest ja mis pole omane kaugeltki mitte ainult kaitseväelaste terminiaruteludele) osutab selgelt tõigale, et keelekasutajal puudub vajalik metakeel. Sotsiokognitiivse lähenemise järgi peaks terminoloogia kui uurimisvaldkonna eesmärk olemagi juhiste väljatöötamine keeletasandil ilmnevate metafoorsete mudelite kirjeldamiseks, mitte standardimine. Selle taustal on huvitav nentida, et kaitseväelased ise peavad terminoloogiatöö ülesandeks just standardi loomist ja variatiivsuse kõrvaldamist.

Ohvitseride terminipõhjendustest on adutav, et kujundlike võõrmõjuliste terminite taga ei ole tõenäoliselt pelk võõrkeele jäljendamise soov. Selle hüpoteesi kontrollimise tarvis on vaja küll edasise töö käigus hankida lisaandmeid, kuid käsitletud materjal viitab matkimisteooriale kui vaid ühele võimalikule selgitusele. Teatud mõistete või mõistevaldkondade puhul näib kujundlikkus olevat erialainimeste loomulikem ja käepäraseim valik. Seik, et mõistet tähistava lähtekeelse termini taga on sama kujund, viitaks selles valguses inimtunnetuse üldisele seaduspärale: ühe kogemusvaldkonna mõtestamisele teise abil. Nii oleks fraasi *teise keele jäljendamine* kõrval/asemel tihti täpsem kõnelda pigem *tunnetuslikust analoogiast* või *samahäälsusest*.

Käsitletud aines ei anna küll alust kõikehõlmavate järelduste tegemiseks, küll aga järgmiste küsimuste edasiseks käsitlemiseks. Kõigepealt: kust läheb piir toorlaenu (pelga teise keele jäljendamise) ning keeletasandil avalduva ja *oskuskeelekorralduses aktsepteeritava* kontseptuaalse metafoori vahel? Just nimetatud matkimisteooriaga on (oskus)keelekorralduses siiani põhjendatud paljude, teiste seas ka kujundlike terminite sobimatust. Ja edasi: kuidas seda piiri ikkagi kirjeldada? Teisisõnu: kas vastavate kirjeldusvahendite väljatöötamine on võimalik ja kui, siis millised need peaks olema? Kolmandaks väärrib eraldi esiletõstmist metakeele küsimus: termin (?) *toorlaen*, mida sageli kasutatakse terminivaidlustes argumentina ja mis on juba iseenesest hinnanguline, subjektiivne ja loosunglik, on senini defineerimata. Nende küsimuste teadvustamine võiks siinkirjutaja arvates toetada funktsionaalse ja kasutajaskonnas juurdumisvõimelise terminoloogia väljatöötamist ning ehk ka laiemalt terminoloogia kui uurimisvaldkonna arengut. Õeldu tähendab ühtlasi oskuskeelekorralduse aluspõhimõtete ja metakeele kui mitte revideerimist, siis vähemalt teadvustamist ja uuesti läbimõtlemit.

## Viidatud kirjandus

- Black, Max 1962. *Models and Metaphors*. Ithaca: Cornell University Press.
- Boyd, Richard 1998. *Metaphor and theory hange: What is "metaphor" a metaphor for?* – Andrew Ortony (Ed.). *Metaphor and Thought*. Cambridge: Cambridge University Press, 481–532.
- EVS-ISO 2002 = EVS-ISO 1087–1:2002. Terminoloogiatöö. Sõnastik. Osa 1. Teooria ja rakendus. Standardikeskus.
- Erelt, Tiiu 2007. *Terminiõpetus*. Tartu: Tartu Ülikooli Kirjastus.
- Erelt, Tiiu; Kull, Rein 1997. Euroopasse integreerumine ja keeleminetussündroom. – Sõnumileht, 11. veebruar. <http://keeleabi.eki.ee/artiklid2/euroopasse.html> (14.09.2012).
- Faber, Pamela 2011. The dynamics of specialized knowledge representation Simulational reconstruction or the perception-action interface. – *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication*, 17 (1), 9–29. <http://dx.doi.org/10.1075/term.17.1.02fab>
- Felber, Helmut 1984. *Terminology Manual*. Paris, Unesco, Infoterm, XXI.
- Fuertes-Olivera, Pedro A.; Nielsen, Sandro 2011. The dynamics of terms in accounting. What the construction of the accounting dictionaries reveals about metaphorical terms in culture-bound subject fields. – *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication*, 17 (1), 157–180. <http://dx.doi.org/10.1075/term.17.1.09fue>
- Gentner, Dedre; Jeziorski, Michael 1998. The shift from metaphor to analogy in Western science. – Andrew Ortony (Ed.). *Metaphor and Thought*. Cambridge: Cambridge University Press, pp. 447–480.
- Hendrikson, Reet 2002. Keeleteaduse teooriate metafoorikäsitusi. – *Keel ja Kirjandus*, 11, 764–776.
- Hendrikson, Reet 2003. Ilukirjanduslikud liitsõnalised metafoorid eesti keeles kognitivistliku metafooriteooria perspektiivist. Magistritöö. Juhendaja Haldur Õim. Tartu: Tartu Ülikooli eesti keele õppetool.
- Hendrikson, Reet 2012. Keelesõjas sõjakeel? Ohvitseride oskuskeelevaadetest. – *KVÜÕA toimetised*, 16, 15–39.
- Kittay, Eva Feder 1987. *Metaphor: Its Cognitive Force and Linguistic Structure*. Oxford: Clarendon Press.
- Krikmann, Arvo 2003. Kaasaegse metafooriteooria panus parömioloogiasse. <http://www.tlu.ee/~jaanike/loengud/mteooria.pdf> (19.09.2012).
- Kövecses, Zoltan 2002. *Metaphor. A Practical Introduction*. Oxford: Oxford University Press.
- Lakoff, George; Turner, Mark 1989. *More than Cool Reason: A Field Guide to Poetic Metaphor*. Chicago and London: The University of Chicago Press.
- Lotte, Dmitri 1961 = Д. Лотте 1961. *Оснвы построения научно-технической терминологии. Вопросы теории и методики*. Москва: Наука.
- McLaughlin, Thomas 1995. *Figurative language*. – F. Lentricchia, T. McLaughlin (Eds.). *Critical Terms for Literary Study*. Second edition. Chicago and London: The University of Chicago Press, 80–90.
- Mäekivi, Helika 2010. Üheülbalsed eurosõnad. – *Et lugeja võiks tulla. Artiklite kogumik*. Tallinn: Eesti Keele Sihtasutus, 7–39.
- Ots, Heido 1980. Kümme aastat ületuspõhimõtet. – *Keel ja Kirjandus*, 10, 582–587.
- Pullerits, Egle 2012. Ei taha enam elada loomadega rongis! – *Postimees*, 9. jaanuar. <http://arvamus.postimees.ee/696198/egle-pullerits-ei-taha-enam-elada-loomadega-rongis/> (10.01.2012).
- Rohult, Elo 2008. *Conceptualizing the World*. – N.-L. Johannesson, D. C. Minugh (Eds.). *Selected papers from the 2006 and 2007 Stockholm Metaphor Festivals*. Stockholm: Stockholm University, 151–164.

- Rosch, Eleanor; Mervis, Carolin B. 1975. Family resemblances: Studies in the internal structure of categories. – *Cognitive Psychology*, 7 (4), 573–605. [http://dx.doi.org/10.1016/0010-0285\(75\)90024-9](http://dx.doi.org/10.1016/0010-0285(75)90024-9)
- Sager, Juan Carlos; Dungworth, David; McDonald, Peter F. 1980. *English Special Languages. Principles and Practice in Science and Technology*. Wiesbaden: Brandstetter.
- Tavast, Arvi 2002. *Onomasioloogia ja semasioloogia vahekorra oskuskeeles*. Magistritöö. Juhendaja Mati Ereht. 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.
- Temmerman, Rita 2000. *Towards New Ways of Terminology Description: the Sociocognitive Approach*. Amsterdam/Philadelphia: John Benjamins Publishing.
- Temmerman, Rita 2002. *Metaphorical Models and the Translation of Scientific Texts*. – *Linguistica Antverpiensia*, 1, 211–226.
- Temmerman, Rita 2009. *Why special language translators need insights in the mechanisms of metaphorical models and figurative denominations*. – B. Lewandovska, M. Thelen (Eds.). *The Meaning of Translation*. Bern: Peter Lang, 351–369.
- Temmerman, Rita 2011a. *Stars, problem children, dogs and cash cows: evocative terminology in multilingual business communication*. – *SYNAPS – A Journal of Professional Communication*, 26, 48–61.
- Temmerman, Rita 2011b. *Researching plurilinguistic competences in specialised domains. Looking for means to prevent exclusion*. – *European Journal of Language Policy*, 3 (2), 163–174. <http://dx.doi.org/10.3828/ejlp.2011.9>
- Tercedor Sánchez, Maribel; Ureña Gómez Moreno, José Manuel; Prieto Velasco, Juan Antonio 2012. *Grasping metaphoric and metonymic processes in terminology*. – *The Journal of Specialised Translation*, 18, July. [http://www.jostrans.org/issue18/art\\_tercedor.php#abstract](http://www.jostrans.org/issue18/art_tercedor.php#abstract) (27.09.2012).
- Wüster, Eugen 1979. *Einführung in die allgemeine Terminologielehre und terminologische Lexikographie*. Wien – New York.

## Allikad

- Afanasjev, Ainar 2011. *Kaitseväe pioneeriterminoloogia korrastamine*. Magistritöö. Juhendaja Reet Hendrikson ja mjr Raul Kütt. Tartu: Kaitseväe Ühendatud Õppeasutused.
- Militerm = Sõjandusterminoloogia andmebaas Militerm. <http://termin.eki.ee/militerm/> (14.09.2012).
- Paimla, Raigo 2007. *Jalaväetaktika terminite ühtlustamine*. Magistritöö. Juhendaja Reet Hendrikson-Pungar ja mjr Martin Herem. Tartu: Kaitseväe Ühendatud Õppeasutused.
- Tameri, Tarmo 2009. *Eesti Kaitseväe logistikaterminoloogia korrastamine*. Magistritöö. Juhendaja Reet Hendrikson ja kpt Erlend Zirk. Tartu: Kaitseväe Ühendatud Õppeasutused.
- VL = Vääri, Eduard; Kleis, Richard; Silvet, Johannes 2006. *Võõrsõnade leksikon*. Seitsmes, parandatud ja täiendatud trükk. Tallinn: Valgus.
- ÕS = Eesti õigekeelsussõnaraamat ÕS 2006. *Toimetanud Tiiu Ereht*. Tallinn: Eesti Keele Instituut.

**Reet Hendrikson** (Kaitseväe Ühendatud Õppeasutused) töötab 2005. aastast alates KVÜÖAs terminoloogina ja keeleteimetajana ning annab magistrkursuse terminoloogialoenguid. Samast aastast alates on Kaitseministeeriumi sõjanduse ning julgeoleku- ja kaitsepoliitika terminoloogia komisjoni liige. Uurimisvaldkonnad: ilukirjanduslik ja erialane kujundkeel ning terminoloogia. [reet.hendrikson@mil.ee](mailto:reet.hendrikson@mil.ee)



# AT THE CROSSROADS OF MILITARY AND FIGURATIVE LANGUAGE

**Reet Hendrikson**

Estonian National Defence College

Until now, the figurativeness of specialised language and metaphorisation at the conceptual level has been treated with insufficient attention, indeed with an ambivalent attitude. If, in theoretical discussions, there is an inclination towards the cognitive view (i.e. the claim that the creative use of language is an essential component of specialised communication), then in the practice of specialised language management the natural state of affairs seems to be something more akin to hesitation, questioning, even denial. However, the first and natural choice among specialists in the field often tends to be the figurative term.

The foundation of this article is the research on figurativeness in military terminology and the thesis that, in order to develop a functional terminology, there is a need to build upon an awareness of the needs of users. Therefore, this article will focus on officers' terminological preferences, asking whether there are any conditions in which figurative terminology is preferred and, the choice having been made, how they justify their preferences.

Such an approach allows us to draw conclusions about the functionality of figurativeness in military terminology and to analyse how the attitudes towards figurativeness can influence the development of military terminology and terminological work in general. As with many specialised fields, in military terminology there are a couple of terms in use which seem to be derived directly from English equivalents and which reflect a metaphorical understanding. Such terms are traditionally considered to be rough loan words and are often rejected by language planners. It turns out, though, that it is precisely these forms which are favoured by military specialists. Considering the choices officers make while selecting terms, and bearing in mind the arguments advanced in favour of one term or the other, it would seem that the following hypothesis is worth careful deliberation: besides purely linguistic imitation, there can be the observed preference for such metaphorical expressions influenced by the human disposition for conceptualising and communicating certain phenomena by means of imagery.

Emerging from this discussion there are two vital questions. Firstly, where are the boundaries between rough loan words and conceptual metaphors acceptable by language planners? And secondly, do the well-defined lines of demarcation actually exist and how can we describe them?

**Keywords:** military terminology, sociocognitive approach to terminology, language management, terminology harmonisation, figurativeness of specialised language



# EESTI VERBNET'I LOOMISE VÕIMALIKKUSEST

Indrek Jentson

**Ülevaade.** Artiklis käsitletakse eesti keele jaoks verbide leksikaalse andmebaasi loomise ideed, lähtudes inglise ja juba ka teiste keelte jaoks loodud (Liu, Chiang 2008) samalaadse andmekogu VerbNet koostamise põhimõtetest ja senistest kogemustest. Vaatluse alla võetakse verbide klassifitseerimine nii nende tähenduse kui ka predikaat-argument-struktuuri järgi ning analüüsitakse võimalusi täiendada VerbNet'i andmestikku eesti keelele omaste tunnustega. Kirjeldatakse leksikaalsete kirjete loomist ühe mitmetähendusliku tegusõna näitel ja esitatakse eksperimendi tulemused sõnatähenduse ühestamise kohta sellise leksikaalsete kirjete abil teostatud semantilise analüüsi käigus.\*

**Võtmesõnad:** keeleressursid, arvutilingvistika, arvutileksikoloogia, semantiline märgendus, eesti keel

## 1. Sissejuhatus

Seoses arvutustehnika jõudsa arenguga viimasel paaril aastakümnel on hakatud pöörama suuremat tähelepanu selliste keeleressursside loomisele, mis oleksid arvutite poolt kasutatavad mitmesuguste loomuliku keele töötlemise probleemide lahendamisel. Olulisteks tugisammasteks on mitmesugused korpused ja mitmed spetsiifilise eesmärgiga loodud teadmusbasisid nagu WordNet (Miller jt 1990, Miller 1995), FrameNet (Baker jt 1998), VerbNet (Kipper Schuler jt 2000) ja PropBank (Kingsbury, Palmer 2002). Nende andmekogude üks eesmärk on pakkuda lähteandmeid loomulikus keeles kirjutatud tekstide semantilise analüüsi võimaldamiseks. Kahjuks peab tõdema, et nimetatud ressursid on inglise keele kesksed ning kui eesti keele kontekstis on samas suunas liigutud ainult WordNet'i loomisel (Kerner jt 2010) ja mingis osas ka FrameNet'i andmekirjete defineerimisel (Õim jt 2010), siis VerbNet'i sarnast andmekogu eesti keele jaoks veel ei ole.

VerbNet'i loomise üheks põhjuseks oli äratundmine, et WordNet'i andmestikust on jäänud välja oluline semantilise info allikas – predikaat-argument-struktuur (Kipper Schuler jt 2000). VerbNet'i koostajad valisid verbide klassifitseerimiseks

\* Artikli valmimist on toetanud Haridus- ja Teadusministeerium (projekt EKT12). Autor tänab anonüümseid retsensente asjalike soovitude ja konstruktiivse kriitika eest.

<sup>1</sup> VerbNet <http://verbs.colorado.edu/~mpalmer/projects/verbnet.html> (15.02.2013).

WordNet'i klassifikatsioonist hoopis erineval põhimõttel koostatud Beth Levini inglise keele verbide käsitluse (1993), kus grupeerimise aluseks on just argumentide struktuuri sarnasus ja verbide käitumine erinevate teisenduste korral. Erinevalt WordNet'ist ja FrameNet'ist on VerbNet'i andmestruktuuris koht ka viidetele, mis võimaldavad luua seoseid eelnimetatud andmekogudega. Shi ja Mihalcea (2005) tõstsid oma töös erinevate andmekogude kombineerimisest semantilise analüüsi parendamiseks esile, et selline andmekogude ühendamine tõstab märgatavalt semantilise analüüsi kvaliteeti. Viimasel ajal ongi näha, et erinevate lähenemiste ühendamisele pööratakse üha enam rõhku (Palmer 2009).

Eesti keeles on verbe analüüsinud ja grupeerinud Huno Rätsep juba 1970-ndate lõpus oma töös eesti keele lihtlausete tüüpidest (1978). Tema tööst nähtub, et eesti keeles kannavad lausete tähenduse ja ülesehituse koha pealt olulist infot ka verbi juurde kuuluvate argumentide käänded. Nende märgendamiseks ei ole inglise keele jaoks koostatud VerbNet'is praegu küll vahendeid, kuid VerbNet'i andmestruktuuris on olemas koht nende lisamiseks.

Selleks, et kasutada predikaat-argument-struktuure eesti keele semantilise analüüsi ülesannetes, nagu näiteks sõnatähenduste ühestamise probleemide lahendamisel, on autor seisukohal, et eesti keele jaoks tuleb luua oma VerbNet, milles on ära kasutatud senise VerbNet'i loomise ja laiendamise kogemusi (Kipper Schuler jt 2006) ning mida on täiendatud eesti keelele omaste nüanssidega. Järgnevalt vaatleme, kuidas seda teha saab.

## 2. Inglise VerbNet'i leksikaalse kirje koostamise põhimõtted

VerbNet'i andmestik koosneb kirjetest, millest igaüks kirjeldab üht verbiklassi. Iga verbiklass hõlmab endas ühesuguse argument-struktuuriga ja sarnase süntaktilise käitumisega verbe – klassi liikmeid – ning võib sisaldada ka alamklasse, moodustades nii klasside hierarhia. Klassi nimetus ja number annavad infot nii klassi liikmete üldisema tüübi, tähenduse kui ka klassi järjenumbri ja hierarhia taseme kohta – näiteks *PUT-9.1* ('panema') või *PUT\_SPATIAL-9.2-1* ('panema kusagile spetsiifilisel viisil', näiteks *riputama*). Klassi kuuluvate verbide argumentide fikseerimiseks on loetletud vastavad semantilised rollid, mille juures võivad olla ka valikukitsendused määratlemaks rollide peenemaid nüansse. Süntaktilise käitumise iseloomustamiseks on igas verbiklassis esitatud üks või mitu süntaktilist freimi. Igas sellises freimis on ära toodud üks näitelause, lausemall ja freimi semantika kirjeldus. Lausemall näitab üht lubatud lauseehitust, esitades verbi, selle semantiliste rollide nimetustega tähistatud argumentide ja muude kohustuslike lauseliikmete järjestuse. Kui on vajadus spetsifitseerida süntaksist tulenevaid kitsendusi, siis lisatakse need lausemalli juurde. Freimi semantika kirjeldamiseks kasutatakse valitud hulka semantilisi predikaate ning ajaline dimensioon tuuakse sisse sündmust tähistava muutujaga ja sellele muutujale rakendatud operaatoritega.

### 3. Eesti ja inglise keele verbide võrdlus

Järgnevalt vaatame, kas ja kuidas on võimalik olemasolevaid inglise VerbNet'i andmeid eesti keele jaoks ära kasutada. Lähtudes seisukohast, et verbi tähendus ja argumentide struktuur ei sõltu kasutatavast keelest (Dorr 1997), võib valida ühe eestikeelse verbi, leida sellele vastav ingliskeelne verb ja püüda teisendada vastavad inglise keele jaoks koostatud VerbNet'i kirjed eesti keelele sobivaks. Erinevate keelte verbide omavahelise vastavuse kindlakstegemiseks on otstarbekas kasutada WordNet'i andmeid, sest paljudel juhtudel on valitaval eesti keele verbil mitmeid eri tähendusi ja iga erinev tähendus viib meid üldjuhul erineva inglise keele verbi juurde.

Valime meid huvitavaks eestikeelseks verbiks *istuma*. Sellel sõnal on Eesti Wordnet'i<sup>2</sup> andmetel 4 tähendust (näide 1).

- (1) *istuma\_1*, istet võtma → ingl *sit* (sit%2:38:00::)<sup>3</sup>  
*istuma\_2*, kinni istuma → ingl *serve* (serve%2:42:00::), *do* (do%2:42:02::)  
*istuma\_3*, sobima, passima, klappima → ingl *fit* (fit%2:42:02::)  
*istuma\_4* 'istuvas asendis olema' → ingl *sit* (sit%2:35:00::)

Leitud ingliskeelsetele vastetele vastavad VerbNet'is<sup>4</sup> tõepoolest erinevad klassid (näide 2).

- (2) *sit* (sit%2:38:00::) → ASSUMING\_POSITION-50 , PUT\_SPATIAL-9.2-1  
*serve* (serve%2:42:00::) → SPEND\_TIME-104  
*fit* (fit%2:42:02::) → FIT-54.3, FUNNEL-9.3-1  
*sit* (sit%2:35:00::) → SPATIAL\_CONFIGURATION-47.6, PUT\_SPATIAL-9.2-1

Võtame lähema vaatluse alla tähendusele 'istuma\_1, istet võtma' leitud vaste klasside kirjeldused ASSUMING\_POSITION-50 'asendit sisse võtma' (näide 3), PUT\_SPATIAL-9.2-1 'panema spetsiifilisel viisil' (näide 4). Analüüsides klasside semantilisi rolle ja lausemalle võib tõdeda, et klass ASSUMING\_POSITION-50 (vt näidet 3) on tähenduslikust küljest õige vaste, kuna AGENT võtab sisse istuva asendi, ja klass PUT\_SPATIAL-9.2-1 (näide 4) ei saa sobida selle tõttu, et eesti keele seisukohast elutud asjad üldjuhul istet ei võta.

- (3) ***assuming\_position-50*** 'asendit sisse võtma'<sup>5</sup>

Liikmed (25):

..., *rise*, *sit*, *sit\_down*, *slouch*, ...

Rollid:

AGENT [+ELUS]

KOHT [+KOHT & -REGIOON]

Freimid:

NP V

Näide: "The dog flopped." 'Koer heitis pikali.'

Süntaks: AGENT V

<sup>2</sup> Eesti Wordnet <http://www.cl.ut.ee/ressursid/teksaurus/> (01.10.2012).

<sup>3</sup> See sümbolite jada on WordNet'i tähendusviide (ingl sense key), mille täpsema kirjelduse leiab WordNet'i dokumentatsioonist veebilehel <http://wordnet.princeton.edu/man/senseidx.5WN.html> (10.12.2012).

<sup>4</sup> VerbNet, vt sit <http://verbs.colorado.edu/verb-index/index/S.php> (15.02.2013).

<sup>5</sup> Näide on eestindatud. Vt VerbNet: [http://verbs.colorado.edu/verb-index/vn/assuming\\_position-50.php#assuming\\_position-50](http://verbs.colorado.edu/verb-index/vn/assuming_position-50.php#assuming_position-50) (15.02.2013).

Semantika: EITUS(POSITSIOON(ALGUS(SÜNDMUS), AGENT, POS))  
LIKUMINE(TOIMUMISEL(SÜNDMUS), AGENT)  
POSITSIOON(LÖPP(SÜNDMUS), AGENT, POS)<sup>6</sup>

NP V PP.KOHT

Näide: “*The dog flopped in the corner.*”  
‘Koer heitis nurgas pikali.’

Süntaks: AGENT V {{+LOC}} KOHT

Semantika: EITUS(POSITSIOON(ALGUS(SÜNDMUS), AGENT, POS))  
LIKUMINE(TOIMUMISEL(SÜNDMUS), AGENT)  
POSITSIOON(LÖPP(SÜNDMUS), AGENT, POS)  
PREP(SÜNDMUS, AGENT, KOHT)

(4) **put\_spatial-9.2-1** ‘panema spetsiifilisel viisil’<sup>7</sup>

Liikmed (8):

..., *rest, sit, stand, ...*

Rollid (pärinevad ülemklassist *put\_spatial-9.2*):

AGENT [+ELUTU]

TEEMA [+KONKREETNE]

SIHT [+KOHT & -REGIOON]

Freimid:

NP V PP.SIHT

Näide: “*The books lean against the shelf.*”  
‘Raamatud toetuvad riiuli vastu.’

Süntaks: TEEMA V {{+loc}} SIHT

Semantika: PREP(SÜNDMUS, TEEMA, SIHT)  
POSITSIOON(SÜNDMUS, TEEMA, POS)

NP V ADVP

Näide: “*The books lean there.*” ‘Raamatud toetuvad seal.’

Süntaks: TEEMA V SIHT <+adv\_loc>

Semantika: PREP(SÜNDMUS, TEEMA, SIHT)  
POSITSIOON(SÜNDMUS, TEEMA, POS)

Järgnevalt vaatleme üht võimalikku süntaktiliste freimide hulka sõnale *istuma* klassis ASSUMING\_POSITION-50 ‘asendit sisse võtma’ (näide 5). Kuna selle klassi alla peaksid kuuluma verbid, mis vastavad sellisele kirjeldusele nagu ‘tegevus, mille tagajärjel tegija võtab sisse tegevusega määratletud positsiooni’, siis verbi *istuma* puhul vaatleme neid lauseid, kus AGENT istub mingisugusele KOHALE<sup>8</sup> (määratud lokaaladverbiaaliga, siin latiivse kohamäärusega) – lause peab vastama küsimusele “istub kuhu?”. Selleks, et eristada tähendust ‘istuma\_1, istet võtma’ tähendusest ‘istuma\_4 (‘istuv asendis olema’), peab freimile lisama kitsenduse, et rollis KOHT olev lauseliige peab olema kas allatiivis või illatiivis või genitiivis koos sobiva adpositsiooniga. Huno Rätsepa järgi on siin kõige õigem kasutada substitutsiooniklassi *intralokaalne direksionaal* (tähistatakse sümboliga “Di”, vt näidet 5), mis katab kõik eespool nimetatud variandid (Rätsep 1978: 45).

<sup>6</sup> Esitatud semantilist formalisatsiooni saab lugeda järgmiselt: sündmuse (antud näites pikali heitmise) alguses ei ole agent (koer) positsioonis POS (pikali); sündmuse toimumise ajal agent liigub; sündmuse lõppedes on agent positsioonis POS.

<sup>7</sup> Näide on eestindatud. Vt VerbNet: [http://verbs.colorado.edu/verb-index/vn/put\\_spatial-9.2.php#put\\_spatial-9.2-1](http://verbs.colorado.edu/verb-index/vn/put_spatial-9.2.php#put_spatial-9.2-1) (15.02.2013).

<sup>8</sup> Antud näiteverbi tähenduse ‘istuma\_1, istet võtma’ puhul oleks põhjendatud semantilise rolli SIHT kasutamine, kuid Inglise VerbNet’i andmetes on vaadeldavas verbiklassis kasutusele võetud üldisema tähendusega semantiline roll KOHT just verbiklassi teisi liikmeid silmas pidades.

On huvitav märkida, et õige käände kasutamine võib olla sõltuvuses rollis KOHT kasutatavast asjast, nagu näiteks: *istus pingile / tugitooli / mootorrattale / autosse*. Ehk siis on olemas asju, kuhu saab ainult sisse istuda ja asju, kuhu saab ainult peale istuda. Samas tuleb tõdeda, et nii peent eristamist ei saa VerbNet'i andmestikku kasutades ette kirjutada, sest see on spetsiifiline iga üksiku verbi puhul ja VerbNet'i kirje keskendub siiski verbide klassile. Selline iga verbi jaoks sobivate atribuutide ja nende süntaktilise käitumise määratlemine vajab täiendavat andmekogu.

(5) NP V PP.location

Näide: “Mees istus pingile.”

Süntaks: AGENT <+nom> **V** KOHT <+Di>

Semantika: EITUS(POSITSIOON(ALGUS(SÜNDMUS), AGENT, POS))

LIKUMINE(TOIMUMISEL(SÜNDMUS), AGENT)

POSITSIOON(LÕPP(SÜNDMUS), AGENT, POS)

PREP(SÜNDMUS, AGENT, KOHT)

Freimis (5) olevas lauses võib kolme lauseliiget järjestada kuuel erineval viisil, mis tähendab, et kõik permutatsioonid on lubatud. Võib muidugi kõik sõnajärjed eraldi freimidena üles lugeda, kuid sellistel juhtudel on otstarbekam täiendada freimi kirjapanemise viisi. Permutatsioon võiks tähistada funktsiooniga “Perm”, nagu näites (6).

(6) Perm(NP V PP.location)

Näide: “Mees istus pingile.”

Süntaks: Perm(AGENT <+nom> **V** KOHT <+Di> )

Semantika: EITUS(POSITSIOON(ALGUS(SÜNDMUS), AGENT, POS))

LIKUMINE(TOIMUMISEL(SÜNDMUS), AGENT)

POSITSIOON(LÕPP(SÜNDMUS), AGENT, POS)

PREP(SÜNDMUS, AGENT, KOHT)

Lähtudes Huno Rätsepa poolt pakutud lihtlausetüübist 2.2.126 (N+nom V<sup>1</sup> (Di) (V<sup>2</sup>+ma)) (Rätsep 1978: 100), saab lisada veel ühe freimi (7):

(7) Perm(NP V PP.location S\_INF)

Näide: “Peeter istus laua taha sööma.”

Süntaks: Perm(AGENT <+nom> **V** KOHT <+Di> EESMÄRK <+v\_ma> )

Semantika: EITUS(POSITSIOON(ALGUS(SÜNDMUS), AGENT, POS))

LIKUMINE(TOIMUMISEL(SÜNDMUS), AGENT)

POSITSIOON(LÕPP(SÜNDMUS), AGENT, POS)

PREP(SÜNDMUS, AGENT, KOHT)

TEGEVUS(PÄRAST(SÜNDMUS), AGENT, EESMÄRK)

Selleks, et fikseerida iga verbiklassi atribuudid, eriti just süntaktilised freimid, tuleb loomulikult leida kõigi klassi liikmete atribuutide ühisosa. Kui leidub selliseid erinevusi, mille alusel saab osasid verbe täiendavalt grupeerida, siis võib kaaluda ka inglise keele VerbNet'ist tuleneva klassihierarhia täiendamist.

## 4. Milliseid ülesandeid saab VerbNet'i abil lahendada?

Alustades lihtlause uurimist morfoloogilise analüüsiga, on üheks järgmiseks võimalikuks sammuks leitud verbi määratlemine VerbNet'is. Selleks tuleb leida kõik uuritava verbi esinemised klassihierarhias ja hinnata iga vastava klassi freimi sobivust uuritava lausega. Kõige paremini sobiva freimi leidmise läbi oleme kindlaks teinud ka verbi viite WordNet'ile ning sisuliselt lahendanud sõnatähenduste ühestamise ülesande (Brown jt 2011).

Kui on kindlaks tehtud lausele vastav VerbNet'i freim, siis sellele infole toetudes on võimalik läbi viia nende lauseliikmete semantiline märgendamine vastavalt verbiklassis loetletud semantiliste rollide kasutusele freimi lausemallis.

Selleks, et saada aimu VerbNet'i andmestikul põhineva sõnatähenduse ühestaja loomisega kaasnevatest probleemidest, viidi läbi katse Keeleveebi ilukirjanduse korpuse andmetega.<sup>9</sup> Katse eesmärgiks oli verbi "istuma" jaoks koostatud VerbNet'i kirje abil selle verbi ühe tähenduse esinemise määratlemine testandmetes. Kuna määratluse korrektsust tuli kontrollida käsitsi, siis sai esialgu piiratud ühe juhuslikult valitud andmefailiga korpusest. Selleks osutus teos "Seitsmes rahukevad", mis sisaldas 110 lauset meid huvitava verbiga *istuma*. Laused olid eelnevalt automaatselt märgendatud morfoloogilise analüsaatoriga ja osalauseste piiride määrajaga.

Tähenduse 'istuma\_1, istet võtma' määratlemiseks kasutati algoritmi, mis kontrollis lausemallile (vt näidet 6) vastavate käänetega sõnade olemasolu kontrollitavas lauses ning arvestas seejuures osalause piiride ja käskiva kõneviisiga. Täheendusega 'istuma\_1' verbe oli 110 lause hulgas 39, neist õigesti määrati 30 ja 9 jäi leidmata. Samas leiti 6 teise tähendusega verbi, mis väideti kuuluvat ka tähenduse 'istuma\_1' alla, aga mis seda tegelikult ei olnud. See annab tehtud katse saagiseks 76,9% ja täpsuseks 83,3%.

Analüüsisides nii vale määratlusega lauseid kui ka lauseid, milles õige tähendus jäi leidmata, selgus, et mitmel juhul oli tegemist andmetöötluse eelmise etapi vajakajäämistega. Esines nii morfoloogilisest analüüsist kui ka osalause piiride määramisest pärit ebatäpsusi, mis tõid kaasa algoritmi töö tulemuste ebakorrektsuse. Sisulisest küljest näitas katse seda, et ettevalmistatud VerbNet'i kirjes olid puudulikult kirja pandud subjekti käänete kitsendused, kuna arvesse jäi võtmata võimalus, et subjektiks on hulgafras. Seega, täiendav töö eesti keele süntaksireeglite kirjapanemisel VerbNet'i lausemallidesse võimaldab luua parema täpsusega töötava rakenduse.

## 5. Tulemused

Eestikeelse verbi jaoks vastavat inglise keele verbi VerbNet'i kirjet muutmata kujul kasutada ei saa, kuid on põhimõtteliselt võimalik ära kasutada verbiklasside hierarhia koos klasside semantiliste rollide, nende valikukitsenduste ning freimi semantika kirjeldustega. Verbiklasside liikmed tuleb asendada samatähenduslike eesti keele verbidega ja need varustada viidetega nii Eesti Wordnetile kui ka inglise VerbNet'ile. Eraldi tähelepanu nõuavad süntaktilised freimid, sest erineva süntaksi ja vabama sõnajärje tõttu võib eesti keeles olla verbidel rohkem lausemalle ning kindlasti on vaja võtta kasutusele käänetega seotud grammatilised kitsendused.



Tehtud katsetuste põhjal võib hinnata, et ühe verbi analüüsimine ja vastavate VerbNet'i kirjade vormistamine võib vastavalt eri tähenduste hulga võtta aega üks kuni kolm päeva. Kogemuste kogunemine kiirendab antud tööd, samas iga järgmise verbi lisamine andmekogusse võib põhjustada vajaduse mõne klassi ümberstruktureerimise järele. Kui võtta eesmärgiks inglise keele VerbNet'iga võrreldavas mahus andmekogu loomine, siis võib hinnata töömahuks ligikaudu 35 inimaastat.<sup>10</sup>

## 6. Kokkuvõte

Artiklis on kirjeldatud VerbNet'i põhimõtteline andmestruktuur, mis on sobiv kasutamiseks ka eesti keele jaoks. Lisatud on VerbNet'i kirjade esitamist illustreerivad näited. Autor on seisukohal, et Eesti VerbNet on kindlasti vajalikuks täienduseks meie keeleressursside kogule, andes omapoolse panuse mitmete keeletehnoloogia alamülesannete lahendamisse. Tuleb rõhutada, et selline andmekogu ei ole valdkonnaspetsiifiline ja on seega kasutatav kõikide valdkondade rakendustes, mis võivad vajada loomuliku keele töötlemisega seotud funktsionaalsust. Kuna verbe ja nende predikaat-argument-struktuure ning tähendusi tuleb analüüsida ning klassifitseerida igapäev individuaalselt, siis on inglise VerbNet'i raames välja pakutud andmestruktuur sobiv verbide semantilise info kirjelduste kogumiseks nii selle keeleressursi võimalikke rakendusi kui ka teistes maades tehtava samalaadse tööga kokkusobivust silmas pidades.

### Viidatud kirjandus

- Baker, Collin F.; Fillmore, Charles J.; Lowe, John B. 1998. The Berkeley FrameNet project. – Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics, Vol. 1, 86–90.
- Brown, Susan Windisch; Dligach, Dmitriy; Palmer, Martha 2011. VerbNet class assignment as a WSD task. – Johan Bos, Stephen Pulman (Eds.). Proceedings of the Ninth International Conference on Computational Semantics. IWCS 2011. Stroudsburg: Association for Computational Linguistics, 85–94.
- Dorr, Bonnie J. 1997. Large-scale dictionary construction for foreign language tutoring and interlingual machine translation. – Machine Translation 12 (4), 271–322.
- Kerner, Kadri; Orav, Heili; Parm, Sirli 2010. Growth and revision of Estonian WordNet. – Principles, Construction and Application of Multilingual Wordnets. Proceedings of the 5th Global Wordnet Conference. Narosa Publishing House, 198–202.
- Kingsbury, Paul; Palmer, Martha 2002. From TreeBank to PropBank. – Proceedings of the 3rd International Conference on Language Resources and Evaluation (LREC-2002). <http://www.lrec-conf.org/lrec2002/> (25.02.2012).
- Kipper Schuler, Karin; Dang, Hoa Trang; Palmer, Martha 2000. Class-based construction of a verb lexicon. – Proceedings of the Seventeenth National Conference on Artificial Intelligence and Twelfth Conference on Innovative Applications of Artificial Intelligence. AAAI Press / The MIT Press, 691–696.
- Kipper Schuler, Karin; Korhonen, Anna; Ryant, Neville; Palmer, Martha 2006. Extending VerbNet with novel verb classes. – Proceedings of 5th International Conference on Language Resources and Evaluation (LREC 2006), No. 2.2.

<sup>10</sup> VerbNet'i versioonis 3.2 on 6280 verbi, 270 verbiklassi ja 206 alamklassi (Unified Verb Index 2013).

- Levin, Beth 1993. *English Verb Classes and Alternations: A Preliminary Investigation*. Chicago: University of Chicago Press.
- Liu, Mei-chun; Chiang, Ting-yi 2008. The construction of Mandarin VerbNet: A frame-based study of statement verbs. – *Language and Linguistics*, 9 (2), 239–270.
- Miller, George A. 1995. WordNet: a lexical database for English. – *Communications of the ACM*, 38 (11), 39–41. <http://dx.doi.org/10.1145/219717.219748>
- Miller, George A.; Beckwith, Richard; Fellbaum, Christiane; Gross, Derek; Miller, Katherine J. 1990. Introduction to WordNet: An on-line lexical database. – *International Journal of Lexicography* 3(4), 235–244. <http://dx.doi.org/10.1093/ijl/3.4.235>
- Palmer, Martha 2009. Semlink: Linking PropBank, VerbNet and FrameNet. – *Fifth International Workshop on Generative Approaches to the Lexicon (GL 2009)*. Pisa, Italy, 9–15.
- Rätsep, Huno 1978. *Eesti keele lihtlausete tüübid*. Tallinn: Valgus.
- Shi, Lei; Mihalcea, Rada 2005. Putting pieces together: Combining FrameNet, VerbNet and WordNet for robust semantic parsing. – Alexander Gelbukh (Ed.). *Computational Linguistics and Intelligent Text Processing*. 6th International Conference, CICLing 2005, Mexico City, Mexico, February 13-19, 2005, Proceedings. Lecture Notes in Computer Science 3406. Berlin, Heidelberg: Springer-Verlag, 100–111.
- Õim, Haldur; Orav, Heili; Kahusk, Neeme; Taremaa, Piia 2010. Semantic analysis of sentences: The Estonian experience. – Inguna Skadiņa, Andrejs Vasiljevs (Eds.). *Human Language Technologies: The Baltic Perspective*. Proceedings of the Fourth International Conference, Baltic HLT 2010. *Frontiers in Artificial Intelligence and Applications* 219. IOS Press, 208–213. <http://dx.doi.org/10.3233/978-1-60750-641-6-208>

## Võrgumaterjalid

- Eesti Wordnet, <http://www.cl.ut.ee/ressursid/teksaurus/> (01.10.2012).
- Keeleveebi ilukirjanduse korpus, <http://www.keeleveeb.ee/> (01.10.2012).
- Unified Verb Index, <http://verbs.colorado.edu/verb-index/> (21.03.2013).
- VerbNet, <http://verbs.colorado.edu/~mpalmer/projects/verbnet.html> (15.02.2013).
- WordNet'i tähendusviite kirjeldus, <http://wordnet.princeton.edu/man/senseidx.5WN.html> (10.12.2012).

**Indrek Jentson** (Tartu Ülikool) on erialalt informaatik. Huvipakkuvaks uurimisvaldkonnaks on semantiline representatsioon.  
[indrek.jentson@ut.ee](mailto:indrek.jentson@ut.ee)

# THE FEASIBILITY OF ESTONIAN VERBNET

**Indrek Jentson**

University of Tartu

The idea of creating a lexical database of verbs for the Estonian language is introduced, based on principles and experience of creating a similar database for the English language – VerbNet. Classification of verbs by meaning and also by predicate-argument structure is placed under focus and the possibilities of improving the data structures with properties distinctive to the Estonian language are analysed. The creation of lexical entries for the database is described using the example of one polysemic verb. The results of an experiment on semantic analysis and sense disambiguation on sentences from the morphologically disambiguated corpus when using defined lexical entries in the database of verbs are described. It is concluded that the possibility of using English VerbNet data structures and verb classification hierarchy for Estonian VerbNet is feasible. The importance of having such a language resource for the Estonian language is stressed.

**Keywords:** language resources, computational linguistics, computational lexicology, semantic labelling, Estonian



# KATSELINE SEMANTIKA: PLANEERIMINE JA TEOSTUS

Anni Jürine, Jane Klavan, Ann Veismann

**Ülevaade.** Artiklis leiab käsitlemist katseliste meetodite kasutamine keeleteaduses, täpsemalt semantikas. Pikemalt peatutakse kahel etapil – katse kavandamise ja läbiviimise etapil. Tuues näiteid katselise semantika valdkonnast, arutletakse muuhulgas järgmiste teemade üle: hüpoteeside püstitamine ja operatsionaliseerimine, uurimuse valiidsus ja reliaablus, mõõdetavad tunnused, valimi esinduslikkus ja juhuslikkus, valimi maht, katseplaani valik, katsematerjali koostamine ja esitus, katsealuste värbamine ja instrueerimine. Artikli eesmärk ei ole pakkuda vastuseid kõigile küsimustele, mis keeleteadlastel katset planeerides tekkida võiks. Paljuski oleneb katse planeerimine ja teostus konkreetsest uurimisküsimusest ja iga katsega kaasneb rida spetsiifilisi küsimusi. Siinses artiklis on peetud oluliseks välja tuua vaid mõningad põhiaspektid, millega iga uurija, kes eksperimentaalset meetodit tähenduse uurimisel rakendab, peaks arvestama.\*

**Võtmesõnad:** katselised meetodid, semantika, kognitiivne keeleteadus

## 1. Sissejuhatus

Käesolev artikkel jätkab teemat, mida alustas varem ilmunud artikkel (Klavan jt 2013). Seal toodi välja põhjuseid, miks võiks katselisi meetodeid kasutada kõrvuti introspektsiooni ja korpusanalüüsiga, ning kirjeldati eri tüüpi katseid (nt vaba moodustuse katse, sunnitud valiku katse, väljamõeldud stiimuliga moodustuskatse, klassifitseerimiskatse, sobituskatse, vastuvõetavuse hindamise katse, ajaliselt mõõdetud verifitseerimine). Kui nimetatud artikli eesmärgiks oli anda ülevaade mõningatest võimalikest katsetest semantikas, täpsemalt peamiselt ruumisemantikas, siis käesoleva artikli eesmärk on arutleda selle üle, kuidas praktiliselt tähendust puudutavaid keelelisi katseid kavandada ja läbi viia. Ruuminappusest sõltuvalt keskendumine ka siinses artiklis vaid semantika uurimiseks kasutatavatele katsetele ning seega jääb

\* Artikkel on valminud sihtfinantseeritava teema SF0180056s08 toel.

siinse kirjutise fookusest välja näiteks loetelukatse kui välimeetod (suuline intervjuu), mille eesmärgiks on välja selgitada teatud valdkonna (põhi)sõnavara ja/või teadmiste struktuuri (nt Sutrop 2001, Vainik 2004, Uusküla 2008 jt).

Keeleteadlastel nagu teistelgi uurijatel on võimalus andmeid koguda kas vaatluse või katse teel. Teaduslikult huvi pakkuvat nähtust võib vaadelda kas selle loomulikus keskkonnas ilma sekkumata või luua katseolukord ja uurida, kuidas mingite tingimuste muutmine mõjutab uuritavat nähtust. Võimalus tingimusi muuta ongi olemuslik erinevus nende kahe meetodi vahel. Vaatlus annab uurijale kahtlemata loomulikumat andmeid, katse muudab loomulikku olukorda ja seda võib ühelt poolt pidada katseliste meetodite puuduseks, teisalt on just see põhjus, miks üldse katseid teha – isoleerida erinevad tegurid ja kontrollida nende mõju.

See omakorda on põhjus, miks katse kavandamisel tuleb olla eriti hoolikas. Iga konkreetse katse ülesehitus oleneb palju uurimisküsimusest ja katse tüübist ning väga üldisi must-valgeid reegleid pole võimalik leida. Siiski on olemas mõningad põhiaspektid, millega iga uurija, kes eksperimentaalset meetodit tähenduse uurimisel rakendab, peaks arvestama. Nende aspektide üle siinses käsitluses pikemalt arutletaksegi keeleteadusliku semantika vaatenurgast.

Kuna katselised meetodid on nii Eesti kui ka rahvusvahelises keeleteaduses suhteliselt uus ja vähelevinud valdkond (kui pidada silmas eelkõige semantikat ja süntaksit, mitte foneetikat, sotsiolingvistikat ja psühholingvistikat), on raske leida selleteemalisi üldkäsitlusi. Paljud keeleteadlased, kes tegelevad süntaksi või semantikaga ja tahaksid kasutada katselisel teel saadud andmeid, tunnevad, nagu neil tuleks endal jalgratast leiutama hakata. Rahvusvahelistes erialastes ajakirjades leidub küll hulgaliselt erinevate (psühho)lingvistiliste katsete tulemusi kirjeldavaid artikleid, kuid põhirõhk on seal arusaadavalt asetatud tulemuste analüüsile ja sisulistele küsimustele ning konkreetsete katsete kavandamise ja läbiviimise etappi on kirjeldatud vaid paari lausega. Autorid leiavad, et selline tühik väärib täitmist ning käesolev artikkel oleks esimeseks sammuks sellel teel. Kõige enam võtabki siinne käsitlus aluseks psühholoogide vastavasisulisi kirjutisi (nt Cobb 1998, Field, Hole 2006, MacLin, Solso 2007, Kantowitz jt 2009). Kuna nii eesti- kui ka ingliskeelsed õpikud katselise semantika alal puuduvad, siis on kasutatud ka erinevaid autorite poolt läbitud loengukursuste ning tööpadade konspekte ja materjale (nt E. Dąbrowska (2010), E. Käärik (2011) ja A. Luuk (2012)).

Mööda ei saa vaadata tõsiasjast, et süntaksi uurimisel on võrdlemisi laialt kasutusel vastuvõetavuse hinnangu laadsed katsed<sup>1</sup>, mille kohta leidub nii monograafiaid (nt Schütze 1996, Cowart 1997) kui ka artikleid (nt Sprouse, Almeida (ilmumas)). Sellele vaatamata võib öelda, et puudub üks terviklik eestikeelne ülevaade võimalustest ja valikutest, kuidas semantika uurimisel oma hüpoteese katselisel kontrollida.

Katset planeerides tuleb läbi mõelda väga paljud erinevad punktid (katsealuste valik, katseühikute hulk, sõltuvad ja sõltumatud muutujad, saadud andmete tüüp, andmeanalüüsi võimalused ja piirangud, kui nimetada vaid mõned). Osa katse ülesehitust puudutavatest küsimustest on väga üldised ja puudutavad igat eksperimenti, teised on spetsiifilisemad ja sõltuvad uurimisküsimusest. Alljärgnevalt kirjeldatakse pikemalt kaht etappi – katse kavandamist ja läbiviimist – ning arutletakse mõningate praktiliste küsimuste üle, mis nende kahe etapiga kaasnevad.

<sup>1</sup> Autorid on tänulikud ühele retsensentidest, kes juhtis tähelepanu, et vastuvõetavuse hindamise katse on olemuselt pigem generatiivse kui kognitiivse keeleteadusega sobiv. Autorid leiavad, et katse tüüp pole siiski seotud kindla koolkonnaga. Generatiivse koolkonnaga sobivalt saab vastuvõetavuse hindamise katset kasutada grammatiliselt korrektsete/vastuvõetavate keelendite kindlaksmääramiseks, samas kui kognitiivse koolkonna esindajaid huvitavad selle katsetüübi puhul kõnelejate eelistused ja kasutuse varieeruvuse ulatus.

## 2. Kavandamise etapp

### 2.1. Probleemi sõnastus ja ülesande püstitus

Enne katse tegemist tuleks hoolega kaaluda, milleks katset teha. Katse tegemine ei saa olla eesmärk iseeneses. Keeleandmete kogumiseks on ka muid võimalusi ja katse tegemist võib pidada üheks vaevarikkamaks ja riskantsemaks neist. Seepärast on vaja põhjalikult läbi mõelda, millist hüpoteesi katse on mõeldud kontrollima. Nagu maailmas üldiselt võib ka keeles märgata mitmesuguseid korrelatsioone. Ent korrelatiivne seos ei tähenda tingimata põhjuslikku seost. Seepärast luuaksegi katseolukord, kus saab kontrollida arvatavat põhjust ja välja selgitada, kas sellel on ka tegelik mõju kaasnevale nähtusele. Seejuures ei tohi siiski unustada, et tegelik elu on enamasti palju keerulisem kui katseolukord.

Hästi läbimõeldud eesmärgid ja konkreetset sõnastatud uurimusküsimused tagavad uurimuse parema kvaliteedi ehk selle suurema usaldusväarsuse. Konkreetset sõnastatud uurimisküsimustest kasvavad omakorda välja **hüpoteesid** ehk oletuslikud põhjendused mingi nähtuse olemuse ja toimumise seletamiseks. Väga oluline samm on **hüpoteeside operatsionaliseerimine** ehk hüpoteeside viimine sellisele kujule, et neid on võimalik mõne konkreetse katsega kontrollida (s.o kinnitada või ümber lükata). See tähendab, et iga teoreetilise mõistega tuleb seada vastavusse mingi konkreetne mõõdetav parameeter (operatsionaliseerimisest semantikas vt ka Geeraerts 2006, Geeraerts 2010, Stefanowitsch 2010). Hüpoteeside püstitamine ja operatsionaliseerimine nõuavad ühelt poolt suurt loomingu- ja teiselt poolt ranget loogilisust, sest iga uurimisküsimuse kohta on võimalik püstitada erinevaid hüpoteese ja hüpoteesi kontrollimisel on võimalikud mitmed operatsionaliseeringud, kuid keeleuurija peab suutma selgelt põhjendada, miks just mingid kindlad valikud on tehtud.

Geeraerts (2010: 72–73) toob välja empiirilise (semantika)uurimuse kuus tunnust: 1) põhineb kogutud andmetel; 2) andmed on saadud kas vaatluse või katse teel; 3) kasutab kvantitatiivset analüüsi; 4) sisaldab korrektselt formuleeritud hüpoteesi; 5) hüpotees on operatsionaliseeritud; 6) empiirilise uurimusega kaasneb empiiriline tsükkel, mis tekib korduvast andmete kogumisest, hüpoteesi testimisest ja tulemuste tõlgendamisest. Nii võib näiteks tuua, et kui tuntud polüseemia võrgustikanalüüsile sõnast *over* (Brugman 1988, Lakoff 1987) on ette heidetud intuitsioonil põhinemist (nt Sandra, Rice 1995), siis võrgustikanalüüsi saaks pidada esimeseks sammuks hüpoteesi püstitamise ja selle operatsionaliseerimise teel semantikas (vt Glynn 2010: 241–242). Näiteks võiks eesti keele sõna *üle* kohta testida, kas keeleteadlased eristavad/tajuvad selle sõna tähendustes uurimuses (Veismann 2009) välja toodud kolme kujundskeemi. Semantika alase katse teeb keeruliseks just objektiivselt mõõdetavate parameetrite ja seoste leidmine. Sorteerimiskatse ja (tähenduse) sarnasuse hindamise katse on üks võimalus kontrollida, milliseid eristusi polüseemse üksuse tähendustes kõnelejad tajuvad (vt Sandra, Rice 1995), kuid pole päris selge, kuidas seada need andmed vastavusse kujundskeemidega (kujundskeemide psühholoogilisest reaalsusest katsetes vt Gibbs, Colston 1995).

Igasuguse teadusliku uurimuse puhul tuleb silmas pidada kahte olulist nüanssi – uurimuse valiidsust ja reliaablust. **Valiidsust** võib defineerida kui kasutatava meetodika paikapidavust, kehtivust või adekvaatsust. Valiidsus näitab,

missugusel määral mõõdab katse seda, mida ta on plaanitud mõõtma. Toome näiteks mittevälidsest katsest sellise vastuvõetavuse hindamise katse, kus katseisikuil tuleb hinnata sõnade *kõrval*, *juures*, *lähedal*, *küljel* sobivust pildi kohta, kus pall on puu kõrval. Selle katse tulemusena saame me küll pildi sobivate keelendite hulgast ning võimalikest kasutuseelistustest, kuid me ei saa järeldada, et kõneleja tegelikult kaaluks kõikide nende kasutamist vastavas olukorras. Kuna kõik need variandid on pildi kohta võimalikud, siis võib oletada, et katsealuste vastustest tuleb esile hoopis suurem varieeruvus väljendites kui see tegelikus keelekasutuses reaalne võiks olla. Seega võime tulemuste põhjal järeldada vaid seda, mis oli juba enne teada – kõik need sõnad on selle situatsiooni kirjeldamiseks võimalikud. Siinse katse mittevälidus tuleneb sellest, et puudub selgesõnaliselt esitatud testitav hüpotees, mis muudab võimatuks ka operatsionaliseerimise. Näiteks võiks siin püstitada hüpoteesi, et objektide kaugus üksteisest ja/või nende suuruse suhe määrab, millist kaassõna kasutada. Sel juhul saame manipuleerida objektide suuruse või omavahelise kaugusega ja kontrollida, kas hüpotees peab paika. Hea näide valiidsest katsest on selline, millel on selgesõnaline eesmärk, kontrollitavad hüpoteesid ja nende operatsionaliseeringud. Näiteks tahtes kontrollida, mille alusel valib kõneleja sünonüümsete sõnade *hulgas* ja *seas* vahel, saame me püstitada hüpoteesi (nt et valik sõltub objektide arvust või ühtsusest) ja vastavat tunnust manipuleerides selgitada kas vastuvõetavuse hindamise või produktsioonikatsega välja, millisel juhul millist keelendit eelistatakse (vt Klavan 2008).

**Reliaablus** on kasutatava meetodika stabiilsus ja järjekindlus. See näitab, kui suure tõenäosusega saadaks katse kordamisel sama tulemus. Nii nagu ei ole olemas ideaalselt objektiivset uurimust, pole olemas ka ideaalse valiiduse ja reliaablusega uurimust. Uurimuse reliaablus ehk usaldusväärsus on suuresti seotud valimi varieeruvusega (vt ptk 2.3). Kuna keeleliste katsete puhul on katsealusteks inimesed ja inimesed on teatavasti väga erinevad, siis ei saa me ilmselt kunagi ühtegi katset korrates täpselt sama tulemust. Rakendades aga adekvaatseid statistilisi andmeanalüüsi meetodeid on võimalik iga katse tulemusi analüüsides juhuslik viga sisse arvestada ja leida, kas valimi põhjal saadud tulemused on üldistatavad populatsioonile ehk suuremale kõnelejaskonnale või mitte. Madala valiidusega uuring, olenemata sellest, kas reliaablus on kõrge või madal, on ühtviisi halb.

## 2.2. Mõõdetavad tunnused

Keeleuurijaid huvitab tihtilugu, kuidas keeles midagi muutub, mis selle muutuse põhjustab ja kuidas seda muutust mõõta. Muutumine ei pea olema mõistetud aja(loo)liselt, nii võib näiteks semantikas uurida, kuidas muutub ümbritsev kontekst vastavalt sellele, kas kasutatakse sünonüümipaari üht või teist liiget (näiteks millal kasutatakse kaassõna *peal* ja millal alalütlevat käänat (vt Klavan 2012)); või seda, millistel tingimustel kalduvad katsealused kirjutama kaassõna kokku sellele eelneva käändsõnaga (vt Jürine 2011); või kuidas määrata polüseemse sõna põhitähendust, millega saaks siduda kõik selle sõna ülejäänud tähendused (vt Veismann 2008, 2009). Kõike seda, mida uurijad mõõdavad, nimetatakse **tunnuseks** ehk **muutujaks** (ingl *variable*)<sup>2</sup>. Tunnus on mõõtmise, küsitluse,

<sup>2</sup> Termineid *muutuja* ja *tunnus* kasutatakse artiklis paralleelselt, esimene on levinud rohkem psühholoogia ja teine statistikaalastes kirjutistes.



katse või vaatluse tulemusena saadud arvuline või mitteamvuline näitaja. Tunnuse väärtus sõltub sellest, mida mõõdetakse või keda küsitletakse, ning mõõtmisel omistatakse igale mõõdetavale objektile mõõdetava tunnuse väärtus (vrd hüpoteeside operatsionaliseerimine). Keeleteadlastele, kelle huviks on tähendus – midagi, mida ei saa käega katsuda – on mõõdetavate tunnuste leidmine kriitilise tähtsusega. Erinevad meetodid, millega keelenähtusi mõõdetakse, võivad erineda selles osas, millist liiki andmeid nende abil saab koguda. Katse koostamisel on väga oluline teadvustada, mis liiki tunnustega on tegemist ja millisel skaalal tunnuste väärtusi katsega mõõdetakse.

Kõigepealt tehakse vahet sõltumatutel ja sõltuvatel muutujatel. **Sõltumatud muutujad** on muutujad, mis on uurija poolt manipuleeritavad; s.t muutujad, mis uurija ise on välja valinud vastavalt püstitatud hüpoteesile ja mille väärtusi on tal endal võimalik määrata. **Sõltuvad muutujad** on need muutujad, mille väärtusi uurija mõõdab, eeldades, et nende mõõtmise tulemustes ilmnev varieeruvus on uurija poolt välja valitud sõltumatu muutuja erinevate väärtuste manipuleerimise tagajärg. Teisisõnu, uurija eeldab, et sõltumatu muutuja väärtuste tahtlik varieerimine uurimuses põhjustab sõltuva muutuja väärtuse varieerumise. Näitena sõltuvast ja sõltumatutest muutujatest võib tuua vastuvõetavuse hindamise katse, mis kontrollib, kas kaassõna *peal* ja alalütleva käände kasutus kombineerituna fraasi pikkuse, sõnajärje ja objekti semantikaga mõjutab statistiliselt oluliselt vastuvõetavust (Klavan 2012). Sellises uurimuses on sõltuvaks muutujaks konkreetne vastuvõetavuse hinnang ja sõltumatuteks muutujateks, mis eeldatavasti vastuvõetavuse hinnangut kuidagi mõjutavad, on konstruktsioon (adessiiv või *peal*), fraasi pikkus, sõnajärg ja objekti semantika.

Muutujaid on võimalik mõõta erinevatel mõõteskaaladel. Üldiselt eristatakse **nelja mõõteskaalat**: nimetuste skaala ehk nominaalskaala, järjestikaskaala ehk ordinaalskaala, vahemikaskaala ehk intervallskaala ja suhteskaala (skaalade kohta vt Tooding 2007: 32–36). Erinevate mõõtmiskaaladega on seotud ka erinevad **tunnusetüübid**. Tehakse vahet **arvulistel** ehk kvantitatiivsetel ja **mittarvulistel** ehk kvalitatiivsetel tunnustel. Arvulistel tunnustel on konkreetne arvuline väärtus ja neid võib omakorda jagada **diskreetseteks** ja **pidevateks** arvtunnusteks. Diskreetsete arvtunnused saadakse loendamisel ja tegemist on lõpliku arvu täisarvulise väärtusega, nt mitu lausete gruppi on katsealune lausete sorteerimise katsel moodustanud. Pidevad arvtunnused saadakse otsesel mõõtmisel ja tegemist on lõpmatu arvu võimalike väärtustega, nt reaktsiooni kiirus mõistmiskatses. Mitteamvulistel tunnustel puudub arvuline väärtus ja neid võib jagada **järjestustunnusteks** ja **nominaaltunnusteks**. Järjestustunnuste väärtuste vahel on võimalik objektiivne järjestus, nt hinnangud etteantud skaalal (hästi sobiv – sobiv – pigem ei sobi – üldse ei sobi) või fraasi pikkus (lühike, keskmine, pikk). Nominaaltunnuste väärtuste vahel aga ei ole sisulist järjestust, nt sõnaliik (nt *üle* võib olla kas kaassõna, määrsõna või abimäärsõna), elus või eluta objekt, semantiline roll.

Igal skaalal ja tunnuseliigil on oma eripära ning mõõtmiskaala ja tunnuse tüüp määravad andmeanalüüsi võimalused – eri tüüpi tunnustele on rakendatavad erinevad meetodid. Arvulistele tunnustele on rakendatav suurem hulk andmetöötlusmeetodeid, mitteamvuliste tunnuste puhul on meetodite hulk piiratum. Näiteks nominaaltunnuste puhul on ainuke mõeldav matemaatiline tehe sageduste lugemine, s.t loetakse kokku, kui mitu liiget igas nominaalses kategoorias esineb ja

võrreldakse erinevaid kategooriaid vastavalt nende suurusele. Järjestustunnuste puhul saab lisaks sagedusele järeltõlge teha ka tunnuste järjekorra järgi. Samas puudub täpne info selle kohta, kui suur on erinevate väärtuste vahe. Viimast liiki infot näitavad arvulised tunnused. Oluline on meeles pidada, et mitteamarvuliste tunnuste väärtuste ümber kodeerimine arvudeks (sobib hästi = 5, üldse ei sobi = 0) ei muuda neid arvulisteks tunnusteks.

Keeleteaduse andmed erinevad paljuski standardsetest, kvantitatiivsete meetodite sissejuhatavates õpikutes esitatud näidisandmestikest. Suur hulk keeleandmeid on mitteamarvulised ja suure variatiivsusega. See omapära seab piirid andmeanalüüsile ja tulemuste tõlgendamisele. Näiteks on korpusanalüüsi meetodeid kasutades saadud andmed enamasti seotud sagedusega. **Sagedus** on aga oma olemuselt mitteamarvuline, kvalitatiivne tunnus – loetletakse, mitu korda mingi keelenähtus mingis kategoorias esines, kuid sellisele andmestikule ei ole võimalik rakendada andmeanalüüsi meetodeid, mis põhinevad arvulistel väärtustel. Ka mitmete katsete meetodite teel kogutud andmestikes esineb sageli just mitteamarvulisi tunnuseid, nt sunnitud valiku katse (mitu korda esines ühte vastust võrreldes teiste vastustega), moodustuskatse (mitu korda moodustati ühte vastust võrreldes teiste vastustega, mitu korda anti õige või vale vastus), klassifitseerimiskatse (millistesse kategooriatesse andmed jagunesid ja/või kui palju üksuseid kuhugi klassifitseeriti). Arvulisi tunnuseid leidub keeleteaduse andmestikes suhteliselt vähe. Selliste tunnuste näideteks on keeleüksuste pikkus sõnades või silpides ja reaktsioonikiirus mingi vastuse andmisel.

## 2.3. Valim

### 2.3.1. Valimi esinduslikkuse ja juhuslikkuse printsiibid

Empiirilise uurimuse puhul on oluline teha vahet kahel kesksel mõistel – üldkogum ehk populatsioon ja valim. **Üldkogum** on “uurimisülesande sisuga piiritletud kõigi uurimisobjektide kogum”, **valim** aga “kindlal viisil eraldatud osa üldkogumis, mida analüüsis kasutatakse üldkogumi asemel” (Tooding 2007: 123). Valimi moodustamisel, s.t nii katsematerjali koostamisel kui ka katses osalevate keekekõnelejate leidmisel, tuleb silmas pidada, et valim oleks uuritava üldkogumi suhtes **esinduslik**. Oluline on selgeks teha, mis on see uuritav nähtus, mille kohta katse koostatakse ja kas katses osalevad isikud esindavad kõnelejaskonda, mida uuritakse. Üldjuhul on soovitatav kasutada juhuslikkuse printsiibil põhinevaid meetodeid valimi moodustamiseks. Rakendades **lihtsa juhuvalimi** (ingl *simple random sample*) meetodit on populatsiooni igal liikmel võrdne tõenäosus saada valimisse kaasatud. Valimi moodustamise meetodid, mis eiravad juhuslikkuse printsiipi, põhjustavad sageli nihkeid, kuna selline valim võib osutada ebatüüpiliseks või mitteesinduslikuks. Seega ei ole selliste valimite puhul võimalik teha statistiliselt korrektseid järeltõlge populatsiooni kohta. **Mugavusvalim** (ingl *convenience sampling*) ja **lumepalli-valim** (ingl *snowball sampling*) on näited mitte-juhuslikest valimitest. Esimese meetodi puhul moodustatakse valim “mugavalt”, näiteks hõlpsa kättesaadavuse, leitavuse, käsitletavuse, koostöövalmiduse alusel (nt üliõpilased). Teise meetodi puhul leitakse vaid mõned inimesed, kes vastavad tingimustele ning igaüks neist

toob järgmised esindajad, kes omakorda peavad kaasama veel mõned. Sellist liiki valimi moodustavad näiteks sotsiaalvõrgustike kaudu leitud katsealused ja sel teel levitatud katsed. Muutujad nagu sugu, vanus, haridustase, sotsiaalne klass jne on soovitatav muuta kontrollitud muutujateks.

### 2.3.2. Valimi maht

Ka valimi mahu määramisel on oluline silmas pidada, milliseid andmeanalüüsi meetodeid on soov hiljem kogutud andmete kirjeldamiseks ja analüüsimiseks kasutada. Valimi mahu määramisel peab lähtuma püstitatud statistilisest hüpoteesist ja selle kontrollimiseks kasutatava teststatistiku kujust<sup>3</sup>. Tabelis 1 on välja toodud mõned sagedamini kasutatavad andmeanalüüsi meetodid ja soovitatav valimi maht (tähistatud  $n$ -ga). Üheks rusikareegliks peetakse, et katses osalevaid isikuid peaks olema vähemalt 15 iga katsetingimuse kohta. Sealjuures soovitatakse alati koguda andmeid rohkematelt isikutelt, kuna uurija peab arvestama ka võimaliku andmete kaoga (Gries 2009: 54). Soovitatud on ka varasemate uurimuste kasutamist eeskujuna. Hilisemat statistilist andmeanalüüsi silmas pidades ei ole ka liiga suur katsealuste hulk soovitatav, sest väga suur andmehulk võib muuta väiksemagi erinevuse statistiliselt oluliseks.

**Tabel 1.** Rusikareeglid valimi mahu määramisel (Wilson Van Voorhis, Morgan 2007: 48)

Meetod	Soovitatav valimi maht
Rühmakeskmiste erinevus	$n = 30$ (rühmas)
Seoste hindamine (korrelatsioon, regressioon)	$n = 50$
$\chi^2$ -test	$n = 20$

### 2.4. Katseplaani

Keeleuurijal on katse koostamisel valida mitme erineva katseplaani vahel. Katseplaani valimisel on taaskord oluline läbi mõelda, mida uuritakse ja milleks. Oluline on ka selgeks teha erinevate plaanide eelised ja kitsaskohad. Kaks peamist katseplaani, mille vahel tuleb otsus langetada, on **osalejatevaheline planeering** (ingl *between-subjects design*) ja **osalejatesisene planeering** (ingl *within-subjects design*). Kõige lihtsama katse malli puhul on uurijal kaks katsealuste rühma – katserühm ja kontrollrühm. Neid kahte rühma koheldakse katse käigus mingil viisil erinevalt, nt katserühma mõjutatakse kuidagi, samas kui kontrollrühma ei mõjutata. Seejärel uuritakse, kuidas ja kas nende kahe rühma tulemused erinevad. Kui erinevus on olemas, võib järeldada, et selle tingis(id) katserühmale avaldatud mõju ehk katses rakendatud sõltumatu(d) muutuja(d). Sellise malli puhul on tegemist osalejatevahelise planeeringuga. Näiteks kui testida hüpoteesi, et pikemas fraasis eelistavad eesti keele kõnelejad kasutada pigem adessiivi kui kaassõna, siis ühele katsealuste rühmale antakse pikad fraasid hinnata või moodustada, teisele lühikesed.

<sup>3</sup> Vabavaralised kalkulaatorid valimi mahu määramiseks: <http://www.danielsooper.com/statcalc> ja <http://biostat.mc.vanderbilt.edu/wiki/bin/view/Main/PowerSampleSize> (12.02.2013).

Levinud on ka osalejatesisene katseplaan, kus sama katsealune osaleb katse ühes osas katsetingimustes (katserühma liikmena) ja teises osas kontrolltingimustes (kontrollrühma liikmena). Peamiseks eeliseks sellise katseplaani juures on see, et uurijal on suurem tõenäosus leida tulemustes erinevusi, kuna juhusliku vea suurust on vähendatud (nt individuaalsed erinevused). Teiseks eeliseks on see, et võrreldes osalejatevahelise katseplaani nõuab osalejatesisene katseplaani katse väiksemat arvu osalejaid. Peamiseks probleemiks on siin aga nn (stiimuli) järjekorra mõju, millest tuleb täpsemalt juttu allpool (vt ptk 2.5). Üldiselt on võimalik järjekorra mõju vähendada, kasutades juhuslikus järjestuses stiimuli esitust, s.t katse- ja kontrolltingimused vahelduvad samas katseseerias osaleja jaoks juhuslikus järjestuses.

Kuna osalejatesisene katseplaan on efektiivsem (osalejatevaheline indiviidi tasandi varieeruvus on vähendatud), siis üldiselt soovitatakse rakendada just sellist katsemalli. Osalejatesisene katseplaan ei sobi, kui katsel on katsealustele püsiv mõju (nt katsealuste treenimine vastavalt kahele erinevale katsetingimusele) või kui katse on väga pikk. Rakendatav on ka nn segakatseplaan (ingl *mixed design*), kus korraga on kasutusel mitu erinevat katseplaani, nt nii osalejatevaheline kui -sisene planeering. Näiteks võib osalejatevaheliseks sõltumatuks muutujaks olla sõnajärg (üks rühm hindab ühe sõnajärgiga ja teine teistsuguse sõnajärgiga stiimulit) ja osalejatesiseseks sõltumatuks muutujaks fraasipikkus (kõik osalejad hindavad nii pikki kui lühikesi fraase).

Lisaks kahele peamisele eelpool nimetatud katseplaanile on uurija käsutuses veel suur hulk erinevaid katseplaanide, nt mitme rühmaga katseplaan (ingl *multiple group design*), faktorplaanid (ingl *factorial design*), kvaasiekperimentaalne katseplaan (ingl *quasi-experiments*), plokk-plaanid (ingl *block design*), ladina ruut (ingl *Latin square*)<sup>4</sup>. Milline katseplaan valida sõltub uurimuse iseloomust, uurimisküsimusest, hüpoteesidest, sõltumatutest ja sõltuvatest muutujatest jne.

## 2.5. Katsematerjali koostamine ja esituse järjekord

Keeleteadusliku katsematerjali ehk **stiimuli** koostamisega seotud praktilisi nõuandeid ja kasulikke näpunäiteid pakuvad nt Schütze (1996: 184–186) ja Cowart (1997: 89–102)<sup>5</sup>. Siinkohal tahame välja tuua mõned põhireeglid: stiimul (nt katsetes kasutatavad laused, fraasid, sõnad jms) tuleks katsealustele esitada juhuslikus järjekorras; uurija peaks stiimulit kontrollima selle leksikaalse sisu, tähenduse, kompleksuse, sageduse, pikkuse ja struktuuri osas (Schütze 1996: 150–168); uurijal tuleks iga katsetingimuse jaoks välja mõelda mitu erinevat stiimulit (Schütze, Sprouse (ilmumas): 13). Kui uurija ei suuda ülalmainitud aspektides stiimulit kontrollida, võib tulemuste tõlgendus osutada keeruliseks, kuna ei saa olla kindel, millised stiimuli aspektid katsealustele mõju avaldasid. Siinjuures on paslik meelde tuletada uurimuse valiidsuse küsimust (vt ptk 2.1). Nii näiteks tekib tulemuste tõlgendamisel probleeme, kui katses kasutatavad keelelised üksused erinevad leksikaalse sisu poolest, aga soovitakse hoopis uurida keelelise üksuse pikkuse mõju konstruktsiooni/lause vastuvõetavusele. Kui näiteks tulemustes tuleb esile oluline erinevus lühikeste ja pikkade üksuste vahel, siis ei ole uurijal võimalik kindlaks teha, kas erinevused tulemustes on tingitud just stiimuli erinevast pikkusest (uurimuse

<sup>4</sup> Mainitud katseplaanide kohta võib lugeda mõnest sotsiaalteadlaste mõeldud katsete planeerimise käsiraamatust (nt Cobb 1998, Field, Hole 2006, MacLin, Solso 2007, Kantowitz jt 2009).

<sup>5</sup> Nii Schütze (1996) kui Cowart (1997) piirduvad ainult vastuvõetavushinnangu katsetega, kuid paljuski kehtivad autorite poolt välja toodud soovitusel ja hoiatused ka teist liiki keeleteaduslike katsete puhul.

seisukohast oluline tulemus) või hoopis stiimuli leksikaalsest sisust (uurimuse seisukohast ebaoluline tulemus, nn kõrvalmuutuja). Üks võimalus selle probleemi vältimiseks on, et katsematerjali tuuakse sisse kontroll-lauseid, kus katsealustel lastakse hinnata sama leksikaalse sisuga lauseid mingis kolmandas lausemallis. Kui erinevate leksikaalse sisuga katsestiimulite vahel ei tule selles kolmandas kontrolltingimuses oluliselt erinevaid tulemusi, võib järeldada, et katsetingimustes esinenud erinevus on tõepoolest tingitud sõltumatust muutujast “pikkus” ja mitte sellest, et laused on erineva leksikaalse sisuga.

Katse pikkus, s.t stiimuli **maht**, oleneb suuresti sellest, kuidas sõltuvat tunnust mõõdetakse (kas nominaalsel, järjestus- või intervallskaalal), millist andmeanalüüsi meetodit plaanitakse hiljem tulemuste analüüsimisel rakendada. Samuti sõltub stiimuli maht valimi suuruselt (vt ptk 2.3). Üldiselt on välja toodud, et nominaalse skaala puhul on optimaalne kasutada vähemalt nelja katseüksust ühe katsetingimuse kohta. Igal juhul tuleb arvestada, et katse ei saaks nii pikk, et katsealune tüdineb vastamast.

Katse ettevalmistamisel on oluline ka küsimus, millisesse **järjekorda** stiimulid (lauseid, visuaalne info vm) seada. Stiimulid ise peavad olema hoolega kontrollitud, kuid nende esitamise järjekord peab olema juhuslikustatud. Interneti teel läbiviidavas katses on võimalik juhuslikustada iga katsealuse katse eraldi. Paberil läbiviidava katse korral võiks kasutusel olla erinevad versioonid katsest (Cowart 1997: 96). Seda seepärast, et mitme stiimuli esitamisel järjest võib eelneva stiimuli vastus mõjutada järgneva stiimuli vastust. Erinevate mõjudena on välja toodud (Schütze 1996: 134–140): harjumine – vastus igale järgnevale stiimulile tuleb kergemini; väsimine – igale järgnevale stiimulile vastamine väsitab katsealust ja ta võib vastamisest tüdineda; ülekanne – eelnev stiimul mõjutab samamoodi vastama järgnevale; tundlikkus – iga järgneva stiimuliga aimab katsealune üha enam, millist vastust talt oodatakse. Nende mõjude vältimiseks on soovitatud (Cowart 1997: 91–92, Schütze 1996: 134–140) lasta katsealustel enne katset harjutada; teha katse nii lühike ja põnev kui võimalik; pikendada aega katsetingimuse vahel (vältides nii ülekanne mõju); koostada katse nii, et katsealune ei märkaks stiimulite varieeruvust (vt ka ptk 3.1); juhuslikustada stiimulite järjekord. Kõige selle juures on oluline jälgida, et katse läbiviimise tingimused ei muutuks.

### 3. Läbiviimise etapp

Peale katse hoolikat ettevalmistamist saab asuda katse läbiviimise juurde. Ka selles etapis on olulisi nüansse, mida silmas pidada. Siinses artiklis on eristatud katse ettevalmistamise ja läbiviimise etappi. Tegelikuses ei ole aga need teemad üksteisest rangelt eristatavad ja nii mõnelegi seigale, millele siinses peatükis viitame, tuleb mõelda juba katse planeerimise ajal. Esiteks tahame toonitada, et olenemata sellest, mis tüüpi katsega on tegu, tuleb igal juhul esmalt koostada pilootkatse. Pilootkatse viiakse läbi väikse hulga katsealustega ning selle eesmärgiks on kontrollida, kas koostatud stiimul töötab. Kui millegagi on probleeme (nt ei ole juhised piisavalt selge või on stiimulis vajakajäämisi), tuleb probleemne koht parandada ning viia läbi uus pilootkatse. Pilootkatsest saab teada ka seda, kui kaua katse keskmiselt aega võtab.

### 3.1. Katsealuste värbamine ja instrueerimine

Kõige enam kasutatud sihtrühm katsealuste värbamisel on üliõpilased, samas märgivad Gonzalez-Marquez jt, et katsealuste leidmiseks on ka teisi võimalusi ja kohti (katsealuste värbamise viiside kohta vt nt 2007: 74). Katsealuste värbamisel tuleb aga kindlasti silmas pidada valimi koostamise põhimõtteid, mida käsitleti ptk-s 2.3. Inimeste motiveeritus katses osaleda sõltub sellest, kuivõrd nad peavad selleks pingutama. Kui katse ja selle läbiviimiseks vajalik tehnika võimaldab uurijal endal katsealuste juurde minna, on osalejaid lihtsam värvata. Samuti võib katsealuste leidmist hõlbustada preemiade olemasolu. (Gonzalez-Marquez jt 2007: 74) Oluline on katsealuste värbamisel ausalt öelda, kui kaua katses osalemine aega võtab. Internetis täidetavate katsete puhul peaks katse tutvustuses olema kirjas aus prognoos katse kestuse kohta. Vastasel juhul võib katsealune vastamisest tüdida, katse pooleli jätta või järgmisel korral osalemast keelduda. Eetilistel põhjustel on oluline rõhutada, et katses osalemine on alati vabatahtlik.

Katsealuste kohta tuleb koguda ka taustaandmeid. Keeleteaduslike katsete puhul on eeskätt tähtsad demograafilised andmed nagu emakeel, teised keeled, päritolu, vanus, sugu, haridus. Need andmed annavad ülevaate sellest, kui homo- või heterogeenne valim on. Katsealuse emakeele olulisus on semantikat puudutavates uurimustes ilmselge, kuid ka teiste keelte oskus võib katsealuse keeletunnetust mõjutada, kui indiviid on teatud keelt õppinud väga pikka aega. Ka sotsiaalsed muutujad nagu vanus, sugu ja haridus, samuti regionaalne päritolu võivad tulemusi teatud suunas mõjutada. (Nende muutujate kontrollitavuse kohta vt ptk 2.3). Üldiselt ei eelda enamik semantika uurimiseks kasutatavatest katsetest isikuandmete (nt nimi) kogumist, vaid piisab demograafilistest andmetest. Kui identifitseerimist võimaldavaid andmeid on mingil põhjusel tarvis koguda, tuleb isikuandmeid sisaldavate failidega hoolikalt ümber käia, sest uurija peab tagama katseisikute anonüümsuse (Field, Hole 2006: 101, Gonzalez-Marquez jt 2007: 74).

Enne katsega alustamist tasub lausuda mõni sissejuhatav lause, et osalejaid motiveerida. Samas kaasneb igasuguse info jagamisega uurimuse kohta alati oht katsealuste vastuseid mingil viisil mõjutada. (Coward 1997: 87) Nimelt on täheldatud, et kui katsealused taipavad, milliseid vastuseid neilt teatud küsimustele oodatakse, vastavadki nad nii, et nende vastused oleksid sotsiaalselt vastuvõetavamad, s.t nad püüavad uurijat aidata (Gries 2009: 49). Vainik (2004: 130) sedastab oma kogemusest, et katsealused on väga vastutulelikud, täites neile antud viimase ülesande innuga ka juhul, kui nad on eelnevast katsest väsinud ja neile pakutakse võimalust lõpetada. Katse läbiviija võib tulemusi mõjutada ka näiteks kehakeele abil katsealustele vihjeid andes. Seetõttu oleks hea kasutada topelt-pimedat meetodit, mil katset viib läbi keegi, kes ei tea, mis on uurija(te) hüpoteesid. Üks võimalus tulemusi mõjutada ehk katseisikuid praimida on teatud keelendi varasem mainimine samas olukorras.

Tulemuste mõjutamise välistamiseks kasutatakse sotsiaal- ja humanitaarteadustes sageli varjamist. See tähendab, et katsealustele ei avaldata olulisi detaile katse eesmärkide kohta või öeldakse, et katse eesmärk on miski muu, kui see tegelikult on (Kirk 2013: 23). Samas tuleb meeles pidada, et katsealuste petmine ja nende eest informatsiooni varjamine on taunitav eetilistel põhjustel. Katses osalejad peavad uurijale andma informeeritud nõusoleku, mis tähendab, et nad mõistavad,

millega nad ennast seovad ja nõustuvad osalema. On aga küsitav, kas informeeritud nõusolekut saab anda ilma (korrektse) infota (Field, Hole 2006: 99). Field ja Hole osutavad, et psühholoogias pole teatud uurimisküsimuste puhul sellest midagi halba, kui katsealune teab, mis on katse eesmärk, ning see ei mõjuta eksperimendi tulemusi soovimatus suunas (2006: 99). Keeleteadusest võib näiteks tuua, kui katsega uuritakse kahe konstruktsiooni varieerumist keeles (nt kaassõna *peal* ja adessiivi varieerumist ruumisuhete väljendamisel (Klavan 2012)). Sellisel juhul, kui ei ole alust arvata, et keelenorming saadavaid tulemusi mingil moel mõjutaks, ei pea eksperimendi eesmärki katsealuste eest varjama ning katsealuseid võib suunata teadlikult kahe konstruktsiooni vahel valima. Kuigi äärmustena on teada näiteid sotsiaalpsühholoogia vallast, kus osalejaid tulebki uurimuse eesmärkide osas eksiteele juhtida, siis üldiselt tuleks otsest katsealuste petmist võimalusel vältida (Field, Hole 2006: 99–100). Kui tahetakse varjata katse otsest eesmärki, võiks katse eesmärgi sõnastada võimalikult üldiselt (nt katses uuritakse keelekasutust või ruumiväljendeid vms). Oluline on, et katsealustele selgitatakse katse läbiviimise protseduuri piisaval määral, et nad teaks, mis juhtuma hakkab (Gonzalez-Marquez jt 2007: 74), samuti tuleb pärast katse sooritamist avaldada katse tegelik eesmärk.<sup>6</sup>

Katse eesmärgi varjamise huvides on kasutusel ka kas täitelauseid (ingl *filler*) või sõltuvalt stiimulist muud täiteüksused (nimetatakse ka distraktoriteks). Need on katseüksused, mis pole seotud uurimisküsimusega (Schütze, Sprouse (ilmumas): 14). Täiteüksuste eesmärgiks ongi katsealustes katse eesmärgi osas segadust tekitada, et katsealune ei aimaks katse konkreetset eesmärki ja uurija hüpoteese ega püüaks vastata uurijale meelepäraselt. Cowarti (1997: 92) järgi peab täiteüksuseid olema vähemalt sama palju kui neid üksusi, mis katse eesmärki täidavad, ideaaljuhul on neid aga 2–3 korda rohkem. Täiteüksusi kasutades tuleb aga silmas pidada, et katse liiga pikaks ei läheks, samuti on tähtis täiteüksuste järjekord (vt ka ptk 1.5). Täiteüksused peaksid olema paigutatud teiste katseüksuste vahele pooljuhuslikus järjekorras selliselt, et katse esimene stiimul on täiteüksus ning et täiteüksus eraldaks kõiki eesmärki täitvaid katseüksusi.

### 3.2. Juhis ja stiimul

Uurija ei tohiks kunagi alahinnata mõju, mida tema instrueerimine katsealustele avaldab (Cowart 1997: 56). Seetõttu tuleb hoolikalt läbi mõelda nii see, mida katsealustele enne katset öelda, kui ka kirjalik tööjuhis, kus on kirjas, mida neilt oodatakse. Praktikas kättesaamatu, kuid eesmärk, mille poole püüelda, on olukord, kus kõik katsealused saavad ühtmoodi õigesti aru, mida nad tegema peavad. Iga väiksemgi erinevus selles, kuidas iga indiviid ülesandele läheneb, suurendab varieeruvust andmetes (Cowart 1997: 89).

Cowarti (1997: 56–57) järgi on üks olulisemaid küsimusi, mis tekib naiivsete keelekasutajate kasutamisel katsealustena, see, missugune juhise toob esile vajalikud otsustused, s.t sellised, mis ei ole mõjutatud keelevälistest faktoritest. Teatud uurimisküsimuste, näiteks lause vastuvõetavuse hindamise puhul võib keelekasutajate vastuseid mõjutada teadmine koolis õpitud grammatikast, kuid uurija

<sup>6</sup> Enne katse läbiviimist soovitatakse ka uurida (Field, Hole 2006: 98), kas institutsiooninormide eetikakomitee, kellele tuleb esitada hindamiseks kavandatava uurimistöö plaan, ning kes peab andma nõusoleku uurimistöö läbiviimiseks. Eestis tegutseb Tartu Ülikooli inimuurimise eetika komitee, kes menetleb nii ravimi kliiniliste uuringute ja biomeditsiiniliste, psühholoogiliste käitumisteaduste alase uurimistöö projekte kui ka üliõpilastööde projekte (<http://www.ut.ee/et/teadus/eetikakomitee>, 12.02.2013).

on huvitatud hoopis tegelikust keelekasutusest. Sellisel juhul võib juhis suunata katsealust ignoreerima õpitud reegleid ja tegema otsustusi ükskõik missugustele muudele kriteeriumidele toetudes. Näiteks uuris Jürine (2011) sunnitud valiku katse abil, kuidas on kaassõnaühendite norminguvastane kokkukirjutamine seotud tähendusega. Kuna kaassõnaühendite kokkukirjutamine on praeguse normingu kohaselt reeglitevastane, suunas juhis katsealuseid valima kirja pilti selle alusel, mis lausesse kõige paremini *sobib*, mitte õpitud reeglitele toetudes. Sellist juhust nimetatakse intuitsiooni-juhiseks, alternatiiv on kasutada preskriptiivset juhust, mille puhul suunatakse katsealust toetuma keelenormile. Samas nendib Cowart (1997: 57–58), et pole piisavalt tõendeid väitmaks, et katsealused suudavad lauseid hinnates rakendada erinevaid standardeid, s.t hinnagud võivad olla samad nii intuitsiooni- kui ka preskriptiivse juhise puhul. Siiski on oluline, et juhis, mis katsealustele antakse, on võimalikult selge ja ühene ning suunab nende tähelepanu sellele ülesandele, mille täitmist neilt oodatakse (Cowart 1997: 91).

Materjali esitamiseks katsealustele on erinevaid võimalusi. Cowarti (1997: 63) järgi võib materjali esitamise viis katse tulemusi teatud aspektides mõjutada. Stiimuli võib esitada nii kirjas, piltidena, reaalse objektidena kui ka auditiivsel kujul. Tavalisim vorm visuaalse stiimuli esitamiseks on prinditud küsitluslehed, kus stiimul on esitatud kirjalikult ja/või piltidena. Näiteks võib vastuvõetavuse hindamise katses, milles uuritakse ruumisemantikast, kasutada ruumi kirjeldamiseks pilti ning esitada katsealustele iga pildi kohta lause (Carlson, Hill 2007: 253). Kirjaliku katse plussina võib nimetada asjaolu, et selle ettevalmistus ja läbiviimine on lihtne ja see võimaldab koguda andmeid suurelt hulgalt katsealustelt korraga. Ka võimaldab kirjalik vorm sama küsitluslehe erinevaid versioone korraga esitada. Sellise katse puhul on miinuseks see, et tulemused sõltuvad inimeste lugemisoskusest, mille tase on üldkogumis varieeruv. Samuti see, et inimeste ootused kirjutatud keelele võivad olla üsna erinevad suulise keelega võrreldes. Näiteks vastuvõetavuse hindamise katse puhul võivad teatud laused suuliselt esitatuna olla vastuvõetavamad kui kirjapanduna. (Cowart 1997: 64) Ainult kujundliku stiimuli puhul kirjaoskusest tulenevaid probleeme ei esine. Küll aga on ka piltide kasutamisel omad miinused, näiteks võivad pildid olla raskesti tõlgendatavad ning anda suurema varieeruvusega andmeid. Vahel kasutatakse visuaalse stiimuli esitamiseks ka reaalseid objekte, sellise katse miinuseks on see, et katse läbiviijal on suurem võimalus katsealuseid kogemata mõjutada, kui katset korraldatakse mitmes järgus. Auditiivne stiimul võib osutada kasulikuks, kui on oluline juhtida katsealuse tähelepanu võimalikult kaugele kirjakeele normeringutest.

Katse stiimuli võib esitada ka arvuti abil. Näiteks kasutatakse sellist esitlusviisi interneti teel levitatavates küsitlustes. Üldiselt on arvuti abil täidetavatel küsitluse tüüpi katsetel samad plussid ja miinused, mis ülalpool esitatud, kuid siin käsitletavat erinevused tulenevad just interneti kui teistsuguse meediumi rollist. Gonzalez-Marquez jt toovad välja, et ideaalis tuleb tagada, et kõik katsealused on võrdses situatsioonis (2007: 74). Internetis levivate katsete puhul puudub aga kontroll katseks kulutatud aja, katsealuste fokuseerituse ja motivatsiooni üle. Olevalt võimalustest, mis internetikatse jaoks kasutatavad tehnilised abivahendid pakuvad, on uurijal siiski teatud juhtudel võimalik tuvastada katse ajaline kestus. Internetikatsetel on ka plussid. Katses osalemine interneti teel on katsealusele palju mugavam mitmel põhjusel. Kuna katse tegemise koht ja aeg ei ole määratud, saab



katsealune sooritada katse talle sobival ajal ja sobivas kohas. See tähendab, et katse situatsioon on loomulikum ja vabam. Peale selle võimaldab interneti teel levitav katse leida hõlpsasti katsealuseid.

#### 4. Kokkuvõte

Artikli eesmärgiks oli anda ülevaade katselise semantika võimalustest ja probleemidest, tuues konkreetseid näiteid seikadest, millega semantikas katseid tehes peaks arvestama. Katse kavandamise juures toodi välja olulised etapid, mis, nagu iga muugi teadusliku katse puhul, tuleb läbida ka tähendust uurides. Probleemi sõnastuse ja ülesande püstituse juures on oluline hüpoteeside operatsionaliseerimine, mis semantika omapära arvestades on üks keerulisemaid ülesandeid tähenduste uurimisel empiiriliselt, kuid just sellest sõltub uurimuse valiidsus ja reliaablus. Lähemalt peatuti ka muutujate/tunnuste valiku küsimusel ja erinevatel tunnuste tüüpidel. Mitteamuliste tunnuste ülekaal ei iseloomusta mitte ainult semantikat, vaid kogu keeleteaduslikku andmestikku (v.a foneetika). Katset ette valmistades tuleb läbi mõelda ka valimi koostamise põhimõtted ja valimi optimaalne suurus. Artiklis püüti näidata, et keeleuurijal on võimalik valida mitme erineva katseplaani vahel ja et konkreetne valik sõltub uurimuse iseloomust. Olulist rolli mängib keelise katse kavandamisel loomulikult ka see, milline on katsematerjal, mille uurija katsesse valib. Sellega seoses toodi välja, mida arvestada katsematerjali koostamisel ja esitamisel. Viimaks kirjeldati probleeme, millega tuleb arvestada katset läbi viies – katsealuste värbamine ja instrueerimine ning sellega seonduvad eetilised küsimused, juhis ja katse stiimuli esitamise viis.

Selge on, et iga uurimuse puhul kerkivad katset tehes üles konkreetselt just vastava uurimusküsimusega seotud probleemid ja siinne ülevaade suutis anda vaid valiku olulisematest ja tavalisematest aspektidest, mida kaaluda. Igasuguse keelise katse puhul jääb aga kehtima vanasõna “üheksa korda mõõda, ükskord lõika”. Selleks, et katselisel teel tähenduse uurimiseks adekvaatseid andmeid keelekõneleajatelt kätte saada, tasub planeerimise peale kulutada proportsionaalselt rohkem ressursi kui teostusele ja andmeanalüüsile. Kui katse ülesehituses on tehtud viga, ei päästa uurimust ei see, kui katse viiakse läbi suure hulga katsealustega, ega see, kui andmeid analüüsitakse mitmekesiste statistiliste andmeanalüüsi meetoditega. Potentsiaalseid puudusi, mida tuleb katse ülesehitamisel ette ka kogenenumatel katsekoostajatel, aitab tuvastada pilootkatse.

#### Viidatud kirjandus

- Brugman, Claudia 1988 [1981]. *The Story of over: Polysemy, Semantics, and the Structure of the Lexicon*. New York: Garland Press.
- Carlson, Laura A.; Hill, Patrick L. 2007. Experimental methods for studying language and space. – Monica Gonzalez-Marquez, Irene Mittelberg, Seana Coulson, Michael Spivey (Eds.). *Methods in Cognitive Linguistics. Human Cognitive Processing 18*. Amsterdam, Philadelphia: John Benjamins, 250–276.
- Cobb, George W. 1998. *Introduction to Design and Analysis of Experiments*. New York: Springer.

- Cowart, Wayne 1997. *Experimental Syntax: Applying Objective Methods to Sentence Judgments*. London, New Delhi: Sage.
- Dąbrowska, Ewa 2010. *Masterclass in Experimental Methods in Linguistics*. Loengumaterjalid.
- Field, Andy; Hole, Graham 2006. *How to Design and Report Experiments*. London: Sage.
- Geeraerts, Dirk 2006. *Methodology in Cognitive Linguistics*. – Gitte Kristiansen, Michel Achard, René Dirven, Francisco J. Ruiz de Mendoza Ibáñez (Eds.). *Cognitive Linguistics: Current Applications and Future Perspectives*. *Applications of Cognitive Linguistics* 1. Berlin and New York: Mouton de Gruyter, 21–49.
- Geeraerts, Dirk 2010. *The Doctor and the Semanticist*. – Dylan Glynn, Kerstin Fischer (Eds.). *Quantitative Methods in Cognitive Semantics: Corpus-Driven Approaches*. *Cognitive Linguistics Research* 46. Berlin, New York: Mouton de Gruyter, 63–78.
- Gibbs, Raymond W. Jr.; Colston, Herbert L. 1995. *The cognitive psychological reality of image schemas and their transformations*. – *Cognitive Linguistics*, 6 (4), 347–378. <http://dx.doi.org/10.1515/cogl.1995.6.4.347>
- Glynn, Dylan 2010. *Testing the hypothesis. Objectivity and verification in usage-based Cognitive Semantics*. – Dylan Glynn, Kerstin Fischer (Eds.). *Quantitative Cognitive Semantics: Corpus-Driven Approaches*. Berlin: Mouton de Gruyter, 239–270.
- Gonzalez-Marquez, Monica; Becker, Raymond B.; Cutting, James E. 2007. *An introduction to experimental methods for language researchers*. – Monica Gonzalez-Marquez, Irene Mittelberg, Seana Coulson, Michael Spivey (Eds.). *Methods in Cognitive Linguistics*. *Human Cognitive Processing* 18. Amsterdam, Philadelphia: John Benjamins, 53–86.
- Gries, Stefan Th. 2009. *Statistics for Linguistics with R: A Practical Introduction*. Berlin, New York: Mouton de Gruyter. <http://dx.doi.org/10.1515/9783110216042>
- Jürine, Anni 2011. *Vormierinevused tähenduserinevuse ilminguna. Kaassõnähendite grammatiseerumine konstruktsioonides*. – *Keel ja Kirjandus*, 12, 899–916.
- Kantowitz, Barry H.; Roediger, Henry L., III; Elmes, David G. 2009. *Experimental Psychology*. 9th edition. Belmont: Wadsworth, Cengage Learning.
- Kirk, Roger E. 2013. *Experimental Design: Procedures for the Behavioral Sciences*. 4th ed. Thousand Oaks: Sage.
- Klavan, Jane 2008. *Medial Region Adpositions in English and Estonian: A Cognitive Grammar Perspective*. *Magistritöö*. Tartu: Tartu Ülikooli Kirjastus.
- Klavan, Jane 2012. *Evidence in Linguistics: Corpus-Linguistic and Experimental Methods for Studying Grammatical Synonymy*. *Dissertationes linguisticae Universitatis Tartuensis* 15. Tartu: Tartu Ülikooli Kirjastus.
- Klavan, Jane; Veismann, Ann; Jürine, Anni 2013. *Katselised meetodid tähenduse uurimisel*. – *ESUKA / JEFUL*, 4 (1), 17–34.
- Käärik, Ene 2011. *Empiiriline andmeanalüüs*. Loengumaterjalid.
- Lakoff, George 1987. *Women, Fire and Dangerous Things: What Categories Reveal about the Mind*. Chicago: University of Chicago Press.
- Luuk, Aavo 2012. *Eksperimentaalne psühholoogia*. Loengumaterjalid.
- MacLin, M. Kimberly; Solso, Robert L. 2007. *Experimental Psychology*. 8th ed. Boston: Pearson.
- Sandra, Dominiek; Rice, Sally 1995. *Network analyses of prepositional meaning: Mirroring whose mind – the linguist's or the language user's?* – *Cognitive Linguistics*, 6 (1), 89–130. <http://dx.doi.org/10.1515/cogl.1995.6.1.89>
- Schütze, Carson T. 1996. *The Empirical Base of Linguistics: Grammaticality Judgments and Linguistic Methodology*. Chicago: University of Chicago Press.
- Schütze, Carson T.; Sprouse, Jon (ilmumas). *Judgment Data*. – Devyani Sharma, Rob Podesva (Eds.). *Research Methods in Linguistics*. Cambridge: Cambridge University Press.
- Sprouse, Jon; Almeida, Diogo (ilmumas). *Power in Acceptability Judgment Experiments and the Reliability of Data in Syntax*.

- Stefanowitsch, Anatol 2010. Empirical cognitive semantics: Some thoughts. – Dylan Glynn, Kerstin Fischer (Eds.). *Quantitative Methods in Cognitive Semantics: Corpus-Driven Approaches*. *Cognitive Linguistics Research* 46. Berlin: Mouton de Gruyter, 355–381.
- Sutrop, Urmas 2001. The list task and a cognitive salience index. – *Field Methods*, 13 (3), 263–276. <http://dx.doi.org/10.1177/1525822X0101300303>
- Tooding, Liina-Mai 2007. *Andmete analüüs ja tõlgendamine sotsiaalteadustes*. Tartu: Tartu Ülikooli Kirjastus.
- Uusküla, Mari 2008. *Basic Colour Terms in Finno-Ugric and Slavonic Languages: Myths and Facts*. *Dissertationes linguisticae Universitatis Tartuensis* 9. Tartu: Tartu Ülikooli Kirjastus.
- Vainik, Ene 2004. *Lexical Knowledge of Emotions: The Structure, Variability and Semantics of the Estonian Emotion Vocabulary*. *Dissertationes linguisticae Universitatis Tartuensis* 5. Tartu: Tartu Ülikooli Kirjastus.
- Veismann, Ann 2008. *Eesti kaassõnade tähenduste kirjeldamise probleemidest*. – *Keel ja Kirjandus*, 5, 335–352.
- Veismann, Ann 2009. *Eesti keele kaas- ja määrsõnade semantika võimalusi*. *Dissertationes linguisticae Universitatis Tartuensis* 11. Tartu: Tartu Ülikooli Kirjastus.
- Wilson Van Voorhis, Carmen; Morgan, Betsy 2007. Understanding power and rules of thumb for determining sample sizes. – *Tutorials in Quantitative Methods for Psychology*, 3 (2), 43–50.

## Võrgumaterjalid

- Tartu Ülikooli Inimuuringute Eetikakomitee. <http://www.ut.ee/et/teadus/eetikakomitee> (12.02.2013).
- Daniel S. Soperi statistika kalkulaator. <http://www.danielsoper.com/statcalc> (12.02.2013).
- William D. Duponti ja Walton D. Plummer Jr-i statistika kalkulaator. <http://biostat.mc.vanderbilt.edu/twiki/bin/view/Main/PowerSampleSize> (12.02.2013).

**Anni Jürine** (Tartu Ülikool) on uurinud morfosüntaktilisi protsesse keeles, erialased huvid seotud eelkõige tänapäeva eesti keele ja keele muutumisega.  
[anni.jyrine@ut.ee](mailto:anni.jyrine@ut.ee)

**Jane Klavan** (Tartu Ülikool) on uurinud eesti keele kohakaassõnade ja kohakäänete paralleelset kasutust. Peamiseks keeleteaduslikuks huviks on muuhulgas korpusandmete ja katselisel teel saadud keeleandmete võrdlus.  
[jane.klavan@ut.ee](mailto:jane.klavan@ut.ee)

**Ann Veismann** (Tartu Ülikool) tegeleb kognitiivse (ruumi)semantikaga, polüseemiaga, grammatiliste üksuste tähendustega. On uurinud eesti keele kaas- ja määrsõnade semantikat.  
[ann.veismann@ut.ee](mailto:ann.veismann@ut.ee)

# EXPERIMENTAL SEMANTICS: DESIGN AND IMPLEMENTATION

Anni Jürine, Jane Klavan, Ann Veismann

University of Tartu

In semantics, conducting corpus linguistic analyses is much more common than conducting experiments. There are various reasons for this bias in empirical methods, including, among other things, the conviction that conducting an experiment is very difficult and risky. The aim of the present article is to discuss some of the practical aspects involved in experimental semantics. The focus is on two crucial stages; planning and conducting linguistic experiments. The article is concerned, *inter alia*, with the following topics: postulating and operationalising hypotheses, the validity and reliability of an empirical quantitative study, different types of variables, the representativeness of the sample, simple random sample, sample size, choosing an experimental design, constructing and presenting experimental materials, recruiting and instructing participants, and ethical considerations in experimental research. The aim here is not to give a complete list of all the possible practicalities one may encounter when conducting linguistic experiments. The specific design and implementation of an experiment depends largely on the specific research question. The focus is instead on some of the basic aspects that every researcher needs to consider when doing experimental semantics.

One of the crucial points in designing any empirical study is the postulation and operationalisation of hypotheses and variables. Taking into account the complex nature of semantic studies, this aspect requires both creativeness as well as vast experience. At the same time, the quality of an experiment (its validity and representativeness) depends largely on the decisions the researcher makes in these early stages of research planning. Although it may seem tedious to spend too much energy and resources on detailed planning, it is worth paying meticulous attention to the crucial step of experimental design. Neither large quantities of data nor advanced statistical methods will save a linguistic experiment that suffers from dubious validity.

**Keywords:** experimental methods, semantics, cognitive linguistics

# KUIDAS TUVASTADA DIALOOGI STRUKTUURI: KORPUSE ANALÜÜS

Mare Koit

**Ülevaade.** Artiklis uuritakse dialoogi kui erilist sidustekstide liiki, eesmärgiga määrata dialoogi struktuur, mille moodustavad mikrotaseme üksused – dialoogiaktid. Empiiriliseks materjaliks on väike telefoni-müügivestluste alamkorpus. Analüüsitavates kõnedes suhtlevad omavahel kaks ametiisikut: helistaja (koolitusfirma konsultant) pakub koolituskursusi, mida eeldatavasti vajab vastuvõtja (teise asutuse esindaja). Dialoogisalvestused on litereeritud, kasutades vestlusanalüüsi transkriptsiooni. Korpuses on märgendatud dialoogiaktid. Artiklis otsitakse reegleid, mis võimaldaksid automaatselt tuvastada müügivestluse erinevaid faase, kasutades selleks dialoogiaktide järjendeid ja nende positsiooni dialoogis. Edasine eesmärk on luua tarkvara dialoogide automaatseks pragmaatiliseks analüüsiks, sh dialoogiaktide, dialoogide lineaarse struktuuri ja alamdialoogide tuvastamiseks.\*

**Võtmesõnad:** dialoog, müügivestlus, dialoogi struktuur, dialoogiakt, eesti keel

## 1. Sissejuhatus

Sidusteksti pragmaatiline analüüs on üks keele automaattöötamise ülesandeid. Lausungite ja nende tähenduste vaheliste seoste ning lausungi eesmärgi leidmiseks saab kasutada erinevat informatsiooni. Kui eelnevalt on läbi viidud teksti morfoloogiline, süntaktiline ja/või semantiline analüüs, siis on nende etappide väljund rikkalikuks sisendiks pragmaatilisele analüsaatorile (Barton 2004). Sellisel juhul saab sõnastada reeglid, mis võtavad arvesse eelnenud analüüsi-etappide infot. Teisalt, kui on olemas mahukas dialoogikorpus, siis saab kasutada andmepõhiseid meetodeid, et määrata diskursuse struktuuri ja uurida suhtlemisel kasutatavat keelt.

Käesolevas artiklis uuritakse sidustekstide eriliiki – dialooge. Dialoogi analüüsimiseks tuleb see jaotada väiksemateks struktuuri- ja funktsionaalseteks üksusteks

\* Artikli valmimist on toetanud Euroopa Regionaalarengufond Eesti Arvutiteaduse Tippkeskuse kaudu, Haridus- ja Teadusministeerium (projektid SF0180078s08 ja EKT5) ning Eesti Teadusagentuur (grandid ETF9124 ja ETF8558). Autor tänab Riina Kasterpalu asjatundlike märkuste eest.

(Gibbon jt 2000). Siinseks eesmärgiks on määrata dialoogi struktuur, mille moodustavad mikrotaseme üksused – dialoogiaktid ja nende järjendid.

Dialoogi struktuuri on vaja teada inimesega loomulikus keeles suhtlevate dialoogsüsteemide arendamisel. Dialoogihalduril peab olema info suhtluse ülesehitusest, selleks et mõista kasutaja lausungeid ja kavandada oma vastuseid (Lester jt 2004). Teiselt poolt võib dialoogi struktuuri tuvastamist käsitleda ka omaette ülesandena, mille lahendamise tulemusi saab kasutada erinevates rakendustes, nt sisukokkuvõtete koostamine, infootsing jne.

Artiklis uuritakse institutsionaalsete dialoogide ühte liiki – autentseid müügi-vestlusi. Ühe koolitusfirma konsultandid helistavad võimalikele klientidele – teiste asutuste juhtidele või personalispetsialistidele – ja pakuvad asutuse töötajatele koolituskursusi. Eesmärgiks on tutvustada koolitusfirmat, koguda infot kliendi kohta, äratada temas huvi kursuste vastu, põhjendada kursuste kasulikkust ja toetada klienti positiivse otsuse tegemisel (s.o tellida koolitusfirmast mõni kursus).

Artikli ülesehitus on järgmine. Jaotises 2 antakse ülevaade empiirilisest materjalist – Eesti dialoogikorpusest ja korpuse märgendamisel kasutatavast dialoogiaktide tüpoloogiast. Jaotis 3 on pühendatud korpuse analüüsile. Vaadeldakse müügi-vestluse erinevaid faase. Jaotises 4 arutletakse korpuse analüüsi tulemuste üle ja jaotises 5 tehakse kokkuvõtteid.

## 2. Korpuse ja dialoogiaktide tüpoloogia

Uurimus põhineb Eesti dialoogikorpusel (EDiK) (Hennoste jt 2009). Selle suurima osa moodustavad inimestevahelised dialoogid, mis on võetud Eesti suulise keele korpusest – 1137 telefonikõnet (infotelefonile, reisibüroodesse, müügi-vestlused jne) ja 45 silmast-silma vestlust, kokku 1182 transkribeeritud teksti mahuga 250 000 tekstisõna.

Teise osa EDiK-ist moodustavad umbes 200 kirjalikku dialoogi, mis on kogutud arvutisimulatsioonides võlur Ozi meetodil.

EDiK-i kolmandat osa (inimese ja arvuti vahelised vestlused) kogutakse kahe veebipõhise dialoogsüsteemi vahendusel, mis suhtlevad kasutajaga eesti keeles. Üks neist annab infot Tartu kinodes linastuvate filmide kohta ja teine hambaraviinfot.

Korpuses on märgendatud dialoogiaktid (DA), kasutades Tartu Ülikoolis välja-töötatud aktitüpoloogiat (Hennoste, Rääbis 2004). Tüpoloogia põhiosa on sarnane üldtuntud tüpoloogiatega (nt DAMSL ja SWBD-DAMSL). Tüpoloogia põhineb vestlusanalüüsist pärit suhtluse organiseerimise printsiipidel (Hutchby, Wooffitt 1998), mille kohaselt peab partner alati reageerima eelmisele voorule, sõltumata sellest, millised on tema oma plaanid või strateegiad. Mõned DA-de klassid moodustavad kokkuleppelisi paare, kus esimese akti lausumine teeb relevantseks teise akti. Selliseid seotud aktipaare nimetatakse naabruspaarideks (nt küsimus–vastus). Tüpoloogias on aktid jaotatud kahte klassi: naabruspaariaktid ja üksikaktid. Igal naabruspaariaktil on esiliige ja järelliige (vrd edasisaavatav ja tagasisaavatav funktsioon DAMSL-is).

Kõik DA-d on jaotatud infoaktideks (küsimused, direktiivid jms) ja suhtluse juhtimise aktideks (rituaalsed aktid, tagasiside jms). Suhtluse juhtimise aktid on omakorda jaotatud kahte alamrühma: aktid, mis juhivad sujuvat suhtlust, ja suhtlusprobleemide lahendamise aktid (e parandusaktid).

Tüpoloogias on 12 rühma; DA-de koguarv on 126.

DA-de nimed koosnevad kahest osast, mis on teineteisest eraldatud kooloniga: a) esimesed kaks tähte annavad aktirühma nime, nt KY – küsimused, VR – vabatahtlikud reaktsioonid. Kolmas täht esineb ainult naabruspaariaktide nimes – esiliige (E) või järelliige (J); b) akti pärisnimi, nt KYE: AVATUD, KYJ: INFO ANDMINE, VR: VASTUVÕTUTEADE (vt lähemalt Hennoste, Rääbis 2004).

Dialoogiakte on seni märgendatud käsitsi, kasutades tarkvara, mis hõlbustab tekstide valikut korpusest ja DA-de valikut menüüst. Kaks lingvisti märgendavad sama teksti teineteisest sõltumatult ja kolmas ühtlustab märgenduse. Praeguseks on kasutamiskvaliteet ka veebipõhine tarkvara DA-de poolautomaatseks märgendamiseks (Aller 2012). Programmi sisendiks on dialoogi tekst, programm tükeldab selle lausungiteks ja määrab igale lausungile kuni viis aktimärgendit, kasutades Bayesi naiivset klassifitseerijat. Seejärel saab inimene nende hulgast valida õige(d) aktimärgendi(d) või lisada menüüst uue(d) märgendi(d), kui programmi leitud märgendite hulgas ei olnud õiget.

Järgnevas dialoogikatkendis (1) on osalejateks klient (A) ja ametnik (B). Telefonikõne on lindistatud ja litereeritud, kasutades vestlusanalüüsi transkriptsiooni. DA-d on märgendatud.

- (1) A: tere? | RIE: TERVITUS |  
kas te ütleksite (.) doktor `Kullassepa `telefoninumbri. | KYE: JUTUSTAV KAS |  
(...)  
B: kaheksakend `neli kuus seitse kaheksa. | KYJ: INFO ANDMINE |  
A: kaheksakend neli kuus seitse kaheksa | KYE: VASTUST PAKKUV |  
| PPE: ÜLEKÜSIMINE |  
[ai]tääh | RIE: TÄNAN |  
B: [jah] | KYJ: JAH | | PPJ: PARANDUS |  
palun? | RIJ: PALUN |

Käesoleva uurimuse jaoks valiti EDiK-i suletud alamkorpus, mis koosneb telefoni teel peetud müügistlustest (Rääbis, Vellerind 2000). Telefonikõnedes suhtlevad kaks ametiisikut, kellest üks, helistaja (konsultant), esindab koolitusfirmat (näidetes on firma muudetud nimi Tiritamm), mis pakub koolitust neljas valdkonnas: juhtimine, müük, turundus ja teenindus. Pakutavad kursused on eeldatavasti vajalikud teisele ametiasutusele, kelle esindajat nimetatakse siin kliendiks. Dialoogid on korpuse kinnises osas, vastavalt kokkuleppele koolitusfirmaga. Publitseerida on lubatud üksnes dialoogide katkendeid, maksimaalse pikkusega neli voo. Kogu korpuses on 48 telefonikõnet (kõned on salvestanud ja litereerinud Riina Vellerind-Kasterpalu ning DA-d on märgendatud TÜ suulise keele rühmas). Käesolevas artiklis on müügistluste hulgast juhuslikult valitud analüüsimiseks 9 kõnet.

Korpuses on küll märgendatud DA-d, kuid see on morfoloogiliselt ja süntaktiliselt analüüsimata. Siiski osutub, et ka DA-de märgendid annavad piisavaid märguandeid dialoogi struktuuri tuvastamiseks.

## 3. Müügivestluste analüüs

### 3.1. Vestluse üldine struktuur

Kõik korpuse müügivestlused kuuluvad läbirääkimiste algusetappi. Analüüsitud kõnedes helistab konsultant võimalikule kliendile pärast seda, kui on saatnud talle koolitusfirma kataloogi, mis sisaldab pakutavate kursuste ja koolitusprogrammide nimestiku ning hinna. Telefonikõne kestel annab helistaja infot oma firma kohta, et tekitada ja/või suurendada kliendi huvi kursuste vastu. Saamaks infot kliendi vajadustest, esitab helistaja talle palju küsimusi. Reeglina on konsultandil vaja pidada mitu kõnelust, enne kui klient teeb kursuste kohta otsuse. Otsus võib olla nii positiivne (s.t tellida mõni kursus) kui ka negatiivne (s.t lükata tagasi konsultandi ettepanekud). Suur osa kõnedest lõpeb siiski kokkuleppega hoida kontakti, mis tähendab, et kõne jooksul ei jõua klient küll otsusele, aga on valmis vestlema samal teemal edaspidi.

Müügivestlus nagu iga teinegi dialoog kujutab endast nn transaktsiooni, milles esinevad sellised hierarhilised tasemed nagu vooruvahetused, mis omakorda koosnevad voorudest ja DA-dest. Müügivestluse üldine lineaarne struktuur (s.o vestluse kui voorude ja lausungite järjendi liigendus osadeks) on sarnane teiste ametlike telefonikõnedega omaga. Kõne algab rituaalse osaga, kus osalejad tervitavad ja tutvustavad ennast. Sellele järgneb põhiosa, milles toimuvad läbirääkimised. Kõne lõpeb jälle rituaalse osaga, kus vestlejad tänavad ja jätavad hüvasti.

Konsultandid on läbinud enne läbirääkimiste alustamist koolituse, kus on kasutatud firmas väljatöötatud materjale. Läbirääkimiste protsessis nähakse ette järgmised faasid: 1) ettevalmistus, 2) müügiläbirääkimiste avamine, 3) kliendi kaardistamine, 4) argumenteerimine, 5) otsuse saavutamine, 6) jätkutegevus (Vellerind 1998, Rääbis, Vellerind 2000).

Esimene, ettevalmistav faas leiab aset enne esimest telefonikõnet potentsiaalsetele kliendile. Konsultant püüab hankida võimalikult palju infot kliendi kohta: asutuse taust, finantsolukord, töötajate arv jms. Viimane faas (jätkutegevus) toimub pärast seda, kui klient on juba osalenud treeningkursusel: siis helistab konsultant kliendile uuesti ja uurib tema hinnangut kursuse kohta. Võimalusel püüab ta äratada kliendi huvi teistegi kursuste vastu. Siin kasutatud korpuses ei leidu selliseid telefonikõnesid, mis käsitleksid jätkutegevusi. Seega saab dialoogides tuvastada faase 2 kuni 5.

Avamisfaasis esitleb konsultant ennast ja oma firmat. Ta püüab luua positiivset atmosfääri, motiveerida klienti suhtlema, äratada/suurendada tema huvi koolituskursuste vastu.

Kliendi kaardistamise faasis esitab konsultant rea küsimusi, et koguda infot kliendi tegelike vajaduste kohta. Konsultant käitub väga viisakalt ja kannatlikult, laseb kliendil rääkida, talle vahele segamata.

Argumenteerimisfaasis esitab konsultant mõjukaid argumente, mis võtavad arvesse kliendi vajadusi. Kasutatav strateegia on järgmine. Kõigepealt määrab konsultant kliendi jaoks olulised tegurid ja toob need esile, et mõjutada klienti otsustama teatava kursuse kasuks. Kui klient nõustub nende teguritega, siis näitab ja tõestab konsultant, et pakutav kursus lahendab kliendi probleemi. Lõpuks peaks klient nõustuma konsultandi tõestuskäiguga.



Otsuse saavutamise faasis on konsultandi eesmärk, et klient jõuaks otsusele koolituskursuste suhtes. Otsus võib olla nii positiivne kui ka negatiivne, samuti võib otsustamise edasi lükata. Viimasel juhul lepivad vestlejad kokku järgmise kontakti aja.

### 3.2. Müügivestluse lineaarse struktuuri tuvastamine

Nagu eespool mainitud, on käesolevas artiklis vaatluse all 9 müügivestlust. Üks vestlus jäeti edasisest analüüsist välja, sest selles selgus, et konsultant helistas valel numbril. Ülejäänud 8 kõnet saab jaotada kahte rühma: konsultant ja klient 1) suhtlevad esmakordselt (üks kõne), 2) on olnud kontaktis juba varem (seitse kõnet). Mõlemal juhul on tegu pikkade vestlustega, kõne keskmine pikkus on 93 vooru.

Uurimuse eesmärgiks on leida markerid müügivestluse erinevate faaside tuvastamiseks, et sõnastada reeglid nende automaatseks määramiseks dialoogides, kus on märgendatud DA-d. Selgub, et üldiselt saab kõneluste struktuuri tuvastada DA-de naabruspaaride alusel.

#### 3.2.1. Avamine

Iga kõnelus algab rituaalse osaga, kus osalejad tervitavad teineteist (näide 2; A – konsultant, B – klient; kasutatud on vestlusanalüüsi transkriptsiooni; dialoogides on märgendatud DA-d).

- (2) A: te:re päevast. | RIE: TERVITUS |  
B: £ tere? £ | RIJ: TERVITUS |

Konsultant esitleb ennast ja koolitusfirmat (3).

- (3) A: siin on `Asta=Topelt (.) `Tiritammest. | RY: ESITLUS |  
((mikrofoni surin)) `koolitus ja konsultatsiooni`firmast. | IL: TÄPSUSTAMINE |  
B: jah? | VR: JÄTKAJA: NEUTRAALNE |

Konsultant identifitseerib vastuvõtja (4) ja/või avaldab soovi rääkida teatud isikuga, kasutades kas direktiivi või küsimust (5). Sellised DA-de järjendid on telefonivestlustes üsna tüüpilised (Kasterpalu 2006).

- (4) A: .h Minna: Patrak kuuleb. | KYE: VASTUST PAKKUV | | RIE: IDENTIFITSEERIMINE |  
B: £ j:ah? £ | KYJ: JAH | | RIJ: IDENTIFITSEERIMINE |
- (5) A: ja kas ma `saaksin Aigar `Juur- `Juur:seljaga rääkida. | KYE: SULETUD |  
| RIE: SOOV RÄÄKIDA |  
B: jaa | KYJ: JAH |  
ma `kuulen. | RIJ: NÕUSTUMINE |

Müügivestlustes võib esineda veel üks rituaalne tegevus – viitamine eelnevale kõnelusele. Konsultant meenutab kliendile eelmist telefonikõnet (6).

- (6) A: mt vestlesime teiega:: (...) mt tetsembri lõpul, mt | SEE: VÄIDE |  
 (.)  
 B: täiesti õige. | SEJ: NÕUSTUMINE |  
 A: jaa. (.) lubasin teile saata ka:: (.) m:aterjalid, mt | SEE: VÄIDE |  
 B: jaa. | SEJ: NÕUSTUMINE |

A esimene dialoogiakt SEE: VÄIDE pärast tervitusrituaale viitab eelmisele kõnelusele.

### 3.2.2. Kliendi vajaduste kaardistamine

Kaardistamise faasis esitab konsultant rea küsimusi, selleks et koguda infot kliendi vajaduste kohta (7, 8). Konsultant käitub äärmiselt viisakalt ja vaoshoitult, innustades klienti rohkem rääkima.

- (7) A: /---/  
 .h mis teil muidu `sellel aastal need `peamised `suunad on. | KYE: AVATUD |  
 (0.5)  
 B: £ no meil mi- `enne meil oli ainult ütleme `metallimüük aga nüüd meil juba siss=ee (.) ämn=ä `kauplus ja `jää ja `hulgimüük ee `tööriistad ja siss ee (0.5) see on nagu jälle noh `uus `suund meil. £ | KYJ: INFO ANDMINE |
- (8) A: /---/  
 (3.0) .mt ja `palju teil `on inimesi firmas. | KYE: AVATUD |  
 B: ehh he `hetkel (2.2) kuus. | KYJ: INFO ANDMINE |

Selles faasis leiab aset info jagamine, A esitab küsimusi ja B vastab, andes infot oma asutuse kohta.

### 3.2.3. Argumenteerimine

Argumenteerimisfaasis esitab konsultant võimalikult mõjukaid argumente koolituskursuste kasuks, võttes arvesse kliendi vajadusi. Argumendid ja vastuargumendid on tavaliselt vormistatud kui väited, mida suhtluspartner võib kas aktsepteerida või tagasi lükata (9).

- (9) A: /---/  
 (1.0) .hh sest loomu`likult et=ee `töökogemuste kaudu õpib ka: alati aga .hh a `sageli ongi just `see (0.5) mt ee `kursused pakuvad sellise `võimaluse kus saab siis `teiste .hh oma hh `ala `spetsia`listidega samuti `kokku=ja `rääkida nendest `ühistest prob`leemidest ja samas siis ka .hh ee `mõtteid ja `ideid ee hh ee=`Tiritamme poolt sinna `juurde. | SEE: VÄIDE |  
 (.)  
 B: £ `jah? | SEJ: NÕUSTUMINE |

Paljudel juhtudel toimub argumenteerimine siiski üksikaktide abil (YA: INFO ANDMINE). Konsultant esitab fakte oma koolitusfirma kohta, mis osutuvad argumentideks kursuste kasuks, ja klient annab tagasisidet (10).

(10) A: /---/

(0.5) .hh et=et `siis loomu`likult=et .hh ee `tegelikult=ee: see ee mil-  
lised `valdkonnad nüüd `Tiritamm `pakub on `ka: > .h me oleme just  
orien`teeritud sellistele `lühikestele. < | YA: INFO ANDMINE |

(0.5) [mt] | YA: MUU |

B: [ £ jah? £ ] | VR: JÄTKAJA: NEUTRAALNE |

A: inten`siiv`kursustele [et] kus saab siis võimalikult palju | YA: INFO  
ANDMINE |

B: [ £ jah, £ ] | VR: JÄTKAJA: NEUTRAALNE |

### 3.2.4. Otsusele jõudmine

Otsuse saavutamise faasis on konsultandi eesmärk, et klient teeks otsuse kursuste kohta. Otsus võib olla kas positiivne või negatiivne, kuid kõige sagedamini lükkub analüüsitud kõnedes kursuse tellimine edasi ja osalejad lepivad kokku suhelda hiljem. Konsultant kasutab avatud küsimust või pakkumist, mis sisaldab ajaväljendi (nt *millal, selle nädala lõpus, mõne aja pärast*, näited 11, 12).

(11) A: .hh aga: (.) `kudas te `arvate=et **millal**=ee saame `rahulikult ee `rää-  
kida nüüd hh ee `personali poole pealt. | KYE: AVATUD |

(12) A: mt ee ja siis `võtan juba=m mnee `**mõne aja pärast** uuesti `ühendust  
kui te olete saanud `tutvuda:. | DIE: PAKKUMINE |

B: £ j:ah. | DIJ: NÕUSTUMINE |

(.) võib `olla `jah. £ | DIJ: PIIRATUD NÕUSTUMINE |

Analüüsitud dialoogides on kliendi otsus ühel juhul negatiivne ja ülejäänud juhtudel lepitakse kokku suhelda sel teemal edaspidi.

### 3.2.5. Sulgemine

A lõpetab telefonikõne, soovides partnerile B kõike head (13), ning osalejad jäta-  
vasti (14). Selle osa automaatseks tuvastamiseks saab tugineda rituaalsetele  
DA-dele.

(13) A: /---/

.hh ja `seniks=ee `soovin teile mt (0.2) ee `kena kevadkuu `jätku=ja:  
nüüd on `ilmad läinud ka: (0.5) ilusamaks=et | RIE: SOOVIMINE |

B: £ \$ suur tänu? \$ £ heh | RIJ: TÄNAN |

(14) A: `jääme `kuulmiseni. | RIE: TERVITUS |

B: > kõike ead. < | RIJ: TERVITUS |

### 3.3. Parandussekventsides tuvastamine

Dialoogi põhiosas võivad esineda kõrvalekalded “pealiinist” (nn alamdialoogid). Üks alamdialoogide liik on parandused, mida vestlusanalüüs nimetab partneri algatatud parandusteks: üks osaleja algatab paranduse ja tema partner viib selle läbi. Märgeandatud dialoogides on sellistel lausungitel topeltmärgeandid: üks märgeandustase sisaldab spetsiaalseid akte paranduse algatamiseks (PPE: ÜMBERSÕNASTAMINE, PPE: ÜLEKÜSIMINE, PPE: MITTEMÕISTMINE), paranduse läbiviimiseks (PPJ: PARANDUS) ja (ainult mõnikord esinevaks) paranduse hindamiseks (üksikakt VR: PARANDUSE HINDAMINE). Nende märgeandite abil on parandussekventsse hõlbus tuvastada (15, 16).

- (15) A: /---/  
nii=et .hh (.) ainult siis `inna `kauplemine käib. | KYE: VASTUST PAKKUV |  
| PPE: ÜMBERSÕNASTAMINE |  
(.)  
B: `täpselt nii. | KYJ: JAH | | PPJ: PARANDUS |
- (16) A: `inseneritöö on teil kõige `pri[maarsem] | KYE: VASTUST PAKKUV | | PPE:  
ÜLEKÜSIMINE |  
B: [ `igal] `juhul. | KYJ: JAH | | PPJ: PARANDUS |  
A: jajaa. | VR: VASTUVÕTUTEADE: NEUTRAALNE | | VR: PARANDUSE HINDAMINE |

## 4. Arutelu

Korpuse analüüsi tulemusel võib väita, et müügivestluse erinevate faaside tuvastamise põhilisteks markeriteks saavad olla DA-de naabruspaarid.

Rituaalne algus ja rituaalne lõpp on tuvastatavad rituaalsete DA-de naabruspaaride ja üksikakti RY: ESITLEMINE abil vastavalt dialoogi alguses ja lõpus. Konsultandi A esimene seisukohavõtuakt SEE: VÄIDE vahetult pärast avamisfaasi viitab eelmisele kõnele samade osalejate vahel (juhul kui see on toimunud).

Kliendi vajaduste selgitamine leiab enamasti aset müügivestluste varases staadiumis. Konsultant esitab küsimusi (tüüpiliselt on need avatud küsimused) ja kogub sel viisil infot kliendi asutuse kohta. Küsimuste-vastuste ahel võib olla pikem kui üks naabruspaar, sõltuvalt sellest, kui kaugele on juba arenenud läbirääkimised eelmistes kõnelustes. Alusetapil esitab konsultant rohkem küsimusi, sest siis ta ei tea kliendi vajadustest veel eriti palju. Selle faasi lõpus teeb ta kliendile ettepaneku valida teatud kursus.

Müügivestluse tähtsaim osa on argumenteerimine. Analüüs näitab, et konsultandid käituvad väljapeetult ja piirduvad enamasti info andmisega koolitusfirma kohta. Siiski võib ka sellisel suhtelisel neutraalsel viisil edastatud infot vaadelda kui argumenteerimist teatavate kursuste kasuks. See osa ei sisalda tüüpiliselt DA-de naabruspaare, vaid konsultant annab infot üksikaktide YA: INFO ANDMINE abil ja klient vastab kas vastuvõtuteate või jätkajaga. Mõnel juhul on konsultandi argumendid esitatud kui väited (SEE: VÄIDE), millele klient vastab kas nõustumise või mittenõustumisega (SEJ: NÕUSTUMINE, SEJ: MITTENÕUSTUMINE). Argumente saab eristada eelmisele vestlusele viitamisest nende asukoha alusel dialoogis. Võib täheldada, et analüüsitud dialoogides kliendid enamasti nõustuvad konsultandi väidetega, mis

tähendab, et need suudavad leida “õigeid” väiteid. Argumenteerimine on müügi-dialoogi kõige pikem faas.

Kliendi vajaduste uurimine ja koolituskursuste kasuks argumenteerimine võivad toimuda ka vaheldumisi. Ettepaneku tellida kursus teeb konsultant kliendile enamasti kaudsel viisil, vältimaks negatiivset reaktsiooni, nt *kas te oleksite huvitatud nendest kursustest*. Oodatud ja loomulik vastus on *jah*.

Tüüpiliselt ei võta klient otsustusfaasis vastu koolituskursuse ostuotsust, vaid osalejad lepivad kokku edasise suhtlemise osas. See pole üllatav, sest nagu öeldud, kuuluvad korpuses leiduvad dialoogid läbirääkimiste algusetappi. Analüüsitud dialoogidest lõpeb ainult üks kliendi resolootse keeldumisega nii kursustest kui ka edasisest kontaktist. See osa eelneb vahetult kõneluse sulgemisele ja algab tüüpiliselt konsultandi A avatud küsimusega tulevase kontakti kohta.

Pärast mõningate DA-de destilleerimist (Jönsson, Dahlbäck 2000) võib müügi-vestluse lineaarse struktuuri esitada järgmisel viisil.

#### [(1) Avamine]

A: RIE: TERVITUS B: RIJ: TERVITUS [A ja B tervitavad]

A: RY: ESITLEMINE [A esitleb ennast]

A: RIE: IDENTIFITSEERIMINE B: RIJ: IDENTIFITSEERIMINE / A: RIE: SOOV RÄÄKIDA B: RIJ: NÕUSTUMINE [A identifitseerib B]

A: SEE: VÄIDE B: SEJ: NÕUSTUMINE [A viitab eelmisele kontaktile (juhul kui see on toimunud)]

#### [(2) Kliendi vajaduste kaardistamine]

A: KYE: AVATUD B: KYJ: INFO ANDMINE [A esitab küsimusi ja B vastab (võib korduda)]

A: KYE: SULETUD / DIE: ETTEPANEK [A teeb B-le ettepaneku võtta kursus]

#### [(3) Argumenteerimine]

A: YA: INFO ANDMINE / A: SEE: VÄIDE [A annab infot või väidab midagi koolitusfirma kohta (võib korduda)]

#### [(4) Otsuse saavutamine]

A: KYE: AVATUD B: KYJ: INFO ANDMINE / A: DIE: PAKKUMINE B: DIJ: NÕUSTUMINE [A ja B lepivad kokku jätkata suhtlemist]

#### [(5) Sulgemine]

A: RIE: SOOVIMINE B: RIJ: TÄNAN

A: RIE: TERVITUS B: RIJ: TERVITUS [hüvastijätt]

Partneri algatatud parandused võivad esineda kõneluse põhiosa suvalises kohas (2. kuni 4. faasis). Parandusi võib algatada nii konsultant kui ka klient.

## 5. Kokkuvõte ja edasine töö

Artikli eesmärk on uurida dialoogi struktuuri, mille moodustavad mikrotaseme üksused – dialoogiaktid ja nende järjendid. Empiirilise materjalina kasutatakse ühte juhuliselt valitud osa müügivestluste korpusest (9 dialoogi). Suhtlejateks on koolitusfirma konsultant ja teise asutuse esindaja (nn klient), kelle kompetentsi kuulub oma asutuse töötajate täiendõpe. Konsultandi eesmärgiks on, et klient otsustaks tellida mõne koolitusfirma pakutavatest kursustest. Dialoogid on salvestatud

autentses situatsioonis ja litereeritud, kasutades vestlusanalüüsi transkriptsiooni. Tekstides on märgendatud dialoogiaktid, samas ei ole läbi viidud morfoloogilist ega süntaktilist analüüsi. Seetõttu otsitakse dialoogides DA-de järjendeid, mis markeerivad müügivestluse erinevaid faase: 1) avamine, kus konsultant esitleb ennast ja oma firmat, 2) kliendi vajaduste kaardistamine, mille käigus toimub küsimus-vastusdialoog konsultandi ja kliendi vahel, 3) argumenteerimine, kus konsultant esitab mitmesuguseid argumente oma firma kursuste kasuks, võttes seejuures arvesse kliendi tegelikke vajadusi, 4) otsuse saavutamine, 5) sulgemine. Dialoogide analüüsi tulemusel saab sõnastada reeglid müügivestluste erinevate faaside tuvastamiseks, võttes arvesse DA-de naabruspaare ja nende järjendeid ning positsiooni dialoogis. Tulevane töö seisneb automaatse pragmaatilise analüüsi tarkvara arendamises, mis võimaldaks tuvastada dialoogides nii dialoogiakte kui ka aktijärjendeid moodustuvat lineaarset struktuuri ja alamdialooge.

## Viidatud kirjandus

- Aller, Sven 2012. Dialoogiaktide tuvastamine Eesti dialoogikorpuses: ülevaade ressursidest ja tarkvaraarendus. Magistritöö. Tartu Ülikooli arvutiteaduse instituut. [http://comserv.cs.ut.ee/forms/ati\\_report/](http://comserv.cs.ut.ee/forms/ati_report/) (27.09.2012).
- Barton, Ellen 2004. Linguistic discourse analysis: How the language in texts works. – Charles Bazerman, Paul Prior (Eds.). What writing does and how it does it: An introduction to analyzing texts and textual practices. Mahwah N.J.: Lawrence Erlbaum Associates, 57–82.
- Gibbon, Dafydd; Mertins, Inge; Moore, Roger (Eds.) 2000. Handbook of Multimodal and Spoken Dialogue Systems: Resources, Terminology and Product Evaluation. Kluwer Academic Publishers.
- Hennoste, Tiit; Rääbis, Andriela 2004. Dialoogiaktid eesti infodialoogides: tüpologia ja analüüs. Tartu: TÜ Kirjastus. <http://dSPACE.utlib.ee/dSPACE/handle/10062/18995> (10.12.2012).
- Hennoste, Tiit; Gerassimenko, Olga; Kasterpalu, Riina; Koit, Mare; Rääbis, Andriela; Strandson, Krista 2009. Suulise eesti keele korpus ja inimese suhtlus arvutiga. – Eesti Raken- duslingvistika Ühingu aastaraamat, 5, 111–130. <http://dx.doi.org/10.5128/ERYa5.07>
- Hutchby, Ian; Wooffitt, Robin 1998. Conversation Analysis. Cambridge: Polity Press.
- Jönsson, Arne; Dahlbäck, Nils 2000. Distilling dialogues – A method using natural dialogue corpora for dialogue systems development. – Proceedings of the 6th Applied Natural Language Processing Conference, Seattle, 44–51.
- Kasterpalu, Riina 2006. Algusrituaalid müügikõnedes. – Mare Koit, Renate Pajusalu, Haldur Õim (toim.). Keel ja arvuti. Tartu Ülikooli üldkeeleteaduse õppetooli toimetised 6. Tartu: Tartu Ülikooli Kirjastus, 156–169.
- Lester, J.; Branting, K.; Mott, B. 2004. Conversational Agents. – Munindar P. Singh (Ed.). The Practical Handbook of Internet Computing. Chapman & Hall, 220–240.
- Rääbis, Andriela; Vellerind, Riina 2000. Müügiläbirääkimised kui institutsionaalne dialoog. – Tiit Hennoste (Toim.). Eesti keele allkeeled. Tartu Ülikooli eesti keele õppetooli toimetised 16. Tartu: Tartu Ülikooli Kirjastus, 148–173.
- Vellerind, Riina 1998. Suulise vestluse struktuur telefoni teel peetud müügiläbirääkimiste näitel. Bakalaureusetöö. Tartu Ülikooli üldkeeleteaduse õppetool.

**Mare Koit** (Tartu Ülikool) on uurinud dialoogi modelleerimist arvuti. [mare.koit@ut.ee](mailto:mare.koit@ut.ee)

# HOW TO RECOGNISE THE DIALOGUE STRUCTURE: CORPUS ANALYSIS

**Mare Koit**

University of Tartu

The aim of this study is to determine the dialogue structure formed by microlevel units – dialogue acts. The empirical material of the study is a small part (nine human-human dialogues) of a corpus consisting of telemarketing calls. In the calls communication takes place between business representatives on a win-win basis, which means that the caller (a sales clerk of an educational company) is offering a competence, supposedly needed by the receiver (a customer). Dialogues are recorded in authentic situations and then transliterated using the transcription of conversation analysis. Dialogue acts (DA) are annotated in the transcripts but neither syntactic nor semantic analysis is performed. The study investigates DA markers which would help to identify the phases of a telemarketing call: 1) the opening, where a sales clerk introduces himself and the educational company, 2) explanation of the needs of a customer where question-answer dialogue takes place between a clerk and a customer, 3) argumentation where several arguments for taking a course are presented by a clerk, taking into account the actual needs of the customer, 4) coming to a decision. Rules for recognition of the structure of telemarketing calls will be formulated on the basis of sequences of (adjacency pairs of) dialogue acts and their position in a dialogue. A further aim is to implement software for automatic pragmatic analysis of the calls in order to recognise dialogue acts, sub-dialogues, and the linear structure of dialogues.

**Keywords:** dialogue, telemarketing call, dialogue structure, dialogue act, Estonian





## DEVELOPING INTERVIEWER PROFICIENCY: A SELF-PERCEPTION SURVEY

Riin Kont-Kontson, Ene Alas, Suliko Liiv

**Abstract.** The article reports the findings of a survey among 440 teachers of English in Estonia who participate in the oral proficiency interview of the national examination in the English language as interviewers. Examining the results of the questionnaire statement by statement, the study reveals that interviewer behavioural patterns emerge during the interview, some of them threatening its validity. The understanding of interviewer competence is very diverse among the interviewers and the degree to which they adhere to the standards defined for the interviewers of national examination in the English language in Estonia is sometimes quite low. The interviewers are often unable to separate their role as an interviewer from being a teacher; they transfer their teaching behaviour to the interviewing situation. The teachers have difficulties with identification of appropriate accommodation strategies as well as controlling the extent of interaction between the interviewer and the assessor. The authors suggest a number of interviewer training strategies to combat the problems.\*

**Keywords:** validity, reliability, interviewer, interviewer training, oral proficiency interview

### 1. Introduction

In 2008, a substantial change was introduced in the oral proficiency interview (OPI) of the national examination in the English language in Estonia, the aim of which was to increase the level of standardisation of the process. The new format envisaged all interviewers using an interview script, which specified the amount of support the interviewer was obliged to give to the candidate during the process of the speaking test and limited the amount of improvisation allowed by the interlocutor. This change was undertaken in order to provide uniform conditions for all candidates during the OPI, thus increasing the validity of the test. In 2009, a survey was carried

---

\* This research has partly been supported by the Estonian Science Foundation Grant No 9037.

out involving 81 teachers who had all participated in the national examination as interviewers, to investigate their responses to the new format. The main finding of the study was a disparity among the interviewers in terms of understanding the interviewer's role during the OPI as well as their confidence level, their attitude to the interviewer script and their overall professionalism during the interviewing process (cf. Alas 2010: 109, for the results see Alas 2010: 75–109). A clear outcome of the study was a need for further interviewer training. As the 2009 survey was carried out barely a year after the new format was introduced, a new study was deemed necessary to make more informed decisions about training needs. Since the interviewers' perception of their own interviewing practices is an important part of those decisions, the current study aims to investigate the interviewers' self-evaluation. It also involves significantly more people (440) in order for the results to be more generalizable and reliable. Interviewer proficiency can be defined in terms of the interviewer's language proficiency, which may vary as the target group is comprised of non-native speakers with diverse foreign language learning experience and exposure to the English language. This aspect of interviewer proficiency was not investigated in the current study. Instead, interviewer proficiency was interpreted as the participants' ability to conduct OPIs in a systematic and efficient manner, adhering to the interviewer script and general guidelines for interviewer behaviour which all participants had been trained to do.

## 2. Materials and methods

The research was conducted using both quantitative and qualitative methods of analysis: descriptive, analytical, comparative and statistical methods of study were applied. The survey was compiled in co-operation with the Estonian National Examinations and Qualifications Centre (currently Foundation Innove) and was designed to focus on further training needs for OPI interviewers, i.e. the target group of investigation were the teachers who had already gone through obligatory initial interviewer training and had acted as interviewers during the OPIs of the national examination in the English language. The teachers were asked to respond to eight statements: five open-ended statements and three multiple choice items with an opportunity to add comments.

The first statement (*I think I am a competent interviewer because...*) is an open-ended question and was intended to disclose the respondent's understanding of what it entails to be a competent interviewer, and accounting for their own estimated level of professionalism. The second statement (*I think that proper interviewing atmosphere is... and I achieve this by...*), also an open-ended question, may be interpreted as an expectation to comment on either the macroclimate in the examination room over which teachers have little control, e.g. the classroom is too warm or there is too much noise coming from outside, which is often beyond the interviewer's control yet can have an effect on the candidates' performance (construct irrelevant variance). Alternatively, they could comment on the atmosphere created by relations between the interviewer and the student, i.e. microclimate – the level of friendliness and rapport, for instance. It was expected that respondents would be aware of both angles of the issue, yet it was still useful to find out if this was so and which of the two was more important.

Open-ended statements 3 (*During the interview, I struggle most with...*) and 4 (*What annoys me about interviewing is that...*) were more tied to the specific problem areas for interviewers and so as not to restrict the respondent, no options were proposed. Previous research had suggested a number of stumbling blocks: interviewer's own language ability (speed of speech, accent), the interview materials (the script, cue cards), time keeping and students' responses to prompts or follow-up questions (cf. Alas 2010). In addition to that, depending on the personality of the interviewer, they might be annoyed by students who do not say much or who ask questions, or by the procedure itself, because OPI is an unnatural situation to some extent – there is repetition of the same routine irrespective of the number of students or time of the day, which makes interviewing tedious (ibid.). Another aspect related to this question was whether interviewers were able to separate teaching from testing, because it is natural for a teacher to help the student in the classroom, but during the OPI the teacher, now an interviewer, has to behave differently. He/she is deprived of the right to assist with prompting questions unless specified by the script. Question number 7 (*During the interview I interact with the assessor in terms of...*) was about the interviewer proficiency and professionalism. The question investigated the extent to which the interviewers were aware of the norms of interaction with the examiner, and what format the co-operation took.

There were three multiple-choice statements in the interviewer questionnaire. With statement number 5 (*When the student does not understand me, I use the following accommodation strategies*), the interviewers had to circle either a 'yes' or a 'no' after the options. The accommodation options listed were some of the most common accommodation strategies listed in OPI research (cf. Ross, Bervick 1992, Lazarton 1996). The respondents could also add any other techniques if they wished. In statement 6 (*The script is...*), the respondents needed to comment on the quality and usefulness of the script (*helpful/unhelpful, clear/unclear, too detailed/detailed enough/not detailed enough, easy to use/confusing, other comments*). They had to select the appropriate ones from the given options and add qualities if necessary. The final statement, number 8 (*Useful formats for further interviewer training would be ...*), investigates the usefulness of particular interviewer training formats. A choice was given between *evaluating the interviews of other interviewers, listening to and evaluating the respondent's own interviews, getting comments about the respondent's interviews from a trainer, and getting comments about the respondent's interviews from a colleague*, with a request to specify any other.

In the final part of the questionnaire, the interviewers had to give some information about themselves; their workplace, length of service as an interviewer and as an English teacher, and their gender. Inserting one's name was optional in the hope that by remaining anonymous the respondents would be more honest and straightforward with their answers.

The results were analysed by first numbering the questionnaires, after which a Microsoft Excel file was created, where the results of each interview were entered. The results were processed, presented as descriptive statistics (including absolute frequencies and mean value) and summarised in tables and graphs.

### 3. Results

A total of 487 questionnaires were handed out to the English language teachers of Estonia, 485 responded to these questionnaires, and for the purpose of this research 440 answer sheets could be used. The completed questionnaires qualified as data if the respondent had been an interviewer for at least one year and had answered at least two questions. Forty-five questionnaires were discarded because they did not meet these criteria.

Analysing the respondents' experience as interviewers and teachers, it can be said that the longest service was by an interviewer who had worked as a teacher for 50 years, the shortest for one year, and the mean was 21 years. The longest service as an interviewer was as long as the OPI has been used in the English language national examination, i.e. 17 years, the shortest was one year, the mean was nine years. The majority of respondents were female non-native speakers. There were only 22 male respondents out of 440, and due to this imbalance the results were not analysed from a gender perspective. Also, many people gave more than one answer to the statements. This meant that one examiner could be found in two or more results for those statements, e.g. if a statement with four possible options has been answered by 440 people and each of them had marked every given option, then there were altogether 1760 responses, i.e. more than the number of the interviewers involved in the study.

Next, the survey results will be discussed statement by statement. For the sake of convenience, the statements are numbered and referred to in an abbreviated form Q1, Q2, Q3, etc. The results are expressed both in the form of figures and discussion.

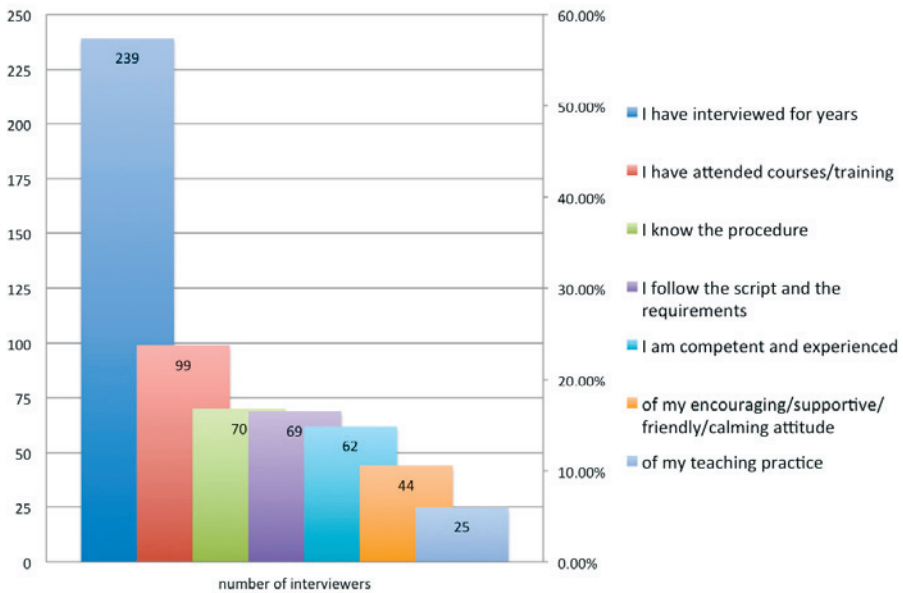
The information in Figures 1 to 5 and 11 should be read as follows: the vertical axis of the figure shows the potential number of participants, the horizontal axis shows the actual number of particular participant answers. The numbers on the columns show how many people gave the same answer. The percentage on the right margin shows the proportion of the participants. The comments on the right are the most popular answers of the current statement.

#### 3.1. Statement Q1. I think I am a competent interviewer because...

The responses given to the first statement are summarised in Figure 1.

Statement Q1 received responses from 433 teachers out of 440. The most common result, given by 55% of the teachers, is that they are competent interviewers because they have been interviewing for years. The next largest group (23%) considered attending training courses the reason for being a professional examiner; 16% see the key to competence in knowing the procedure, following the script and the exam requirements, and 14% in being competent and experienced, giving a rather circular answer, which is hard to interpret. A further 10% of the interviewers believe that their encouraging, supportive, friendly and calming attitude makes them competent interviewers, which is an important aspect but insufficient without procedural rigour. Just 25 teachers (6%) consider their teaching experience enough to be competent when interviewing. This implies that teachers sometimes equate teaching skills with testing skills.

## I think I am a competent interviewer because



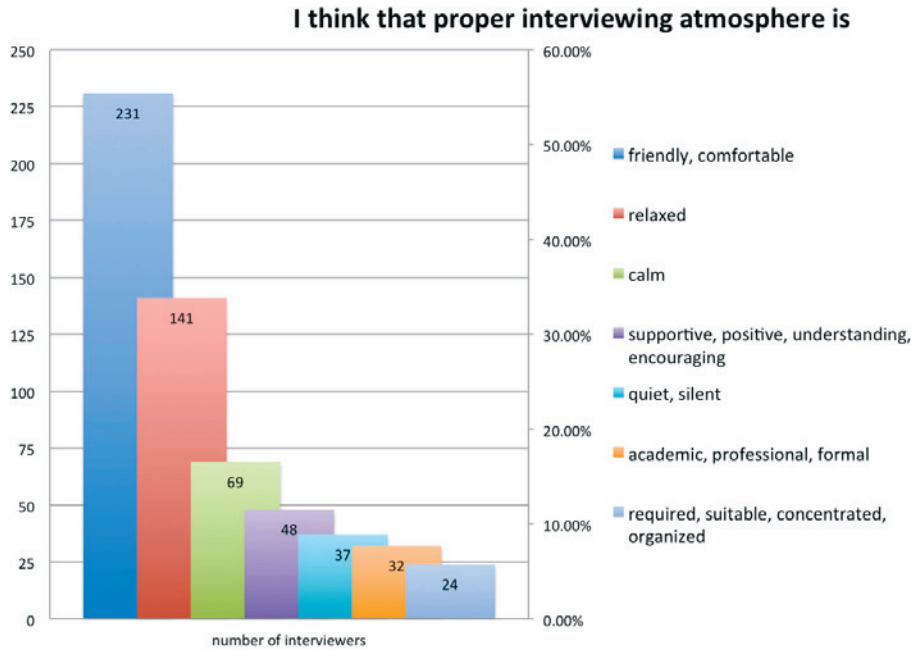
**Figure 1.** Responses to Q1

There are some other opinions given which illustrate the respondents' concept of interviewer competence, but are too few to be considered significant although they are quite revealing. Five teachers consider themselves competent interviewers because they analyse their mistakes afterwards and try to be as objective as possible treating everyone in the same way; one respondent thinks that he/she is competent because he/she has seen other teachers interviewing, and worked in the team who prepared the first speaking exam.

The survey reveals that if people have been interviewing for a long time, e.g. since the introduction of the OPI in the speaking part of the English language national exam and they know and follow the exam procedure, they believe that it qualifies them as competent interviewers. Both Brown (2003) and Fulcher (2010) seem to suggest that is only partly so: time does give experience, but there is more to language testing and to interviewing skills than spending time interviewing and knowing the theory. However many times one has interviewed, if it is not done meticulously and in a standardised manner, time alone is not going to guarantee that one would be a more competent interviewer. Brown says that the styles of the interviewers may become so diverse over time that they are leading to different outcomes for the students (2003: 17), therefore regular training is needed. But training alone is not enough – the knowledge that one gains through different courses and workshops should be actively practised in order to acquire the skills needed to become a more experienced and competent interviewer.

### 3.2. Statement Q2.1. I think that proper interviewing atmosphere is...

Figure 2 presents the results given to the first part of the second statement:



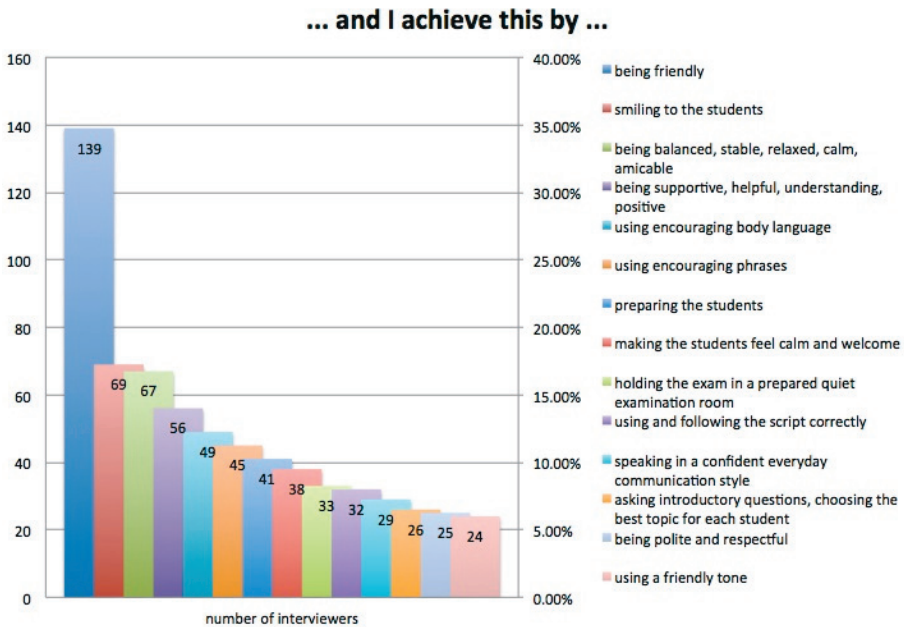
**Figure 2.** Responses to Q2.1

Q2.1 was answered by 415 respondents. Of these, 56% believe that the atmosphere has to be friendly and comfortable; 34% suggested that it should be relaxed and 17% that it should be calm. A further 12% consider a proper interviewing atmosphere is supportive, positive, understanding and encouraging; 9% quiet and silent; 7% find the proper interviewing atmosphere to be academic, professional and formal. The remaining 6% use fairly vague terms; required, suitable, organized and concentrated. The term ‘required’ is used to form a rather ambiguous response. It seems to suggest that the teachers are aware that there are requirements to the atmosphere, but they do not specify what they are.

The responses to Q2 seem to focus mainly on the micro-climate, rather than commenting on the atmosphere that is independent of them as interviewers. The outcome demonstrates the teachers’ preoccupation with providing an atmosphere where the student can relax and perform in an environment which is as stress-free as possible, so that they can display their best skill level. This is an admirable ambition. At the same time, not emphasising the seriousness of the OPI may lead to invalid results, because if the interviewer spends too much energy on creating a relaxed and positive environment it is possible that the students will not take the process seriously enough and not show their real proficiency levels. Nevertheless, there is still some justification in creating an atmosphere which is not over-formal and academic, which would result in the candidate becoming overly anxious.

### 3.3. Statement Q2.2. And I achieve this by...

The results given to the second part of the second statement can be seen in Figure 3.



**Figure 3.** Responses to Q2.2

This statement received 412 responses. About one-third (34%) think that to achieve the right testing atmosphere they need to act in a friendly manner and 17% mention smiling to the test-taker. Also, teachers emphasise their own role as models to achieve an anxiety-free atmosphere by saying that they have to be balanced, stable, relaxed, and calm (16%), and supportive, helpful, understanding and positive (14%) in order for the atmosphere to be appropriate for the OPI. Some teachers find it very important to use encouraging body language (12%) and encouraging phrases (11%), without specifying, however, what they are. Some of the interviewers achieve the appropriate atmosphere by preparing the students (10%), by making the student feel calm and welcome (9%), by having the examination in a quiet classroom, and using and following the script correctly (8%). Speaking in an everyday communication style is considered a key factor for achieving the right atmosphere by 7%, and choosing the best topic for each student, being polite and using a friendly tone by 6% of the interviewers.

The results of Q2.2 revealed a variety of different strategies ranging from different aspects of body language and adherence to the examination procedure to securing a proper macroclimate in the test venue. It seems that the interviewers concentrated mostly on their own role, although some of the comments – choosing the best topic for a particular candidate or finding a good contact with every test-taker – would need follow-up interviews with the particular respondent to find out what exactly was meant.

### 3.4. Statement Q3. During the interview, I struggle most with...

The answers given to the third statement are summarised in Figure 4 below.

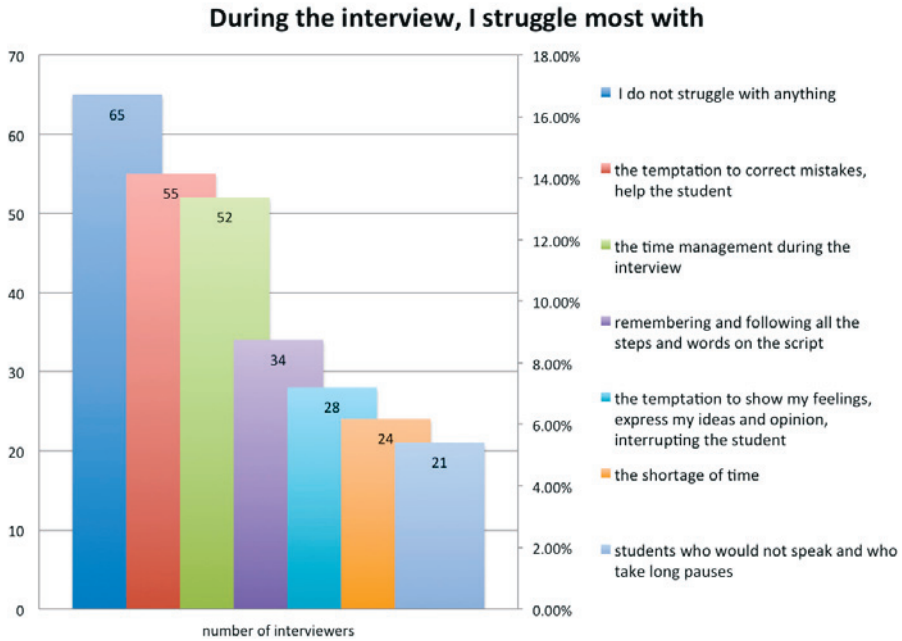


Figure 4. Responses to Q3

There were 396 responses to Q3. Of these, 16% claimed not to struggle with anything; 14% admitted to struggling with the temptation to help the student and correct their mistakes during the OPI; 13% noted trouble with time management during the interview, and 9% had difficulties with remembering and following all the steps and phrases in the script. Some of the interviewers fight with the temptation to show feelings (e.g. irritation), express ideas and opinions and interrupt the student (7%). Shortage of time (6%) and students who do not speak and take long pauses (5%) are also matters causing the interviewers to struggle.

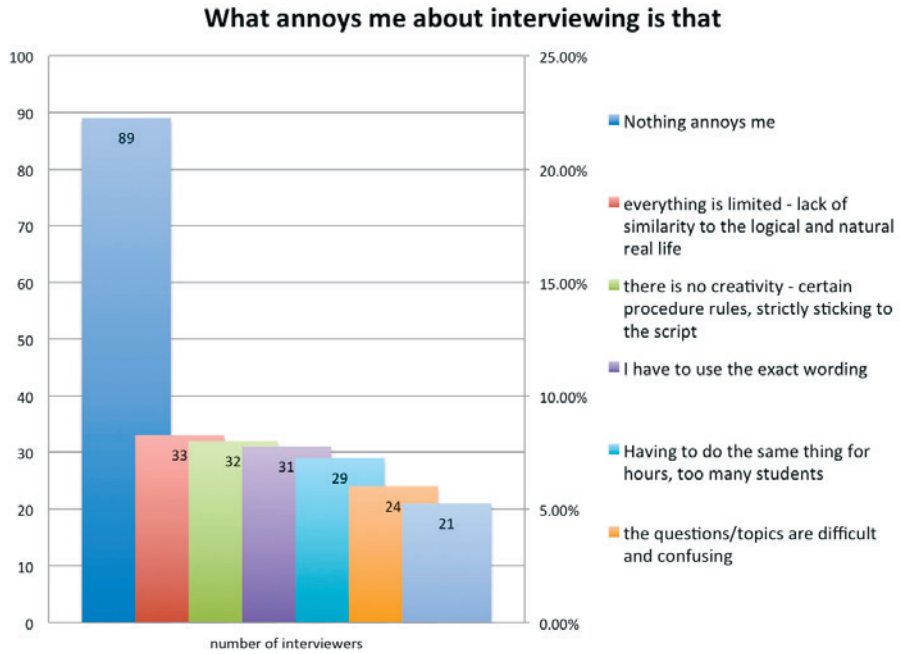
The above outcome seems to indicate that the teachers are aware of the difference between their roles in the classroom and in the examination room but they find separating the two sometimes quite hard. Time management and using the scripts also seem to need further practice.

### 3.5. Statement Q4. What annoys me about interviewing is that...

Q4 yielded 391 responses. Of these, 23% claim not to be annoyed by anything, 8% dislike the lack of similarity between the OPI and real-life conversation, as everything in the OPI is prescribed, and there is no creativity. Some interviewers say that having to go through the same procedure over and over again for hours is annoying (7%), especially when there are very many students to interview. Some



consider the questions and the topics too difficult and confusing, in some cases even too personal (6%). Some claim that they are annoyed by the fact that they cannot help the student in any way during the OPI (5%).



**Figure 5.** Responses to Q4

The results seem to indicate that although almost a quarter of the respondents are unperturbed by interviewing, but three-quarters of them find something that annoys them about it. Mainly, the respondents were dissatisfied with the script and the procedure itself, not so much with student behaviour or their own behaviour. It is important to sensitise the interviewers to the need to maintaining the examination routine from one interview to the next for the sake of the exam validity.

**3.6. Statement Q5. When the student does not understand me, I use the following accommodation strategies...**

Q5 was answered by 438 respondents. The majority (87%) of the interviewers said that they use repetition of the question or statement as an accommodation strategy when the students do not understand them. Less frequently used options are pronouncing the question more clearly (79%), repeating and slowing down the tempo (73%), giving a synonym for a difficult word (55%), and paraphrasing the question (54%). It is interesting that 12 interviewers out of 438 use translating into the candidate’s first language (L1) as an accommodation strategy (3%). This is noteworthy given that translation is an accommodation technique not allowed by the OPI rules. In order to retain fairness in the OPI procedure (Fulcher 2003: 79), accommodation strategies should be the focus of in-service training. The fact that

none of the 438 respondents could name any additional accommodation techniques in Q5 as requested makes it even more necessary.

### 3.7. Statement Q6. The script is...

Q6 concerned the script. The respondents had to circle particular characteristics – six pairs of opposites – and they could add other characteristics if they wished. There were 412 responses to Q6. The answers are summarised in the pie charts below, one characteristic at a time. Every slice shows how many respondents chose a particular answer. The first number shows the exact number of people who chose it, and the second indicates the percentage.

First, the interviewers had to decide if the script was helpful/unhelpful (see Fig. 6):

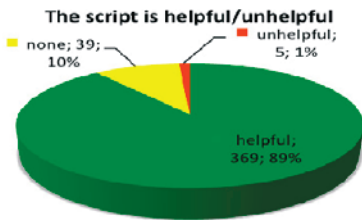


Figure 6. Responses to Q6 – statement a

An overwhelming majority (89%) found the script to be helpful, 1% find it to be unhelpful and 10% cannot decide. It is safe to say that a great majority of interviewers welcome the presence of the script probably because simplifies the OPI procedure.

The second characteristic was clarity (see Fig. 7):

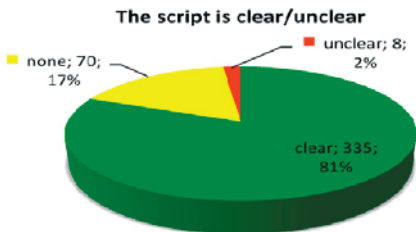
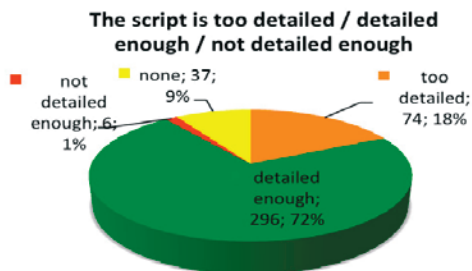


Figure 7. Responses to Q6 – statement b

Again an overwhelming majority (81%) report the script to be clear and 2% claim it to be unclear with 17% undecided. It becomes evident that the majority can understand the script and use it for its intended purpose, but it would be still interesting to know what the 2% find unclear and why 70 of the respondents could not make up their minds.

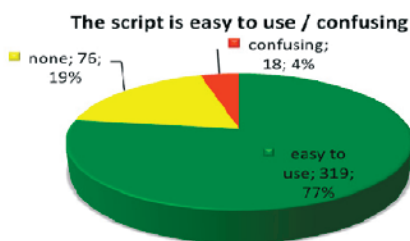
The third characteristic investigated is the degree of detail of the script (see Fig. 8):



**Figure 8.** Responses to Q6 – statement c

Nearly three-quarters of respondents to Q6c (72%) consider the script to be detailed enough, 1% do not find it detailed enough, 18% say that it is too detailed and 9% were undecided. One might speculate here that the issue of detail is perhaps connected to the proficiency and time management skills, because if an interviewer is struggling with keeping time, there may actually be too much information to grasp. It can also be connected to wishing for freedom of action during the OPI. The six people who see it as not detailed enough may need additional help interpreting and managing the script.

The fourth characteristic pertains to the ease of use (see Fig. 9):



**Figure 9.** Responses to Q6 – statement d

The majority (77%) find the script easy to use, 4% note that it is confusing and 19% of the interviewers have not responded to this statement. One can only speculate as to what the confusing aspects might be: script wording, script formatting, cue card content. Some light is shed in the final section of the statement where the respondents have provided comments: there have been grammar and logic mistakes on the teachers' cards, the script is too restricting for the interviewer, which means that the situation is unnatural, some topics are unsuitable for children who live in the countryside (e.g. various places for eating out). All the above suggest that practice is needed in working with the script. Another idea might be to recruit more assistance from teachers with item writing.

### 3.8. Statement Q7. During the interview I interact with the assessor in terms of...

The answers are summarised in Figure 10.

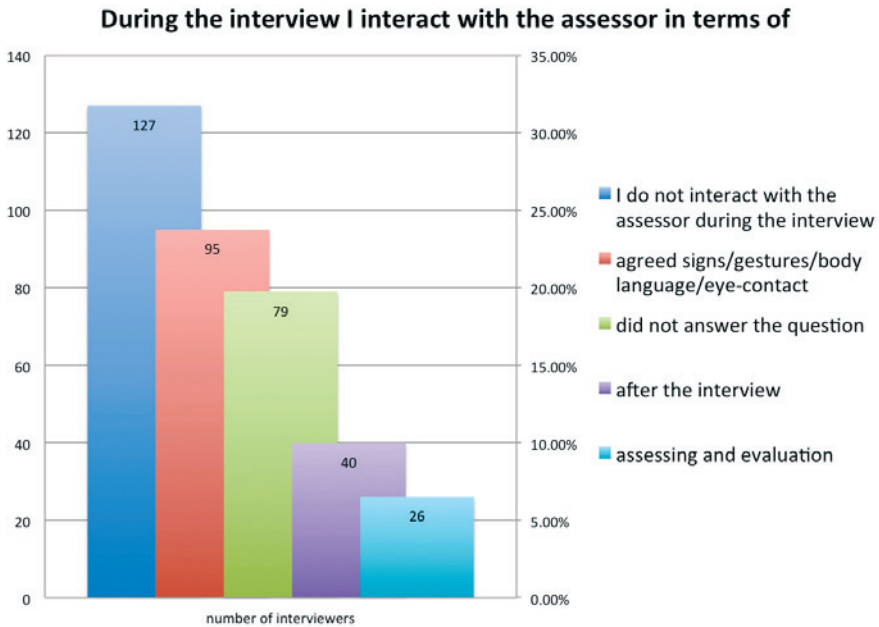


Figure 10. Responses to Q7

Perhaps one of the most intriguing sections of the questionnaire is Q7, requiring information about the interviewer-assessor relationship during the OPI. This is because the script of the OPI does not envisage the interviewer engaging in any discussion or talking to the assessor at all, but the answers to Q7 point to a different situation. Out of the 361 interviewers who answered Q7: 35% claim not to interact with the assessor during the interview; 26% say they interact using agreed signs, gestures, body language, and eye contact, without specifying the reasons; 7% simply report that they interact during assessment. This is a threat to the validity of the result of the OPI, because although the interviewer may know the proficiency level of the test-taker, his/her task as an interviewer detracts from the marking process. The interviewer has to ensure a smooth flow of the OPI, s/he has neither the marking scale nor the time to concentrate on evaluation. This is why there is a third party (assessor) added to the procedure of the OPI. A large number of interviewers (79; 18%) did not provide an answer to Q7. One can only speculate about the reason for this. If the respondents left it blank due to the fact that they do not interact with the assessor during the OPI, and they just did not feel it necessary to indicate it, then it would be a positive result. But if Q7 is not answered because they do not want to reveal that they interact with assessors, then therein lies a threat to the test validity as discussed above.

There were other serious deviations described involving interacting with the assessor. They were too few to be significant but are still interesting from the point of

view of the current study: eight interviewers said they gave the general impression to the assessor and discussed the student's mistakes; four reported giving background information about the student's proficiency in English. One interviewer asked the assessor if he or she would need more data to evaluate the student's performance and offered to prolong the interview, others interacted with the assessor by suggesting extra help that can be offered to the student. These strategies seriously undermine the validity of assessment and these interviewers should be eliminated from the process of measuring the students' oral proficiency. Further training is needed to focus on this problem fairly urgently.

### **3.9. Statement Q8. Useful formats for further interviewer training would be...**

This question asked the interviewers to point out useful formats from the given options and 433 did so. The vast majority (86%) consider evaluating the interviews of other interviewers to be the most useful format, testifying to their reluctance of being subjected to criticism themselves; 62% would like to get comments about their interview from colleagues; 53% would like to get feedback from a trainer. Less than half (44%) believe that a useful format for further training would be listening to and evaluating their own interviews, and 33% say they would not like that. The final part of Q8 asked the respondents to specify other useful formats, and it became evident that listening to, commenting on and evaluating sample videos would be the best training methods. The respondents also suggest sending training materials to schools and getting comments from the students. But also there are a few respondents who cannot understand why obligatory courses for interviewers take place so often, claiming that the teachers of the English language are already feeling over-trained. What the interviewers may fail to understand is that, as the field of language testing is fairly young, improvements are frequently offered and everyone connected to the field needs to be aware of the changes. Also, the effects of training are not very long-lasting and thus need refreshing (cf. Fulcher 2003: 260).

The data demonstrates a notable deviation among the interviewers regarding acceptable interviewer behaviour as well as expectations regarding further training. The following section will try to make some suggestions for the latter in order to reduce the variation.

## **4. Implications of the survey for interviewer training**

The analysis of the survey responses exposes a number of issues that interviewer training needs to address. First, the understanding of the concept of interviewer competence is very diverse among the teachers. In future training, it seems to be important to focus on the discussion of the essence of professionalism, what its different aspects are and how interviewers interpret them. Thus the training needs to give a theoretical background to interviewers' behaviour.

Next, interviewers appeared to struggle with a variety of aspects of OPI, especially with the temptation to assist students. This implies that the interviewers struggle to differentiate between their role as an OPI interviewer and that of a

language teacher. In terms of training, it is again necessary first to further explain the differences between these two roles, together with watching videos of good and bad interviewer practice. Checklists of critical points of behaviour could be implemented as a while-watching tool, which would make both the while-watching evaluation and also the post-watching discussion more focused. The discussion could then be followed by video recording the trainees' own behaviour during a simulation OPI with volunteer student participants and analysing it afterwards, utilising the same checklists. This could be arranged either as a peer-to-peer activity or as a feedback session from an interviewer trainer.

A sub-problem, frequently encountered in Estonia in the context of testing English language speaking skills, related to the urge to help students. This arises when the student does not understand the interviewer and is silent or gives too brief an answer. Here, the interviewer has to master the strategies available to him/her that can be used to provoke a student to say more, e.g. making sure s/he allows for a 5-second silence and does not 'jump in' too fast; asking the student explicitly if s/he has anything else to say, etc.). If the silence on the part of the student continues, the interviewer will have to refrain from attempts at further assistance (which can be quite frustrating) and move on to the next question/task.

A problem that needs continuous attention is the respondents' dissatisfaction and struggle with the script and the interviewer's cue card and the interviewing procedure itself, more precisely its rigid nature. In training, the interviewers may need to be reminded of the reasons why the OPI uses a standard procedure, and why deviations pose a threat to validity. The script would need a more thorough explanation from the point of view of how it functions and what may happen if the interviewees receive different treatment, i.e. when there are deviations from the script. This means in-service training in the principles of language testing. Understanding of the mechanism of cue card functioning can be trained by involving teachers in item writing, which in the Estonian context would mean developing statements for the monologue as well as role plays.

Responses to Q5 suggest that it is necessary to build the interviewers' awareness about accommodation strategies – to explain to them what accommodation strategies are and what kind of variety there is for the teachers to use. To begin with, it is important to give a theoretical background to accommodating behaviours and explain what types of accommodating behaviours people usually engage in interaction. It is also necessary to discuss accommodation in a teaching and a testing situation. There are particular accommodation techniques that the OPI procedural rules do not allow. For example, slowing down the tempo, reiteration and paraphrase of a word are acceptable techniques, whereas translation into one's first language is not. As the OPI is attempting to replicate a foreign language interaction situation in a foreign setting, where it would be fairly unlikely to find a person who could translate the unknown word/expression into Estonian, the interlocutors should be able to use other accommodation techniques. So when an interviewer uses translation, s/he might be giving certain students an advantage, which is not fair to other students who are taking the same examination with a different teacher who does not use translation as an accommodation technique.

A very common concern was time management problems during the OPI. Differences in the amount of time given for preparation for tasks as well as task

performance will affect the comparability of the exam results and consequently exam reliability/validity. The responsibility for training should be assumed by the teachers themselves. Practicing interviewing outside examination time in the classroom gives teachers the possibility to improve their skill of interviewing. It would also give the students necessary experience regarding the time available for task preparation and task delivery on the examination day. Recording one's own interviews and analysing them can give a better idea of one's behaviour in this respect during the OPI as well.

It would be beneficial to go through the requirements set for the interviewers (a document provided by the examining body), especially where the dos and don'ts have been explained in detail. The teachers might benefit from being aware of the fact that studies by Morton et al. (1997) and McNamara and Lumley (1997) both found that the assessors inflate their rating of the candidate when they perceived the interviewer to be less than competent during the interview, i.e. the interviewer's behaviour has an impact on the reliability/validity of the exam results.

Teachers need feedback on their own behaviour as interviewers; without that, no improvement can take place in the quality of the OPI. As was pointed out above, feedback can be obtained at least in two different ways: the teachers could record their own interviews, not only those taking place during the examination session, but during examination preparation time as well. Another option would be arranging training courses in order for the recordings to be analysed by trainers. Getting an outsider's perspective of the conduct of interviews can be quite useful in revealing practices that the interviewers may have failed to notice themselves.

Also, recording of the OPI should be compulsory and not, as is the case hitherto, a choice for the student to make. It is certainly important from the point of view of a more valid assessment, but from the interviewers' point of view, it may have a disciplining influence because when the interviewers are aware that there is a chance that their work may be checked, they will be more meticulous and abide by the requirements more seriously. This will mean a more standardised approach towards the process and consequently a more valid exam.

## **5. Conclusion**

The aim of this paper was to study interviewers' perceptions of their roles during the management of the OPI with a view to deciding potential training needs for interviewers. In order to increase the validity of the OPI through improving interviewer behaviour, the current study, having pinpointed problematic areas that need to be dealt with in interviewer training, established general directions for further training. The main goals here should be awareness building and practice of various aspects of the OPI.

A considerable amount of work needs to be done in training the interviewers to understand what aspects of their conduct need improvement, and what they themselves can do in order to increase their professionalism as interviewers.

## 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.
- Brown, Annie 2003. Interviewer variation and the co-construction of speaking proficiency. – Language Testing, 20 (1), 1–25. <http://dx.doi.org/10.1191/0265532203lt2420a>
- Fulcher, Glenn 2003. Testing Second Language Speaking. London: Pearson.
- Fulcher, Glenn 2010. Practical Language Testing. London: Hodder Education.
- Lazarton, Anne 1996. Interlocutor support in oral proficiency interviews: The case of CASE. – Language Testing, 13 (2), 151–172. <http://dx.doi.org/10.1177/026553229601300202>
- McNamara, Tim; Lumley, Tom 1997. The effect of interlocutor and assessment mode variables in overseas assessment of speaking skills in occupational settings. – Language Testing, 14 (2), 140–156. <http://dx.doi.org/10.1177/026553229701400202>
- Morton, Janne; Wigglesworth, Gillian; Williams, Donna 1997. Approaches to the evaluation of the interviewer performance in oral interaction tests. – G. Brindley, G. Wigglesworth (Eds.). Access: Issues in English Language Test Design and Delivery. Sidney: National Centre for English Language Teaching and Research, 175–196.
- Ross, Steven; Berwick, Richard 1992. The discourse of accommodation in oral proficiency interviews. – Studies in Second Language Acquisition, 14 (2), 159–176. <http://dx.doi.org/10.1017/S0272263100010809>

**Riin Kont-Kontson** on tegevõpetaja ja inglise keele riigieksami intervjuerija.  
[riinkk@gmail.com](mailto:riinkk@gmail.com)

**Ene Alase** (Tallinna Ülikool) teadushuvid on keeletestimine, testide koostamine ja nende kvaliteedi hindamine, õpetajakoolitus, õppekirjanduse hindamine.  
[ene.alas@tlu.ee](mailto:ene.alas@tlu.ee)

**Suliko Liivi** (Tallinna Ülikool) uurimisvaldkonnad on kontrastiivuuringud, kultuuridevaheline suhtluspädevus, keelepoliitika, võõrkeelte õpetamise meetodika, testimine.  
[liiv@tlu.ee](mailto:liiv@tlu.ee)



# **INTERVJUEERIJAJA KUTSEMEISTERLIKKUST ARENDADES: ANKEETKÜSITLUSE TULEMUSTE ANALÜÜS**

**Riin Kont-Kontson, Ene Alas, Suliko Liiv**

Tallinna Ülikool

Uurimuse käigus küsitleti 440 inglise keele õpetajat. Ankeetküsitluse eesmärgiks oli välja selgitada suulist keelepädevustesti läbiviivate intervjuerijate sellised käitumismustrid, mis võiksid vähendada inglise keele riigieksami tulemuste valiidsust ning mille põhjal oleks võimalik intervjuerijate koolitusi planeerida. Uurimus näitas, et intervjuerijate arusaam oma kutsestandardist varieerub märkimisväärselt. Intervjuerijad ei suuda sageli eristada enda kui õpetaja ja enda kui intervjuerija rolle, ei valda sobivaid kohandumustehnikaid ega tunne intervjuerija ja hindaja vahelise suhtlemise protokollid. Nende probleemide lahendamiseks esitatakse artiklis mitmeid vajalikke koolitusstrateegiaid.

**Võtmesõnad:** testimine, valiidsus, intervjuerija kutsemeisterlikkus



# KEELE KESTLIKKUSE JA KAO KONSTRUEERIMINE POOLAVALIKES DISKURSUSTES

Kadri Koreinik, Kristiina Praakli

**Ülevaade.** Artikli peaesmärk on analüüsida keele kestlikkuse ja kao diskursiivset konstrueerimist Saksamaa ja Soome eestlaste ning setu ja võru keele kasutajate poolstruktureeritud intervjuude näitel. Intervjuuandmestik pärineb koostööprojektist *European Language Diversity for All* (ELDIA), mille keskmes on väikese kõnelejaskonnaga soome-ugri keeled tänapäeva mitmekeelsetes ühiskondades. Saksamaal ja Soomes elavad eestlased esindavad allohtoonseid keelerühmi; setu- ja võrukeelsed moodustavad autohtoonsed Eesti keelekogukonnad, mille piiritlemine pole probleemitu. Kõigi nimetatud kogukondade puhul on tegu mitmekeelsete eestlastega, kes on suuremal või vähemal määral säilitanud oma vähemuskeele enamuskeelses, vastavalt saksa-, soome- ja eestikeelses keeleruumis. Diskursusanalüüsi kõrval vaadeldakse põgusalt, kuidas hindavad oma keeleoskust erinevas vanuses inimesed. Kogu keelekestlikkuse ja -kao diskursus on vaadeldav legitimaatsioonidiskursusena, kus osalised argumenteerivad vähemuskeelte funktsioonide ja tuleviku üle. Ühe põhitulemusena võib esile tuua, et kõik vaadeldud keelekogukonnad püsivad keelelise või kultuurilise rikastamise diskursiivses raamistikus.

**Võtmesõnad:** vähemuskeel, enamuskeel, vahendajakeel, keelekogukonnad, diskursusuuringud, keeleideoloogia, keeleökoloogia, eesti keel, inglise keel

## 1. Sissejuhatus

Artiklil on kaks eesmärki. Peaesmärk on analüüsida keele kestlikkuse ja kao diskursiivset konstrueerimist Saksamaa ja Soome eestlaste ning setu ja võru keele kasutajate poolstruktureeritud intervjuudes. Teine eesmärk on tutvustada põgusalt rahvusvahelist valdkondadevahelist projekti ELDIA<sup>1</sup> ja selle raames saadud andmeid.

<sup>1</sup> Vt <http://www.eldia-project.org> (03.03.2013).

Keelelise mitmekesisuse põhjustena tuuakse sageli välja just imperialismi/ kolonialismi, kommunikatsiooni ja rände kasvu, identiteedi fragmenteerumist, vähemuskeelte (taas)sündi ning hariduse, ennekõike võõrkeelte õppe levikut (Myers-Scotton 2002, Cenoz jt 2011). Mõned neist nähtustest on vastutavad ka teise globaalse ilmingu, inglise keele leviku eest. *Global English* või *World Englishes* domineerimine on viinud sinnamaani, kus seni vähemuskeeltele kavandatud keelehoiumeetmete järele tuntakse vajadust ka riigikeelte, ennekõike väikekeelte puhul (Ehala 2010). Sageli kasutatakse analoogiat biosfääriga: kui keel ei suuda kohaneda tänapäeva muutuva keskkonnaga, võib teda pidada ohustatud keeleks (Maffi 2000, Mufwene 2008). Globaalsed muutused on teisendanud keelte ülesandeid nii riikide kui etnoste vahel ja sees, lisades suhtlusesse ja kontaktsituatsioonidesse uusi dimensioone, mis varem või hiljem panevad proovile keelte, eriti väikekeelte kestlikkuse. Nii puutub üha suurem osa inimesi kokku uute keeltega ja on sunnitud tegema keerulisi valikuid, näiteks kas hoida või vahetada keelt. Laste eest teevad paraku keelevelikuid lapsevanemad, kel on keeleülekandes võtmeroll (Fishman 2001). Keelte vastastikune toimimine, aga ka poliitika ja majanduse mõju keelekasutusele on nii sotsiolingvistide kui keeleökoloogia ja -poliitika uurijate järjest suureneva tähelepanu all. Viimased on võtnud kasutusele näiteks etnograafilisi ja ka diskursusanalüüsist tuntud meetodeid, avardades nii keelepoliitika horisonti (Johnson 2009, Hult 2010, McCarty 2011).

Artikli ülesanne ongi näidata, kuidas konstrueerivad keele kestlikkust ja kadu, s.t millisena näevad vähemus-, enamus- ja vahendajakeelte tähtsust ja tulevikku Saksamaal ja Soomes elavad eestlased ning setu ja võru keele kasutajad Eestis<sup>2</sup>. Artikkel on loogiline jätk autorite varasemale uurimistöölle, kus vaatluse all on olnud väliseesti keel, mitmekeelne keelekasutus, eesti keele kestlikkus Tampere eesti kogukonna näitel (Praagli 2007, 2009, 2013a, 2013b) ja keeleideoloogiad avalikus diskursuses (Koreinik 2011a).

## 2. Uuritavatest keelekogukondadest ELDIA raamistikus

ELDIA kõige laiem ning ühtlasi ambitsioonikam eesmärk on individuaalse ja sotsiaalse mitmekeelsuse ümberhindamine, ümbermõtestamine ja edendamine. Projekti keskmes on väikese kõnelejaskonnaga soome-ugri keeled tänapäeva mitmekeelsetes ühiskondades. Kaheksat Euroopa ülikooli siduva võrdlusuuringu peamine eesmärk on piirkonna- ja rahvuskeelte ning rahvusvaheliste vahendajakeelte – enamikel juhtudel inglise keele – vastastikuse toimimise, aktiivse keelekasutuse ja keelevelikute kirjeldamine ning analüüsimine Euroopa erinevates keelekeskkondades soome-ugri keelte kontaktsituatsioonide näitel. Nendeks keelteks on vepsa ja karjala keel Venemaal, ungari keel Austrias ja Sloveenias, meä ja rootsisoomi keel Rootsis, kveeni ja põhjasaami keel Norras, võru ja setu keel Eestis ning eesti keel Soomes ja Saksamaal. Uuritavad soome-ugri keeled kirjeldavad hästi väikekeelte globaalset seisundit, aga ka nende keelte kontaktsituatsioonide ja kõnelejate mitmekeelsust. Ühtlasi esindavad need erinevaid poliitilisi ja majanduslik-ühiskondlikke olusid. Uuritavate keelte hulgas on nii põliseid vähemuskeeli nagu karjala või vepsa

<sup>2</sup> ELDIA kontekstis on vähemuskeeled eesti keel Saksamaal ja Soomes, setu ja võru keel Eestis, enamuskeeled on saksa ja soome keel Saksamaal ja Soomes, eesti keel Eestis. Vahendajakeel (ingl *vehicular language*) on kõikides sihtmaades inglise keel.

keel kui ka alles kujunemisel olevaid, hilistekkelisi kogukondi, mida esindavad eesti kogukonnad Soomes ja Saksamaal.

Järgnevalt anname põgusa ülevaate uuritavatest keelekogukondadest. Kogukondade detailsema kirjelduse ning viited kirjandusele võib leida ELDIA uurimisraportitest (Koreinik 2013a, 2013b, Praakli 2013a, 2013b).

Saksamaal ja Soomes elavad eestlased esindavad allohtoonseid ehk väljaspool asukohamaa geograafilist ruumi kujunenud keelekogukondi. Kummagi riigi eestikeelsete kogukondade näol pole tegu põliste või pikkade traditsioonidega keelerühmadega, vaid eri põhjustel kujunenud n-ö hilistekkeliste sisserännanute rühmadega, kelle tekkelugu saab alguse Teise maailmasõja aastatest (eeskätt Saksamaa puhul) või Eesti Vabariigi taasiseseisvumisest. Eestikeelsed kogukonnad neis riikides erinevad nii kõnelejaskonna sotsiodemograafilise tausta, (võõr)keeleoskuse kui väljarände motiivide poolest.

Saksamaa Statistikaameti andmetel elas 2011. a lõpus riigis 4840 inimest, kelle päritoluriigiks on Eesti Vabariik.<sup>3</sup> Soome eestikeelse elanikkonna arv on paarikümne aastaga suurenenud paarikümnekordselt. Kui 1990. a elas Soomes 1394 eesti keelt emakeelena rääkijat, siis 2004. aastaks oli nende arv tõusnud juba 13784ni (vt Kyntäjä 1997, Laanekask 2006, Praakli 2009). Eestlaste eeskätt töömotiividest johtunud väljaränne Soome intensiivistus Euroopa Liidu idalaienemise järel. Soome Statistikaameti uuematel andmetel elab Soomes juba ligi 33076 emakeelena eesti keele kõnelejat.<sup>4</sup> Kui Soomes moodustavad emakeelena eesti keele kõnelejad venekeelse uusvähemuse järel suuruselt teise vähemusrühma, siis Saksamaa kirevas keeleruumis on eesti kogukond üks väiksematest.<sup>5</sup>

Enamik nii Soomes kui Saksamaal elavatest eestlastest on sisserännanute rühmadele tüüpiliselt koondunud suurlinnadesse või nende lähiümbrusesse (Soomes Helsingi, Vantaa, Espoo, Tampere, Turku; Saksamaal Berliin, Hamburg, München). Saksamaal domineerivad Eesti kodanike seas naised vanuses 25–40 eluaastat. Ka Soomes on naiste osakaal praegu veel suurem, kuid nooremates vanuserühmades (25–40) on ülekaalus mehed (Praakli 2013a, 2013b).

Setu ja võru keele kasutajaid võib pidada põlisteks Eesti keelekogukondadeks, kuid setude etniline määratlemine pole probleemitu. Mõlemal juhul on siiski tegu kakskeelsete eestlastega, kes eesti keele standardiseerumise edenedes ja ühiskeele tugevnedes on säilitanud oma kohapealse kõnepruugi ning suuremal või vähemal määral identifitseerivad ennast lokaalselt (Eichenbaum, Pajusalu 2001). Lõunaeesti murded on kõige paremini säilinud just Kagu-Eesti perifeerias (Pajusalu jt 1999), mida võib pidada Eesti rikkaimaks kultuuriperifeeriaks (Krikmann 2000).

Ajalooliselt on setud elanud Eesti kaguosas Eesti-Vene piiril, nende keel liigitatakse traditsiooniliselt Võru murde hulka (Pajusalu jt 2002, Pajusalu 2007). Uusajal on setude asuala rahvastik kõikunud; ränne tänapäeva Eesti alale ja alalt ning piirimuutused pole setude (identiteedi) kujunemisloos olnud sugugi väheolulised (Saar 2003). Kui Eestis ei peeta setusid vähemusrahvuseks, siis Vene Föderatsioonis on nad hiljuti arvatud Venemaa põliste väikerahvaste ühendnimistusse.<sup>6</sup> 2011. aasta rahva ja eluruumide loenduse andmed näitasid, et Setu murraku kasutajaid on üle 12500 (REL07). Seega peaks oluline osa setu keele oskajaid elama väljaspool Setumaad. Alates 1990datest on setud olnud poliitiliselt aktiivsed. Nn setu liikumise

<sup>3</sup> <http://de.statista.com/statistik/daten/studie/1221/umfrage/anzahl-der-auslaender-in-deutschland-nach-herkunftsland/> (06.07.2012).

<sup>4</sup> [www.stat.fi](http://www.stat.fi) (30.09.2012).

<sup>5</sup> [https://www.destatis.de/DE/PresseService/Presse/Pressemitteilungen/2011/09/PD11\\_355\\_122.html](https://www.destatis.de/DE/PresseService/Presse/Pressemitteilungen/2011/09/PD11_355_122.html) (06.07.2012).

<sup>6</sup> Pravitel'stvo Rossijskoj Federacii. Postanovlenie ot 24 marta 2000 g. N 255. O edinom perečne korenyh malocislennyh narodov Rossii. <http://www.minnac.ru/minnac/info/13884.html> (03.03.2013).

poliitilisi eesmärged – ühtset Setumaad Eesti Vabariigi koosseisus ja setu kultuuri kestlikkust ja arengut – on erinevad aktivistid erinevalt tähtsustanud (Eichenbaum 1998). Samas on setu aktivistide tegevus traditsioonilise identiteedi, kultuuri ja keele säilitamiseks panustanud hegemooniliste kultuuripraktikate konstrueerimisse (vt Annist 2009).

Ka võru keelt peetakse konventsionaalselt eesti keele murdeks, võrukeelseid ei positsioneeritud vähemusena ei kasutajad ise ega ka uurijad. Nii on keelekogukonna suuruse hindamine selgete ja konsensuslike definitsioonide puudumisel keeruline. Viimase rahvaloenduse andmetel on ennast Võru murde (sh Setu murraku) kasutajatena määratlenud üle 87000 eesti emakeelega püsielaniku. Võrokeste keelevahetuse kõrghetk jääb ilmselt 1960.–1980. aastatesse (Org jt 1994).

Alates 1980. aastate lõpust on üritatud keelevahetust pöörata: muuhulgas on võru ja setu aktivistid püüdnud oma keelekujudele keele staatuse saamiseks korraldada, kuid tulutult keeleseadust muuta. Muudatusettepanekute tagasilükkamise põhjuseid võib otsida eesti keele ohustatuse diskursusest (Koreinik 2011b), mille keskmes representeeritakse eesti keelt peamiselt vene keele, aga ka inglise keele tõttu ohustatuna.

### **3. Teoreetilis-metodoloogiline kontekst: keeleideoloogiad ja -ökoloogia ning diskursusanalüüs**

Käsitame keeleideoloogiat kui keelt ja keelepraktikaid hierarhiliselt positsioneerivate diskursiivsete praktikate tulemust (nt Silverstein 1979, Blackledge 2005). Keeleideoloogiad on keelepoliitikast ja keelepraktikatest makrotasandi sotsiolingvistikas varasemalt selgelt lahus hoitud (Ferguson 1977), teisalt on keeleideoloogiaid ja keelepraktikaid peetud ka keelepoliitika objektideks (Spolsky, Shohamy 2000). Milline on aga täpselt nende kolme nähtuse vahelise seose suund, on raske öelda, ent koos varjatud keelepoliitika mõiste kasutuselevõtu ja keelepoliitika etnograafilise uurimismeetodi levimisega on kõik kolm võrdselt olulised ja suureneva tähelepanu all. Thomas Ricento ja Nancy Hornbergeri (1996: 420) arvates on indiviidide igapäevane suhtlus nende viis väljendada, töötada välja, vaidlustada, tõlgendada ja analüüsida keelepoliitikat. Keelepoliitikaks peavad nad sedagi, millise keele indiviidid suhtluses valivad, kuidas suhtluspartneri poole pöörduetakse, milline on indiviidide kehakeel ja millest räägitakse. Ka keeleökoloogia uurimissuunana (mitte -metoodikana) analüüsib seda, mil moel keeled ühiskonnas positsioneeruvad. Vähem tuntud on keelepoliitika ökoloogiline analüüs, mille üks keskne küsimus on keelepoliitikate suhe indiviidide keelekasutuse kogemuste ja uskumustega keelte suhtes (Hornberger, Hult 2008).

Diskursusuuringutel (ingl *discourse studies*) on mitmeid suundi (Jørgensen, Phillips 2002). Artiklis oleme intervjuutulemuste analüüsis kasutanud tuttavamat, kriitilist diskursusanalüüsi (Fairclough 1989, van Dijk 2005, van Leeuwen, Wodak 1999). Viimane eristab diskursustes nelja tüüpi makrostrateegiaid: konstruktiivsed (*constructive*), alalhoidvad-õigustavad (*strategies of perpetuation and justification*), transformeerivad (*strategies of transformation*) ja destruktiivsed (*destructive*). Alalhoidvad-õigustavad strateegiad toetavad ja taastoodavad näiteks rahvusidentiteeti (de Cillia jt 1999). Kuigi tänapäevalgi on mitmekeelsus pigem reegel

kui erand, ei ole ligi pooled eurooplased aktiivselt mitmekeelsed, vaatamata koolis õpitud võõrkeeltele (üldjuhul inglise keel) (Cenoz jt 2011, Special Eurobarometer 386). Vähemuskeelte kasutajatele aga on inglise keel kolmas või neljas keel (Cenoz, Gorter 2005). Kuna keeltele omistatakse erinev tähendus, tunnetavad vähemus- ja väikekeelte kasutajad enamuskeelses keskkonnas, et nad peavad oma erinevust põhjendama või koguni õigustama. Seega on keelekestlikkuse ja -kao konstrueerimine intervjuudes vaadeldav (mitte küll spontaanse) legitimeerimiskursusena (vrd van Dijk 2005).

Kõige tavalisem intervjuude uurimise meetod on informantide poolt öeldu taandamine sisukategooriateks ja informatsiooni koondamine erinevate teemade alla (Abell, Myers 2008). Teine võimalus on näha intervjuudes ennekõike interaktsiooni osaliste, intervjuueerija ja informandi strateegilist vooruvahetust. Oleme siin valinud esimese tee, keskendudes rohkem sellele, mida intervjuudes öeldi, mitte niivõrd sellele, kuidas seda tehti. Poolavaliku diskursuse uurimine võimaldab jälgida erinevate, poliitikast, haridussüsteemist ja meediast pärinevate kontseptsioonide ja narratiivide rekontekstualiseerimist ja teisendamist, aga ka tähendusloomet laiemas mõttes (de Cillia jt 1999). Peale poolavaliku, fookusrühmade intervjuukorpusest lähtuva diskursuse oleme analüüsinud, kuidas hindavad oma keeleoskust erinevas vanuses inimesed. Keele kestlikkuse seisukohast on põlvkondlik keeleülekanne ülioluline. Vanuselised või koguni põlvkondlikud erinevused hinnangutes keeleoskusele viitavad kas tugevamale või nõrgemale keeleülekandele või koguni keelevahetusele ja -kaole. Küsitlusandmete foonil näitame, kuidas intervjuueeritavad seletavad keelekestlikkust ja -kadu, milliste teiste teemadega need teemad seostuvad.

## 4. Küsitluste ja intervjuude andmed

Kõigis neljas vaatluse all olevas keelekogukonnas toimusid välitööd üldjuhul samadel põhimõtetel ning sarnase metoodika alusel 2010. aasta kevadtalvel. Välitööde korraldamise põhimõtteid ja metoodikat on kirjeldatud kogukonnapõhistes raportites (Koreinik 2013a, 2013b, Praakli 2013a, 2013b).

Eestis Võru- ja Setumaal ning Soomes küsitleti rahvastikuregistrist tellitud esinduslikku valimisse juhuvaliku teel sattunud elanikke. Saksamaal raskendasid küsitluse korraldamist mitmed tegurid nagu Saksamaa range andmekaitse seadus, mis välistas projekti uurijatel saada rahvastikuregistripõhine terviklik ülevaade eesti keelt emakeelena rääkivate inimeste arvu, nende vanuse, soo ning paiknemise kohta. Seetõttu kasutati nii küsitluses kui intervjuudes osalejate leidmiseks nn lumepallimeetodit: kohalike eesti seltside ja sotsiaalmeedia abi ning uurijate isiklikke kontakte. Hoolimata mittetõenäosusliku valimi kasutamisest, annab Saksamaa valim üsna hea läbilõike Saksamaa eestlastest. Kuigi küsitluse korraldamise viisides oli kogukondade lõikes erinevusi (Soomes ja Saksamaal teostati postküsitlus, Setu- ja Võrumaal kasutati küsitlejate abi), ei tohiks mõneti erineva metoodika abil saadud tulemustes suuri erinevusi olla (vt nt de Leeuw 2005). Eestis toimus küsitlus 1783.–1920. aastal eksisteerinud Võru maakonna ligikaudsetes piirides<sup>7</sup> ja neljas setu elanikkonnaga tänapäeva vallas<sup>8</sup>. Soomes läbiviidud küsitlus hõlmas kogu Soomet tervikuna; Saksamaa puhul oli metoodikast johtuvalt enamik informante pärit

<sup>7</sup> Tänapäeva Antsla, Haanja, Kanepi, Karula, Kõlleste, Laheda, Lasva, Meeksi, Mooste, Misso, Mõniste, Orava, Põlva, Räpina, Rõuge, Sõmerpalu, Taheva, Urvaste, Valgjärve, Vastseliina, Varstu, Veriora, Võru vald, Põlva linn ja Võru linn.

<sup>8</sup> Meremäel, Mikitamäel, Missos ja Värskas. Missos küsitleti endise Pankjavitsa/Rootova valla külade elanikke.

suurlinnadest ning eesti kogukondade keskustest. Saksamaal osales küsitluses 70, Soomes 171, potentsiaalsetest setu- ja võrukeelsetest vastavalt 294 ja 296 inimest.

Välitööde teiseks etapiks olid ühtsetel põhimõtetel läbiviidud temaatilised individuaal- ja fookusrühmaintervjuud. Kõikides uuritavates kogukondades toimusid intervjuud vanuserühmades 18–29, 30–49, 50–64 ning 65+ ning täiendavad intervjuud vähemusrühma aktivistide, enamusmeedia esindajate ning enamusrühma poliitikutega. Individuaalintervjuud koosnesid u 40, fookusrühmaintervjuud u 20 küsimusest.

Peatume artiklis detailsemalt nendel fookusrühma intervjuu küsimustel, mis on keelekestlikkuse ja -kao konstrueerimisel relevantseimad: kas vähemuskeelt on kaasaja ühiskonnas vaja; kes peaks vähemuskeele eest vastutama, ja miks; milline võiks olla vähemuskeele tulevik kümne aasta pärast?

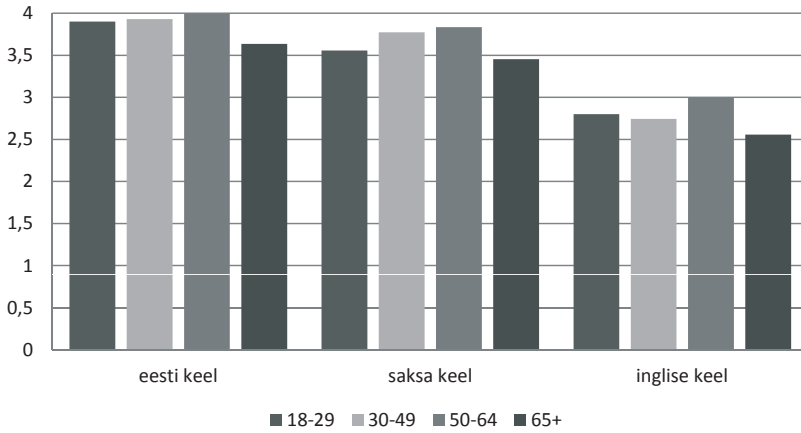
## 5. Tulemused

### 5.1. Küsitluste tulemused

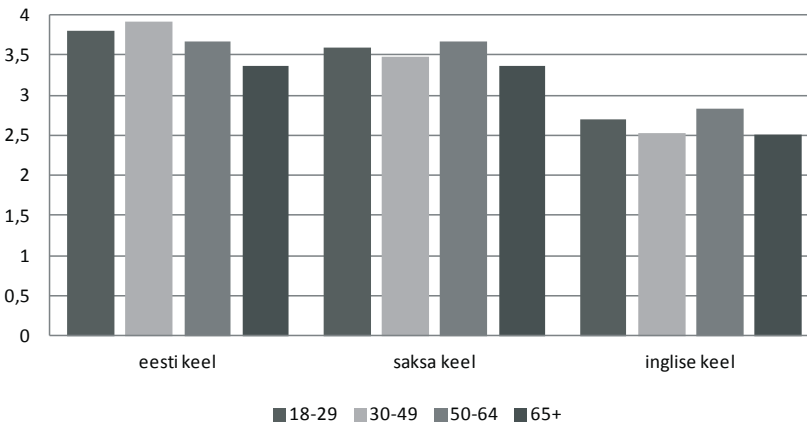
Keelekestlikkuse ja -kao diskursuse ilmestamiseks kasutame Saksamaa ja Soome eestlaste ning setu- ja võru keele kasutajate antud hinnanguid vastavate vähemus- ja enamuskeltele ning inglise keele kui vahendajakeele rääkimis- ja kirjutamisoskuste kohta (vt joonised 1–6). Rääkimis- ja kirjutamisoskust võib keele kestlikkuse ja kao seisukohast pidada olulisemaks keelekompetentsiks kui arusaamis- ja lugemisoskust, mida küsitluse käigus samuti hinnati. Keele osaoskusi mõõdeti viiepalliskaala (vabalt–hästi–mõnevõrra–halvasti–üldse mitte) abil. Respondentidel paluti märkida iga keele iga osaoskuse puhul, kuidas nad oma keeleoskust hindaksid. Selleks, et hinnata põlvkondlikke erinevusi osaoskuste lõikes, arvutati statistika tarkvara *Statistical Analysis Software* (SAS) abil aritmeetilised keskmised ja võrreldi neid eri valimite eri vanuserühmades. Dispersioonanalüüsi abil tehti kindlaks, kas vanuserühmade sisemine, s.t põlvkonnasisene osaoskuste hinnangute väärtuste hajuvus on seletatav valimi, s.t Saksamaa ja Soome eestlaste, võru ja setu keele kasutajate vastuste väärtuste üldise hajuvusega või on vanuserühmade vahel statistiliselt olulised ( $p = 0,05$ ) erinevused. Joonistel oleme esitanud nelja vanuserühma hinnangute aritmeetilised keskmised. Erinevad üldkogumeid (ja valimeid) statistiliselt ei võrreldud.

Joonistelt 1–6 on näha, et eesti keele rääkimisoskust hinnatakse sihtrühmades enam-vähem sarnaselt. Ootuspäraselt hinnatakse paremaks suulist oskust. Saksamaa eestlastel puudusid põlvkondlikud erinevused mõlemas inglise keele osaoskuses. Soome eestlastel on olulised erinevused inglise keele rääkimises ja kirjutamises. Ühelt poolt on ilmselt selle põhjuseks sihtrühmade erinevad sotsiaalmajanduslikud näitajad (elukoht, vanus, sugu, haridus, sissetulek) ja küllap ka väljarände motiivid. Teisalt peab silmas pidama ka võorkeeleõppe poliitikat Eestis. Kui näiteks soome keele aktiivse oskuse osakaal on Eesti elanikkonna hulgas kasvanud, siis saksa keele oskajate osakaal on alates 1934. a kuni millenniumivahetuseni püsinud stabiilne (Tender 2010). Võib oletada, et Saksamaale on välja rännanud pigem hea saksa ja inglise keele oskusega inimesed, Soome on mindud tööplaanidega ja ametitesse, kus vahendajakeelt ei nõuta. Noorema vanuserühma inglise keele oskuse hinnangud on üsna sarnased.

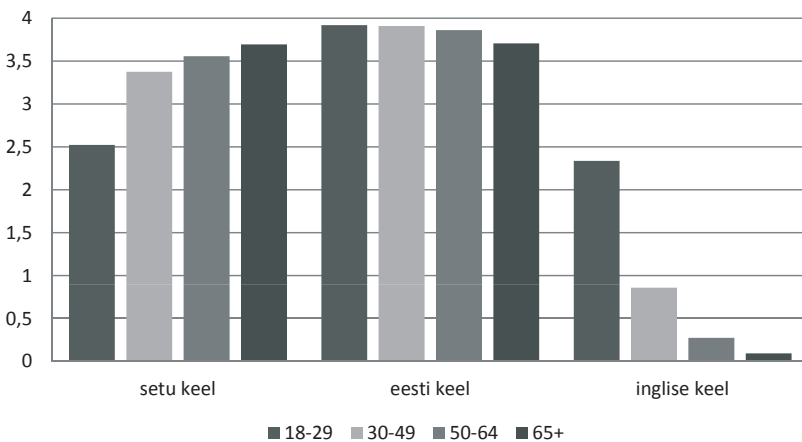




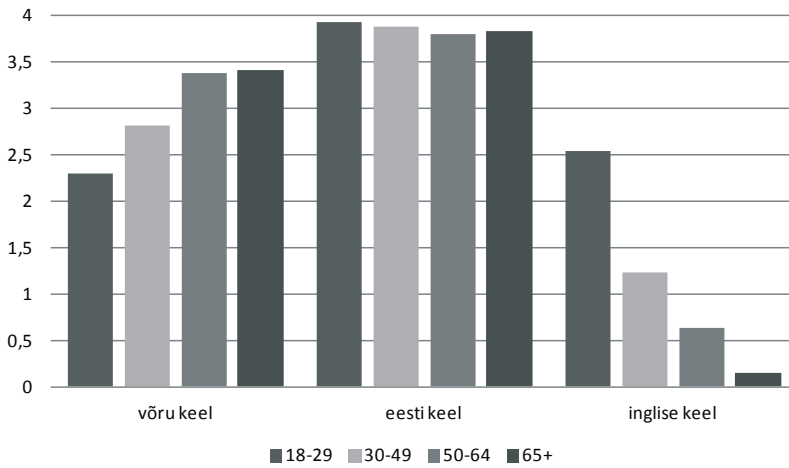
**Joonis 1.** Saksamaa eestlaste subjektiivne hinnang rääkimisoskuse kohta



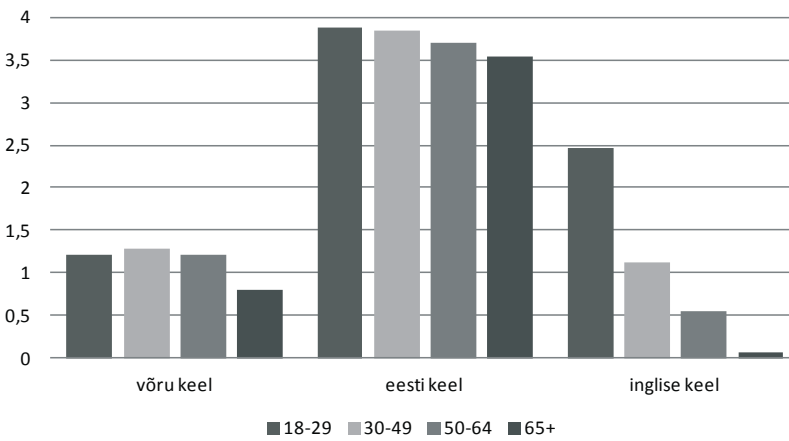
**Joonis 2.** Soome eestlaste subjektiivne hinnang rääkimisoskuse kohta



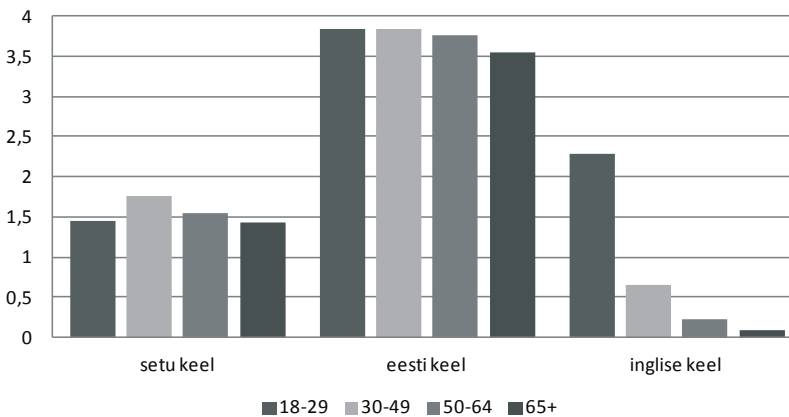
**Joonis 3.** Setu keele kasutajate subjektiivne hinnang rääkimisoskuse kohta



**Joonis 4.** Võru keele kasutajate subjektiivne hinnang rääkimisoskuse kohta



**Joonis 5.** Võru keele kasutajate subjektiivne hinnang kirjutamisoskuse kohta



**Joonis 6.** Setu keele kasutajate subjektiivne hinnang kirjutamisoskuse kohta

Setu ja võru keele rääkimisoskus on põlvkonniti erinev – noortel kehvem ja vanematel parem. Eesti keel on domineeriv keel (vt nt joonis 3 ja 4). Kirjutamisoskuses põlvkondlikke erinevusi pole, see osaoskus on ühtviisi kehv kõigil. Põhjusi võib otsida nii vähemus- kui enamuskeele funktsionaalses (diglossaalses?) eristamises, aga ühtlasi ka setu ja võru keele standardi piiratud levikus ning paraku ka kirja- viisivaidlustes. Ka inglise keele osaoskustes on selged põlvkondlikud erinevused: kõige noorem võru keele kasutajate grupp hindab inglise keele rääkimisoskust kõrgemalt kui võru keele oma.

## 5.2. Intervjuude tulemused

Avalikus diskursuses on tihti peale teemaks olnud eesti keele ja kultuuri tugevdamine ja rikastamine (Koreinik 2011a). Selliseid mõttekäike leidub sageli ka poolavalikes diskursustes. Vastustest küsimusele, kas vähemuskeelt on kaasaja ühiskonnas üldse vaja, aimub ratsionaliseerivaid, põhjust ja tagajärge demonstreerivaid argumente (ingl *rationalization*). Seesugune argumenteerimisstrateegia johtub ühiskonna enamuse heakskiidetud normidest.

Enamuskeelte (saksa, soome, eesti) suhtes vähemuskeele positsioonis olevate keelte (vastavalt eesti, setu ja võru keel), aga ka laiemalt mitmekeelsuse funktsiooni nähakse suuresti eesti keele ja kultuuri ning Eesti majanduse, riigi ja ühiskonna rikastamises:

- (1) --- noh ma arva et tuu tege eesti keelt kimmämbäs (VRO-AKTn)
- (2) kui inemine kasus niüteldä esi hindäst --- kiil om seotu identiteediga ja ku ta nigu saa-ai ennast avada --- vaba ja sääne loominguline inemine om ühiskonnale nigu (...) vajalik (SETO-AKTn)

Keele kestlikkust seotakse ka Eesti põlisasustuse ja/või -rahvastiku püsimisega. Mainitakse ka lojaalsust riigile ja Euroopa Liidu abi. Keelega seotud (kohalik) identiteet või keeleoskus võib osutada Eesti kui terviku jaoks funktsionaalseks:

- (3) jah tuu kiil hoit siin et sa ei lähe Austraaliahe (VRO-5064n)
- (4) --- et euroopa raha tulõ selleperäst et mi jääsi üks siiä paika är ei lääsi sinna kos [!] noh lääneeuroopa põhjaeuroopa et olömi siin visatas raha näpu-otsaga --- ja tantske sääl oma kaerajaani või lööge leelot jääge üks umma kodokanti seepärast et noh lääneeuroopa om rahvast täüs (SETO-3049m)

Veel üks märkimisväärne ja ühtlasi rikastamisega seotud teema on keelelise mitmekesisuse otsene sidumine keelte turuväärtuse ja kaubastumisega (ingl *commodification*). Vähemuskeel kui erinevus võib kanda sümbolset väärtust, kuid sellel võib olla ka instrumentaalne väärtus (5).

- (5) --- noh me tahame riigile ka lihtsalt kultuurilise kasu poolest (.) siis ka majanduslik kasu et mida rohkem omapära välja toome (.) seda suuremaks muutub ka turistide arv siia kanti --- (VRO-3049n)

Leidub ka teistsuguseid ratsionaliseerivaid argumente. Kollektiivi, ühiskonna ja/või riigi asemel on argumentatsiooni keskmes indiviid ning mitmekeelsuse

funktsionaalsus, aga ka vähemus- või vahendajakeel indiviidi seisukohast. Näide (7) illustreerib ühtlasi Eesti väiksuse diskursust.

- (6) --- see lihtsalt rikastab inimest (.) see on ikkagi nagu mõnes mõttes selline kahe tausta olemasolu kahe- (...) kultuuri (.) olemasolu (GER-3049n)
- (7) --- ma tean et kui need keeled ((inglise ja saksa)) on juba lapsest peast suus sa võid (.) maailmas nagu igal pool läbi lüüa et just see et eesti on nii väike et see ei ole lissalt praktiline selles suhtes (GER-1829n)
- (8) no mina mõtlen puht pragmaatiliselt --- see oleks nagu selles mõttes kingitus et see on nagu (.) üks aken veel mis on sinu ees lahti kui perfekt-selt sa seda sis valdad --- ma lähtun iseendast et (.) mina tahaksin olla mitmekeelne (GER-3049n)
- (9) aga piässi toda kuigi nigu soodustama või miilde tulõtama õks või vana-vanõmbitele ka et et et no kotun kasutagõ toda kiilt ja tuu rikastas latsõlatsi mitte ei tii näid ullikõsõst (SETO-5064m)

Küsimusele *Kes peaks üldse vastutama vähemuskeelte eest?* antud vastused jagunevad ootuspäraselt kaheks: ühed asetavad keelte kestlikkuse vastutuse kogukonnale, ühiskonnale ja/või riigile (11–18), teised näevad vähemuskeelt kasutavat indiviidi ennast vastutajana (10, 17–18).

- (10) mina isiklikult ütleks alati et see = n täitsa inimese (.) ((natuke naerdes)) sa = i saa sundida? inimest rääkima ema- emakeeles kui ta = i taha ((kui ei taha)) siis ärgu rääkigu (GER-65+n)
- (11) --- eks ikka eest- eestlased ise (FIN-1829m)
- (12) aga perele ma arvan peaks toeks olema siis see emariik (FIN-5064n)
- (13) --- riigi ülesanne peaks olema see ((setu)) et see ära ei kaoks et see ikka nagu jääks püsima siia (SETO-1829m)
- (14) no riik üleüldse üks vastutas uma elanike iist ja selle kultuuri iist mis tan om sis õks järelikult (.) kuna kuna mi olõmi ka üts osa sellest riigist (VRO-AKTn)
- (15) mina arvan et eesti (.) ee vata (.) kui eesti ise tunneb huvi (.) oma (.) aa nende alluvate vastu siis on ee ja ta (.) ta tunneb huvi ja ta tahab et ee kõigil eestlastel m- laias ilmas hea on siis ma arvan see keel iseenesest säilib (.) ma arvan et ee vata kui seal kus on ü- (.) hea soe ja mu- mugav (.) ja mõnus olla s- sinna tahad tagasi ja ma arvan k- keelega on täpselt samamoodi kui (.) eestis on hea ja mugav ja mõnus olla si = se keel tahab kõik sinna tagasi ja sis ta seal säilib kah (GER-3049m)
- (16) --- see on tõesti eesti rahva enda teha kas ta tahab jääda elujõuliseks (.) nii nagu ta on (.) aastasadu olnud või tahab ta lihtsalt alla minna [!] ja kaotada oma identiteedi --- (GER-65+m)
- (17) aga mi olõmigi mi olõmigi tuusama ütiskund (.) mi olõmigi tuu kes kes toda kes tuu vastutusõ omgi võtnu et mi kõnõlõmõ võru kiilt ja tuu et mu latsõlatse ja ma kõnõlõ nä kõnõlõsõ muga kirjakeelen a mina kõnõlõ

neidega [!] aja ikka [!] umma joru ja ja na saava väega häste arvu ja latsõ omma ka nakanu nüüd kolmekümne aastaselt omma nakanu nüüd rohkõmp võru kiilt pruukma (VRO-5064m)

- (18) eestlasi võib-olla siis juba (.) eesti poolt see eesti instituut kes sis organiseerib kes on selline nagu natuke ametlikum (...) organisatsioon kes võib-olla korraldab üritusi ja toob inimesed kokku aga (.) kui eestlased ise ei räägi sis ega eesti instituut ka sinu eest rääkima ei akka (FIN-1829n)

On ka neid informante, kes näevad keeleökoloogilisi protsesse, sh keele kestlikkust ja kadu kui iseeneslikku ja paratamatut protsessi, kus puuduvad toimijad või keelekasutajal puudub agentsus (vrd Koreinik 2011b):

- (19) tiiat külh et hulka pallo suurõmbit kiili om lännu nisamute (VRO-5064m)
- (20) --- ma arvan et see keel nagu (...) püsib ise siin ((Soomes)) ja ja nagu ise reguleerib end (FIN-1829m)
- (21) eriti siis kui elatakse väga kaua välismaal sa tunned ennast mõnikord ku- tud- kui sa tunned et teised keeled (...) saavad ikka (.) dominantsemaks aga sinu enda keel (...) lissalt (.) natukene (.) lissalt (.) saab ikka väiksemaks nii-öelda (GER-3049n)
- (22) --- ma arvan et kõige suurem nagu probleem ongi see et siis kui tuleb nagu erinevaid igasuguseid uusi keeli peale kas või mingi inglise ja vene noh mis iganes siis nagu kaovadki need murded ära ja ja võibolla lõpuks hajub see eesti keelgi ära kui siin kõik räägivad hoopis teist asja (SETO-1829n)

Keelte ja etnoste olemuslikku (ingl *essentialist*) sisu ja seost rõhutavad järgmised seletused (*explanation*). Informantide hinnangul tagab etnose püsimine ja vähe-  
musrahvuse kasv asukohariigis automaatselt ka keele kestlikkuse (23–25).

- (23) tuleb (.) ästi palju eestlasi teha (FIN-1829n)
- (24) --- ma arvan et ee eesti keel eks ta säilib ikkagi nii kaua (.) kui siin eestlasi on --- (GER-AKTn)
- (25) no nikavva neid ku om setu nikavva om ka kiil (SETO-65+n)

Ühtlasi seostatakse keelevahetust mobiilsuse, sotsiaalse hierarhia ja prestiižiga (26).

- (26) ku tuu üle pruuvi (.) nii perrä mõtõlda et määne tuu aig oll sääl määnegi pääle sõta kohkilgi vai et nuu kiä lätsi säält küläst vällä kohegi kavvõmbahe nuu ope määndsegi aameti (.) korssiva kuulsust ja ku nuu tulli küllä tagasi (.) sis näil oll tuu võõra keele mõotuse mano jäänü (.) ja noile kiä sis sääl iks lehmaperse takah sitta vissiva (.) nisast piimä tõmpsiva noile tundu tuu oi kae tuu om tark miis kae kus tä kõnõlös ja sis nu väiku tatika nuu sändse jah kaiva vot vaja perrä tetä (VRO-5064m)

Mis juhtub keelega kümne aasta pärast? Arvamusi on seinast seinä. Keele kestlikkust ja kadu põhjendatakse mitmesuguseid diskursiivseid skeeme, näiteks võrdlust (27–29) ja seletust (30–31) kasutades. Viimane (31) on ühtlasi näide keele ja etnose olemuslikust seostamisest.

- (27) mina arvan et rohkem aina ära kaob ja rohkem ja rohkem ee (.) ee lapsed hakkavad soomestuma --- siis sellel keelel mingit tulevikku siin ei ole (FIN-AKTn)
- (28) ja kui kaos ära tegelikult tuu põlvkond kes omma parhilla pensionääri et kui tuu põlvkond ära häos ja ja mi pääle kasume et kas mi oleme siis nagu võro kiilt kõnõleva vanaimä ja vanaisä [!] (VRO-1829n)
- (29) õõ (...) selle võrra näiteks et kui nüüd eestis kümne aasta pärast õõ majandus tõusma akkab et = sis näiteks (.) eestlased lähevad rohkem sinna tagasi ja ja jääb siia vähem näiteks õõ (FIN-1829m)
- (30) ta jääss viil aastakümnetest kindlästi seepäst et siin kiräsõnah om väega pallo seda ja ja taa avitas veidkene edesi venütädä ja tekis iks säändsit kes tahtva oppi toda ja ja ja edesi kanda ja nii et ta kandus edesi aga mitte suurõ ei suurõnõ (...) tsutt väiksembäst [!] iks jääs (SETO-5064m)
- (31) ((olen)) ennast eestlaseks pidanud ja soomlaseks samal ajal kolgend aastat laululaval laulnud = ja ma arvan et see ei kao kuskile --- (FIN-5064n)

Hoolimata sellest, et mitmed informandid näevad peamise vastutajana riiki, mainiti mitmel juhul põlvkondlikku keeleülekannet või selle puudumist (32–35).

- (32) --- väga palju on tegelikult inimesi ju (.) noh (.) tändab meie vanuse grupist kel on lapsed ja nad räägivad oma lastega (...) saksa keeles (GER-3049n)
- (33) --- mul klassisõsar tuu ütäl ka et et et noh om kõik seto lauluimä ja kõik mis tä om aga kotun ütles et ma kõnõlõ küll jah seto keelen aga na kõnõlõsõ mullõ õks eesti keelen vasta (.) latsõlatsõga (SETO-5064m)
- (34) et see (.) see ongi ju nagu see mida vanemad laps- (.) lastele edasi annavad et see on see et sa tead kus = sa tuled ja kus = sa pärit oled ja mis on nagu peres tähtis --- ma = i kujuta jälle ette et = et et kui on ee kaks erinevat ee kultuuri need saavad kokku ja siis akatakse elama ainult nagu ühe kultuuri traditsioonide järgi et see oleks (...) noh (.) selgelt teise (.) teise jah nagu maha surumine (GER-3049n)
- (35) latsõlõ või tulla tuu om rassõ asi ku lats ei kõnõlõ tedä ja sa panet talle määndsegi lisakoormuse sinnä otsa viil tolle võru keele --- (VRO-5064m)

Keele tulevikust rääkides peatuti inglise keele domineerimisel – seda mainisid mitmel korral Setu informandid. Muudest teemadest võib esile tõsta kaht: sooline aspekt põlvkondlikus keeleülekandes ning purism. Kui sooline aspekt tõstatub eeskätt Soome ja Saksamaa eestlaste diskursuses, kus keele peamise säilitajana nähakse naist/ema, ei saa ühe meeskeelejuhi (Saksamaa) hinnangul mööda vaadata ka meeste rollist peresiseste keelevalikute puhul.

## 6. Kokkuvõte

Keele kestlikkuse ja kao diskursuses peegelduvad ühiskonnas eksisteerivad keelega seotud representatsioonid, keeleideoloogiad. Keeleideoloogiad omavad suuremat tähtsust (ennekõike poliitilist) kui keelelised faktid (Cameron 2006). Keelelistest representatsioonidest moodustub nn hierarhia-diskursus (Hult 2010), kus erinevaid keeli ja keelekasutajaid positioneeritakse erinevat väärtust omavatena. Siinses uurimuses analüüsisime keelekestlikkuse ja -kao diskursust kvalitatiivselt, nii pole võimalik välja tuua keelekogukondade vahelisi arvulisi erinevusi. Siiski joonistuvad välja mõned tüüpilised teemad ja argumendid, mida kasutasid ühe keelekogukonna informandid, teised aga mitte.

Vähemuskeele funktsiooni nähakse peamiselt eesti keele ja kultuuri, aga ka ühiskonna, riigi ja majanduse rikastajana. Uuringus osalenud informantidest tõstsid viimase argumendi esile eeskätt Soomes elavad eestlased. Ka setu ja võru keele kasutajad nägid oma keelte rolli eesti keele ja kultuuri rikastamises, ent majanduslik argument ei jäänud mainimata neilgi. Nemad rõhutasid ka tugeva (keelelise) identiteediga seotud paiksuse väärtuslikkust Eesti riigi (asustuse, rahvastiku kestlikkus) ja eesti kultuuri seisukohast.

Kõnekaks erinevuseks Soome ja Saksamaa eestlaste vahel on mitmekeelsuse (ka vahendajakeele) roll indiviidi seisukohast. Küllap on see seletatav keelekogukonna kujunemise tausta ning keelekogukonna eripäradega laiemas mõttes. Kuigi uuringus osalenud Soome ja Saksamaa eestlasi võib nimetada esimese põlvkonna sisserännanuteks, kes esindavad mõlemas riigis suhteliselt uut ja noort keelelist vähemusrühma, on kogukonnad ELEDIA tulemuste põhjal sotsiodemograafiliselt vägagi erinevad. Saksamaa eestlased on ülekaalukalt kõrgharidusega naised, kel on hea ühiskondlik positsioon ning erialased eneseteostusvõimalused. Soome eestlaskond on tunduvalt heterogeensem. Saksamaa eestlasi iseloomustab selge mitmekeelne identiteet, kus eesti keele kõrval peetakse oluliseks vähemalt kahte keelt – uue asukohariigi enamuskeelt ehk saksa keelt ning inglise keelt kui vahendajakeelt (see nähtub ka informantide subjektiivsetest hinnangutest saksa ja inglise keele oskuse, kasutusala ja -sageduse kohta). Informantide hoiakus inglise keele suhtes võib näha teatavat pragmaatilisust: inglise keel on maailmakeel, mille valdamist peetakse enesestmõistetavaks, sest seda vajatakse etnostevahelises suhtluses ning tööülesannete täitmisel. Samas ei nähta inglise keelt võimaliku kodukeelena eesti keele asemel või ühe keelena eesti keele kõrval.

Kõigis keelekogukondades nähakse keele kestlikkuse eest vastutajana üksikindiviidi ja riiki. Soomes ja Saksamaal elavate eestlaste silmis on nende uue asukohariigi roll eesti keele kestlikkuse osas siiski marginaalne. Olukord on mõneti erinev Soome puhul, kus eestlaste jätkuvale intensiivsele töörandele viidates lasub keelejuhtide sõnul eesti keele kestlikkuse eest vastutamise kohustus ka Soome riigil. Üldistavalt öeldes: kui Eesti riigi rollis nähakse eeskätt moraalselt ning finantsiliselt toetust, siis Soomelt oodatakse toetust eestikeelsete haridusvõimaluste loomiseks.

Mis saab nendes kogukondades eesti, võru ja setu keelest lähitulevikus? Püüdes kokku võtta keelekogukondade kohta varem avaldatu ning küsitlus- ja intervjuutulemused, võib teha järgmised ettevaatlikud järeldused. Setu ja võru keelekogukonnad on jõudnud keelevahetuse lõppfaasi. Setu ja võru keele kirjutamisest on hinnatud madalalt kõigis vanuserühmades. Uus nn Võro-Seto standard pole kuidagi

suutnud eesti kirjakeele positsiooni kõigutada, mõjudes kirjutamisel ilmselt nii ideoloogilise kui tehnoloogilise barjäärina. Tõenäoliselt on standardi levikule oma jälje jätnud ka kirjaviisivaidlused. Säilinud on kõnekeele oskus, kuigi nooremad põlvkonnad annavad sellele madalama hinde. Keelekestlikkuse ja -kao diskursustes nähakse võru ja setu keele rolli tüüpiliselt eesti keele ja kultuuri rikastamises (vrd Koreinik 2011a). Üldjuhul ei omistata vähemuskeeltele utilitaarset turuväärtust, mistõttu nägid vaid üksikud keelejuhid neid indiviidi rikkuse, mitmekeelsuse osana. Keele kestlikkus või kadu on nii riigi/ühiskonna (organisatsioonide) kui keelekasutaja (perekonna) vastutusel. Keelekadu representeeritakse ka ilma toimijateta, s.t kui midagi iseeneslikku, sageli kui paratamatut (vrd Koreinik 2011b). Mainimata ei saa jätta puristlikke seisukohti, mis siin-seal esile tõusid (vrd Raag 1999).

Saksamaa ja Soome eestlaste puhul on keelevahetusest vara rääkida, kuna mõlema keelekogukonna näol on tegu kujunemisjärgus oleva rühmaga. Oluline on rõhutada, et nendegi argumentatsioon jääb keelelise ja kultuurilise rikastamise raamistikku. Tähelepanu väärivad ühtlasi keele ja etnose olemuslikke seoseid näitavad diskursiivsed praktikad. Nii näiteks püsib eesti keel Saksamaal ja Soomes seni, kuni kestab eestlaste väljaränne nimetatud riikidesse. Inglise keele võimalikus domineerimises eesti keele tulevikule ohtu ei nähta. Eestlaste keelelisi õigusi Saksamaal ja Soomes ei mainitud. Eesti keele toetamist nähakse Eesti riigi ja indiviidi enda kohustusena.

Toetudes Jan Blommaerti, Jan Collinsi ja Stef Slembroucki (2005: 213) mõttele, mille järgi pole mitmekeelsus indiviidide omadus või selle puudumine, vaid keskkond, mis struktureerib indiviidide keelelisi valikuid ja interaktsioone, võib väita, et keeleideoloogiatel on selle keskkonna loomisel oluline roll. See, kas ühiskond ja keelepoliitika toetavad mitmekeelsust ja millised keeled on igas ühiskonnas defineeritud mitmekeelsusena, määrab muude faktorite kõrval vähemuskeelte kestlikkuse.

## Lühendid ja sümbolid

---	väljajätt
(O)	täpsustavad kommentaarid
(.)	lühike paus
(...)	pikk paus
?	tõusev intonatsioon
[!]	keelenäidetele viitamine
1829	18–29-aastaste fookusrühm
3049	30–49-aastaste fookusrühm
5064	50–64-aastaste fookusrühm
65+	65-aastaste ja vanemate fookusrühm
AKT	aktivistid
FIN	Soome eestlased
GER	Saksamaa eestlased
m	meesinformant
n	naisinformant
saks-	pooleli jäänud sõna
SETU	setu keele kasutajad
VRO	võru keele kasutajad



## Viidatud kirjandus

- Abell, Jackie; Myers, Greg 2008. Analyzing research interviews. – Ruth Wodak, Michal Krzyżanowski (Eds.). *Qualitative Discourse Analysis in the Social Sciences*. New York: Palgrave, Macmillan.
- Annist, Aet 2009. Outsourcing culture: Establishing heritage hegemony by funding cultural life in South Eastern Estonia. – *Lietuvos etnologija – Socialines antropologija i etnologija studij*, 9, 117–138.
- Blackledge, Adrian 2005. *Discourse and Power in a Multilingual World*. Philadelphia: John Benjamins Publishing Company.
- Blommaert, Jan; Collins, Jan; Slembrouck, Stef 2005. Spaces of multilingualism. – *Language & Communication*, 25 (3), 197–216. <http://dx.doi.org/10.1016/j.langcom.2005.05.002>
- Cameron, Deborah 2006. Ideology and language. – *Journal of Political Ideologies*, 11 (2), 141–52. <http://dx.doi.org/10.1080/13569310600687916>
- Cenoz, Jasone; Gorter, Durk 2005. Trilingualism and minority languages in Europe. – *International Journal of the Sociology of Language*, 171, 1–5. <http://dx.doi.org/10.1515/ijsl.2005.2005.171.1>
- Cenoz, Jasone; Gorter, Durk; Heugh, Kathleen 2011. Linguistic diversity. – Dafne C. Reymen, Steven Knotter (Eds.). *Glossary on Diversity*. Sussex: Sussex Academic Press, 83–98.
- de Cillia, Robert; Reisigl, Martin; Wodak, Ruth 1999. The discursive construction of national identities. – *Discourse & Society*, 10 (2), 149–173. <http://dx.doi.org/10.1177/0957926599010002002>
- de Leeuw, Edith D. 2005. To mix or not to mix data collection modes in surveys. – *Journal of Official Statistics*, 21 (2), 233–255.
- Ehala, Martin 2010. Refining the notion of ethnolinguistic vitality. – *International Journal of Multilingualism*, 7 (4), 363–378. <http://dx.doi.org/10.1080/14790711003637094>
- Eichenbaum, Külli 1998. Ku kavvas Setomaalõ seto rahvast jakkus? – *Võro Instituudi toimetused/Võro Instituudi Toimõndusõq 2*. Võro: Võro Instituut, 11–15
- Eichenbaum, Külli; Pajusalu, Karl 2001. Setode ja võrokeste keelehoiakutest ja identiteedist. – *Keel ja Kirjandus*, 7, 483–489.
- Fairclough, Norman. 1989. *Language and Power*. London: Longman.
- Ferguson, Charles A. 1977. Sociolinguistic settings of language planning. – Joan Rubin, Bjorn H. Jernudd, Jyotirindra Das Gupta, Joshua A. Fishman, Charles A. Ferguson (Eds.). *Language Planning Processes*. The Hague: Mouton, 9–30.
- Fishman, Joshua A. 2001. Reversing Language Shift. *Theoretical and Empirical Foundations of Assistance to Threatened Languages*. Clevedon: Multilingual Matters.
- Hornberger, Nancy; Hult, Francis M. 2008. Ecological language education policy. – Bernard Spolsky, Francis M. Hult (Eds.). *Handbook of Educational Linguistics*. Malden, MA: Blackwell, 280–296.
- Hult, Francis M. 2010. Analysis of language policy discourses across the scales of space and time. – *International Journal of the Sociology of Language*, 202, 7–24. <http://dx.doi.org/10.1515/ijsl.2010.011>
- Johnson, David C. 2009. Ethnography of language policy. – *Language Policy*, 8 (2), 139–159. <http://dx.doi.org/10.1007/s10993-009-9136-9>
- Jørgensen, Marianne; Phillips, Louise 2002. *Discourse Analysis as Theory and Method*. London, Thousands Oaks, New Delhi: Sage Publications.
- Koreinik, Kadri 2011a. Language ideologies in the contemporary Estonian public discourse: With a focus on South Estonian. *Dissertationes sociologicae Universitatis Tartuensis* 5. Tartu: Tartu Ülikooli Kirjastus.
- Koreinik, Kadri 2011b. Agency lost in the discourse of language endangerment: Nominalization in discourse about South Estonian. – *Eesti Rakenduslingvistika Ühingu aastaraamat*, 7, 77–94. <http://dx.doi.org/10.5128/ERYa7.05>

- Koreinik, Kadri (ilmumas 2013a). Case-specific report on the Seto language. With contributions by Eva Kühhirt, Anneli Sarhima, Sia Spiliopoulou-Åkermark, and Reetta Toivanen. ELDIA Working Papers.
- Koreinik, Kadri (ilmumas 2013b). Case-specific report on the Võro language. With contributions by Eva Kühhirt, Anneli Sarhima, Sia Spiliopoulou-Åkermark, and Reetta Toivanen. ELDIA Working Papers.
- Krikmann, Arvo 2000. Tere teele, tere meele, tere egalõ talõlõ. Valik lõunaeesti mõistatusi. Tartu: Eesti Kirjandusmuuseum.
- Kynttjä, Eve 1997. Eestlased Soomes – akulturatsioonipsühholoogia ja etniline identiteet. – Hill Kulu, Katrin Metsis, Tiit Tammaru (Toim.). Eestlane olla ... Eesti keele ja kultuuri perspektiivid. Tartu: Tartu Ülikooli Kirjastus, 60–71.
- Laanekask, Heli 2006. Virolaiset Suomessa suomenvirolaisiksi. – Olga Haurinen, Helena Sulkala (Toim.). Tutkielmia vähemmistökielistä Jäämereltä Liivnantaan. Vähemmistökielten tutkimus- ja koulutusverkoston raportti V. Universitatis Ouluensis, Acta B, Humaniora B 69. Oulu: Oulun yliopisto, 55–73.
- Maffi, Luisa 2000. Language preservation vs. language maintenance and revitalization: Assessing concepts, approaches and implications for language sciences. – International Journal of the Sociology of Language, 142, 175–190. <http://dx.doi.org/10.1515/ijsl.2000.142.175>
- McCarty, Theresa (Ed.) 2011. Ethnography and Language Policy. New York, London: Routledge.
- Mufwene, Salikoko S. 2008. Language Evolution: Contact, Competition and Change. London, New York: Continuum.
- Myers-Scotton, Carol 2002. Contact Linguistics. Bilingual Encounters and Grammatical Outcomes. Oxford: Oxford University Press.
- Org, Ervin; Reimann, Nele; Uind, Katrin; Velsker, Eva; Pajusalu, Karl 1994. Vastseliina Sute küla elanike murdepruugist ja keelehoiakutest. – Keel ja Kirjandus, 4, 203–209.
- Pajusalu, Karl 2007. Estonian Dialects. – Mati Ereht (Ed.). Estonian Language. 2nd ed. Tallinn: Estonian Academy Publishers, 231–272.
- Pajusalu, Karl; Velsker, Eva; Org, Ervin 1999. On recent changes in South Estonian: Dynamics in the formation of the inessive. – International Journal of the Sociology of Language, 139, 87–103. <http://dx.doi.org/10.1515/ijsl.1999.139.87>
- Pajusalu, Karl; Hennoste, Tiit; Niit, Ellen; Päll, Peeter; Viikberg, Jüri 2002. Eesti murded ja kohanimed. Tallinn: Eesti Keele Sihtasutus.
- Praakli, Kristiina 2007. Eesti keele kasutusväljad Tampere eesti kogukonna näitel. – Pirkko Muikku-Werner, Ossi Kokko, Hannu Remes (Toim.). Suomalais-ugrilaisia kohdekieliä ja kontakteja. VIRSU III. Studies in Languages 42. Joensuu: Joensuun yliopisto, 91–105.
- Praakli, Kristiina 2009. Esimese põlvkonna Soome eestlaste kakskeelne keelekasutus ja koodikopeerimine. Dissertationes philologiae estonicae Universitatis Tartuensis 24. Tartu: Tartu Ülikooli Kirjastus.
- Praakli, Kristiina 2013a (ilmumas). Case-specific report on Estonian language in Germany. With contributions by Eva Kühhirt, Anneli Sarhima, Sia Spiliopoulou-Åkermark, and Reetta Toivanen. ELDIA Working Papers.
- Praakli, Kristiina 2013b (ilmumas). Case-specific report on Estonian language in Finland. With contributions by Eva Kühhirt, Anneli Sarhima, Sia Spiliopoulou-Åkermark, and Reetta Toivanen. ELDIA Working Papers.
- Raag, Raimo 1999. One plus one equals one: The forging of Estonian Standard. – International Journal of the Sociology of Language, 139, 17–38. <http://dx.doi.org/10.1515/ijsl.1999.139.17>
- Ricento, Thomas; Hornberger, Nancy 1996. Unpeeling the onion: language planning and policy and the ELT professional. – TESOL Quarterly, 30 (3), 401–428. <http://dx.doi.org/10.2307/3587691>

- Saar, Evar 2003. Setos – Eine südeestnische Sprache und Kultur an der Estnisch-Russischen Grenze. – Wolfgang Brücher (Hrsg.). Grenzverschiebungen. Interdisziplinäre Beiträge zu einem zeitlosen Phänomen. Internationales Symposium des interdisziplinären Forschungsschwerpunktes “Grenzregionen und Interferenzräume” der Philosophischen Fakultäten der Universität des Saarlandes in Saarbrücken und Forbach 10.-12. Mai 2001. Röhrig Universitätsverlag, 103–118.
- Silverstein, Michael 1979. Language structure and linguistic ideology. – Paul R. Clyne, William F. Hanks, Carol F. Hofbauer (Eds.). The Elements: A Parasession on Linguistic Units and Levels. Chicago: Chicago Linguistic Society, 193–247.
- Special Eurobarometer 386 = Europeans and Their Languages. Report. June 2012. [http://ec.europa.eu/public\\_opinion/archives/ebs/ebs\\_386\\_en.pdf](http://ec.europa.eu/public_opinion/archives/ebs/ebs_386_en.pdf) (03.01.2013).
- Spolsky, Bernard; Shohamy, Elana 2000. Language practice, language ideology, and language policy. – Richard D. Lambert, Elana Shohamy (Eds.). Language Policy and Pedagogy. Essays in honor of A. Ronald Walton. Amsterdam, Philadelphia: John Benjamins Publishing Company, 1–42.
- Tender, Tõnu 2010. Mitmekeelsus Eestis Euroopa Liidu mitmekeelsuse ideaali taustal. Dissertationes de mediis et communicationibus Universitatis Tartuensis 9. Tartu: Tartu Ülikooli Kirjastus.
- van Dijk, Teun A. 2005 [1998]. Ideoloogia. Multidistsiplinaarne käsitus. Tartu: Tartu Ülikooli Kirjastus.
- van Leeuwen, Theo J.; Wodak, Ruth 1999. Legitimizing immigration control: a discourse-historical analysis. – Discourse Studies, 1 (1), 83–119. <http://dx.doi.org/10.1177/1461445699001001005>

## Võrgumaterjalid

- DESTATIS. Statistisches Bundesamt. Pressemitteilung Nr. 355 vom 26.09.2011. [https://www.destatis.de/DE/PresseService/Presse/Pressemitteilungen/2011/09/PD11\\_355\\_122.html](https://www.destatis.de/DE/PresseService/Presse/Pressemitteilungen/2011/09/PD11_355_122.html) (06.07.2012).
- ELDIA. European Language Diversity for All. <http://www.eldia-project.org> (18.02.2013).
- Pravitel'stvo Rossijskoj Federacii. Postanovlenie ot 24 marta 2000 g. N 255. O edinom perečne korennyh maločislennyh narodov Rossii. <http://www.minnac.ru/minnac/info/13884.html> (03.03.2013).
- REL07 = Eesti emakeelegra püsielanikud, 31. detsember 2011 --- murdekeele oskus ning sugu. <http://pub.stat.ee/px-web.2001/Dialog/varval.asp?ma = RLE07&lang = 2> (06.09. 2012).
- Statista. <http://de.statista.com/statistik/daten/studie/1221/umfrage/anzahl-der-auslaender-in-deutschland-nach-herkunftsland/> (06.07.2012).
- Tilastokeskus. [www.stat.fi](http://www.stat.fi) (30.09.2012).

**Kadri Koreinik** (Tartu Ülikool, Võru Instituut) on sotsioloog, kelle peamine uurimishuvi on keeleideoloogiad ja diskursusuuringud.  
[kadri.koreinik@ut.ee](mailto:kadri.koreinik@ut.ee)

**Kristiina Praakli** (Tartu Ülikool) on erialalt eesti filoloog. Tema peamised uurimisteemad on väliseesti keel ja keelekontaktid.  
[kristiina.praakli@ut.ee](mailto:kristiina.praakli@ut.ee)

# THE CONSTRUCTION OF LANGUAGE MAINTENANCE AND LOSS IN SEMI-PUBLIC DISCOURSES

Kadri Koreinik<sup>1,2</sup>, Kristiina Praakli<sup>1</sup>

<sup>1</sup>University of Tartu, <sup>2</sup>Võro Institute

The paper presents an analysis of the construction of language maintenance and loss in semi-public discourses of Estonians living in Germany and Finland, and of Setu and Võru speakers. The data was gathered by the international interdisciplinary project European Language Diversity for All, which focuses on speakers of small or less used Finno-Ugric languages in contemporary multilingual societies. Estonians living in Germany and Finland represent allochthonous language communities, whereas Setu and Võru speakers form autochthonous Estonian speech communities; the demarcation between the two groups is somewhat problematic. All of the speech communities consist of multilingual Estonians who have managed to maintain their minority languages to a lesser or greater extent in majority language contexts, i.e. in German, Finnish and Estonian language environments. Besides discourse analysis, the self-reported language competences of people of different ages are briefly analysed. The discourse of language maintenance and loss can be seen as a discourse of legitimation, where arguments about the functions and futures of minority languages are expressed. One of the explicit results is that all informants remain within the framework of linguistic, cultural or economic enrichment, without considering linguistic rights.

**Keywords:** minority language, majority language, vehicular language, language communities, discourse studies, language ideology, language ecology, Estonian, English

# HOW TO CREATE ORDER IN LARGE CLOSED SUBSETS OF WORDNET-TYPE DICTIONARIES

Ahti Lohk, Ottokar Tilk, Leo Vöhandu

**Abstract.** This article presents a new two-step method to handle and study large closed subsets of WordNet-type dictionaries with the goal of finding possible structural inconsistencies. The notion of closed subset is explained using a WordNet tree. A novel and very fast method to order large relational systems is described and compared with some other fast methods. All the presented methods have been tested using Estonian<sup>1</sup> and Princeton WordNet<sup>2</sup> largest closed sets.

**Keywords:** thesaurus, closed set, seriation, Power Iteration Clustering (PIC), reducing number of crossings, WordNet

## 1. Introduction

There are more than 60 WordNets in the world<sup>3</sup>. The main idea and basic design of all these lexical resources came from Princeton WordNet (more in Miller et al. 1990). Each WordNet is structured along the same lines: synonyms (sharing the same meaning) are grouped into synonym sets (synsets). Synsets are connected to each other by semantic relations, like hyperonymy (IS-A) and meronymy (IS-PART-OF). In this article only hyperonymy-hyponymy relations are considered as objects of analysis. Of course, it is easy to extend the analysis over different word classes and different semantic relations.

WordNet has been used for a number of different purposes in information systems, including word sense disambiguation (Li et al. 1995), information retrieval (Rila et al. 1998), automatic text classification and structuring (Morato et al. 2004), automatic text summarization, natural language generation (Jing et al. 1998), machine translation (Khan et al. 2009) and even language teaching applications (Morato et al. 2004). A description of the Estonian WordNet and its properties has been given by Orav et al. (2011).

In applications where WordNet usage is considerable, the quality of the result depends on the quality of the WordNet used. Our analysis shows clearly that many

<sup>1</sup> Estonian WordNet: <http://www.ci.ut.ee/ressursid/teksaurus/test/estwn.cgi.et> (08.01.2013).

<sup>2</sup> Princeton WordNet: <http://wordnet.princeton.edu/> (08.01.2013).

<sup>3</sup> The Global WordNet Association: [http://www.globalwordnet.org/gwa/wordnet\\_table.html](http://www.globalwordnet.org/gwa/wordnet_table.html) (08.01.2013).

WordNet-type dictionaries have a large closed subset (Table 1) caused by such semantic relations where one synset has connections to more than one supersynset. Liu et al. analyse mistakes in WordNet structures that arise particularly in cases where a synset has more than one supersynsets (Liu et al. 2004). Richens extends the ideas of Liu et al. and presents a list of anomalies in the WordNet verb hierarchy and methods for finding them (Richens 2008)<sup>4</sup>. Vider (2001) proposes that in the best case every synset has only one supersynset. Closed subsets with more than one supersynset refer to possible causes of errors (Richens 2008, Lohk et al. 2012a). We present a convenient tool to study the possible structural inconsistencies of such large separated subsets.

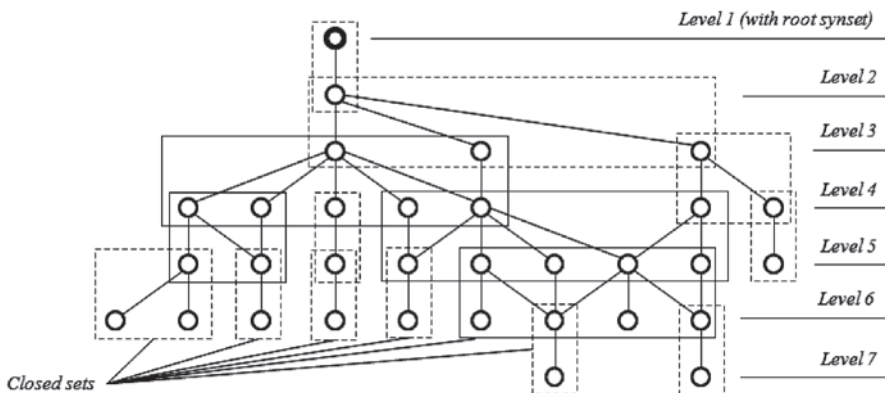
For every synset in WordNet we have a matrix representation, as in Figure 3, upper level on the right. As we have to deal with very big matrices one needs a well ordered final representation of such a matrix to understand its hidden structure. Our goal is to reorder that matrix into the form of Figure 3, lower level on the right. That representation corresponds to the so-called Multidimensional Scale representation in psycholinguistics.

In the next section we explain the content of a closed set (Lohk et al. 2012b).

## 2. Closed sets

The synsets of WordNet-type dictionaries have as semantic connections hierarchy creating ones (*has\_hyponym*, *has\_meronym*, etc.) as well nonhierarchical ones (*near\_synonym*, *be\_in\_state*, etc.). Using hierarchical connections makes WordNet to be a set of trees, whereby part of those trees are threaded (That is a fact from authors' analysis). The vertices of trees are synsets and edges are always some semantic connections.

Such tree has always a notion (synset) on the highest level (so-called root vertice) and other vertices on different levels. In given context we call root vertice also a root synset.



**Figure 1.** Natural tree of the WordNet with closed sets

<sup>4</sup> Results tables relevant to 'Anomalies in the WordNet Verb Hierarchy', paper delivered to COLING 2008, Manchester, UK, August 2008: <http://www.rockhouse.me.uk/Linguistics/> (08.01.2013).

We have an invented example of such a WordNet tree in Figure 1. The synsets of the given tree (vertices) can be divided into seven levels. On the first level is the most general semantic synset – the root synset, and on the last levels (level 6 and level 7) synsets with a possibly concrete meaning. For example, based on semantic connection *has\_hyponym* Princeton WordNet (version 3.0) has 346 root synsets (= trees) and Estonian WordNet (version 64) 204 synsets.

In order to understand closed subsets (Lohk et al. 2012b) in Figure 1 we have to consider only any two neighbouring levels. Let us take for example levels 3 and 4. If we separate those levels with their vertices, one can see that the connections between vertices create two closed sets of vertices. To recognise possible errors it is important to study such sets, where subsynset has a semantic connection with at least two different supersynsets. Such sets are presented in Figure 1 with thick lines and there are four of them. (The number of all closed subsets in Figure 1 is 15). For example, Estonian WordNet (version 64) has as a maximal closed set with dimensions 4,945 x 457. In the language of Figure 1, this closed set has 4,945 vertices in the lower level and 457 vertices in the upper level.

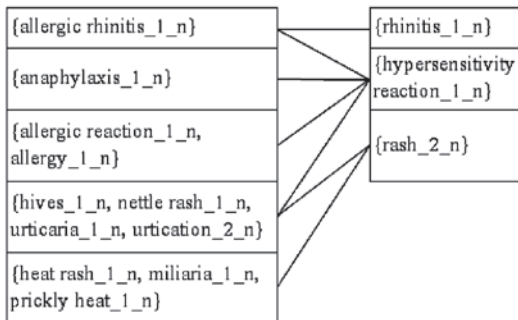
The following table presents an overview of maximal closed sets in the WordNet-type dictionaries that we have analysed to date.

**Table 1.** Dimensions of the maximal closed set in a WordNet

Seq. No.	Name and version of the WordNet	Number of the synsets	Dimensions of the maximal closed set
1	Polish WordNet 1.7	105 074	28 279 x 3 595
2	Cornetto, 1.3	70 492	10 418 x 556
3	Estonian WordNet, 64	54 078	4 945 x 457
4	Princeton WordNet, 3.0	117 659	1 333 x 167
5	Finnish WordNet, 1.1.2	117 659	1 248 x 165
6	Catalan WordNet, 3.0	99 253	1 007 x 91
7	Slovenian WordNet, 3.0	42 919	248 x 3

The number of closed subsets separated using the semantic relation *has\_hyponym* for all those WordNets remains between 4000 and 20 000.

A very suitable algorithm to separate closed subsets is given by Flannery et al. (2009). An example of a closed subset with real data is presented in Figure 2.

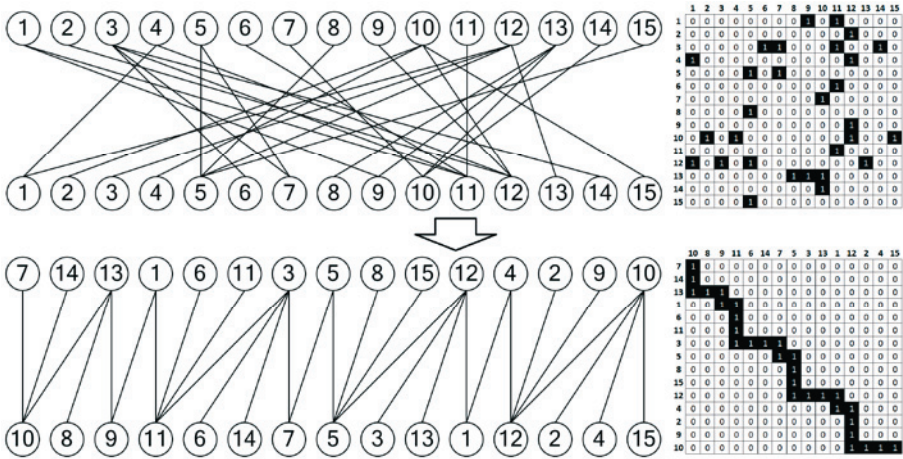


**Figure 2.** Real example of a closed subset (Princeton WordNet, version 3.0), rotated 90 degrees

The next section is dedicated to the study of such maximal closed sets.

### 3. Improving identification of mistakes by reducing the number of crossings

To visually identify possible mistakes in the connections between the synsets of a closed subset of a WordNet it is necessary to visualize the connections as clearly and with as little clutter as possible. One way to achieve this goal is to reduce the number of crossings in the graph representation of the WordNet by reordering vertices as shown in Figure 3.



**Figure 3.** Graph (its corresponding adjacency matrix plotted on the right) with different permutations of vertices, illustrating how a good reordering can reduce the number of crossings in the bipartite graph (from 206 to 0 in this example)

There are many algorithms for this task, with different approaches – such as genetic algorithms (Mäkinen, Sieranta 1994), heuristic algorithms (e.g. barycenter (Sugiyama et al. 1981), median (Eades, Wormald 1994)) and for small graphs even exact methods (Jünger, Mutzel 1997). In the same paper in which Jünger and Mutzel introduced their exact method, they also compared different heuristic algorithms on larger graphs for which the exact method is not viable. They concluded that the iterated barycenter method was clearly the best choice for both its speed and solution quality.

#### 3.1. The two-step method for reducing the number of crossings

In this work we introduce a novel technique which outperforms other widely used methods including barycenter heuristic. Our method consists of two steps:

1. Power iteration seriation;
2. Median heuristic.

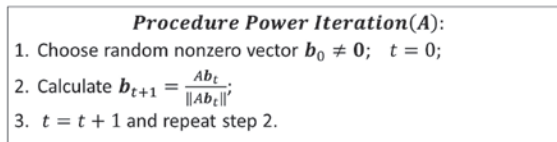
First let us focus on the second step – the median heuristic by Eades and Wormald (1994). The median heuristic is a well known method for crossing minimization. To



reduce crossings, this method finds the median of the positions of adjacent vertices for every vertex and then ranks the vertices in a layer according to these values. Technically it is very similar to the barycenter heuristic (Sugiyama et al. 1981), the only difference being that the latter uses mean values of the positions of adjacent vertices instead of median values. This method is often applied iteratively, fixing one layer and reordering the other in turns, until there is no change in the order of vertices. The final outcome of the median (and also barycenter) heuristic depends on the initial state of the graph. To gain better results one can restart the algorithm a number of times with different random initial orderings and choose the best result but, as Jünger and Mutzel (1997) concluded, for bigger graphs the results improve only slightly. The purpose of the first step of our method is not to rely on random ordering, but to preprocess the graph with the aim of providing as good a starting point for the median heuristic as possible.

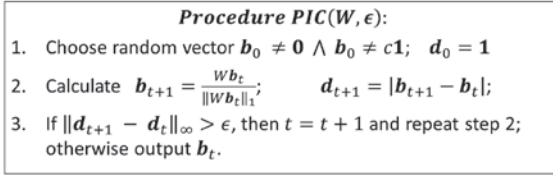
The first step of our method is a custom modification of a very fast (approximately linear to the input size), effective and simple clustering algorithm, Power Iteration Clustering (PIC) by Lin and Cohen (2010). The name of their method comes from the power iteration eigenvalue algorithm on which it is based.

The power iteration algorithm is used to find the dominant eigenvalue  $\lambda_1$  (assuming there is one i.e.  $|\lambda_1| > |\lambda_2| \geq |\lambda_3| \geq \dots \geq |\lambda_n|$ ) and eigenvector  $\mathbf{v}_1$  of a matrix  $A$ . The algorithm takes the steps described in Figure 4. After a sufficient amount of iterations  $\mathbf{b}_t$  converges to  $\mathbf{v}_1$  of  $A$ .



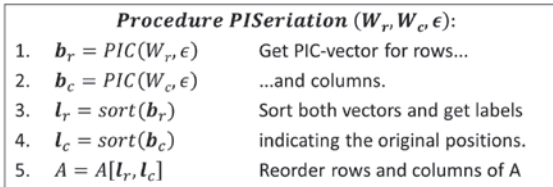
**Figure 4.** Description of power iteration eigenvalue algorithm

The PIC algorithm applies power iteration to a row-normalised (all elements in a row sum up to 1) similarity matrix  $W$ . Since the dominant eigenvector of  $W$  is a constant vector, it is useless for clustering (that's also the reason for additional constraint:  $\mathbf{b}_0 \neq \mathbf{c}\mathbf{1}$ , i.e. initial vector must not be a constant vector). Therefore, to turn the power iteration eigenvalue algorithm into a clustering algorithm, Lin and Cohen augmented it with a stopping criterion which stops the process before converging to the constant dominant eigenvector. As a result we get a vector  $\mathbf{b}_t$  (PIC-vector) which is an eigenvalue-weighted linear combination of all the eigenvectors of  $W$  and turns out to be a good clustering indicator. The main procedure of PIC is described in Figure 5, where  $\epsilon$  is a small number (e.g.  $10^{-5}$ ) used as a parameter for stopping criterion and  $\mathbf{d}_t$  is a vector describing the changes (compared to previous iteration) in the values of the elements of vector  $\mathbf{b}_t$ . The algorithm is stopped when for two consecutive iterations  $\mathbf{d}_t$  has remained almost constant i.e. none of the absolute differences of changes are larger than  $\epsilon$ . Lin and Cohen used  $k$ -means on the PIC-vector to obtain the final result in the form of clusters.



**Figure 5.** Description of the main subroutine of PIC algorithm

Our own work has shown that PIC-vector can also be successfully used for seriation. To do that, we first calculate two PIC-vectors – one for rows and the other for columns. Then we reorder the rows and columns of the matrix according to the ascending or descending order of the values in the corresponding PIC-vector. The exact procedure is shown in Figure 6 where  $W_r$  and  $W_c$  are normalised similarity matrices,  $\mathbf{b}_r$  and  $\mathbf{b}_c$  PIC-vectors and  $\mathbf{l}_r$  and  $\mathbf{l}_c$  labels for reordering. Lower indices  $r$  and  $c$  denote rows and columns respectively.



**Figure 6.** Description of seriation procedure using PIC algorithm

For the first step of the crossing minimisation method we use the power iteration seriation with a very simple symmetric similarity function where the similarity  $s(x_i, x_j) = s(x_j, x_i)$  between two vertices  $x_i$  and  $x_j$  from the same layer is equal to the number of their common neighbours in the opposite layer:  $s(x_i, x_j) = s(x_j, x_i) = |n(x_i) \cap n(x_j)|$  (where  $n(x)$  denotes the set of neighbours of  $x$ ). If we represent the bipartite graph as an adjacency matrix  $A$  and  $n$ th row of  $A$  as  $A(n)$ , then we can rewrite the function as follows:

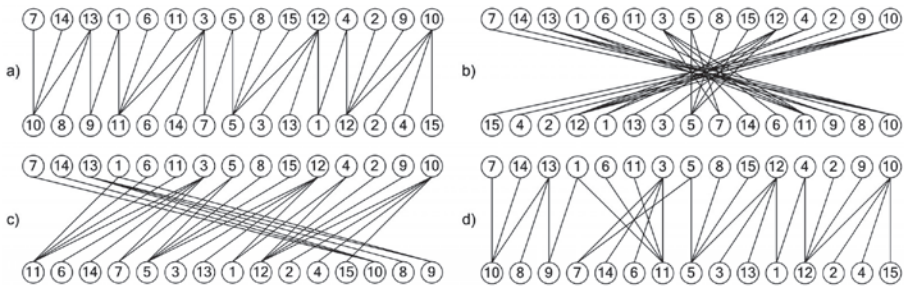
$$s(x_i, x_j) = s(x_j, x_i) = \begin{cases} A(i) \cdot A(j), & \text{if } x_i \text{ and } x_j \text{ are upper layer vertices} \\ A^T(i) \cdot A^T(j), & \text{if } x_i \text{ and } x_j \text{ are lower layer vertices} \end{cases}$$

The similarity matrix of upper layer vertices  $S_r$  (or row similarity matrix of  $A$ ) where element  $S_r(i, j) = s(x_i, x_j)$  can then be calculated as  $S_r = AA^T$  and the similarity matrix of lower layer vertices (or column similarity matrix of  $A$ ) can be calculated as  $S_c = A^T A$ . Both matrices have to be normalised before using power iteration seriation on them.

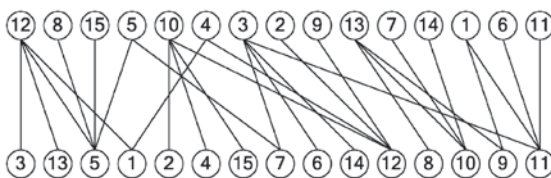
Since the elements of PIC-vector corresponding to similar objects (vertices in our case) tend to obtain similar values, the positions of vertices after seriation also tend to correlate with the number of common neighbours. As a result subsets of vertices with many common neighbours clump together after processing with power iteration seriation in the first step of our method. This kind of approach alone does not always provide very good results in terms of the number of crossings. For example, it is possible that one layer has to be reversed, because power iteration seriation can produce results where the band of ones in the adjacency matrix runs from top-right

to bottom-left (Figure 7b) instead of top-left to bottom-right (Figure 7a), which is not good from the crossing number perspective. In some conditions (when the graph consists of more than one connected component or even when there are multiple components which are weakly connected to each other; with ‘noise’; suboptimal  $\epsilon$ , etc.) it is possible that some subset of similar vertices will be positioned too far away from their common neighbours (Figure 7c). Additionally, there is a risk that some subsets of vertices within layers could be in reverse order (Figure 7d). Even worse, very often multiple problems come up simultaneously.

All the problems mentioned above are solved by applying a median heuristic to the result of power iteration seriation. The median heuristic is not just compensating for the weaknesses of power iteration seriation, but the output of the latter is also a very good initial permutation for the former, enabling it to achieve much better results than some random permutation would. For example: on the initial graph from Figure 1, the iterative median heuristic could only reduce the number of crossings to 27 (Figure 8), while power iteration seriation in conjunction with median heuristic reduced the number of crossings to 0 (Barycenter heuristic reduced the number of crossings to 28).



**Figure 7.** Example graph from Figure 1 illustrating some problems with power iteration seriation: a) one of the optimal permutations of vertices; b) lower layer in reverse order; c) subset of similar vertices in one layer are too far from their common neighbours; d) Subset of vertices in one layer is in reverse order



**Figure 8.** Result of iterative median heuristic on the initial graph from Figure 3

This kind of two-step method not only provides a much smaller number of crossings but may also provide these results while being faster than the iterative barycenter or median algorithm. This is possible because after preprocessing the graph with power iteration seriation, only one iteration of the median heuristic is sufficient to produce a superior result than the iterative barycenter or median method alone. If time is not crucial, then additional iterations of the median heuristic may be applied to polish the result further. Some additional improvement can also be found by trying different values for PIC’s stopping criterion parameter  $\epsilon$  ( $10^{-5} - 10^{-7}$  divided by

number of rows in similarity matrix was usually optimal for us). Next we will give some examples how this method performed on the real WordNet graphs.

### 3.2. Experiments on WordNet graphs

In this section we give an overview of our tests on the largest closed sets of synsets from Estonian and Princeton WordNets. The largest closed set from Estonian WordNet can be represented as a 4,945 by 457 matrix (see Table 1). In the case of the Princeton WordNet the matrix size is 1 333 x 167 (Table 1).

We ran our tests on a PC with 6 GB of RAM and an Intel® Core™ i7-870 Processor and compared three different methods: iterative barycenter, iterative median and our two-step method. In the two-step method we used only one iteration of median heuristic and for PIC's stopping criterion parameter  $\epsilon$  we chose  $10^{-5}$  divided by number of rows in similarity matrix. All 3 methods were run on the same initial permutation of vertices. The results are shown in Table 2.

**Table 2.** Results of three methods compared with a random permutation on the largest closed sets of two WordNets

	Estonian WordNet (v 64)		Princeton WordNet (v 3.0)	
	Time (s)	Crossings	Time (s)	Crossings
Initial	–	2 349 957	–	265 940
Median	4.1	904 629	0.9	36 862
Barycenter	16.9	308 444	1.5	22 927
2-step method	0.4	84 884	0.1	5 484

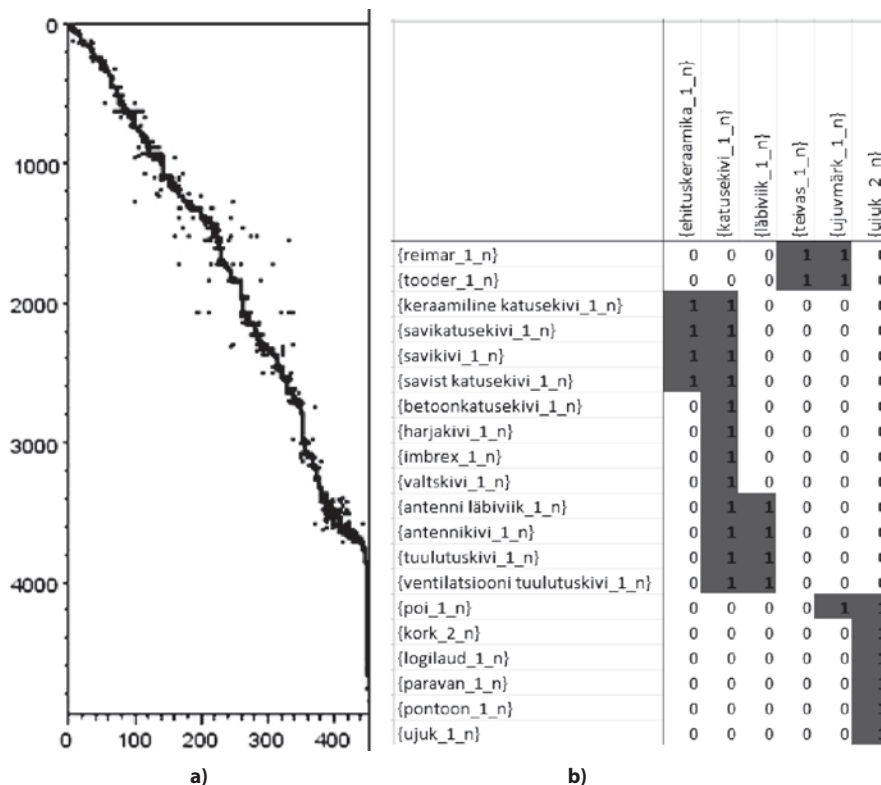
The two-step method turned out to be roughly 9–42 times faster than compared methods while producing more than 3–10 times fewer crossings. From these results we can conclude that our two-step method is the best choice for minimising the number of crossings in WordNet graphs.

Some possible ways of using the Minimal Crossing method to detect inconsistencies in WordNet structures is given by the author and others (Lohk et al. 2012a, 2012b). The detailed handling of those and others inconsistencies would require a separate article.

## 4. Looking at the results

As a result of using our two-step method we did get an ordered matrix (Figure 9a).

By converting this matrix with the labels of synsets into a MS Excel worksheet we have the possibility of studying the large closed subset more methodically. To make it easier to understand the result it is useful to freeze the headings of rows and columns. That makes it possible to move around in the table so that the synsets on both levels are always visible. To find possible errors one has to study such places in that table where conceptual synsets in rows and columns are conspicuously different. Usually such an occasion happens when one concept has several parents. The decision about a possible error will be naturally made by the lexicographer.



**Figure 9.** The biggest closed subset of Estonia WordNet: a) closed subset after ordering; b) closed subset for the investigation, converted for Excel

## 5. Conclusion

Wordnet as a lexical and semantical database is widely used in different language technology applications. Therefore it is important to ensure the quality of any Wordnet used. Previous study has shown that WordNet with its hierarchical structures consists of many relations which quite easily cause errors in the Wordnet structure (Lohk et al. 2012). In this paper we propose a formal way to detect and study possible inconsistencies using closed subsets. The notion of a closed subset has been explained using the WordNet tree. Separated closed subsets are represented as matrices and a new and fast two-step method reorders such sparse relational systems into an easily visible and understandable view. Our method has been compared with other fast reordering methods and tested on Estonian and Princeton WordNets. As a final suggestion we transform the subsets with correct syntactic labels into an Excel spreadsheet to enable convenient study of places where the structural connections of concepts (synonym synsets) are suspicious.

## References

- Eades, Peter; Wormald, Nicholas C. 1994. Edge crossings in drawings of bipartite graphs. – *Algorithmica*, 379–403.
- Flannery, P. B.; Press, H. W.; Teukolsky, A. S.; Vetterling, T. W. 2009. *Numerical Recipes in C. The Art of Scientific Computing*. South Asia: Cambridge University Press India.
- Jing, H. 1998. Usage of WordNet in natural language generation. – *Proceedings of the Workshop Usage of WordNet in Natural Language Processing Systems: COLING-ACL 1998*; August 16, Montreal, Quebec, Canada, 128–134.
- Jünger, Michael; Mutzel, Petra 1997. 2-Layer Straightline Crossing Minimization: Performance of exact and heuristic algorithms. – *Journal of Graph Algorithms and Applications*, 1–25.
- Li, X.; Szpakowicz, S.; Matwin, S. 1995. A WordNet-based algorithm for word sense disambiguation. – *Proceedings of IJCAI 1995*. Morgan Kaufmann Publishers, 1368–1374.
- Lin, Frank; Cohen, William W. 2010. Power iteration clustering. – *Proceeding of the 27th International Conference on Machine Learning*, June 21–24, 2010, Haifa, Israel. Omnipress, 655–662.
- Liu, Y.; Jiangsheng, Y.; Zhengshan, W.; Shiwen, Y. 2004. Two kinds of hypernymy faults in Word-Net: the cases of ring and isolator. – Petr Sojka, Karel Pala, Pavel Smrz, Christine Fellbaum, Piek Vossen (Eds.). *Proceedings of the Second Global WordNet Conference*. Brno, Czech Republic, 20–23 January 2004. Masaryk University, 347–351.
- Lohk, Ahti; Võhandu, Leo 2012. Eesti Wordnet'i struktuuri analüüsist. – *Eesti Rakenduslingvistika Ühingu aastaraamat*, 8, 139–151. <http://dx.doi.org/10.5128/ERYa8.09>
- Lohk, Ahti; Vare, Kadri; Võhandu, Leo 2012a. First steps in checking and comparing Princeton WordNet and Estonian WordNet. – Miriam Butt et al. (Eds.). *Proceedings of the EACL 2012 Joint Workshop of LINGVIS & UNCLH*. April 23–24 2012, Avignon, France. Association for Computational Linguistics, 25–29.
- Lohk, Ahti; Vare, Kadri; Võhandu, Leo 2012b. Visual Study of Estonian WordNet using Bipartite Graphs and Minimal Crossing algorithm. – *Proceedings of 6th International Global WordNet Conference*, Matsue, Japan, 2012, 167–173.
- Miller, G.; Beckwith, R.; Fellbaum, C.; Gross, D.; Miller, K. 1990. Introduction to WordNet: An on-line lexical database. – *International Journal of Lexicography* 3, 235–312.
- Morato, J.; Marzal, M. Á.; Lloréns, J.; Moreiro, J. 2004. WordNet applications. – Petr Sojka, Karel Pala, Pavel Smrz, Christine Fellbaum, Piek Vossen (Eds.). *Proceedings of the Second Global WordNet Conference*. Brno, Czech Republic, 20–23 January 2004, 270–278.
- Mäkinen, Erkki; Sieranta, Mika 1994. Genetic algorithms for drawing bipartite graphs. – *International Journal of Computer Mathematics*, 53 (3–4), 157–166. <http://dx.doi.org/10.1080/00207169408804322>
- Orav, Heili; Kerner, Kadri; Parm, Sirli 2011. Eesti Wordnet'i hetkeseisust. – *Keel ja Kirjandus*, 2, 96–106.
- Richens, Tom 2008. Anomalies in the WordNET verb hierarchy. – *Proceedings of the 22nd International Conference on Computational Linguistics: COLING-ACL 2008*, August, Manchester, UK, 729–736.
- Rila, M.; Tokunaga, T.; Tanaka, H. 1998. The use of WordNet in information retrieval. – *Proceedings of the Workshop Usage of WordNet in Natural Language Processing Systems: COLING-ACL 1998*, August 16, Montreal, Quebec, Canada, 31–37.
- Salam, Khan Md Anwarus; Khan, Mumit; Nishino, Tetsuro 2009. Example based English-Bengali machine translation using WordNet. – *Proceedings of the Triangle Symposium on Advanced ICT 2009 (TriSAI 2009)*, October 28–30, 2009. Tokyo, Japan.
- Sugiyama, Kozo; Tagawa, Shojiro; Toda, Mitsuhiko 1981. Methods for Visual Understanding of Hierarchical System Structures. – *IEEE Transactions on Systems, Man and Cybernetics*, 11 (2), 109–125. <http://dx.doi.org/10.1109/TSMC.1981.4308636>

Vider, Kadri 2001. Eesti keele teaurus – teooria ja tegelikkus. – Margit Langemets (Toim.). Leksikograafiaseminar “Sõna tänapäeva maailmas” / Leksikografinen seminaari “Sanat nykymaailmassa”. Ettekannete kogumik. Eesti Keele Instituudi toimetised 9. Tallinn: Eesti Keele Sihtasutus, 134–156.

## **Web References**

The Global WordNet Association. [http://www.globalwordnet.org/gwa/wordnet\\_table.html](http://www.globalwordnet.org/gwa/wordnet_table.html) (08.01.2013).  
Results tables relevant to “Anomalies in the WordNet Verb Hierarchy” paper delivered to COLING 2008. Manchester August 2008. <http://www.rockhouse.me.uk/Linguistics/> (08.01.2013).  
Princeton WordNet (version 3.0, 3.1). <http://wordnet.princeton.edu/> (08.01.2013).  
Estonian WordNet (version 65). <http://www.cl.ut.ee/ressursid/teksaurus/test/estwn.cgi.et> (08.01.2013).

**Ahti Lohk** (Tallinn University of Technology), main research interests are in the field of data analysis.  
[ahti.lohk@ttu.ee](mailto:ahti.lohk@ttu.ee)

**Ottokar Tilk** (Tallinn University of Technology), main research interests are data analysis and machine learning algorithms.  
[ottokar.tilk@ttu.ee](mailto:ottokar.tilk@ttu.ee)

**Leo Võhandu** (Tallinn University of Technology), main research interests are in the field of data analysis.  
[leo.vohandu@ttu.ee](mailto:leo.vohandu@ttu.ee)

# KUIDAS LUUA KORDA WORDNET'I TÜÜPI SÕNARAAMATUTE SUURTES KINNISTES ALAMHULKADES

**Ahti Lohk, Ottokar Tilk, Leo Vöhandu**

Tallinna Tehnikaülikool

WordNet kui leksikaalsemantiline andmebaas leiab laialdast kasutust keele- tehnoloogia rakendustes, mistõttu on ilmne, et tulemuse kvaliteet sõltub paljuski *wordnet*'i enda kvaliteedist. Varasemad uurimused on näidanud, et *wordnet*'i hierarhiat tekitavates puudes esineb seoseid, mis põhjustavad tema struktuuris vigu (Lohk, Vöhandu 2012). Ühe võimalusena pakutakse artiklis taolisi kõrvale- kaldeid uurida ja avastada kinniste alamhulkade kaudu, mida esitatakse maatrik- sina ja millele rakendatakse autorite pakutud uudset kahesammulist meetodit. Kinniseid alamhulki selgitati tehislikult koostatud *wordnet*'i puu alusel. Pakutud kahesammulist meetodit, mis sobib suurte relatsiooniliste süsteemide korrasta- miseks, kõrvutati teiste kiirete varasemate meetoditega (raskuskeskme meetod ja mediaanmeetod). Jõuti järeldusele, et kahesammuline meetod pakub tulemuseks nii paremat ristumiste arvu kui ka kiiremat algoritmi kui varasemad meetodid. Meetodit testiti Eesti ja Princetoni *wordnet*'idel. Maatriksina saadud tulemusi soovitati koos sünohulkade nimedega konverteerida tabelarvutusprogrammi, lii- kuda mööda korrastatud maatriksil olevat lairiba ning uurida ridades ja veergudes olevaid sünohulkade neid kohti, kus mõisted silmatorkavalt erinevad.

**Võtmesõnad:** tesaurus, suletud hulgad, järjestamine, klasterdamine iteratiivse astendamisega, ristumiste arvu vähendamine, WordNet



# ÜLEMINEK RIIGIKEELSELE GÜMNAASIUMIÕPPELE: VILISTLASTE SEISUKOHTADE Q-METODOLOOGILINE DISKURSUSANALÜÜS

Anu Masso, Katrin Kello, Valeria Jakobson

**Ülevaade.** Artikkel keskendub diskursustele, mis on seotud riigikeelsele gümnaasiumiõppele üleminekuga senistes vene õppekeelelega gümnaasiumites. Empiiriliseks ülesandeks on analüüsida teiskeelse õppega seotud positsioone venekeelsete koolide vilistlaste kogemuste põhjal. Metodoloogiliseks eesmärgiks on selgitada kvantitatiivse Q-metodoloogilise analüüsi võimalusi peamiselt kvalitatiivse diskursusanalüüsi kontekstis. Uuring põhineb Q-metodoloogilistel intervjuudel, mis viidi 2010. aastal läbi neljas piirkonnas (Tallinn, Ida-Virumaa, Tartu, Läti) vene õppekeelelega gümnaasiumite vilistlaste seas ( $n = 42$ ). Selgitatakse nelja riigikeelsele õppele üleminekut iseloomustavat diskursust, mis analüüsis eristusi: “keeleliste valikuvõimaluste tähtsustamine”, “rahulolu oma keeleoskuse ja integreeritusega”, “vastuseis homogeniseerimisele” ning “keeleline ja kultuuriline suletus”.\*

**Võtmesõnad:** keelepoliitika, hariduspoliitika, rahvussuhted, riigikeelsele õppele üleminek, Q-metodoloogia, diskursusanalüüs, vene keel, eesti keel, läti keel

## 1. Sissejuhatus

Eesti ja Läti taasiseseisvumisjärgse keelepoliitika ühisosaks võib pidada lähteko-haks olnud etnilist koosseisu ning nõukogudeaegset venestuspoliitikat. Varasema uuringu (Siiner 2006) hinnangul on neis riikides viljeletud tugevat ja kontrollile suunatud keelepoliitikat, lähtudes normatiivsest ning kaitsvast hoiakust riigikeele suhtes. Sealhulgas kuulub olulisemate algatuste hulka seniste vene õppekeelelega koolide üleminek riigikeelsele õppele. Lätis toimus üleminek veidi varem – juba 2004/05. õppeaasta alguseks pidid üldkeskhariduse asutuste õppekavad sisaldama

\* Artiklis analüüsitud empiirilised andmed on kogutud TÜ haridusuuringute ja õppekavaarenduse keskuse poolt läbiviidud ja Haridus- ja Teadusministeeriumilt tellitud uurimisprojekti “Vene õpilane venekeelse üldhariduskooli eestikeelses õppes” (2008–2011) raames. Artikli valmimist on toetanud ETFi grant “Meedia ja kommunikatsiooni geograafiad siirdeühiskonnas” (ETF9308) ning Euroopa Liit läbi Euroopa Regionaalarengu Fondi (Kultuuriteooria tippkeskus) ja sihtfinantseeritav teema SF0180002s07. Täname retsensente ja toimetajaid põhjalike ja asjatundlike kommentaaride eest.

vähemalt 60% ainete õpetamist riigikeeles. Eestis on ülemineku tähtaegsid korduvalt korrigeeritud – eestikeelsele õppele ülemineku alguseks määratud 2000. aasta asendati hiljem 2007. aastaga. Õppeaasta 2011/2012 oli esimene, mil oli ette nähtud, et 10. klassi astuvad õpilased – välja arvatud spetsiifilised erandjuhud – peaksid õppima gümnaasiumiastme jooksul vähemalt 60% minimaalsest kohustuslikust õppemahust eesti keeles (RT 2007, 61, 392).

Varasemad uuringud on näidanud (Masso, Kello 2010, 2012), et riigikeelsele õppele ülemineku õnnestumine sõltub praktilise toetuse (nt õpetajakoolituste, õppematerjalide) kvaliteedi ja kättesaadavuse kõrval oluliselt ka osapoolte (ministeeriumi ja sidusrühmade) vastastikusest mõistmisest. Eesti Haridus- ja Teadusministeerium on haridusreformi peamise ametliku eesmärgina sõnastanud vene kodukeelega noorte konkurentsivõime tõstmise nende keeleoskuse paranemise kaudu (vt nt KKK 2012). Lisaks on eestikeelsele õppele üleminekut põhjendavates sõnumites kõlanud pragmaatilisemad vajadused, nagu haridussüsteemi integreerimine ning riiklike rahaliste ressursside säästlik majandamine (nt kulutused õppematerjalidele), arvestades õpilaste arvu kahanemist. Samas, nagu osutavad mitmed uuringud (vt Masso 2010, Masso, Kello 2012), võivad erinevate ühiskonnarühmade huvid ja identiteedid olla saanud suhteliselt väiksema tähelepanu osaks.

Käesolevas artiklis keskendume riigikeelsele õppele ülemineku seotud diskursuste ja avalike sõnumite tõlgenduste analüüsimisele erineva keelelis-kultuurilise kogemusega vilistlaste perspektiivist. Uurime, 1) millised vilistlaste keelelis-kultuurilisest kogemusest tulenevad individuaalsete positsioonide muustrid eristuvad ning 2) millised on vilistlaste tõlgendusmustreid ühendavad ja eristavad jooned. Artikli empiirilise allikana kasutame Q-metodoloogilisi intervjuusid, mis viidi läbi neljas sotsiolingvistilise keskkonna poolest eristuvast piirkonnast (Tallinn, Ida-Virumaa, Tartu, Läti).

## 2. Q-metodoloogiline diskursusanalüüs

Traditsioonilistele klassifitseerivatele analüüsimetoditele, nagu faktor- ja klasteranalüüs, on ette heidetud võimetust selgitada hilismodernsuses toimuvaid kultuurilisi muutusi, kuna transformatsioonid avalduvad üha enam diskursustes või on nende diskursuste poolt loodud (nt van Dijk 1985, Bailey 1994, Chouliaraki, Fairclough 1999). Käesolev uuring põhineb kvalitatiivse ja kvantitatiivse lähenemise kombinatsioonil – Q-metodoloogilisel diskursusanalüüsil.

Q-metodoloogiat<sup>1</sup> võib defineerida kui subjektiivsete positsioonide uurimist statistiliste tehnikate ja diskursiivsete muustrite leidmise abil. Rõhuasetus subjektiivsusele on peamine erinevus algset R- ehk faktoranalüüsi edasi arendanud Cyril Burti ja William Stephensoni lähenemisest (vt nt Cattell 1973). Burt (1937) analüüsis indiviidide korrelatsioone nn ümberpööratud andmematriksi abil ja keskendus seega eelkõige uuritava nähtuse objektiivsetele parameetritele. Käesolevas uuringus on eeskujuks võetud eelkõige Stephensoni (1953), aga ka hilisem Steven Browne (1980) lähenemine, mis fokuseerus subjektiivsetele aspektidele. Subjektiivsuse analüüsi metodoloogiliseks eeliseks on suurem tähelepanu pööramine kommunikatiivsusele ja operantsusele<sup>2</sup>. Teisisõnu, subjektiivsus ehk indiviidi positsioon on enesekohane

<sup>1</sup> Nimetame lähenemist metodoloogiks, kuna püüame käsitleda nii klassifitseerivate meetodite tehnilisi üksikasju kui ka subjektiivsuse uurimisele omaseid protseduurilisi, teoreetilisi ja filosoofilisi põhimõtteid (vt nt Stephenson 1953).

<sup>2</sup> Terminit *operantsus* oleme kasutanud ingliskeelse *operant* eeskujul. Terminia laenas Stephenson Ameerika biheivioristilt B. F. Skinnerilt, kes kasutas mõistet viitamaks käitumisele, mis pole põhjustatud välise stiimuli poolt (vrdl eestikeelne mõiste *operantne tingimine*) (vt McKeown, Thomas 1988).

ja edasiantav (kommunikatiivsus) ning võrreldes formaliseeritud küsitlusega sõltub vähem mõõtmisvahenditest (operantsus). Selliselt sarnaneb Q-metodoloogia kvalitatiivsele diskursuseanalüüsile. Lisaks võimalusele analüüsida subjektiivsust on käsitlused lähedased ka analüüsiprotsessi süstemaatilise (nt Fairclough 2003), dünaamilisuse ja konteksti rõhutamise poolest (vrd nt Wodak 1996).

Käesoleva uuringu intervjueeritavad leiti kvalitatiivse strateegilise valimi (vt nt Trost 1986) põhimõtetest lähtuvalt ( $n = 42$ ). Strateegilisele valimile omaselt oli representatiivsuse taotluse asemel eesmärgiks diferentseerida valimit sotsiolingvistiliste kogemuste poolest ning seeläbi tagada subjektiivsete positsioonide mitmekesisus. Vilistlaste valimisse kuulusid noored, kes olid lõpetanud venekeelse gümnaasiumi aastal 2007 või 2008, mil Eestis algas sihipärane eestikeelsele õppele üleminek. Vilistlaste positsioonide mitmekesisuse tagamiseks, aga ka selgituste mõistmiseks eri kontekstides, viidi intervjuud läbi neljast sotsiolingvistilise keskkonna<sup>3</sup> poolest eristuvast piirkonnast – Tallinnast (10), Ida-Virumaalt (12), Tartust (10) ning Lätist (10) – pärit noortega. Sotsiolingvistilise keskkonna poolest on uuritav Läti rühm mõnevõrra heterogeensem (praegune elukoht Riias, hariduse omandamise koht erinev), Eesti vilistlastest eristab aga institutsionaalne kontekst (nt riigikeelse õppe korraldus). Eesmärgiks on analüüsida eri sotsiolingvistiliste kogemustega vilistlaste tõlgendusi, mitte leida eraldi diskursused valimi alarühmade lõikes. Uuritavate vilistlaste kogemused riigikeelse aineõppega gümnaasiumis olid mitmekesised, ulatudes üksikutest enamiku aineteni. Tervikvalim kujunes tegevusala lõikes suhteliselt homogeenseks (pea kõik vilistlased õppisid kõrgkoolis, erandiks kaks noort, kes käisid tööl ja/või õppisid kutsekoolis), mistõttu on tegemist spetsiifilise, õppimises pigem edukate vilistlaste valimiga. Enamik intervjueerituid omas (ning puudutas ka intervjuudes) kokkupuuteid riigikeelse õppega ka kõrgkoolis. Andmed koguti 2010. a sügisel suulise individuaalintervjuu vormis. Enamiku intervjuude kestus jäi ühe ja pooleteise tunni vahele.

Riigikeelsele õppele üleminekuga seotud subjektiivsete tõlgenduste uurimise instrumendina kasutasime nii avatud küsimusi kui ka eelnevalt sõnastatud 63 väidet, mis olid valitud algse väidete koondkogumi<sup>4</sup> ehk ligi 200 konstrueeritud väite hulgast. Koondkogum oli sõnastatud mitmekesiste allikate põhjal (nt meediatekstit, ametlikud dokumendid, akadeemilised uuringud) ning kajastas riigikeelse õppe ning keeleliste ja kultuuriliste identiteetidega seotud arusaamade mitmekesisust. Väidete mitmekesisuse tagamiseks ning arvuliseks tasakaalustamiseks valdkondade lõikes kasutasime varasemate uuringute (vt nt Masso, Kello 2010, 2012) põhjal tuletatud järgmisi analüütilisi dimensioone ja nende põhjal moodustunud 9-väljalist maatriksit (vt joonis 1): üldistusaste (riigikeelne õpe indiviidi, kooli ja klassiruumi ning ühiskonna/riigi tasandil) ning positsioonide fookus (isiklikud kogemused ja praktikad, enesekohased hoiakud ning väliste tingimustele suunatud hoiakud). Väidete konstrueerimisel jälgisime, et nimetatud sisulised valdkonnad oleksid kaetud. Väidete jaotus eri väljade vahel polnud alati üks-üheselt määratav, kuna meetodile omaselt võis üks väide sisaldada samaaegselt mitut eri aspekti.

<sup>3</sup> Rannuti (2005) uuringu eeskujul oleme mõistnud sotsiolingvistilise keskkonna all nii suhtlusvõrgustikku makrotasandil (nt kooli eripärad) kui ka isikutevahelist suhtlusvõrgustikku mikrotasandil. Ida-Virumaa, Tartu ning Tallinna kui erinevate sotsiolingvistiliste keskkondade eristumist on näidanud varasemad kvantitatiivsed uuringud (Masso, Kello 2010).

<sup>4</sup> Esinduslike küsitlusuuringute *üldkogumi* asemel eelistame väidete *koondkogumi* (ingl *concourse*) mõistet, rõhutamaks kvalitatiivuuringle omast valimi moodustamist *küllastumise* põhimõttel. Valimi mõiste on Q-metodoloogias tavapärasest mõnevõrra laiem – eristatakse Q-valimit (väited) ning R-valimit (indiviidid).

Positsioonide fookus	Üldistusaste		
	Eestikeelne õpe indiviidi tasandil (nt mõju valikutele, võimalustele, hoiakutele)	Eestikeelne õpe kooli ja klassiruumi kontekstis (nt toetavad tegurid, probleemid ja takistajad)	Eestikeelne õpe ühiskonna ja riigi kontekstis
<b>Isiklikud kogemused ja tegevused</b> (nt kuidas olen ise hakkama saanud)	1. Minu kogemused (keeled ja toimetulek)	2. Minu kogemused (eestikeelne õpe koolis)	3. Minu kogemused (ühiskondlik sidusus ja kultuur)
<b>Enesekohased hoiakud</b> (nt eesti keele õppimine on minu meelest oluline)	4. Minu hoiakud (mis on minu enda jaoks oluline)	5. Minu hoiakud (mis on koolis oluline)	6. Minu hoiakud (mis on ühiskonnas oluline)
<b>Väliste tingimustele suunatud hoiakud</b> (nt riigi tegevusele, teiste ühiskonnaliikmete hoiakutele ja praktikatele)	7. Minu hoiakud (seoses individuaalsete valikutega)	8. Minu hoiakud (seoses kooli valikutega üldiselt)	9. Minu hoiakud (ühiskonna ja riigi valikud)

Joonis 1. Väidete konstrueerimise analüütiline raamistik

Konstrueeritud 63 väidet paluti uuritavatel sorteerida etteantud nõusolekuskaala (-5...+5) ning fikseeritud sümmeetrilise sagedusjaotuse piires (joonis 2). Skaala koostamisel lähtusime nii soovitusel kasutada normaaljaotusele lähedase kujuga väidete ladumise vormi (Stephenson 1936) kui ka eeldusest, et huvi pakuvad eeskätt tekkivad väidete mustrid (Brown 1980). Väidete sorteerimine võimaldas vilistlastel väljendada oma subjektiivseid positsioone erinevate avalikult levinud arusaamade suhtes. Lihtsustamaks väidete ladumise protseduuri, palusime uuritavatel esmalt jagada väited 3 rühma (väited, millega nõustutakse, millega ei nõustuta või mille suhtes ollakse neutraalsel seisukohal). Alles pärast esmast sorteerimist tuli väited laduda etteantud hinnanguskaalale. Hinnangute subjektiivsus oli tagatud skaala absoluutse nullpunkti puudumise ning avatud kommentaaride lisamise võimalusega. Näiteks, kui konkreetne uuritav soovis nõustuda etteantud kvoodist suurema arvu väidetega, siis toimus skaala algse neutraalse keskpunkti mõtteline nihutamine selliselt, et kõigile etteantud väidetele leiduks skaalal koht. Kõigile uuritavatele esitati täiendavaid avatud küsimusi nende tausta (nt varasema koolikogemuse), keeleoskuse jms kohta. Samuti paluti kõigil uuritavatel kommenteerida väiteid, millele nad olid andnud äärmisi hinnanguid (-5, -4, -3 ja +3, +4, +5). Iga uuritava poolt väidetele antud numbrilised hinnangud fikseeris intervjuuerija kirjalikult. Intervjuu helisalvestati ning salvestustest tehti sõna-sõnaline üleskirjutus. Venekeelse intervjuu jaoks konstrueeritud väited ning artikli tulemuste osas esitatud tsitaadid on tõlgitud eesti keelde.

**Pole üldse nõus**

**Täiesti nõus**

	-5	-4	-3	-2	-1	0	1	2	3	4	5
(3)											(3)
	(4)									(4)	
<b>Nõus</b> Arv ...		(5)							(5)		
<b>Neutraalne</b> Arv ...			(6)					(6)			
<b>Pole nõus</b> Arv ...				(8)			(8)				
											(11)

**Joonis 2.** Väidete sorteerimisel ette antud sagedusjaotus

Analüüsis kasutasime nii kvalitatiivseid tekstilisi andmeid kui ka väidete kvantitatiivseid hinnanguid. Intervjueeritavate poolt sorteeritud väidete analüüs keskendub kolmele statistilisele protseduurile – korrelatsioonide, faktorlaadungite ja faktorskooride arvutamisele. Esimene kasutab traditsioonilise faktoranalüüsiga sarnast arvutuskäiku, mille käigus leitakse üksikute Q-sorteerimiste vahelised seosed<sup>5</sup>. Kuid faktortelgede pööramise asemel kasutasime subjektiivsete tõlgenduste mustrite leidmiseks faktoranalüüsi tsentroidide meetodit ehk lihtsa summeerimise meetodit<sup>6</sup>. Kuigi peakomponentide meetodit on peetud matemaatiliselt täpsemaks (Cattell 1973), eelistatakse Q-metodoloogias tsentroidide meetodit teoreetilistel kaalutlustel (Stephenson 1953) – piiramatult hulga erinevate faktorlahendite hulgast ühe struktuuri väljavalimine võimaldab arvestada uuritavate subjektiivsete tõlgendusruumidega<sup>7</sup>. Konstrueeritud väidetele antud hinnangute kvantitatiivne analüüs tsentroidide meetodil võimaldab esile tuua omavahel seotud väidete “taga” olevaid subjektiivseid diskursuseid või tõlgendusmustreid<sup>8</sup>. Sellisel kombineeriti analüüsis nii kvantitatiivset – subjektiivsete positsioonide mustrite leidmine faktoranalüüsi meetodi abil – kui ka kvalitatiivset lähenemist, s.t kasutatud tsentroidide meetod

<sup>5</sup> Q-lähenemist on ekslikult peetud ka ‘ümberpööratud faktoranalüüsiks’, kus omaduste asemel analüüsitakse seoseid indiviidide vahel. Ümberpööratud andmematriksi ei pruugi aga vastata korrelatsiooni ning faktoranalüüsi lineaarsuse nõuetele (vt nt Brown 1980, McKeown, Thomas 1988). Lineaarsuse nõue on Stephensoni hinnangul (1953) täidetud juhtudel, mil arvestatakse mõõtmisühikute enesekohasusega (ingl *self-referent*), s.t andmematriksi tunnused on ükshaaval koondatud psühholoogilise olulisuse keskväärte ümber.

<sup>6</sup> Kuigi ka tsentroidide meetodil põhineva faktoranalüüsi korral kasutatakse faktorite arvu määramisel statistilisi kriteeriume (nt Kaiseri omaväärtuste ja Catelli joondiagrammi meetodid) (vt Hill, Lewicki 2006), on olulisemad siiski teoreetilised kaalutlused.

<sup>7</sup> Tsentroidide võib mõista kui keskväärte, mis iseloomustab seoseid kõikide ladumiste vahel, s.t tsentroid esindab vektorit, mis ulatub kõige kaugemale dimensioonini mitmemõõtmelises andmeruumis (vt nt Child 2006).

<sup>8</sup> Algselt nimetati Q-metodoloogias (Stephenson 1953, Brown 1980) analüüsi tulemusi faktoriteks. Hiljem on lähenemist kombineeritud kvalitatiivse diskursusanalüüsi põhimõtetega (vt nt Dryzek 2004) või on kvalitatiivse lähenemise eeskujul väidete koondkogumit nimetatud avalikuks diskursuseks ning leitavaid faktoreid individuaalseteks diskursusteks (vt nt Aalto 2003). Sellist rõhuasetuse muutust Q-metodoloogias on nimetatud nihkeks konstruktivismist (nt Stephensoni keskendumine uuritavate enesekohastele positsioonidele) konstruktsionismile (domineerivate sotsiaalsete tõlgenduste määratlemine konkreetses valdkonnas) (Watts, Stenner 2012: 42).

võimaldas uurijatel lisaks statistilistele kriteeriumitele arvestada kvalitatiivsete andmetega üksikute individuaalsete positsioonide omavahelisel võrdlemisel.

Käesolev analüüs on tehtud PQMethod tarkvara abil ning faktorstruktuuri valik on tehtud faktorite paaritise võrdleva pööramise teel. Lisaks faktorlaadungitele<sup>9</sup>, mis võimaldavad määratleda intervjueeritavate jaotumist diskursuste vahel, arvutatakse järgnevas analüüsis ka faktorskoorid<sup>10</sup>, mis võimaldavad iseloomustada faktori koosseisu määravaid väiteid skaala algsetes väärtustes<sup>11</sup>. Lisaks on artiklis kvalitatiivselt analüüsitud väidetele vabas vormis antud kommentaare ning vastuseid avatud küsimustele. Tekstiliste andmete analüüsimisel kasutasime tarkvara Maxqda. Kvalitatiivse analüüsi käigus keskendusime ühe alateema või väite tähendusliku mitmekesisuse uurimisele ning subjektiivsete positsioonide rühmade erinevuste ja sarnasuste selgitamisele.

### 3. Tulemused

Järgnevalt selgitame vilistlaste subjektiivsete positsioonide diskursuseid nii kvantitatiivselt – diskursuseid statistiliselt oluliselt eristavate väidete faktorskooride abil – kui ka kvalitatiivselt, toetudes intervjueeritavate poolt lisatud kommentaaridele.

Tabelis 1 on esitatud uuringus osalejate seotus iga leitud diskursusega kvantitatiivselt. Indiviidi seotust diskursusega näitab faktorlaadung – mida suurem see on, seda tugevamalt on indiviid diskursusega seotud. Mõne indiviidi arvamusi iseloomustab samaaegselt mitu diskursust (kõrge faktorlaadung mitmes faktoris, tabelis märgitud kursiivis). Kahe uuritava faktorkuuluvus on küll määratletav (kõrgeim faktorlaadung “keeleliste valikuvõimaluste tähtsustamise” faktoris), kuid pole hinnangute põhjal selgelt eristuv (statistiliselt madalad faktorlaadungid).

Järgnevates, diskursuste sisu analüüsivates tabelites (2–5) on esitatud ainult need väited, millele antud hinnangud eristasid diskursust statistiliselt oluliselt (kokku 36). Väljajäetud väidetest osa puudutas küsimusi, mille suhtes vilistlastel puudusid selged seisukohad (nt kutseharidus, põhikoolist väljalangemine (väide 35); elus vedamine võrreldes eakaaslastega (2); eestikeelsele õppele üleminekuga seotud võltsparaadlikkus (50)). Teiste väljajäetud väidete osas olid hinnangud kas suhteliselt üksmeelselt pigem nõustuvad (nt vene keele suuremale kasutusmugavusele osutamine (väide 12) või mure oma vene keele oskuse (33), etnilise identiteedi (46) ja keelelise diskrimineerimise pärast (56)) või pigem mitte-nõustuvad (nt rahvustevaheliste suhete paranemise küsimus (19), eesti ja inglise keele õppimise keerukuse võrdlus (54)).

<sup>9</sup> Faktorlaadungid on väärtused, mis väljendavad iga ladumise seoseid tsentroidiga ehk vektorite keskmisega.

<sup>10</sup> Arvutuslikult on faktorskooride (ingl *factor array*) korral tegemist faktorlaadungite kaalude summaga ridade lõikes, mis on omistatud igale analüüsiühikule indiviidi poolt, kes esindab antud faktorit:  $w = \frac{f}{1 - f_e}$ , kus  $f$  on faktorlaadung ning  $w$  on iga uuritava kaal (vt nt Spearman 1927). Faktorskoorid arvutatakse standardiseerituna ( $z$ -skooridena), kuid võimaldamaks võrdluseid eri diskursuste vahel, konverteeritakse täisarvudesse (-5...+5). Faktorskoore ei tule mõista mitte üksnes teatud ideaalse sorteerimisena (vt nt Aalto 2003: 81), vaid leiame, et need võimaldavad sisuliselt välja tuua ka erinevusi/sarnasusi ühe diskursuse sees ja mitme diskursuse vahel. Kuigi mitmetes uuringutes (vt nt Snelling 1999) on soovitatud keskenduda tõlgendustes eelkõige faktorskooride äärmuslikele väärtustele (nt +5, +4 ning -5 ja -4), on käesolevas uuringus arvestatud kõigi statistiliselt oluliselt eristuvate väidetega.

<sup>11</sup> Erinevalt Burti (1937) kasutatud indiviidide korrelatsioonide analüüsist keskendutakse käesolevalt Stephensoni (1953) eeskujul nii indiviidide kui ka uurimisinstrumendina kasutatud väidete samaaegsele analüüsile. Stephensoni Q-metodoloogia põhimõtteid on varemalt enam kasutatud poliitilise kommunikatsiooni (nt Aalto 2003) ja kommunikatsiooniteooria valdkondades (nt Grosswiler 1990). Vähem on lähenemist kasutatud meedia ja kommunikatsiooni (nt Singer 1998/1999) ning keelelise kommunikatsiooni uuringutes (nt Barchak, Marshall 1999).

**Tabel 1.** Faktorlaadungid eristuvate diskursuste ja uuritavate lõikes\*

Sugu	Piirkond	Diskursus			
		Vastuseis homogeniseerimisele	Rahulolu oma keeleoskuse ja integreeritusega	Keeleliste valikuvõimaluste tähtsustamine	Keeleline ja kultuuriline suletus
N	Eesti, Ida-Viru	<b>0.536</b>	0.051	0.487	0.131
M	Eesti, Tallinn	<b>0.531</b>	0.247	0.343	-0.321
N	Eesti, Tartu	<b>0.404</b>	0.026	0.191	0.078
N	Eesti, Tartu	<b>0.382</b>	0.296	0.300	0.012
N	Läti	0.018	<b>0.837</b>	0.088	0.012
N	Eesti, Tartu	-0.019	<b>0.766</b>	-0.029	0.098
N	Eesti, Tartu	0.227	<b>0.765</b>	-0.053	-0.021
N	Eesti, Tartu	0.139	<b>0.691</b>	0.248	-0.071
N	Eesti, Ida-Viru	0.203	<b>0.688</b>	0.129	0.015
M	Läti	0.014	<b>0.639</b>	0.170	-0.077
M	Läti	0.148	<b>0.631</b>	0.084	0.235
N	Eesti, Ida-Viru	-0.097	<b>0.611</b>	0.127	0.154
N	Läti	0.115	<b>0.609</b>	0.232	-0.146
M	Läti	0.144	<b>0.568</b>	0.163	0.072
N	Läti	0.087	<b>0.500</b>	0.430	-0.077
M	Eesti, Ida-Viru	-0.102	<b>0.488</b>	0.375	-0.036
M	Eesti, Tallinn	0.184	<b>0.424</b>	0.252	0.148
M	Läti	0.032	0.157	<b>0.683</b>	-0.019
M	Eesti, Ida-Viru	-0.037	0.255	<b>0.682</b>	-0.108
N	Eesti, Tartu	0.081	0.167	<b>0.618</b>	0.018
N	Eesti, Tartu	0.218	0.333	<b>0.608</b>	0.154
M	Eesti, Ida-Viru	-0.031	0.200	<b>0.588</b>	-0.136
M	Eesti, Ida-Viru	-0.034	0.343	<b>0.587</b>	-0.106
N	Eesti, Ida-Viru	-0.015	0.230	<b>0.584</b>	-0.107
N	Eesti, Tallinn	-0.020	0.305	<b>0.581</b>	0.120
M	Eesti, Ida-Viru	0.038	0.408	<b>0.561</b>	0.100
M	Eesti, Tartu	0.105	0.294	<b>0.530</b>	0.107
M	Eesti, Tallinn	0.001	0.488	<b>0.527</b>	0.155
N	Eesti, Tartu	-0.001	0.104	<b>0.523</b>	0.296
N	Eesti, Tallinn	0.089	0.262	<b>0.480</b>	-0.194
N	Läti	0.416	0.076	<b>0.476</b>	0.009
N	Läti	0.405	0.109	<b>0.469</b>	0.041
N	Eesti, Ida-Viru	0.209	-0.220	<b>0.469</b>	-0.034
N	Eesti, Tartu	0.386	-0.420	<b>0.443</b>	-0.122
N	Läti	0.376	0.246	<b>0.440</b>	0.235
N	Eesti, Tallinn	0.139	0.394	<b>0.383</b>	0.186
N	Eesti, Ida-Viru	-0.225	0.226	<b>0.376</b>	0.015
N	Eesti, Tallinn	0.006	0.096	<b>0.337</b>	-0.143
M	Eesti, Tallinn	0.164	0.203	0.298	0.131
N	Eesti, Tallinn	-0.020	0.043	0.267	0.153
N	Eesti, Tallinn	0.120	-0.134	0.204	<b>0.462</b>
M	Eesti, Ida-Viru	0.177	0.346	0.306	<b>-0.439</b>

\* Statistiliselt olulised faktorlaadungid on märgitud tumedas kirjas (antud juhul faktorlaadungid, mis ületavad väärtuse 0,325 [ $p > 0.01$ ]). Mitmes diskursuses statistiliselt olulised faktorlaadungid on märgitud kursiivis.

Lisaks väidetele antud hinnangute ja indiviidide faktorkuuluvuse kvantitatiivsele analüüsile tsiteerime diskursust iseloomustavaid kommentaare, mida uuritavad vabas vormis lisasid. Tsitaatides on nurksulgudes märgitud kommenteeritud väite number. Samuti on iga tsitaadi juures märgitud uuritava sotsiolingvistiline kuuluvus.

### 3.1. Keeleliste valikuvõimaluste tähtsustamine

See diskursus on uuritavate poolest kõige arvukam, selle defineerimises osales 21 intervjueeritut (vt tabel 1). Ühtlasi iseloomustab diskursust regionaalne mitmekesisus – intervjueeritavad olid võrreldaval määral pärit Ida-Virumaalt (7), Tallinnast (5), Tartust (5) ning Lätist (4). Tabelis 2 on esitatud diskursust eristavate väidete standardiseeritud faktorskoorid.

**Tabel 2.** Diskursust “keeleliste valikuvõimaluste tähtsustamine” eristavad väited\*

Väite nr	Väide	Faktorskoor
59	<i>Paljud probleemid oleksid lahendatud, kui Eestis (Lätis)<sup>1</sup> hakkaks olema kaks riigikeelt – eesti (läti) ja vene keel</i>	5
61	<i>Minu arust on inglise keel venelastele palju vajalikum kui eesti (läti) keel</i>	3
52	<i>Ma arvan, et kui minust saab mingi valdkonna väga hea spetsialist, siis jõuan elus edasi ka ilma eesti (läti) keeleta</i>	2
44	<i>Peaks olema nii, et kui soovid, võid valida 60% aineid eesti (läti) keeles, aga kui ei soovi, võid jätkata õppimist emakeeles</i>	1
63	<i>Tänu sellele, et valdan nii vene kui ka eesti (läti) keelt, on mul rohkem võimalusi saavutada edu kui nendel minu eakaaslastel (eestlastel (lätlastel) või venelastel), kes valdavad ainult üht keelt</i>	1
7	<i>Eesti (läti) keele õppimisel aitas mind väga eesti (läti) meedia kasutamine ning võimalus kasutada eesti (läti) keelt erinevates olmesituatsioonides</i>	1
58	<i>Nendes ainetes, mida õppisin koolis eesti (läti) keeles, pühendati liiga palju tähelepanu keelele ja liiga vähe ainele</i>	-2

\* Tabelites 2–5 on esitatud vaid statistiliselt oluliselt eristuvad väited olulisuseniivool  $p \leq .05$ .

Eristuvate väidete analüüs näitab, et selle diskursuse eripäraks on **keeleliste valikuvõimaluste tähtsustamine**. Viimane väljendub eelkõige nõusolekus kahe riigikeele – eesti/läti ja vene – vajalikkuse väitega (59) ning väitega inglise keele suuremast instrumentaalsest kasulikkusest võrreldes riigikeelega (61). Kahe riigikeele kasulikkusega nõustunud intervjueeritavate hinnangul oleks kahe riigikeele kehtestamine õiglane ning võimaldaks tagada rahvuste võrdsuse hariduses ja töösfääris, aga muudaks vähem tähtsaks ka ülemineku riigikeelsele aineõppele. Seejuures ei eitatud praeguse riigikeele oskamise tähtsust, nagu näites (1).

- (1) [59] --- Ma arvan, et jah, paljud probleemid oleksid lahendatud --- need-samad reformid. Ma arvan, et võiks kuidagi --- sellele mitte tähelepanu pöörata. --- paljudes riikides on kaks riigikeelt, isegi rohkem, kolm – kus vähemuste protsentuaalne osakaal on palju väiksem kui meil Eestis. Nii et ma arvan, et see küsimus on väga aktuaalne. Kaks riigikeelt --- lahendaks

<sup>12</sup> Läti vilistlastele esitatud väidetes olid Eestile või eesti keelele osutavad sõnad asendatud Lätile või läti keelele osutavatega.



väga palju mingeid selliseid probleeme, arusaamatuid olukordi. --- Kõikjal on muidugi vaja eesti keele oskust ---. (Narva)

Ühe noore hinnangul vähendanuks kaks riigikeelt eestlaste ja venelaste vahelisi pingeid – kuid samas ta mõõnis, et üks riigikeel võib olla mõistetav riikliku ühtekuuluvustunde tagajana (näide 2).

- (2) [59] --- probleemid oleksid tõepoolest lahendatud, vähemalt ei oleks võib-olla sellist otse konfrontatsiooni ses mõttes, et venelased ja eestlased. Aga sellisel juhul peab venelaste poolt olema mõistmine, et nad elavad Eestis, see on eraldi riik, oma poliitika, oma väärtuste, oma kultuuriga. Ja mida ma väga sageli tähendan venelaste poolt, [on] selline suhtumine, et pidevalt surutakse Venemaa poole. Venelased peavad mõistma, et nii ei tule teha. Riik ja eestlased võiksid ise valida, mida nad tahavad teha ja mida nad tahavad, et riigis oleks. See tähendab, [see on] võib-olla peamine probleem, miks praegu ei saa lubada seda, et riigis on kaks riigikeelt. (Tartu)

Kui riiklikest valikutest rääkides toodi ühtviisi välja nii instrumentaalseid kui ka integratiivseid argumente, siis riiklike valikute seostamisel omaene keeleliste väärtustega oli fookus pigem instrumentaalne (näited 3 ja 4).

- (3) [61] --- Ma arvan, et inimene peab ise valima [kas tema jaoks on tähtsam inglise või eesti keel]. Ma arvan, mis minusse puutub, siis antud hetkel on eesti keel tähtsam kui inglise. Kuna ma töötan, töö on vaja eesti keelt, ja ma elan siin. (Narva)
- (4) [61] --- Igal juhul, kui inimene tegeleb tõsise asjaga, siis tal on vaja inglise keele oskust, ilma selleta ei sõida sa kuskile. Jällegi, kõige tavalisem info-otsing Internetis --- (Tallinn)

Ühtlasi oli sellele diskursusele iseloomulik **kriitilisus kooli riigikeelsele aineõppele ülemineku suhtes, rõhutades õpilaste valikuvõimaluse tähtsust** (väide 44). Samas ei nõustunud ohuga, et riigikeelses õppes pööratakse liiga suurt tähelepanu keelele ja liiga väikest ainele (väide 58). Mittenõustumine selle väitega oli osal juhtudel seotud vähese riigikeelse aineõppe kogemusega koolis ja ülikoolis (st väide vastajat otseselt ei puudutanud), osal juhtudel aga ei nõustunud väite endaga. Seega oli vilistlaste argumentatsioonide rõhuasetus siin pigem õpilaste keelelis-kultuurilistel valikuvõimalustel, mitte aga konkreetsetel teiskeelse aineõppe meetodikaga seotud valikutel, nagu näites (5).

- (5) [44] Ma leian, et muidugi peab valik jääma õpilase teha. Kui ta tahab, et need ained oleksid, et neid õpetataks eesti keeles, siis las ta väljendab oma soovi, et ma tahan, ja las siis õpib. – Aga kui tal ei ole soovi? – Siis las jätkab õppimist emakeeles, vene keeles. (Narva)

Väitega 44 nõustumine on selles diskursuses sarnane allpool peatükis 3.3 esitatud diskursusega “vastuseis homogeniseerimisele”. See-eest erinevalt “keelelise ja kultuurilise suletuse” diskursusest, kus väljendati muuhulgas vähest valmidust kasutada formaalharidussüsteemiväliseid võimalusi riigikeele õppimisel (nt meedia, olmelised kontaktid), ilmnis selles diskursuses ka vastavate alternatiivide nägemine riigikeele õppimisel.

### 3.2. Rahulolu oma keeleoskuse ja integreeritusega

See diskursus on seda defineerivate vilistlaste arvu poolest teisel kohal (13 uuritavat). Diskursuse koosseisus on ülekaalus Läti noored (6), kuid esindatud on ka Tartu (3), Ida-Viru (3) ja Tallinna (1) vilistlased.

**Tabel 3.** Diskursust “rahulolu oma keeleoskuse ja integreeritusega” eristavad väited

Väite nr	Väide	Faktorskoor
55	<i>Tänu eesti (läti) keele oskusele tunnen, et olen Eesti (Läti) ühiskonda integreeritud paremini kui need minu eakaaslased, kes oskavad keelt halvasti</i>	5
28	<i>Olen rõõmus, et sain koolis mitte-emakeeles õppimise kogemuse</i>	4
10	<i>Tänu eesti (läti) keele oskusele avaneb noortel võimalus osaleda aktiivselt ühis(konna)elus</i>	4
53	<i>Eesti (läti) keele omandamise seisukohalt eestikeelne (lätikeelne) aineõpe täiendas keeletunde</i>	3
13	<i>Eesti (läti) keele oskus on minu jaoks väga tähtis</i>	2
48	<i>Nende ainete eestikeelne (lätikeelne) õpetamine, mida õppisin omal ajal koolis eesti (läti) keeles, oli täiesti põhjendatud</i>	2
30	<i>Üleminek eesti (läti) õppekeelele gümnaasiumis ei olnud minu jaoks nii keeruline, kuna mind toetas perekond ja head õpetajad</i>	2
36	<i>Õpilase enda huvi ja pere toetus on tähtsamad kui varustatud klassiruum ja hea õpetaja</i>	0
14	<i>Eestikeelsele (lätikeelsele) õppele ülemineku tulemusel vähenevad pinged eestlaste (lätlaste) ja venelaste vahel</i>	0
59	<i>Paljud probleemid oleksid lahendatud, kui Eestis (Lätis) hakkaks olema kaks riigikeelt – eesti (läti) ja vene keel</i>	0
51	<i>Üleminek eestikeelsele (lätikeelsele) õppele võib olla vajalik, kuid mitte sellise kava alusel ja sellisel viisil, nagu seda tehakse praegu</i>	-1
3	<i>Paljudel ei lähe eesti (läti) keelt pärast kooli enam tarvis</i>	-2
60	<i>Eestikeelsele (lätikeelsele) õppele ülemineku sunniviisilisus ei too kellelegi head, vaid pigem tekitab pingeid ja vastuseisu noorte seas</i>	-3
58	<i>Nendes ainetes, mida õppisin koolis eesti (läti) keeles, pühendati liiga palju tähelepanu keelele ja liiga vähe ainele</i>	-3
9	<i>Väide, et ainete õpetamine eesti (läti) keeles tõstab õpilaste konkurentsivõimet, on absurdne</i>	-4
37	<i>Eesti (läti) keele sildi all pööratakse vene lapsi eesti (läti) usku – vene mentaliteedi asemel sunnitakse neile peale eesti (läti) oma</i>	-5

Tabelis 3 esitatud diskursust eristavate väidete faktorskooridest ilmneb, et diskursuse iseloomulikuks jooneks on **avalike/riiklike keelepoliitiliste rõhuasetustega nõustumine**. Argumentatsioonides tuginetakse isiklikele positiivsetele kogemustele. Näiteks nõustusid seda diskursust defineerivad vilistlased riigikeele tähtsusega nii üldiselt (nähes riigikeele oskuses koostegemise ja osalemise võimaldajat, näide 6) kui ka iseendasse puutuvalt (öeldes, et riigikeele oskus on toetanud nende integreerumist ühiskonda, näide 7).

- (6) [10] Ma olen täiesti nõus sellega, et nad saavad võtta osa mingitest üritustest, mis on ainult eesti keeles --- st suhelda inimestega, mitte ainult oma ringkonnaga, ütleme maailma tundma õppida. (Tartu)

- (7) [55] Jah, ma olen rohkem integreeritud eesti ühiskonda, kuna ma suhtlen valdavalt eestlastega. Need minu sõbrad, kes räägivad eesti keelt halvasti, justkui oskavad seda gümnaasiumi tasemel, või ütleme, selle kursuse [tasemel], millel nad õpivad, nad tulevad töödega suurepäraselt toime, kuid neil ei ole suhtlemist, ning sellest tulenevalt on neil raske ületada suhtlusbarjääri. Ja seetõttu nad ei ole eesti ühiskonda nii hästi integreeritud. Eestikeelne õpe [ülikoolis], see ei tähenda, et me oleme juba Eesti ühiskonnas. (Kiviõli)

Ühtlasi nõustusid seda diskursust defineerivad noored riigikeeleoskuse positiivsete instrumentaalsete mõjudega individuaalse elutee, s.t kas ametialase või muu kasutusvajaduse seisukohalt (mittenõustumine väitega 3). Sarnaselt teiste diskursustega seostati keeleoskuse tähtsust ka siin indiviidi tulevase elukohamaaga (näide 8).

- (8) [3] --- Praegu sõidavad inimesed väga sageli välismaale. --- Mul väga paljud minu sõbrad, klassikaaslased on sõitnud õppima välismaale, peamiselt Inglismaale. Ja no seal nagu väga ei ole läti keelt tarvis, kuigi niipalju kui ma tean, nad loevad sedasama Delfit, et olla kogu aeg kursis. (Läti)

Teine sellele diskursusele iseloomulik joon on **positiivsete kogemuste rõhutamine seoses riigikeeles õppimisega koolis**. Rahulolu väljendati nii üldiselt (nõustumine väitega 28) kui ka konkreetsemalt seoses riigikeelsele õppele üleminekuga (vastavalt nõustumine väitega 53 ja mittenõustumine väidetega 9, 51). Üldist rahulolu toetas nii vastanute koolikogemus tasakaalustatud keele- ja aineõppega (näide 9; mittenõustumine väitega 58) kui ka usk kultuuriliste eripärade arvestamise teiskeelses õppes (näide 10; mittenõustumine väitega 37).

- (9) [58] See on täielik lollus. Õpetajad, kes neid aineid õpetavad, ja need, kes kogu selle süsteemi, ütleme, välja mõtlevad, mulle tundub, paigutavad keeleõppe ja, ütleme, aineõppe, võrdsetele positsioonidele. Seal ei ole sellist ülekaalu, et kas rohkem tähelepanu läheb keelele või ainele. Seda õpitakse kõike koos, mulle tundub. (Tallinn)
- (10) [37] Siin ma kindlasti ei ole nõus. Pööratakse eesti usku... Mis mentaliteet on inimesel... ma ei saa sulle öelda... Siiski esimesed seitse aastat... ta kasvab ju vene perekonnas. Ja see mentaliteet on tal seitsme aastaga juba välja kujunenud. --- (Kiviõli)

Positiivset suhtumist seoses riigikeelse õppega soosis nii koduse kui ka õpetajapoolse toe olemasolu (vt väide 30); vabas vormis lisatud kommentaarides rõhutati aga eelkõige õpetajatepoolse toetuse tähtsust, kuna erinevalt perekonnast omatakse õppetöö läbiviimiseks vajalikke metodoloogilisi teadmisi (näide 11).

- (11) [30] Sellega ma võin nõustuda. Jah, tõepoolest, õpetajad aitasid sellega, et nad ei esinenud, et vat, peab olema 80% läti keeles. Minutite mõõtmist, seda ei olnud. Õpetajad ju, nad mõistavad, mõistavad siiski õpilaste huve. Ja formaalselt nad ei lähenenud sellele küsimusele, nagu näiteks ametnikud, kes selle kõik välja mõtlesid. Meil on õpetajad efektiivsemad, muidugi, kuigi neil ei ole metodoloogiat, nagu tavaliselt, aga [neil] on pea, nad teavad, mida teha. Just seetõttu üleminek ei olnud, noh, nii raske, nagu oleks võinud olla. No aga see, et perekond --- perekonnal ei

ole sellist funktsiooni – aidata õppes, eriti, kui [õpilane on] juba piisavalt suur – 8., 9., 10. klass. --- Pere on muu, mitte õppimise jaoks. (Läti)

Seda diskursust võib pidada üheks paremini eristuvaks, kuna teiste diskursustega kattuvaid ja statistiliselt oluliselt eristuvaid väiteid on vaid üks (55): sarnaselt diskursusega “vastuseis homogeniseerimisele”, nõustuti ka siin (kuid veelgi tugevamat) väitega, et tänu riigikeele oskusele ollakse ühiskonda paremini integreeritud kui keelt mitteoskavad eakaaslased. Kahte diskursust eristab väite 55 osas riikliku integratsioonipõhimõtte tõlgendus – kui selles diskursuses on keeleoskus ja rahulolu isikliku kooli- ja elukogemusega viinud integreeritusetundeni, siis diskursuses “vastuseis homogeniseerimisele” väljendati pigem vastupanu põhimõtte “keel kui integratsiooni eeltingimus” suhtes.

### 3.3. Vastuseis homogeniseerimisele

Selle diskursuse defineerimises osalenute hulk on kahe eelneva diskursusega võrreldes tagasihoidlikum (4), kuid samas on selles esindatud kõik kolm Eesti sotsiolingvistilist piirkonda (kaks esindajat Tartust, üks Tallinnast, üks Ida-Virumaalt).

**Tabel 4.** Diskursust “vastuseis homogeniseerimisele” eristavad väited

Väite nr	Väide	Faktorskoor
18	<i>On raske mõista, mida halba oli praeguses Eesti (Läti) haridussüsteemis, kus külg külje kõrval on olemas nii eesti (läti) kui ka vene õppekeelega koolid</i>	5
44	<i>Peaks olema nii, et kui soovid, võid valida 60% aineid eesti (läti) keeles, aga kui ei soovi, võid jätkata õppimist emakeeles</i>	4
55	<i>Tänu eesti (läti) keele oskusele tunnen, et olen Eesti (Läti) ühiskonda integreeritud paremini kui need minu eakaaslased, kes oskavad keelt halvasti</i>	2
6	<i>Eestikeelse (lätikeelse) ainete õppimise tagajärjel minu aineteadmised kannatasid</i>	0
1	<i>Kui inimene elab Eestis (Lätis), peab ta oskama eesti (läti) keelt</i>	0
29	<i>Igapäeva elus ei ole mul läinud tarvis seda sõnavara, mille omandasin koolis, õppides aineid eesti (läti) keeles</i>	0
52	<i>Ma arvan, et kui minust saab mingi valdkonna väga hea spetsialist, siis jõuan elus edasi ka ilma eesti (läti) keeleta</i>	-1
26	<i>Koolis ma kartsin eestikeelset (lätikeelset) aineõpet ja minu hirm osutus põhjendatuks</i>	-1
31	<i>Kõige tähtsam on see, et aine oleks huvitav – kui aine ja teema on huvitavad, muutub õppekeel teisejärguliseks</i>	-3
20	<i>Koolis ma alguses natuke kartsin õppida aineid eesti (läti) keeles, kuid minu hirm osutus põhjendatuks</i>	-4

Seda diskursust eristab teistest fokuseeritum **kriitilisus haridussüsteemis toimuva ehk eestikeelsele aineõppele ülemineku suhtes**. Kriitikat riigikeelse aineõpetuse suhtes argumenteeriti näiteks asjaoluga, et teises keeles aine õppimine võib veelgi vähendada õpilaste niigi tagasihoidlikku õpimotivatsiooni (näide 12). Rõhutati õpilaste passiivsust õppeprotsessis tervikuna ning seeläbi suurema

interaktiivsuse vajadust eestikeelses aineõppes. Viimasest tulenevalt pooldasid sellesse diskursusse kuuluvad Eesti vilistlased suhteliselt üksmeelselt olemasoleva koolisüsteemi säilimist (näide 13).

- (12) [31] No lastel pole nii võimalik, et kõik lihtsalt on huvitav, et nad ise kuidagi kõike õpiksid, kõike teaksid, ei ole midagi sellist. Mitte keegi ei õpi, mitte keegi ei taha õppida, kõik kirjutavad maha, ja üldse. (Tartu)
- (13) [18] No jah, raske mõista, miks nad lihtsalt ei jäta seda küsimust rahule, ja las olla nagu viis aastat tagasi, venelased vene koolis, eestlased eesti omades. Need, kes tahavad minna õppima eesti keeles, tee on lahti, kui on selline soov. Ei oleks sellist probleemi, et vene õpetajad peavad koolist lahkuma. Oleks neil jälle nii koht kui ka töö. (Narva)

Skeptilisus eestikeelse aineõpetuse suhtes on sellele diskursusele omane vaatamata diskursust defineerivate vilistlaste suhteliselt kõrgele hinnangule oma riigikeeleoskuse suhtes. Tänu oma riigikeele oskusele peeti omaenda ühiskondlikku integreeritust suhteliselt heaks (väite 55 faktorskoor 2). Siiski väärtustati riigikeele oskuse asemel kõrgemalt ametialaseid teadmisi (väite 52 faktorskoor -1). Keeleoskuse negatiivse hindamise taga võisid olla ka hirmud seoses eestikeelse aineõpetusega (väite 20 faktorskoor -4) ning kahtlused ainealase sõnavara vajalikkuses pärast kooli (väite 29 faktorskoor 0; näide 14). Selliselt on keeleoskuse argumentatsioonide fookus selles diskursuses eelkõige integratiivsetel aspektidel. Viimast kinnitab ka asjaolu, et instrumentaalset ainealast sõnavara ning asukohamaast tuleva keeleõppimise kohustust puudutavate väidete osas puudusid selle diskursuse liikmetel selged seisukohad (väidete 1, 6 faktorskoor 0).

- (14) [29] Ei, geograafiline sõnavara, mida ma sellega veterinaarias, vabandage väga? (Tartu)

Ka sellel diskursusel oli kahe teise diskursusega kattuvaid eristavaid väiteid. Sarnaselt diskursusega “keeleliste valikuvõimaluste tähtsustamise” (kuid veelgi tugevamalt) väljendati vastuseisu eestikeelse aineõppe üldkohustuslikkusele (nõustumine väitega 44, faktorskoor 4). Sarnaselt diskursusega “rahulolu oma keeleoskuse ja integreeritusega” nõustuti (küll mõnevõrra nõrgemalt) oma riigikeele oskuse ja ühiskondliku integreerituse seosega (väite 55 faktorskoor 2). Kokkuvõtteks võib öelda, et seda diskursust defineerivad vilistlased tähtsustasid suuremat paindlikkust nii keelepoliitikas üldiselt (keeleoskuse nõude üleüldisuse kahtluse alla seadmine) kui ka riigikeelsele aineõppele ülemineku osas (suuremate valikuvõimaluste soovimine koolidele ja õpilastele). Kuigi põhimõtteliselt nõustuti integratsiooni kui sellisega, kaheldi riigikeelsele aineõppele ülemineku praeguse korralduse sobivuses.

### 3.4. Keeleline ja kultuuriline suletus

Viimane analüüsis eristuv diskursus on arvuliselt väikseim, sisaldades ühte uuritavat Tallinnast ja ühte Ida-Virust. Kuigi kvantitatiivselt oli Ida-Viru intervjueeritav selle diskursusega seotud negatiivselt (vt tabel 1), erinesid nende kommentaarides ilmnevad seisukohad üksnes osaliselt. Vaatamata seda diskursust defineerivate indiviidide vähesele hulga on statistiliselt eristuvate väidete hulk üks suuremaid.

**Tabel 5.** Diskursust “keeleline ja kultuuriline suletus” eristavad väited

Väite nr	Väide	Faktorskoor
38	<i>Alguses ma tahtsin õppida aineid eesti (läti) keeles koolis, aga pärast pettusin</i>	5
6	<i>Eestikeelse (lätikeelse) ainete õppimise tagajärjel minu aineteadmised kannatasid</i>	4
37	<i>Eesti (läti) keele sildi all pööratakse vene lapsi eesti (läti) usku – vene mentaliteedi asemel sunnitakse neile peale eesti (läti) oma</i>	4
3	<i>Paljudel ei lähe eesti (läti) keelt pärast kooli enam tarvis</i>	4
29	<i>Igapäevaelus ei ole mul läinud tarvis seda sõnavara, mille omandasin koolis, õppides aineid eesti (läti) keeles</i>	3
40	<i>Üleminek eestikeelsele (lätikeelsele) õppekeelele lähtus nii koolide kui ka riigi poolt peamiselt majanduslikest põhjustest</i>	3
26	<i>Koolis ma kartsin eestikeelset (lätikeelset) aineõpet, ja minu hirm osutus põhjendatuks</i>	2
34	<i>Tahaksin õppida erinevaid keeli, kuid mitte eesti (läti) keelt. Eesti (läti) keele suhtes tunnen antipaatiat</i>	0
27	<i>Mul ei ole erilisi eeldusi keelte omandamiseks, seetõttu läheb mul ka eesti (läti) keele õppimine raskustega</i>	-1
43	<i>Pean ennast kultuuriliselt avatuks ja sallivaks inimeseks, st minu jaoks ei kujuta erinevast rahvusest inimestega suhtlemine erilist raskust</i>	-2
15	<i>Oskan veel mõnda keelt (näiteks inglise keelt) piisaval tasemel, et kasutada seda igapäevaelus</i>	-3
32	<i>Eesti (läti) keele õppimine andis mulle võimaluse kokkupuuteks eesti (läti) kultuuriga. Iga haritud inimene peab teadma selle riigi kultuuri, kus ta elab</i>	-3
7	<i>Eesti (läti) keele õppimisel aitas mind väga eesti (läti) meedia kasutamine ning võimalus kasutada eesti (läti) keelt erinevates olmesituatsioonides</i>	-4
53	<i>Eesti (läti) keele omandamise seisukohalt, eestikeelne (lätikeelne) aineõpe täiendas keeletunde</i>	-4
21	<i>Olen valmis suhtlema nii vene kui ka eesti (läti) keeles, ning minu sõprade ja tuttavate hulgas on nii venelased kui eestlased (lätlased)</i>	-5
63	<i>Tänu sellele, et valdan nii vene kui ka eesti (läti) keelt, on mul rohkem võimalusi saavutada edu kui nendel minu eakaaslastel (eestlastel (lätlastel) või venelastel), kes valdavad ainult üht keelt</i>	-5

Peamise erinevusena võrreldes teiste diskursustega seostub **eestikeelsele õppele ülemineku vastasus** siin teatud **isikliku ilmajäetustundega**. Negatiivseid hinnanguid seoses riigikeelse õppega põhjendati nii isikliku pettumuse (nõustumine väitega 38) kui ka konkreetsete aine- ja keelealaste teadmiste saamatajäämisega (nõustumine väitega 6, väite 53 eitamine; näited 14 ja 15).

- (14) [6] Jah, nii oli. Sest alguses ma ei saanud millestki aru. --- st ei saanud täielikult aru. Lihtsalt raske oli süveneda nendesse eestikeelsetesse sõnadesse. (Narva)
- (15) [6] Seesama kunstiõpetus – meile anti mingid küsimused, me tõlkisime need ära, siis panime kokku vastused ja tõlkisime jälle, nii et mis teadmised seal. Me panime suurema rõhu keelele ja, otse öeldes, kirjutasime kõik õpikust maha. Selle tulemusena kannatasid minu aineteadmised. (Tallinn)

Ühtlasi olid need vilistlased kriitilised omandatud keeleoskuse ja erialase sõnavara hilisemate rakendusvõimaluste suhtes (väite 29 faktorskoor 3). Lisaks kooliga seotud isiklike kogemustele on nende kahe noore keelelis-kultuuriline suletus selgitatav ka eestikeelsele õppele ülemineku eesmärkide üldise tajumisega – nad leidsid, et õppekelevahetus ei arvesta õpilaste keelelis-kultuurilise mentaliteediga (nõustumine väitega 37) ning üleminek põhineb pigem formaalsetel ja majanduslikel kaalutlustel (nõustumine väitega 40).

Teiseks selle diskursuse eristajaks on **suhteliselt tagasihoidlik individuaalne keelelis-kultuuriline avatus**. Väiteid vabas vormis kommenteerides pidasid vilistlased üldiselt (normatiivselt) kultuurilist avatust ja sallivust oluliseks, kuid personaalset kogemust selgitades eelistati ühekeelseid ja -kultuurilisi valikuid (näide 16; mittenõustumine väidetega 21, 43). Vähese keelelis-kultuurilise avatuse taga näib antud juhul olevat nii isiklike keeleliste ressursside piiratus (väide 15) kui ka teatud hirm võõrkeelse suhtluse ees (näide 16). Hirmu teises keeles rääkimise ees väljendati vaatamata eri keeltes rääkimise varasemale kogemusele. Samas nõustuti keeleõppimise võimalusega keelekeskkonda sattudes, vaatamata algselt tajutud ebakindlusele.

- (16) [21] Ei, ma pigem eelistan suhelda vene keeles kui eesti keeles, sellepärast ei nõustu. Vististi hirm – samamoodi nagu mul oli ka inglise keelega. Küllap ei ole keel mulle väga tuttav. Näiteks ma käisin Rootsis tööl, ja seal tuli mul suhelda inglise keeles, kuigi siin ma seda keelt ei rääkinud. Selle tulemusena hakkasin suhtlema. Kui praegu satuksin keskkonda, kus vene keeles ei räägita, hakkaksin samamoodi rääkima. (Narva)

Lisaks vahetute teiskeelsete kontaktide vähesusele iseloomustavad seda diskursust ka suhteliselt piiratud teiskeelsed meediakontaktid (mittenõustumine väitega 7). Individuaalse suletuse ja eestikeelses õppes pettumuse taga võis olla ka isiklikest pragmaatilistest põhjendustest tulenev eripärane olukorrataju (nt ühel juhul ebaõnnestumine eestikeelsetes õpingutes ning eesseisnud Venemaale-kolimine ning teisel juhul kogemus tööintervjuul, kus oli ilmnunud keeleoskuste eelistamine muudele pädevustele).

Lisaks eelkirjeldatud väidetele on sellel diskursusel teiste diskursustega sarnasel määral statistiliselt eristuvaid väiteid, mille nõusoleku suund oli aga erinev. Näiteks erinevalt diskursusest “vastuseis homogeniseerumisele” nõustuti tugevamalt väidetega, mis puudutavad negatiivset kogemust eestikeelse õppega (väited 6, 26, 29). Võrreldes diskursusega “rahulolu oma keeleoskuse ja integreeritusega”, kus ohtu kultuurilisele identiteedile ei tajutud, nõustusid selle diskursuse liikmed väitega 37. Kui “keeleliste valikuvõimaluste tähtsustamise” diskursuses peeti igapäevaseid vahetuid ja vahendatud (meedia)kontakte keeleõppimist toetavaks (väide 7), siis antud diskursuses sellega ei nõustunud.

## 4. Kokkuvõte ja arutelu

Käesolev artikkel keskendus riigikeelse õppega seotud positsioonide selgitamisele venekeelsete gümnaasiumite vilistlaste hulgas. Analüüsis eristus neli riigikeelsele õppele üleminekut iseloomustavat diskursust: “keeleliste valikuvõimaluste tähtsustamine”, “rahulolu oma keeleoskuse ja integreeritusega”, “vastuseis homogeniseerimisele” ning “keeleline ja kultuuriline suletus”.

Artiklis kasutatud Q-metodoloogia võimaldas analüüsida vilistlaste sotsiolingvistilise konteksti erinevate aspektide rolli subjektiivsete positsioonide kujunemisel. Analüüs toetab varasemate uuringute tulemusi (Masso, Soll 2013), mille kohaselt traditsioonilise kvantitatiivse lähenemise ohuks võib olla kontekstuaalsete tegurite alahindamine riigikeelsele õppele ülemineku valmisoleku hindamisel. Käesolevas analüüsis ilmnis kolme diskursuse ühisjoonena (“rahuolu oma integreeritusega”, “vastuseis homogeniseerimisele”, “keeleliste valikuvõimaluste tähtsustamine”) fokuseerumine riiklikke integratsioonipõhimõtteid ja keelepoliitika ühiskondlikke väljundeid puudutavatele argumentidele. Seevastu “keelelise ja kultuurilise suletuse” diskursus keskendus kooli ja indiviidiga seotud väidetele. Analüüsis ilmnis, et igapäevane sotsiolingvistiline keskkond võib soosida teatud orientatsioone – näiteks kakskeelse olmekeskonnaga harjunud Tallinna ning Ida-Virumaa valdavalt venekeelses keskkonnas üleskasvanud vilistlastele olid olulised keelelised valikuvõimalused. Samuti kaldusid Läti vilistlased väljendama rahulolu oma keeleoskuse ja integreeritusega. Siiski ei pruugi sotsiolingvistiline keskkond ning individuaalsed valikud olla alati üheses seoses – rahulolu enda Eesti ühiskonda integreeritusega väljendasid erineva sotsiolingvistilise kogemusega noored, käesolevas analüüsis nt Tartu ja Ida-Virumaa päritoluga vene õppekeelelega gümnaasiumite vilistlased. Avalikkuses suhteliselt homogeense sotsiolingvistilise kontekstina konstrueeritud Ida-Virumaalt pärit vilistlaste positsioonid seoses riigikeelsele õppele üleminekuga jaotusid eri diskursuste vahel kõige mitmekesisemalt.

Analüüs näitas Q-metodoloogilise diskursusanalüüsi mitmekesiseid võimalusi peamiselt kvalitatiivse diskursusanalüüsi kontekstis. Artikli uurimisobjekti – sotsiolingvistiliselt kompleksse teema ehk riigikeelsele õppele ülemineku – puhul seisnis Q-metodoloogilise uurimuse eelis eelkõige selle holistlikus lähenemises (vt nt Watts, Stenner 2012). Puht kvantitatiivses analüüsis võivad indiviidide arvu poolest vähe esindatud, kuid tähenduslikult eristuvad individuaalsete positsioonide mustrid (nt “keeleline ja kultuuriline suletus”) jääda tähelepanuta. Q-metodoloogia kvalitatiivne komponent ehk konstrueeritud 63 väitele spontaanselt lisatud kommentaaride analüüs võimaldas iga indiviidi subjektiivseid tõlgendusi tähenduslikult täpsustada. Üksikute väidete ja nende seoste, aga ka indiviidide sotsiolingvistiliste kogemuste vaatlemine (nii eraldiseisvalt kui ka seostatuna) võimaldas analüüsida eestikeelsele õppele üleminekuga seotud individuaalseid positsioone, keskendudes nii üksikutele detailidele kui ka tervikule.



## Viidatud kirjandus

- Aalto, Pami 2003. *Constructing Post-Soviet Geopolitics in Estonia*. Routledge Studies in Geopolitics. London, New York: Routledge.
- Bailey, Kenneth D. 1994. *Typologies and Taxonomies: An Introduction to Classification Techniques*. Quantitative Applications in the Social Sciences 102. Thousand Oaks: Sage Publications.
- Barchak, Leonard J.; Marshall, Russell (Gene) 1998/1999. What do they know of English, who only English know? Ascertaining attitudes toward foreign language study using Q methodology in a public relations framework. – *Operant Subjectivity*, 22 (1/2), 31–48.
- Brown, Steven R. 1980. *Political Subjectivity: Applications of Q Methodology in Political Science*. New Haven: Yale University Press.
- Burt, Cyril 1937. Correlations between persons. – *British Journal of Psychology*, 28 (1), 59–96. <http://dx.doi.org/10.1111/j.2044-8295.1937.tb00862.x>
- Cattell, Raymond 1973. *Personality and Mood by Questionnaire*. San Francisco: Jossey-Bass.
- Child, Dennis 2006. *The Essentials of Factor Analysis*. 3rd ed. London, New York: Continuum.
- Chouliaraki, Lillie; Fairclough, Norman 1999. *Discourse in Late Modernity: Rethinking Critical Discourse Analysis*. Edinburgh: Edinburgh University Press.
- Fairclough, Norman 2003. *Analysing Discourse: Textual Analysis for Social Research*. London: Routledge.
- Dryzek, John S. 2004 [1990]. *Discursive Democracy: Politics, Policy, and Political Science*. Cambridge: Cambridge University Press.
- Grosswiler, Paul 1994. The convergence of William Stephenson's and Marshall McLuhan's communication theories. – *Operant Subjectivity*, 17 (3/4), 2–16.
- Hill, Thomas; Lewicki, Pawel 2006. *Statistics: Methods and Applications: A Comprehensive Reference for Science, Industry, and Data Mining*. Tulsa: StatSoft.
- Housen, Alex 2002. Contextual, output and operational variables in bilingual education in Latvia. – *Intercultural Education*, 13 (4), 391–408. <http://dx.doi.org/10.1080/1467598022000023849>
- KKK 2012 = Korduma kippuvad küsimused. Haridus- ja Teadusministeerium. <http://www.hm.ee/index.php?047810> (12.12.2012).
- Masso, Anu 2010. Geographical perspective on identity construction: Identification strategies of Russian youth in Estonia. – *International Journal of Interdisciplinary Social Sciences*, 5 (6), 51–62.
- Masso, Anu; Kello, Katrin 2010. 2009. a kevadel vene õppekeele gümnaasiumite õpetajate ja 11. klassi õpilaste seas läbi viidud küsitluse aruanne. – Uurimisprojekti “Vene õpilane venekeelse üldhariduskooli eestikeelses õppes” aruanne. Tartu: Tartu Ülikool.
- Masso, Anu; Kello, Katrin 2012. 2011. a läbiviidud ekspertküsitluse aruanne. – Uurimisprojekti “Vene õpilane venekeelse üldhariduskooli eestikeelses õppes” aruanne. Tartu: Tartu Ülikool.
- Masso, Anu; Soll, Maie 2013. The transition to Estonian as the language of instruction in upper secondary schools: Readiness of pupils and teachers in Estonia. – *Journal of Baltic Studies*, avaldamiseks vastu võetud käsikiri.
- McKeown, Bruce; Thomas, Dan 1988. *Q Methodology*. Newbury Park: Sage Publications.
- Rannut, Ülle 2005. Keelekeskkonna mõju vene õpilaste eesti keele omandamisele ja integratsioonile Eestis. Tallinn: TLÜ Kirjastus.
- Siiner, Maarja 2006. Planning language practice: A sociolinguistic analysis of language policy in post-communist Estonia. – *Language Policy*, 5 (2), 161–186. <http://dx.doi.org/10.1007/s10993-006-9004-9>
- Singer, Jane B. 1998/1999. Pairing Q methodology and case studies to explore journalists' attitudes toward new media forms. – *Operant Subjectivity*, 22 (1/2), 1–19.

- Snelling, Susan J. 1999. Women's perspectives on feminism: A Q-methodological study. – *Psychology of Women Quarterly*, 23 (2), 247–266. <http://dx.doi.org/10.1111/j.1471-6402.1999.tb00357.x>
- Spearman, Charles 1927. *The Abilities of Man: Their Nature and Measurement*. New York: Macmillan.
- Stephenson, William 1936. The foundations of psychometry: Four factor systems. – *Psychometrika*, 1 (3), 195–209. <http://dx.doi.org/10.1007/BF02288366>
- Stephenson, William 1953. *The Study of Behavior: Q-technique and its Methodology*. Chicago: University of Chicago Press.
- Trost, Jan 1986. Statistically nonrepresentative stratified sampling: A sampling technique for qualitative studies. – *Qualitative Sociology*, 9 (1), 54–57. <http://dx.doi.org/10.1007/BF00988249>
- RT 2007, 61, 392 = Vabariigi Valitsuse 25. jaanuari 2002. a määruse nr 56 “Põhikooli ja gümnaasiumi riiklik õppekava” muutmine. Vastu võetud 23.11.2007 nr 235. Riigi Teataja I 2007, 61, 392.
- van Dijk, Teun A. 1985. *Ideoloogia: multidistsiplinaarne käsitlus*. Tartu: Tartu Ülikooli Kirjastus.
- Watts, Simon; Stenner, Paul 2012. *Doing Q Methodological Research. Theory, Method and Interpretation*. Los Angeles: Sage.
- Wodak, Ruth 1996. *Disorders of Discourse*. London: Longman.

**Anu Masso** (Tartu Ülikool) on uurinud personaalseid kultuurilisi distantse, geo-kultuurilist mobiilsust ja keelelist kommunikatsiooni.

[Anu.Masso@ut.ee](mailto:Anu.Masso@ut.ee)

**Katrin Kello** (Tartu Ülikool) on tegelenud vene õppekeelega gümnaasiumite eestikeelsele aineõppele ülemineku küsimustega uurimisprojekti “Vene laps venekeelse üldhariduskooli eestikeelses õppes”

(2008–2011) raames.

[Katrin.Kello@ut.ee](mailto:Katrin.Kello@ut.ee)

**Valeria Jakobson** (Tartu Ülikool) on uurinud rahvusvähemuste meediat, selle rolli kollektiivsete identiteetide ja rahvuslike tähtpäevade kujunemisel ning rahvusvähemuste meedia jälgimise harjumusi.

[valerijakobson@hotmail.com](mailto:valerijakobson@hotmail.com)

# **TRANSITION TO STATE-LANGUAGE INSTRUCTION IN UPPER SECONDARY SCHOOLS ACCORDING TO ALUMNI: Q-METHODOLOGICAL DISCOURSE ANALYSIS OF INDIVIDUAL POSITIONS**

**Anu Masso, Katrin Kello, Valeria Jakobson**

University of Tartu

This article deals with discourses related to the transition to Estonian as the language of instruction in the currently Russian-medium upper secondary schools, based on Q-methodological semi-formalised interviews with alumni (42) of Russian-medium schools in three Estonian regions and, for a comparison, in Latvia. Since Q-methodology combines qualitative and quantitative approaches it enabled the authors to conceive subjective meanings as well as patterns therein (discourses) as they relate to this complex topic. The analysis distinguished four discourses based on the interviewees' agreement with the 63 statements that they were asked to sort (evaluate) and on which they were asked to comment selectively.

The most represented discourse (21 individuals representing all the four regions) was 'valuing the availability of linguistic alternatives'. This title refers to the position that two state languages would eliminate problems rather than create them; agreeing that English was instrumentally more advantageous than Estonian; and valuing the students' right to choose their language of instruction. Somewhat similar was the 'resistance to homogenisation' discourse, represented by four individuals from Estonia, although within this discourse, the critique focused more on instruction (in Estonian) itself (e.g., as detrimental to the student motivation). However, the general necessity for the Russian-speaking population to learn Estonian was also doubted.

The 'linguistic and cultural distanciation' discourse (two representatives from Estonia) additionally expressed personal pessimism both regarding the state policies and as related to personal mono-linguistic and mono-cultural preferences. Opposed to the previous was the 'satisfaction with one's language proficiency and own integration' discourse, represented by the second largest number of individuals (13), six of whom were from Latvia. This pattern of positions expressed personal optimism as well as agreement with official arguments on the benefits of state language skills and a change of instruction language.

**Keywords:** language politics, education politics, interethnic relationships, transition to state language instruction, Q-methodology, discourse analysis, Russian language, Estonian language, Latvian language



## LAENSÕNADE TÄHENDUSSUHETEST

Vilja Oja, Iris Metsmägi

**Ülevaade.** Eesti keelde laenatud sõna ei kopeeri alati laenuallika tähendust või tähendusi, vaid omandab sageli sellest kitsama või laiema või hoopis uue tähenduse. Selle põhjuseks on kultuuride erinevus, paikondlikud etnograafilised iseärasused, eri taime- ja loomaliikide levik, tabu jm. Rahvaetümolooogia mõjul võib sõna tähendus muutuda nii, et seos laenuallikaga puudub. Täendusmuutus on sageli toimunud hiljem, mitte laenamise käigus, ning tingitud olude või esemete muutumisest. Mõnel juhul on sama sõna laenatud korduvalt, eri aegadel erinevas tähenduses. Sõnade ja nende tähenduste levik murretes aitab määrata laenude liikumise teid ja ühtlasi avastada murdealati eri suunast tulnud laene. Teiselt poolt aitab eesti sõna päritolu täpsustada võrdlus sarnase keelendi kasutusega sugulaskeelte ja kontaktkeelte murretes. Sõnade tähendussuhteid tuleb kindlasti arvestada nii keeleõppes ja tõlketöös kui ka sõnade päritolu määratlemisel.\*

**Võtmesõnad:** etümolooogia, keelekontaktid, murdesõnavara, sõnalevik, tähendusmuutus, eesti keel, läänemeresoome keeled

### 1. Sissejuhatus

Sõnade päritolu on sageli püütud tõestada üksnes või eelkõige keeleajalooliste häälikumuutuste varal, ehkki termin “etümolooogia” lähtub mõistest ‘tähendus’, kreeka *etymon*. Ka tõlkimisel on oht samastada laensõna häälikujult sarnase võõrkeelse vastega, süvenemata tähendusnüanssidesse. Käesolev artikkel käsitleb laenude semantilist külge. Tähelepanu keskmes on eesti sõnade tähenduserinevused võrreldes doonorkeeltega, kuid selle kõrval näidatakse ka erinevusi murretes ja läänemeresoome keeleruumis. Uurimuses analüüsitakse muutuste liike ja põhjusi ning võimalikke laenude kulgemise teid. Järelduste tegemisel arvestatakse sõna ja selle tähenduste levikut murretes, lähisugukeeltes ja kontaktkeeltes. Näited on valitud põhiliselt uuemate laenude hulgast, mis on eesti keelde tulnud pärast 12. sajandit. Mõne nähtuse puhul kõrvutatakse uuemaid ja vanu laene. Sõnade

\* Uurimust on osaliselt rahastanud ETF9367 ja SF0050037s10.

päritolu osas on aluseks “Eesti etümoloogiasõnaraamat” (Metsmägi jt 2012), teistele allikatele viidatakse tekstis.

Tähenduserinevusi on laias laastus kolme liiki. Laensõna võib originaaliga osaliselt kattuda, märkides kas 1) kitsamat või 2) laiemat mõistet kui doonorkeeles, või 3) omandada uue tähenduse. Nimetagem olukorda, kus laensõna tähendus on kitsam kui laenuallikal, tähenduse kitsenemiseks, vastupidist situatsiooni tähenduse laienemiseks ning täielikku erinevust tähenduse teisenemiseks. Keeles kodunenud laensõna allub edasistele muutustele ning alati ei õnnestu välja selgitada, kas tähendusmuutus toimus laenamisel või hiljem.

## 2. Tähenduse kitsenemine

Laensõna tähendus saab kitseneda üldjuhul siis, kui originaalkeele sõna on kas mitmetähenduslik või märgib suhteliselt laia mõistet. Paljudel juhtudel on doonorkeeles mitmetähenduslik sõna laenatud eesti keelde tähistamiseks vaid ühte või osa originaali tähendustest. Näiteks eesti *kleit* < sks *Kleid* ‘kleit, rõivas, rüü’; *nööp* < asks *knōp* ‘nööp; nupp; pung; sõlm’; *pood* ‘kauplus’ < asks *bode* ‘väike käsitööliste jt “tähtsusetute” inimeste maja; barakk, telk; käsitööliste müügi- ja tööruum’; *plaaster* ‘nahale asetatav raviotstarbega kate (nt haavaplaaster, sinepiplaaster)’ < asks *plāster* ‘plaaster; sillutis, tänavakate’. Erijuhtudel tuntakse laensõna üksnes ülekantud tähenduses, nt eesti *kelm* < asks *schelm* ‘korjus, raibe’, piltlikult ‘kelm, lurjus’.

Ainsas tähenduses laenatud sõna võib kinnistuda veelgi kitsamat mõistet tähistava terminina, nt eesti *moos* < asks *mōs* ‘kapsas, köögivili; puderjas toit’. Hoopis laiema tähendusega on samast algtüvest lähtuv soome kõnekeelne *muusi* ‘pudi; puder; püree; moos’, mis on laenatud ilmselt rootsi keelest, vrd rts *mos id.* ning verb *musa* ‘pudiks või pudruks teha’ (> soome *muusata id.*) (PS 2: 255). Soome etümoloogilistes sõnaraamatutes see sõna puudub. Ahjulõõri keeratavat (kaanetaolist) siibrit märgib eesti keeles vene laen *kriska* < vn *крышка* ‘kate, kaas; luuk; klapp’, eesti murretes ka *kriiska*, *krõ(õ)ska* jms (vt EMS III: 849, Must 2000: 141). Põhjaeesti keeles ja osalt Mulgi murdes on lihapoodi nimetatud *karn*, gen *karni* (ka *karnits*, *kärn*) < asks *scharne* ‘lett, lettide kogum; ruum, kus letid paiknevad’ (EMS II: 758, 760).

Kitsamas eritähenduses on eesti keelde laenatud ka selliseid sõnu, mis doonorkeeles väljendavad laia üldmõistet, nt *kee* ‘lülidest koosnev kaelahe’ < asks *kede*, *kedene* ‘kett’; *leer* ‘noorte õpetamine ning ettevalmistamine kogudusse vastuvõtmiseks’ < asks *lere* ‘õpetus, õppimine’; *paak* ‘masina, seadmestiku vm vedelikumahuti’ < asks *bak* ‘anum, küna’.

Laensõna tähendus võib kitseneda korduvate keelekontaktide tagajärjel. Näiteks *kang* ‘kitsas käik või läbipääs; võlvialune läbikäigukoht’ on laenatud kas alam- või ülemsaksa keelest, < asks *gank* ‘käimine; rongkäik; käik, tee; komme, harjumus’ või < sks *Gang* ‘kõnnak; (jalutus)käik, teekond; koridor, käik; kang, kangialune; kulg’. Juhul, kui algselt laenati alamsaksa sõna tähenduses ‘käik, tee’, muutus see hiljem saksa mõjul konkreetsemaks. 20. saj eesti keeles oli *kang* üldtuntud võlvialuse läbikäigu nimetusena (*kangi all*, *kangialune*), ent varasem tähendus säilis näiteks Tallinna tänavanimis *Saiakang*.

Üks ja sama laentüvi ei väljenda alati sama tähendust üksiksõnana ja liitsõna osana. Sel juhul on liitsõna tavaliselt eraldi laenatud. Näiteks saksa laen *kreis* (< sks *Kreis*) märgib eesti keeles varasemat haldusüksust Saksa- ja Venemaal ning Eesti maakonda 18.–19. saj. Saksa keeles on selle sõna põhitähendus ‘ring, sõõr’ ning sellest lähtuvad muud tähendused, nagu ‘ringkond, teatud seltskond, grupp; sfäär, tegevusala; haldusüksus, piirkond’. Sama tüvi esineb teises tähenduses liitsõnas *kreissaag* ‘ketassaag’, mis ei ole moodustatud eesti keele baasil, vaid on omaette laen < sks *Kreissäge* id. Verb *leierdama* ‘tüütuseni kordama’ < sks *leiern* ‘vántorelit mängima; kõnek ‘ühtesama jorutama; monotoonselt lugema’ ei seostu eesti keeles otseselt leierkastiga. Liitsõna *leierkast* ‘vántorel’ < sks *Leierkasten* id. on laenatud eraldi, kusjuures liitsõna järelois *kast* on võtnud varasema, alamsaksa laenu kuju (vt ptk 4).

### 3. Tähenduse laienemine

Laienenud tähendust kohtame mitmes laenatud loomanimetuses, kus noorlooma või ühest soost isendi nimetusest on saanud liigi üldnimetus, nt eesti *kits* ‘veislaste hulka kuuluv keskmise suurusega imetaja, kodukits, metskits’ < sks *Kitze* ‘kitsetall, mets- või mägikitse tall’. Eesti *kass* on laenatud vene keelest < vn *кошка*, murdekeeles *коша, кошара, кошюра* ‘emakass’. Koduloomade nimetused on üldse omavahel mitmel viisil segunenud. Niisugused tähenduse varieerumised tulevad hästi esile murdekeeles, eriti kui vaadelda läänemeresoome ainest tervikuna (vt ptk 5).

Algselt ühte mõistet tähistav laensõna võib originaaltähenduse kõrval omandada täiesti uusi lisatähendusi. Laensõnad, mida kirjakeeles kasutatakse originaalset laiemas tähenduses, võivad murdeti olla doonorkeelega tähenduselt identsed. Näiteks eesti põhjarannikul, Risti, Jõelähtme ja Kuusalu murrakutes nimetatakse võrguujukit *klibu* (< eestirootsi *klibu* id.). Sama tüvega sõna on levinud pisut laiemalt põhjaeesti murretes tähenduses ‘lapikud kivikesed; miski väike, väärtusetu, tühine’. Eesti kirjakeeles tuntakse seda sõna just viimases tähenduses ning selle baasil on nii murde- kui kirjakeeles moodustatud tuletised *klibune*, *klibuline*, *klibustik* (EKSS 2: 326, EMS III: 320).

### 4. Tähenduse teisenemine

Ka siis, kui laensõna ja originaali ükski tähendus ei kattu, on tähenduse muutus toimunud mingi seose alusel. Eesti vaste väljendab enamasti laenuallikale lähedast mõistet. Näiteks osmikut ja väikelooma putkat märkiv *kuut* (murdekeeles ka *kuft*, *kuudik*) < läti *kūts* ‘loomalaut’; *kūūt* ‘vedu’ < vana-rts *skiūt* ‘veohobune; mära’; eesti rahvustoiduna tuntud *kama* ‘kaun- ja teraviljajahu segu ning sellest valmistatud toit’ < vn murdesõna *комы* (mitmus) ‘saiakesed, leivad, pannkoogid’. Lööbega nakkushaiguse nimetus *leetrid* on levinud peamiselt põhjaeesti murretes (ka *leedrid*, *leeted* jt häälikuvariandidid) ning mõni keelejuht on selle tähenduseks öelnud ‘tuulerõuged’ (EMS V: 42–43). Etümolooogiasõnaraamatus on see esitatud alamsaksa laenuna (< asks *bladder*, *bledder*, *bledere* ‘vill, rakk’). Eesti sõna mitmuslik vorm näitab, et haiguse nimetus väljendab sel juhul mõistet ‘villikesed, lööve’ vms.

Paljude laensõnade sisulist seost laenuallikaga ei ole tänapäeval enam lihtne tajuda, see ilmneb alles taustteabe ja selgituste varal. Näiteks *laat* on eesti keeles üldtuntud tähenduses 'suur ajutine turg' vms. Laen pärineb alamsaksa sõnast *afflade* 'patukustutus, indulgents' ja seostub eesti sõnaga sel teel, et kirikupühadel kuulutati pattude andeksandmist ja peeti ka laatasid. Vanemat eesti elu kujutavas kirjanduses esineb sõna *vonn*, gen *vonni* 'mõisnik, saks' < sks *von* (eessõna; aadlitiitel). Nimelt hakati Saksamaal 17. saj seest- ja alaltütleva käände tähendust märkiva eessõnaga *von* tähistama aadlipäritolu. Tarbetuid jäätmeid, puru, rämpsua märkiv *prügi* (< asks *brugge* 'sild, tänav', varem 'tamm') tuleneb sellest, et tammi põhja täitena kasutati muuhulgas prahti. Täidlast, tüsedavõitu inimest või looma iseloomustab omadussõna *priske* < asks *vrisch* 'noor, värsk, rõõmus'. Vanemas eesti keeles kasutati seda sõna tähenduses 'elujõuline; värsk', kõnekeeles ka 'suur, tõhus, rikkalik, tugev vms', murretes ka 'ilus', nt Põlva murrakus *ega miiss haard priškit, kestu hirmsat* [neiut] *taht* (EMSUKA, Wiedemann 1973: 876). Samast tüvest on tulnud Leivu murraku *priš's* 'värsk' < läti *prišs id.* (< sks *frisch*) ja liivi Kuramaa murde *priš id.*, mis võib olla laenatud otse saksa keelest või läti keele vahendusel (ALFE 3: 315–318, Suhonen 1973: 181–182).

Kõige kardinaalsemaid tähenduse muutusi tekitab rahvaetümoloogia, kus laensõna ühitatakse häälikuliselt lähedase, keeles juba varem esineva, hoopis teise päritolu ja tähendusega tüvega. Sel viisil võidakse sõna tõlgendada vale mõiste baasil. Näiteks punakas- või pruunikaskollast värvust märkiv eesti *paat*, gen *paadi* on läti laen (< läti *pātis, pāts* 'kahvatu pruunikaskollane, kollane') ning eesti keelest omakorda on laenatud soome ja isuri murretes hobuse värvust kirjeldav *paatti* 'kollakas, helepruun' (Oja 2003: 101–102, Vaba 1997: 146–147). Eesti sõna on mõnikord ekslikult seostatud paekivi nimetusega, määratledes tähenduseks 'paekarva'. Rahvaetümoloogia tõttu võib keelend omandada ka varem tuntud sõna kuju koos tähendusega, nt *orleani kollane* 'oranž' > *ordenikollane* 'kuldkollane'. Rahakoti vananenud nimetusena tuntud *tengelpung* on baltisaksa laen (< baltisaksa *Denkelbuch* 'kirjatasku, taskuraamat, kirjutustahvel', sks *denken* 'mõtlemine; arvamine; kavatsemine; kujutlemine', *Buch* 'raamat'). Eesti keeles on sõna osiseid seostatud juba varem keeles esinenud laensõnadega *teng* 'raha' (< vene *деньга* 'raha') ja *pung* van 'rahakott' (< asks *punge* 'tasku, väike kott' või < rootsi *pung* 'rahakott, kukkur; tubakakott'). (Ariste 1942: 20, Viires 1960b: 158.) Niisuguseid tähendusmuutusi kohtab sageli eesti kohanimes, eriti neis, mis on lähtunud võõrastest, kohalikust pruugist kadunud isikunimedest (vt nt Kallasmaa 2000: 134).

Laenude liigitus tähendusmuutuste alusel on tinglik. Mitmetähenduslikkust eristab üldisest tähendusest mõnikord vaid kategoriseerimise detailsus. Tänapäeva kirjakeeles teisenenud tähendusega sõna on võidud varem või murdekeeles kasutada laenuallikaga samas tähenduses. Laensõna võib originaalist erineda ka sel moel, et osa tähendusi puudub, aga samas on uusi juurde tulnud, nt *kast* (< asks *kast(e)*, *kass* 'hoiukoht või -kamber, mahuti, kast; sörmusekivi pesa; [talus] peenestatud õlgede hoiuruum; vangla; mõõtühik') tähistab eesti keeles mõisteid 'puust vm materjalist (täisnurkne) säilitus-, transportimis- vms vahend; joonistatud nelinurkne kujund'; murdekeeles 'rehealuse kõrval olev ruum, kuhu vili viidi enne ahtmist; sillaalune palkidest tugiehitus; pidemed adrahargi keskosa ümber' (EKSS 2: 138–139, EMS II: 811–812). Wiedemanni sõnaraamatu andmetel on seda sõna kasutatud ka tähenduses 'talupojavangla' (1973: 218).



## 5. Täendusmuutuste põhjused

Sõna laenatakse enamasti koos mingi uuendusega, mistõttu kultuuriuendus ja laensõna tulevad tihti samast allikast. Laen tähistab sellisel juhul konkreetset uut objekti, nt *kriska* 'ahjulõõri siiber', *paak* 'vedelikumahuti', *plaaster*, *kreis-saag*, *leierkast*, kuigi doonorsõna võib olla mitmetähenduslik või tunduvalt laiema tähendusega (vt ptk 2). Sakslaste kaudu tuntuks saanud söögiriist *kahvel* (murdekeeles ka *kahver*) omandas nimetuse alamsaksa keelest < asks *gaffele* 'puust või rauast hark millegi tõstmiseks'. Eesti keeles olid olemas suured tõsteriistad *hang* ja *hark*, niisiis laenati *kahvel* üksnes söögiriista nimena. Purjehargi ja mitmesuguste masinate või sõiduriistade kaheharulise osa nimetusena võib *kahvel* olla hilisem laen samast allikast või hoopis ülemsaksa või rootsi keelest. Merenduse ja kalastusega seotud sõnavara on hollandi (ja alamsaksa) keelest levinud paljudesse Euroopa keeltesse, sealhulgas rootsi ja saksa keelde, ning sageli ei ole võimalik täpselt kindlaks teha, missugusest keelest need sõnad on eesti keelde laenatud. Millegi hargikujulist osa märgib *kahvel* vms vaid põhjaeesti murretes, üldtuntud söögiriista nimetus on levinud kõigis murretes (EMS II: 535–536). Mitmed sõnad on laenatud tähistamaks konkreetseid ühiskonnaelu toiminguid ja uusi nähtusi või riigikorraldust puudutavate terminitena, nt *kreis*, *leer* (vt ptk 2).

Uus, laenuallikale lähedane, kuid sellest erinev tähendus, on mõnel juhul seotud tabuga. Tabu tõttu on näiteks jahiloomade ja kardetavate metsloomade nimetusi püütud vältida, asendades need teiste, sageli laenatud nimetustega. Nii on hundi algläänemeresoome-aegne nimetus *susi* (< \**šunte* < IE *kunto-*) põhjaeesti ja kohati soome murretes asendatud saksa laenuga: eesti (*h*)*unt*, soome *huntti* (< sks *Hund* 'koer'), osas soome murretes vanavene (või vene) laenuga *hurтта*, vrd vana-vn *хѣрмѣ*, vn *хорп* 'jahikoer, hurt' (vt ALFE 3: 319–322, Plöger 1973: 56, SSA 1: 192, Zaitseva 2009: 240 jj).

Huvitavaid tähendusmuutusi on toimunud koduloomade nimetustes (vt ka ptk 3). Näiteks kitse märkivat soome sõna *kili* (murretes ka *kila*, *kili*, *kile*, *kilu*, *kilo* < rts *killa*, *kille*, *kilu* (van) 'kits' < skandinaavia) kasutatakse kohati üldnimetusena, kohati emaslooma või kitsetalle nimetusena (vt ALFE 3: 138–140, SMS 7: 197, 199, 202, 224–225, 241). Isuri keeles ja idavadja murdes on *kili* kodu- ja/või metskitse üldnimetus (Nirvi 1971: 164–165, VKS 2: 162). Eesti murretes esineb sõna kahel kujul: idamurdes ja Viru-Jaagupi murrakus *kile*, Setus lastekeelne *kilõ* '(isane või emane) kitsetall' ja saarte murdes (Karja, Emmaste, Pühalepa murrakutes) *kilu* 'kits; emane kitsetall' (EMS III: 128, 142). Üldnimetuse kokkulangevus emaslooma nimetusega on ilmselt tingitud sellest, et emaseid peeti rohkem. Väikeloomade nimetused levisid sageli hellitusnimedena. Paljud koduloomade nimetused on vanad balti laenud ega kuulu otseselt käesoleva teema raamesse, aga näitavad ilmekalt muutuste põhjust. Väikelooma nimetusest on mõnel juhul kujunenud kohitsemata looma nimetus. Näiteks põhjaeesti ja kohati soome murretes tähistab baltlastelt laenatud lambatalle nimetus isaslooma: eesti *jäär*, soome *jäärä* < balti, vrd preisi *eristian* 'tall', leedu *ėras*, läti *jērs* 'lambatal'. Selle kõrval on läänemeresoomlased laenanud ka isaslooma nimetuse: lõunaeesti ja soome *oinas*, vepsa *oinaz*, liivi (Salatsi) *oin*, *oen* 'oinas', karjala *oinaš* 'oinas, utt', põhjaeesti *oinas* 'kohioinas' < balti, vrd preisi *awins* 'oinas', leedu *avinas*, läti *aus*, murdeti *avēns*, *avins* jm (LVDA: 121–122, ME 2: 114). Omapärane tähendusmuutus on toimunud sõna *hārg*

laenamisel. Traditsiooniliselt on seda peetud laenuks balti keelte hobusenimetusest, vrd leedu *žirgas*, läti *zirgs* 'hobune', preisi *sirgis* 'täkk'. Kõigis läänemeresoome keeltes märgib see veist, põhiliselt isast või kohilooma: eesti *härg* ja liivi *äärga* 'härg', vadja *ärtšä* ja vepsa *hārg* 'härg; pull, sõnn', soome *härkä* 'veis: pull, sõnn; härg', karjala *härkä* 'pull, sõnn' (häälikkujud ja levik vt ALFE 3: 121–124). Karjala ja vepsa murretes on nõnda nimetatud ka isast põtra või põhjapõtra (KKS 1: 392, SVJ: 141). Balti hõimudelt on seega laenatud (põllutöödel kasutatava) veolooma nimetus, mille tähendus muutus vastavalt kohalikele oludele.

Looduslikud erinevused, eri taime- ja loomaliikide levik, põllumajanduslikud tavad jms on põhjustanud ka hilisemate laensõnade tähendusmuutusi, nii et sarnased laensõnad tähistavad läänemeresoome keeltes erinevaid taimi või loomi. Näiteks samast allikast laenatud *kaal*-tüveline sõna märgib läänemeresoome keeltes kahte erinevat köögivilja: soome ja isuri *kaali* ning lõunakarjala *koali* 'kapsas, *Brassica oleracea*', vrd rts, norra, taani *kål*, islandi *kál id.* (ALE 1, 7: kaart 76), aga eesti *kaal*, *kaalikas* '*Brassica napus rapifera*' < vana-rts *kaal*, *kal* 'kapsas'. Eesti keelega samas tähenduses esineb see sõna balti keeltes: läti *kālis* 'kaalikas' (< rts või asks), mis on levinud Vidzeme ja Zemgale murretes ning Kurzeme kirdeosas, ning leedu põhjaosas *kolis* (? < läti) (BVA 2009: 111, 154–155, LVDA: 116, kaart 47). Mõlemad juurviljad on aretatud looduslikest Vahemere-äärse kapsasrohu (*Brassica*) liikidest (ENE 4: 181, 310). Taimede kaugem originaalnimetus vanakreeka *kaulós* lähtub mõistest 'vars' (SAOB s.v. *kål*; Wharton 1974: 66). Sellised tähenduserinevused võivad viidata erinevale laenamise ajale.

Laensõna allub hiljem tavaliselt eesti keele harilikele tuletusreeglitele, mille tulemusena tekivad sama tüvega uued sõnad. Näiteks *kratt* 'pisuhänd, tulihänd ehk puuk' (< eestirootsi *skrat* 'kummitus, kurat') esineb põhjaeesti läänepoolsemates murretes (Jõelähtme, Kuusalu, Hiiu murrakud, läänemurre, keskmurde põhjaosa), sellest tuletatud verb *krattima* '(korduvalt vähehaaval) varastama' on registreeritud vaid Hiiumaalt Emmaste ja Reigi murrakutest (EMS III: 817–818). Kirjakeeles on nii noomen kui verb saanud laialt tuntuks, kusjuures nimisõna kasutatakse ka piltlikus tähenduses vallatu lapse kohta, nt *poisikratt* (EKSS 4: 345).

Sageli on sõna tähendus muutunud hiljem, mitte laenamise käigus. Näiteks Viljandimaa lõunaosa elanike nimetuse *mulk* (< läti *mulķis*, *mulķe* 'narr, lollpea') kohta on arvatud, et algselt laenati see sõna Mulgi murdesse läti keelega sarnases tähenduses. Teistelt murdealadelt naabrite hulgas sai see mulkide hüüdnimeks, millest hiljem kujunes hõimunimetus (Vasmer 1920: 345 ja 1922: 13). Läti keelega sarnases tähenduses kasutatakse sõna *mulķ* Saaremaal Jāmaja murrakus ja *mu'iki* liivi keeles, mis on eraldi laenud samast läti sõnast (Kettunen 1938: 233, Suhonen 1973: 160, Vaba 1997: 136). Saksa *Krebs* 'vähktõbi, kartsinoom (pahaloomuline kasvaja); vähk (veeloom)' on eesti keelde laenatud esimeses tähenduses *kreeps*, gen *kreepsu* (van) 'vähktõbi (pahaloomuline kasvaja)', mis kõnekeeles on hiljem teisenenud, märkimaks mõistet 'rabandus; vapustus'.

Sõna hilisemad tähendusmuutused on sageli tingitud olude või esemete muutumisest. Kui mingi tööriist või muu konkreetse otstarbega ese asendatakse täiuslikuma, sama või eelmisele lähedast funktsiooni täitva esemega, on endise nimetuse kasutamine küllalt tavaline. Näiteks hoorattaga ketrusriista märkiv *vokk* < asks *wocke*, sks *Wocke* 'kedervars; koonal; koonlalaud' võis olla laenatud esmalt kedervarre nimetusena ning sai uue tähenduse 17. saj, kui vokk Eestis laiemalt tuntuks sai (Ariste 1980a: 371–372, Viires 1960: 184).

## 6. Laensõnade ja nende tähenduste levik abiks etümologiseerimisel

Laensõnade tähendusmuutuste tuvastamine ja põhjendamine on sageli keerukas ja aeganõudev, sest see eeldab ulatusliku keeleainese (sealhulgas murdeainese) analüüsimist semantilise aspektist ja taustteabe kogumist. Teiselt poolt pakub sõnade päritolu uurimine mõnigi kord ootamatuid avastusi ja rõõmsat äratundmist. Näiteks siis, kui saadakse kinnitust kaheldavale etümoloogiale või leitakse varem omatüveks (sageli deskriptiivseks) peetud sõnale vasteid sugulas- või kontaktkeeltest. Etümoloogiat aitab täpsustada võrdlus sarnase sõnaga kontaktkeelte murretes, pidades seejuures silmas nii levikut kui ka tähenduslikku külge. Näiteks Võru murdesõna *villan* ~ *villañ* 'naiste sõba; muistne villane seelik' tundub olevat tavaline eestikeelne *ne*-sufiksiga tuletis sõnast *vill*, tegelikult on aga läti laen, vrd läti *villane, vilnaine* 'villane suurrätt, sõba' (Vaba 1997: 242). Sageli on laensõna laiialt levinud just algsest erinevas, muutunud tähenduses, kuid säilitanud algse tähenduse murdekeeles. Ala, kus laensõna esineb doonorkeelsega samas tähenduses, võib olla laenu lähtekoht, nt eesti *klibu* on tõenäoliselt hakanud levima põhjaranniku läänepoolsetest murretest (vt ptk 3).

Mõnigi kord on lisateave kujundanud sõna etümoloogia suhtes varasemast erineva seisukoha. Endisaegset kodukootud villast triibulist või ruudulist seelikut nimetati Pärnumaa ja Läänemaa lõunaosa murrakutes *kört*, gen *kördi* (Saareste 1938: kaart 9). A. Saareste on omal ajal pidanud seda rootsi laenuks < rts *skört* 'seeliku eesmine osa', vrd ka skandinaavia *skyrta* 'naiste põll' (1924: 165–166). Eesti murdesõnavara kogu põhjal esineb sõna samas tähenduses ka Muhu, Suure-Jaani ja Viljandi murrakus ning Mulgi murdes (EMS IV: 618). Eesti etümoloogiasõnaraamatu andmeil on eesti sõna laenatud alamsaksa keelest < asks *schorte*, mis tundub usutav, sest sõna levik ei kattu tüüpilise rootsi laenude alaga. Eesti keelest laenatud liivi (Salatsi) murdesõna *kört* ~ *kürt* 'naiste seelik' on seostatud sama alamsaksa sõnaga (Kettunen 1938: 179, SSA 1: 485, SLW: 103). Soome murretes märgib *körtti* kohati seelikut, jämedakoelist värvilist alusseelikut või põlle, kohati jaki vöökohast allapoole ulatuvat osa ning see on tunnistanud rootsi laenuks < rts *skört*, van *skörte* (SSA 1: 485).

Viimastel aastakümnetel avaldatud keeleatlased (ALFE, ALE, LVDA jt) pakuvad võimalusi eesti murdesõnade ja nende tähenduste leviku võrdlemiseks läänemeresoome keeleruumis ja teiste Euroopas kõneldavate keelte murdematerjaliga. Laensõnade ja nende tähenduste levik on osutanud murdealati või keeliti eri suunast tulnud laenudele. Sageli on ühise indoeuroopa tüvega sõnad läänemeresoome murretesse laenatud erinevate keelte kaudu, nt germaani algupäraga *klee-/klöö*-tüvelised ristikehina nimetused: eesti *klee*, vrd sks *Klee*; soome (*k*)*lööveri*, vrd rts *klöver*; muud variandid, nagu (*k*)*leever*, *kliever* jt, vrd vn *клевер* (ALFE 2: 57–62, Oja 2011: 131, 135). Rootsi ja saksa laenud tähistavad tõenäoliselt kultuurristikut, vene laenud looduslikku taime.

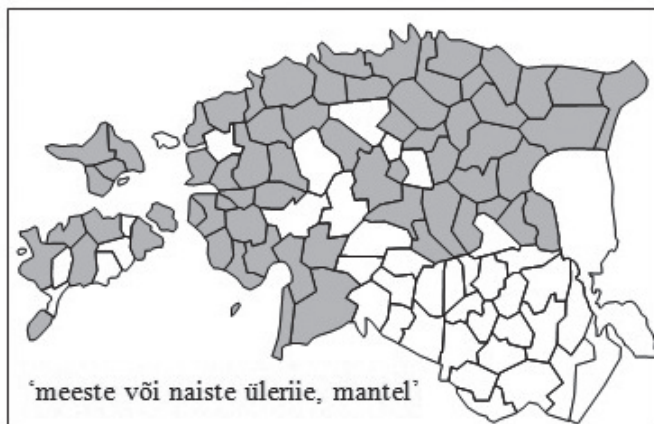
Indoeuroopa keeltes mitmes tähenduses kasutatav sõna võib läänemeresoome murretesse olla laenatud erinevail aegadel eri tähenduses. Jättes kõrvale kirumisõna funktsiooni, märgib noomen *pakana* läänemeresoome keeltes peamiselt kaht teineteisest erinevat mõistet: eesti, vadja, liivi ja kohati soome ja karjala murretes 'mittekristlane' ning karjala, vepsa, vadja, liivi, isuri ja mõnedes soome murretes

'halb, kuri (inimene või loom); räpane, kõlbmatu, väärtusetu (asi, nähtus vms)' (vt ALFE 2: 253–257). Viimases tähenduses loetakse sõna üldiselt vanavene laenuks, mille kaugem allikas on ladina *paganus* 'külaelanik, talupoeg', adj 'maaelule omane', kusjuures idapoolsetes läänemeresoome keeltes on oletatud hilisemaid vene mõjutusi, vrd vn *погань* 'jälkus, ilgus, rämps', *поганый* 'mittesöödav; vastik; vilets, rämps' (SSA 2: 294). Hiljem on vene keelest eesti Setu murrakusse laenatud *o*-lised variandid: noomen *pogan, pogañ* 'kõlbmatu, vastik, räpane, paha jms', sama tüvega adverb *poganahe* 'räpaselt' ja verb *pogandat* 'solkima, mustaks tegema' (ALFE 2: 256, EMSUKA). Sõnad tähenduses 'mittekristlane' pärinevad hilisemast ladina keelest ning on laenatud tõenäoliselt eesti ja liivi keelde saksa ning soome keelde rootsi keele vahendusel. Siit omakorda on need levinud vadja, isuri ja karjala (peamiselt Aunuse) murretesse (ALFE 2: 256, Viitso 2006: 894–897).

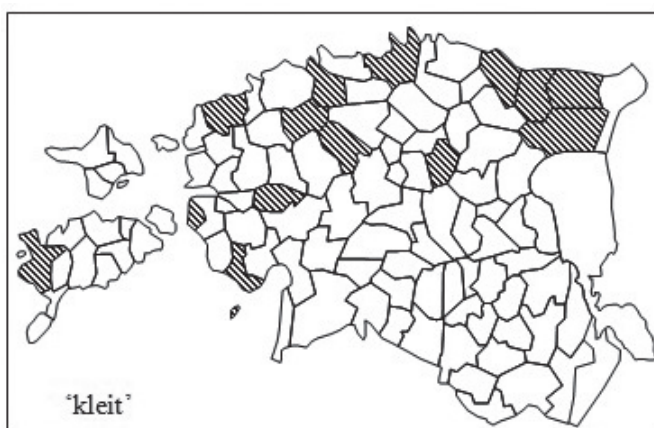
Samatüveliste sõnade kasutus erinevail Eesti murdealadel erinevas tähenduses viitab sellele, et siingi võivad sõnad olla saadud eri teid pidi, nt kitse nimetused *kile, kilõ* Ida-Eestis ja *kilu* läänesaartel (vt ptk 3); *mulk* Jämaja murrakus ja mujal (vt ptk 5). Tänapäeva eesti keeles meeste pintsakut märkivat sõna *kuub* on kasutatud vanemas keeles ja murretes mitme riietuseseme nimetusena. Sõna on levinud peamiselt põhjaeesti murretes, lõunaeesti murretest (Mulgi, Tartu) on harvad uuemad kirjapanekud (EMS IV: 138–139). Kõige laiemalt tunti seda tähenduses 'pikk villane meeste või naiste üleriie' (kaart 1: a), millest ilmselt tulenebki uus, praegune sõnakasutus. Vana tähendus on tänapäeval tuntud Mulgi meeste rahvarõiva pikk-kuue nimetuses (selle varasem nimetus oli *vammus*) ning märgib pikka üleriidet liitsõnades nagu *mungakuub, preestrikuub, vihmakuub* jt. Põhjarannikul ja kohati Läänemaal oli *kuub* laste (nii tüdrukute kui ka väikeste poiste) särgitaoline riietusesee või kleit (kaart 1: b; vt ka Saareste 1924: 171). Saaremaal ja Hiiumaal oli *kuub* naiste seeliku põhিনিmetus (kaart 1: c). Üksikteateid on *seeliku* või seda märkivate murdesõnade paralleelnimetuseks ka teistest põhjaeesti murretest, osalt tähenduses 'alusseelik' (Saareste 1924: 165, 167 ja 1938: kaart 9). Mandril on *kuub* seeliku nimetusena uuem kui saartel. Peale selle on Hargla murrakust kirja pandud *kuup* : *kuuba* 'kuub', liitsõna *kuubarõõvass* ja deminutiiv *kuubakõnõ* (EMS IV: 139, 166). Sõna *kuub* peetakse alamsaksa laenuks, ehkki sobivat doonorsõna pole sealt leitud. Tõenäoline keskülemsaksa vaste *schūbe, schoube* 'lihtne pikk üleriie' (sks *Schaube* 'naiste seelik või mantel') arvatakse tulenevat itaalia sõnast *giubba*, mis pärineb arabiakeelsest pikkade varrukatega üleriide nimetusest (Vasmer 4: 482). Levikut arvestades võiks seelikut tähistava sõna puhul kõne alla tulla ka teiste germaani keelte mõju.

Ühest algallikast pärineva, kuid eesti keeles mitut üsna erinevat mõistet märkiva sõna puhul võib tegemist olla eriaegsete laenudega. Nimisõnal *kudrus* on eesti kirjakeeles ja murretes kaks tähendust. Põhjaeesti murretes on sõna laialt levinud tähenduses 'väike klaas- või kivihelmes' (Hiiumaal ja Kirbla murrakus *kudru*) ja hajalevikuga tähenduses 'lõnga- või villakruss, juuksekihar' (vt kaart 2). Vastseliina ja Setu murrakuis tuntakse variante *kudrä, kudra*, mis tähendavad 'juukselokk, kihar', adj 'kähar, lokiline' ning verbi *kudrätämä* 'lokkima, kähardama'. Teise tähenduse alusel on moodustatud mäarsõnad *kudrus, kudris* 'krässus, krussis' ja rida omadussõnu, esimese baasil liitsõnad *kudruskael* murdekeeles 'kirju kaelaga lind', rahvaluules 'neiu' ja *kudrustükk* 'pärlitikandiga seelikualane' (EKSS 2: 524, EMS III: 940–941, Wiedemann 1973: 398).

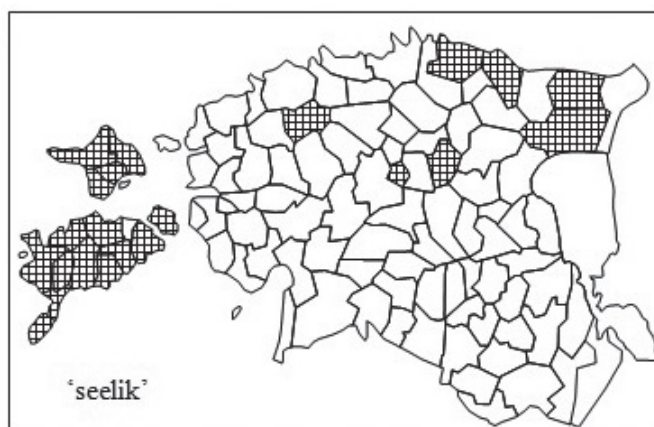
a.



b.

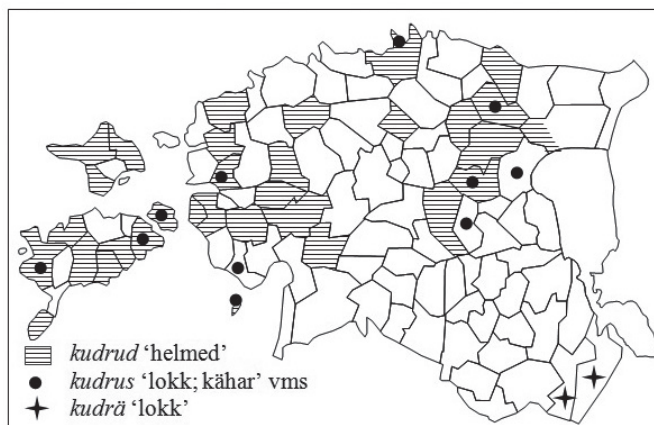


c.



**Kaart 1.** *kuub*, tähendused eesti murretes

Võru murdesõna on suhteliselt hiline vene laen < vn *кудерь, кудря* 'lokk, juuksekihar', mitmus *кудри*. Põhjaeesti sõna arvatakse olevat laenatud varem, samast allikast või vastavast vanavene sõnast, kusjuures tähendusmuutus 'lokk' > 'helmes' on toimunud eesti keeles (Blokland 2009: 140, Must 2000: 149, Mägiste 1962: 17, 1982–1983: 1015). Vene murretes nimetatakse nõnda ka mitmeid taimi, ent tähenduse 'helmed' kohta andmeid pole (Dal 2: 221, SRNG 16: 9). Samatüveline vene laen esineb sarnases tähenduses ka läänemeresoome idarühma keeltes: isuri ja karjala *kutri* 'juuksekihar' (< vn *кудри*), lüüdi *kudräv*, vepsa *kudräv* 'kähar, lokkis' (< vn *кудрявый*). Ilmselt "Kalevala" kaudu on sõna jõudnud ka soome kirjakeelde: *kutrit* 'lokid; lokkis juuksed', adj *-kutrinen* (PS 1: 596, SSA 1: 454).



Kaart 2. *kudr*-tüveliste sõnade tähendused eesti murretes

## 7. Kokkuvõtteks

Eesti keele suurest hulgast laensõnadest on paljud originaaliga häälikuliselt üsna sarnased, kuid tihtilugu ei ole nende tähendus identne ja hästi tuntud tõlkevaste ei pruugi tekstiga sisuliselt sobida. Laenu tähendus võib olla kitsam või laiem kui doonorsõnal või hoopis teisenenud. Laensõnu endid ei saa tähenduserinevuste põhjal alati nii konkretselt rühmitada, kuna tähendussuhted tervikuna on märksa komplitseeritumad. Sõna, mis on laenatud tähistamaks vaid ühte või osa paljudest doonorkeele tähendustest, on mõnikord eesti keeles kinnistunud veelgi kitsamat mõistet tähistava terminina. Teiselt poolt on algselt ühte mõistet tähistav laensõna originaaltähenduse kõrval omandanud täiesti uusi lisatähendusi. On ka selliseid laensõnu, millel võrreldes originaaliga osa tähendusi puudub, aga samas on uusi juurde tulnud. Ka siis, kui laensõna ja originaali ükski tähendus ei kattu, on tähendusmuutus üldjuhul toimunud mingi seose alusel. Paljude laensõnade sisulist seost laenuallikaga ei ole tänapäeval enam lihtne tajuda, see ilmneb alles taustteabe ja selgituste varal.

Laensõnade tähenduse muutusi on põhjustanud kultuuride erinevus, paikkondlikud etnograafilised iseärasused, eri taime- ja loomaliikide levik, tabu jm. Mitmed sõnad on laenatud tähistamaks üksnes konkreetseid ühiskonnaelu toiminguid ja

uusi nähtusi või riigikorraldust puudutavate terminitena. Rahvaetümoloogia mõjul võidakse uus laen ühitada varem tuntud, hoopis teist mõistet märkiva sõnaga, andes sellele tähenduse, millel pole laenuallikaga mingit seost. Sageli on sõna tähendus muutunud hiljem, mitte laenamise käigus, kuid alati ei õnnestu muutuse aega tuvastada. Hilisemad tähendusmuutused on enamasti tingitud olude või esemete muutumisest.

Laenu päritolu aitab täpsustada võrdlus sarnase sõnaga sugulaskeelte ja kontaktkeelte murretes, pidades seejuures silmas nii levikut kui ka tähenduslikku külge. Samatüveliste sõnade kasutus erinevail Eesti murdealadel erinevas tähenduses osutab, et tegemist võib olla eriaegsete või eri teid pidi saadud laenudega. Laensõna levik murretes aitab sageli määrata laensõna ja ühtlasi kultuuriuuduse lähtekohta. Sõnatüve häälikulise kuju ja selle arengute kõrval on laenu semantiline külg koos levikupildiga etümologiseerimisel väga tähtis.

### Kasutatud lühendid

adj	adjektiiv
asks	alamsaksa
gen	genitiiv
<i>id.</i>	<i>idem</i> 'seesama (tähendus)'
rts	rootsi
sks	saksa
<i>s.v.</i>	<i>sub voce</i> (osutab märksõnale)
van	vananenud
vene	vene

### Viidatud kirjandus

- ALE = Atlas Linguarum Europae (ALE). Volume I, 1–7. Commentaires. Cartes. Assen, Maastricht: Van Gorcum 1983–1990; Roma: Istituto Poligrafico e Zecca dello Stato, 1997–2007.
- ALFE = Atlas Linguarum Fennicarum. Itämerensuomalainen kielikartasto. Läänemeresoome keeleatlas. Ostseefinischer Sprachatlas. Лингвистический атлас прибалтийско-финских языков. ALFE 1–3. Suomalaisen Kirjallisuuden Seuran Toimituksia 800/1295. Kotimaisten kielten tutkimuskeskuksen julkaisuja 118/159. Helsinki: Suomalaisen Kirjallisuuden Seura, Kotimaisten kielten tutkimuskeskus, 2004–2010.
- Ariste, Paul 1942. Etümoloogilisi märkmeid I. – ACUT B XLIX, 1, Tartu, 1–26.
- Ariste, Paul 1980. Huvitavat sõnavara rahvakunstiraamatus. – Keel ja Kirjandus XXIII (6), 371–372.
- Blokland, Rogier 2009. The Russian Loanwords in Literary Estonian. Veröffentlichungen der Societas Uralo-Altaica 78. Wiesbaden: Harrassowitz.
- BVA = Baltu valodu atlants/Baltu kalbų atlasas. Prospekts. Latvijas Universitātes Latviešu valodas institūts, Lietuvių kalbos institutas. Anna Stafecka, Danguolė Mikulėnienė (Toim.). Rīga, Vilnius: Latvijas Universitātes Latviešu valodas institūts.
- Dal = Даль, Владимир 1955. Толковый словарь живаго великорусскаго языка, 2. Москва: Государственное издательство Иностранных и национальных словарей. (Толковый словарь В. Даля ON-LINE: <http://vidahl.agava.ru> (22.02.2013).)
- EKSS = Eesti keele seletav sõnaraamat 1–6. “Eesti kirjakeele seletussõnaraamatu” 2., täiendatud ja parandatud trükk. Margit Langemets, Mai Tiits, Tiia Valdre, Leidi Veskis, Ülle Viks, Piret Voll (Toim.). Eesti Keele Instituut. Tallinn: Eesti Keele Sihtasutus, 2009.

- EMS = Eesti murrete sõnaraamat I–V. Anu Haak, Evi Juhkam, Mari-Liis Kalvik, Mari Kendla, Varje Lonn, Helmi Neetar, Ellen Niit, Piret Norvik, Vilja Oja, Eevi Ross, Aldi Sepp, Tiina Tärk-Laansalu, Jüri Viikberg (Toim.). Eesti Teaduste Akadeemia, Eesti Keele Instituut. Tallinn: Eesti Keele Instituut / Eesti Keele Sihtasutus, 1994–2012.
- EMSUKA = Eesti murrete ja soome-ugri keelte arhiiv (Eesti Keele Instituudis, Tallinnas).
- ENE = Eesti nõukogude entsüklopeedia, 4. Tallinn: Valgus 1989.
- Kallasmaa, Marja 2000. Saaremaa kohanimed II. Tallinn: Eesti Keele Instituut.
- Kettunen, Lauri 1938. Livisches Wörterbuch mit grammatischer Einleitung. Lexica Societatis Fenno-Ugricae V. Helsinki: Suomalais-Ugrilainen Seura.
- KKS = Karjalan kielen sanakirja, 1. Lexica Societatis Fenno-Ugricae XVI, 1. Kotimaisten kielten tutkimuskeskuksen julkaisuja 25. Helsinki: Suomalais-Ugrilainen Seura, 1968.
- LVDA = Latviešu valodas dialektu atlants. Leksika. Latvijas Zinātņu akadēmija. Latviešu valodas institūts. Literatūras, folkloras un mākslas institūts. Rīga: Zinātne, 1999.
- Metsmägi, Iris; Sedrik, Meeli; Soosaar, Sven-Erik 2012. Eesti etimoloogiasõnaraamat. Eesti Keele Instituut. Tallinn: Eesti Keele Sihtasutus.
- Must, Mari 2000. Vene laensõnad eesti murretes. Tallinn: Eesti Keele Sihtasutus.
- Mägiste, Julius 1962. Äldre ryska länord i estniskan särskilt i det gamla estniska skriftspråket. Lunds Universitets Årsskrift. Ny följd I, 55 (1). Lund: Gleerup.
- Mägiste, Julius 1982–1983. Estnisches etymologisches Wörterbuch. Helsinki: Finnisch-Ugrische Gesellschaft.
- Nirvi, R. E. 1971. Inkeröismurteiden sanakirja. Lexica Societatis Fenno-Ugricae XVIII. Helsinki: Suomalais-Ugrilainen Seura.
- Oja, Vilja 2003. Hobu oli *võik*, lehm oli *leet*. – Keel ja Kirjandus XLVI (2), 101–107.
- Oja, Vilja 2011. Läänemeresoomlaste keele- ja kultuurikontaktidest laensõnade leviku põhjal. – Eesti Rakenduslingvistika Ühingu aastaraamat, 7, 129–140. <http://dx.doi.org/10.5128/ERYa7.08>
- Plöger, Angela 1973. Die russischen Lehnwörter der finnischen Schriftsprache. Veröffentlichungen der Societas Uralo-altaica 8. Wiesbaden: Otto Harrassowitz.
- PS = Suomen kielen perussanakirja. Kotimaisten kielten tutkimuskeskuksen julkaisuja 55. Helsinki: Edita, Kotimaisten kielten tutkimuskeskus, 1996.
- Saareste, Albert 1924. Leksikaalseist vahekordadest eesti murretes I. – ACUD B VI, 1. Tartu.
- Saareste, Andrus 1938. Eesti murdeatlas. Atlas des parlers estoniens. I vihik. 28 kaarti. Tartu.
- SAOB = Svenska Akademiens ordbok, <http://g3.spraakdata.gu.se/saob/> (22.02.2013).
- SLW = Winkler, Eberhard; Pajusalu, Karl 2009. Salis-livisches Wörterbuch. Linguistica Uralica. Supplementary Series 3. Tallinn: Teaduste Akadeemia Kirjastus.
- SMS = Suomen murteiden sanakirja 7. Kotimaisten kielten tutkimuskeskuksen julkaisuja 36. Helsinki: Suomalaisen Kirjallisuuden Seura, Kotimaisten kielten tutkimuskeskus, 2003.
- SRNG 16 = Словарь русских народных говоров. Выпуск 16. Академия наук СССР. Институт русского языка, словарьный сектор. Ленинград: Наука, 1980.
- SSA = Suomen sanojen alkuperä. Etymologinen sanakirja. Suomalaisen Kirjallisuuden Seuran toimituksia 556. Kotimaisten kielten tutkimuskeskuksen julkaisuja 62. Helsinki: Suomalaisen Kirjallisuuden Seura, Kotimaisten kielten tutkimuskeskus, 1992–2000.
- Suhonen, Seppo 1973. Die jungen lettischen Lehnwörter im Livischen. Suomalais-Ugrilaisen Seuran Toimituksia 154. Helsinki: Suomalais-Ugrilainen Seura.
- SVJ = Зайцева, М. И.; Муллонен, М. И. Словарь вепсского языка. Академия наук СССР. Карельский филиал. Институт языка, литературы и истории. Ленинград: Наука, 1972.
- Zajceva, Nina 2009. Pari pientä kuvaa aineellisesta ja henkisestä elämästä ALFEn peilaamana ('housut', 'susi'). – Marja Kallasmaa, Vilja Oja (Toim.). Kodukeel ja keele kodu / Home language and the home of a language. Pühendusteos Helmi Neetarile



75. sünnipäevaks 29. jaanuaril 2009. Eesti Keele Instituudi toimetised 13. Tallinn: Eesti Keele Sihtasutus, 237–249.
- Vaba, Lembit 1997. Uurimusi läti-eesti keelesuhetest. Tallinn, Tampere: Eesti Keele Instituut, Tampereen yliopiston suomen kielen ja yleisen kielitieteen laitos.
- Vasmer, Max 1920. Indo-germaani laensõnad eesti keeles. – Eesti Kirjandus, 343–346.
- Vasmer, Max 1922. Mõned indo-euroopa laensõnad eesti keeles. – Eesti Keel, 11–15.
- Vasmer = Фасмер, Макс 1986. Этимологический словарь русского языка. Перевод с немецкого и дополнения члена-корреспондента АН СССР О. Н. Трубачева. Под редакцией и с предисловием проф. Б. А. Ларина. Издание второе, стереотипное. Том IV. Москва: Прогресс.
- VKS = Vadja keele sõnaraamat. Eesti Teaduste Akadeemia, Keele ja Kirjanduse Instituut / Eesti Keele Instituut. Tallinn: Signalet / Eesti Keele Instituut / Eesti Keele Sihtasutus, 1990–2011.
- Viires, Ants 1960a. Eesti rahvapärane puutööndus. Ajalooline ülevaade. Tallinn: Eesti Riiklik Kirjastus.
- Viires, Ants 1960b. *hundilaut ja tengelpung*. – Keel ja Kirjandus III (3), 157–158.
- Viitso, Tiit-Rein 2006. *pagan, põrgu ja papp*: kolm kristlikku terminit. – Keel ja Kirjandus, 11, 894–902.
- Wharton, Edward Ross 1974. Etymological Lexicon of Classical Greek. Etyma graeca. Chicago: Ares Publishers Inc. Unchanged reprint of the edition: London 1882.
- Wiedemann, Ferdinand Johann 1973. Eesti-saksa sõnaraamat. Neljas, muutmata trükk teisest, Jakob Hurda redigeeritud väljaandest [1893]. Tallinn: Valgus.

**Vilja Oja** (Eesti Keele Instituut) peamised uurimisvaldkonnad on dialektoloogia, geolingvistika, läänemeresoome keeled. Eesti murrete sõnaraamatu koostajaid ja toimetajaid, Läänemeresoome keeleatlase ja Euroopa keelte atlase autoreid.  
[vilja.oja@eki.ee](mailto:vilja.oja@eki.ee)

**Iris Metsmägi** (Eesti Keele Instituut) peamised uurimisvaldkonnad on olnud uurali keelte verbimorfoloogia, uurali keelte kontaktid, eesti sõnade päritolu. Eesti etümoloogiasõnaraamatu peatoimetaja.  
[iris.metsmagi@eki.ee](mailto:iris.metsmagi@eki.ee)

# SEMANTIC RELATIONS OF LOANWORDS

**Vilja Oja, Iris Metsmägi**

Institute of the Estonian Language

Semantic differences between recent Estonian loanwords and the donor languages are discussed on the background of Estonian dialects and other Finnic languages. The types and reasons for the changes observed, as well as the possible routes of diffusion of the words, are analysed. The data on word origin is based mainly on the Estonian Etymological Dictionary by Metsmägi et al. (2012) if not referred to otherwise.

If the meanings of the borrowing and the source word differ, then the borrowed word may denote (1) a narrower or (2) wider or (3) a totally different concept than the source word. A word borrowed to denote one or part of the source word's meanings may become fixed as a term for a still narrower concept. However, sometimes an originally monosemous loanword may acquire new additional meanings. In other cases, some source meanings are missing, whereas some new ones have been added. Sometimes the source and the loanword share no meanings at all, and the semantic change is detectable only from background information.

The semantic change in loanwords may be caused by cultural differences, local ethnographic specifics, areal distribution of plant and animal species, taboo etc. Folk etymology may change word meaning totally beyond recognition. Often the semantic change takes place later than the time of borrowing. Such later changes are often conditioned by changes of circumstances or objects.

Word origin can be specified by following similar words in the dialects of contact languages, considering their spread and meaning. The variation of word meaning in Estonian dialects may suggest different borrowing times or routes of travel. The areal distribution of loanwords helps to pinpoint the centres of cultural innovations. Thus, an etymologist should consider not only the phonetic shape and developments of a stem, but also its semantic changes, areal distribution of the stem and its meanings in dialects and contact languages.

**Keywords:** etymology, linguistic contacts, areal distribution of words, semantic change, dialect vocabulary, Estonian, Finnic languages

# SPEED IN COGNITIVE TASKS AS AN INDICATOR OF SECOND/FOREIGN LANGUAGE READING AND WRITING SKILLS

Sanna Olkkonen

**Abstract.** In a cross-sectional study 823 Finnish school children were tested to examine the relation between speed of performance in cognitive and linguistic tasks and second/foreign language reading and writing. Participants were Finnish-speakers with English as foreign language and Russian-speakers with Finnish as second language which made it possible to compare the results across these two language groups. The Finnish group was furthermore divided into three groups by age to see how speed develops with age and education. Groups were tested with a number of cognitive instruments that included measures of speed of performance. Overall, performance on the speed measures improved with age; often, the second language learners outperformed their foreign language peers of the same age. Regression analyses indicated that speed measures could predict from 20% to over 40% of performance in second/foreign language reading and writing tasks. Prediction was somewhat stronger for writing than reading. The best predictors were also somewhat different for the foreign and second language learners, as well as for the different age groups.\*

**Keywords:** processing speed, reading, writing, cognitive testing, L2, Finnish, English, Russian

## 1. Introduction

Fluency in language skills can be seen as a combination of speed and accuracy (Grabe 2010, Segalowitz 2010). Speed and accuracy can either be seen as contrasting or complementary skills, and for example dyslexia is considered as a problem with either speed or accuracy. Usually dyslexics encounter problems in one of these skills, and only the most difficult cases have problems in both. In addition to reading problems, dyslexics often have problems with other kinds of automatic constructs, such

---

\* I would like to thank my supervisors Prof. Ari Huhta and Dr. Lea Nieminen for their guidance and feedback. I also would like to thank the two reviewers for their invaluable comments on the manuscript.

as reciting weekdays, months or multiplication tables (Cronin, Carver 1998). Fluent reading requires fast and automatic knowledge of vocabulary. Automatization of language skills is thus an integral part of language learning. When language skills, and especially word retrieval, become automatic, more cognitive capacity becomes available for the higher processes such as holistic text comprehension (Segalowitz 2003). William Grabe (2010: 72) states that: “Effective L1 reading comprehension generally assumes reading fluency – a person reading at a reasonable reading rate, between 250–300 WPM, using very efficient and fast word recognition skills, and combining information from various sources while reading under fairly intense time constraints.” Automaticity and its development have been researched widely in first language (L1) environments, but in second language (L2) acquisition the studies are fewer. Also the role of L1 skills in L2 fluency is a question that has evoked discussion (cf. Van Gelderen et al. 2007). The purpose of the current study is to examine the development of automaticity of vocabulary in both L1 (Finnish or Russian) and L2 (English or Finnish), and their relation to second/foreign language reading and writing skills. This examination was conducted in a cross-sectional design using four different age groups with two different language backgrounds: cross-group comparisons were performed on the speed of performance on the similar cognitive and linguistic tasks in both of their languages.

The capacity of working memory is restricted, so the more automatic word retrieval is, the more capacity there is for the use of higher mental processes (Kirby et al. 2003). Overall, the efficient lower-level processes such as memory, attention and word-decoding skills are a prerequisite for the higher ones. Fluent reading and reading comprehension require fast word retrieval. Automatic access to, and retrieval of, word knowledge are required for reading words as wholes and they are essential for being able to get past grapheme-phoneme method of reading. Automatic processes in general are nonconscious, unintentional, involuntary and effortless. They are also more immune to interference from outside sources (Segalowitz 2003, 2010, Grabe 2010). For example, mother-tongue word recognition is involuntary and unavoidable as the so-called Stroop effect shows. In the Stroop test, when the name of a colour is printed in a colour not denoted by the name (e.g., the word ‘red’ printed in blue ink instead of red ink), naming the colour of the printed word takes longer and is more prone to errors than when the colour of the ink matches the name of the colour (Stroop 1935). As Maryanne Wolf and Patricia Greig Bowers (1999: 418) state, rapid naming can be seen as ‘a microcosm of reading’. Working memory is also responsible for the combination and manipulation of new information. If word retrieval is slow, the beginning of the sentence can disappear from memory and integration of information can be hindered. This might mean that there is a certain threshold level in reading speed that should be attained for adequate comprehension (Kirby et al. 2003). Control of attention and retrieval from long-term memory also rely on working memory capacity. Eye-tracking studies have revealed that good readers can control their attention to some extent and concentrate on the relevant parts of the text while skimming (Kaakinen 2004). Effective word retrieval thus has an effect on all levels of reading. In addition to requirements in L1, fluent reading in another language requires direct connections from concepts to L2 vocabulary, and L2 word recognition would be automatic only when direct connections from the print to meanings have been formed. Connections develop

slowly with exposure to language, and especially at the early stages of learning to read in L2 the L1 vocabulary also becomes activated (e.g. Kroll, Sunderman 2003; Segalowitz 2003; Hulstijn et al. 2009).

In the *Jyväskylä Longitudinal Study of Dyslexia (JLD)*<sup>1</sup>, which started in 1993, the researchers have been following 200 children from birth and have tried to identify the best predictors for diagnosing learning difficulties as early as possible. The results show that phonological awareness and word recognition can, as early as the age of 3.5 years, predict possible reading problems at school age (Puolakanaho et al. 2007). The role of phonological processing as the best predictor for reading proficiency is supported by several other studies (e.g. Torgesen et al. 1997, Bowey et al. 2004). The effect of phonological processing is most notable in languages with very opaque orthography, such as English. In a more transparent language (e.g. Finnish, German or Italian) word recognition has been shown to be a stronger predictor of reading proficiency (Dufva, Voeten 1999, Di Filippo et al. 2005, Landerl, Wimmer 2008). In such languages, word decoding skills develop quickly, and accuracy seems to be close to a maximum level from a very early age. Thus, the ease of word retrieval seems to be an important differentiating factor. Also Grabe (2010: 76) concludes that L2 research, albeit its scarcity, generally supports the importance of fluency in reading comprehension.

The whole process of fluency development in L2 thus needs to be examined more closely. The current study concentrates on the development of fluency in second language (L2) reading and writing. The main interest in the study was to find out whether, in time-pressure tasks, speed as a reflection of a more or less automatic skill could be used to predict L2 reading and writing, which are higher-order skills. This was done by measuring the speed of performance in cognitive and linguistic tasks. These tasks were conducted in both the mother tongue and in the second or foreign language of the participants. This allowed us to compare the possible differences in the second and foreign language learners, as the second language learners live in an authentic language environment and the foreign language is only learned in the school classes. Also the rarer language combinations of Finnish/English and Russian/Finnish can give us new insights on the matter. The cognitive tasks were conducted orally, as Norman Segalowitz (2010), for example, points out that oral fluency is taken to reflect higher rather than lower-level processing, i.e. articulation. The scores from the speed of performance were then correlated with measures of reading and writing in L2 to see if the former could predict performance in the latter. The three different age groups were also compared to see differences in skill development. The design also allowed us to examine whether the speed of performance in L1 had any affect on the L2 skills, which is a much-debated question. The role of L1 for second and foreign language learners is basically different and the differences in the results of this study also reflect this aspect.

The research questions in this study were:

1. Does speed in cognitive tasks develop with age?
2. Can differences in fluency measures be found between second and foreign language learners?
3. Does the speed in time-pressure cognitive and linguistic tasks performed in L1 and L2 predict second/foreign language reading and writing skills?

---

<sup>1</sup> See: [https://www.jyu.fi/ytk/laitokset/psykologia/huippututkimus/en/research/JLD\\_main/JLD](https://www.jyu.fi/ytk/laitokset/psykologia/huippututkimus/en/research/JLD_main/JLD) (21.12.2012).

## 2. Method

### 2.1. Participants

The data were gathered in the DIALUKI project (Diagnosing reading and writing in a second or foreign language<sup>2</sup>) in a cross-sectional setting in 2010–2011. Participants ( $n = 823$ ) were selected from 111 schools in Finland. The first group consisted of Finnish-speaking learners of English and was divided into sub-groups from three different grades. All these children had started learning English as a foreign language in the third grade. The second group consisted of primary school children with Russian as L1 and Finnish as L2 (see Table 1). Russian students were immigrants living in Finland, and their language skills varied widely in both languages. At the time of the study, they either participated in preparatory classes or were already integrated within mainstream education, depending on their language skills and length of residence. For some of them Finnish was even a stronger language than Russian. Agreements for participation in the study were collected from the county, schools, parents and children themselves.

**Table 1.** Participants of the study

Mother tongue	Grade / age	Foreign / second language	
		English	Finnish
Finnish	4th grade / 9–10 years	210	
	8th grade / 13–14 years	208	–
	Gymnasium / 17–18 years	219	
Russian	3rd–6th grades (mean age 10.9 years)	–	186
Total		637	186

### 2.2. Procedure

Tests were administered in the school during school hours. There were two parts in the test battery: reading and writing tasks and the speed-measuring linguistic tasks were performed in groups in the classroom; the individual cognitive tasks were computer-assisted and administered by trained assistants.

### 2.3. Speed-measuring linguistic tasks

*Word segmentation task.* The children were given a short text written without spaces between the words and without punctuation. The task was to segment the text by drawing a pencil stroke on word boundaries (i.e. between the words). The task was performed in both L1 and L2, but the texts were different for the different age groups. The time the children took to complete the task was recorded (to the nearest 30 seconds); also, the number of correctly separated words was counted.

<sup>2</sup> See: <https://www.jyu.fi/hum/laitokset/solki/tutkimus/projektit/dialuki> (21.12.2012).

*Spelling error detection task.* In the spelling error detection task, the children were given a paper with a list of 100 Finnish words each containing a spelling error. The spelling errors were of several different types; either there was a letter missing or extra in the word (*lauvantai* instead of *lauantai*), or there was a wrong letter (*heunäkuu* – *heinäkuu*), or incorrect inflections (*aurinkonkukka* – *auringonkukka*). Children were instructed to mark all the spelling errors they could detect but not to correct them. They were given 3.5 minutes to mark as many as they could. The maximum score was 100 points and the number of correct answers was counted. The task was designed for youths and adults, so only the eighth graders and gymnasium students took this task (Holopainen et al. 2004). The task was conducted only in Finnish.

## 2.4. Cognitive tasks

Students performed the cognitive tasks with the *Cognitive Workshop* computer program developed in the *Jyväskylä Longitudinal Study of Dyslexia*. The test battery consisted of eight different tasks in total, and three of these measured performance speed. These tasks were *Rapidly Presented Words*, *Reading Aloud Word List*, and *Rapid Alternating Stimulus*. The original tasks had been developed and validated in the JLD project; the Russian and English versions of these tasks were developed by the DIALUKI project. Each task is described in more detail below.

*Rapidly Presented Words* (RPW). The task measured reading (and perceiving) a word as a whole. The words appeared one at a time on the computer screen for 80 milliseconds and the child had to read them aloud. There were 14 words in L1 and 8 or 12 words in L2, depending on students' age. All the words were simple, everyday words but in the task they became progressively longer, from two-letter words up to a nine-letter word. A star first flashed on the screen as a prompt, after which the target word appeared. After 80 milliseconds it was replaced by a mask (e.g., B/W#Q?) to erase the visual image from the working memory. The next stimulus was given only after the child had answered or confirmed not having perceived the word. Correct answers were counted and converted to percentages to enable the comparisons of performances on tasks consisting of a different number of stimuli reported later in this article.

*Reading Aloud Word List*. The children were given a list consisting of 105 words. They were instructed to read the list as fast and as accurately as they could in one minute. The test was conducted in both L1 and L2. The Finnish word list was from the Lukilasse test battery developed for diagnosing reading in L1 Finnish in primary school (Häyrinen et al. 1999). The English and Russian word lists were designed by sampling from frequency lists for these two languages.<sup>3</sup> Time was measured with a stopwatch and the last word read by the child was marked. Afterwards, the words were divided into syllables, which makes the word lists across languages more (although not completely) comparable. This was necessary as the words in both Finnish and Russian typically consist of several syllables and, thus, take longer to pronounce, whereas many frequent English words are rather short.

*Rapid Alternating Stimulus* (RAS) and *Rapid Automatic Naming* (RAN) are considered to be very good tests for measuring the automaticity of language skills

<sup>3</sup> See: [http://ucrel.janacs.ac.uk/bncfreq/lists/1\\_2\\_all\\_freq.txt](http://ucrel.janacs.ac.uk/bncfreq/lists/1_2_all_freq.txt) (15.10.2010), <http://corpus.leeds.ac.uk/serge/frqlist/rnc-modern.num.html> (15.10.2010).

(e.g. Cronin, Carver 1998). It is a matrix consisting of letters, numbers, colours or simple objects or a combination of these. The matrix used in this study was a mixed one with numbers, colours and letters or pictures. This made the test more difficult and therefore more prone to mistakes compared with a task consisting of only numbers, for example. The basic matrix consisted of 50 units. For primary school participants it was shorter in L2 (30 units) and letters were replaced by pictures of easily identifiable objects (i.e., pen, car and house). Children were told to read the entire matrix aloud as fast and as accurately as they could. The completion time for the test was measured by a stopwatch. To make comparisons across the 30 and 50-unit versions of the task possible, the overall completion times were converted to time-per-unit values.

## 2.5. Reading and writing measures

The second/foreign language reading and writing tasks that were in the study were standardised and age-appropriate.

*Reading tasks.* In the primary school, reading comprehension tasks in English as a foreign language were three tasks from the Pearson Young Learners test. The older students had four tasks from the Pearson Test of English General test. The Finnish L2 reading comprehension tasks were three tasks from the YKI test (National Certificate for Language Proficiency) and ten tasks from DIALANG. All reading tasks were paper-and-pencil tests, except DIALANG, which is an online diagnostic system, designed to assess proficiency in fourteen different European languages. The reading test scores were analysed with Rasch software (Winsteps)<sup>4</sup>, and reading ability values were used as dependent variables in the analyses focusing on reading.

*Writing tasks.* The writing tasks were paper-and-pencil tasks. The English L2 writing tasks were a letter or e-mail to a friend, a short story and an opinion; there was one task for the fourth graders and three for the older groups. Finnish L2 tasks included a letter to a friend and a short story. The tasks were taken from the above-mentioned Pearson tests and the Topling research project. Two professional raters evaluated the learners' performances against the Common European Framework of Reference (CEFR) 6-point scale<sup>5</sup>. The ratings were analysed with the multi-faceted Rasch programme Facets<sup>4</sup>, and the ability values that take into account both the raters' leniency/severity and the task difficulty were later used as the dependent variables in the regression analyses reported below.

## 3. Results

### 3.1. Analysis of variance

First the speed of performance in the cognitive tasks was compared across the age groups (fourth graders, eighth graders, and gymnasium students). The second and foreign language learners (i.e. the Russian-speaking learners of Finnish and Finnish-speaking learners of English) were also compared to see the possible differences

<sup>4</sup> See: <http://www.winsteps.com> (21.12.2012).

<sup>5</sup> See (21.12.2012): YKI test [http://www.oph.fi/english/mobility/testing\\_language\\_skills/](http://www.oph.fi/english/mobility/testing_language_skills/); DIALANG <http://www.lancs.ac.uk/researchcenterprise/dialang/about>; Topling. Paths in Second Language Acquisition <https://www.jyu.fi/hum/laitokset/kielet/topling/en>.



between these groups. For these, one-way analysis of variance (ANOVA, see Hatch, Farhady 1982) was used.

The results concerning the *Rapidly Presented Words* (RPW) task are presented for L1 and L2, respectively. Tables 1 and 2 present the means (M) and standard deviations (SD) of the percentages of correct words in each group. The results show that automatic recognition of vocabulary increased with age. The RPW task was able to distinguish between different age groups: a particularly clear distinction was found between the primary school and the older children both in the L1 and L2 condition (for L1 Finnish/Russian  $F = 75.38, p < 0.0001, df = 3$ , and for L2 English/Finnish  $F = 145.03, p < .0001, df = 3$ . The effect size  $\eta^2$  for L1 was 0.235 and for L2 0.364). The task was quite easy for the two oldest age groups and a ceiling effect was in fact found for them; thus, the RPW tasks used in the study could not differentiate between the eighth graders and the gymnasium students. Interestingly, in the L1 task, both the Finnish and Russian-speaking children performed equally well (assuming that the different language versions of the task are comparable; see Table 2), whereas in the L2 task, the Russian-speakers outperformed their same age Finnish-speaking peers although they fell far behind the older Finnish-speaking learners (Table 3).

**Table 2.** Rapidly Presented Words in L1 (Finnish / Russian): percentage of correct words out of 14

Group	n	M	SD
4th grade	203	73.65	24.91
Russian primary	153	70.63	31.88
8th grade	194	93.37	6.32
Gymnasium	191	95.10	4.89

**Table 3.** Rapidly Presented Words in L2 (English / Finnish): percentage of correct words out of eight words for primary school groups and twelve words for older groups

Group	n	M	SD
4th grade	202	58.79	27.83
Russian primary	177	66.24	32.33
8th grade	192	93.45	11.06
Gymnasium	192	97.35	8.56

The results from the *Reading Aloud Word List* task are displayed as syllables read per second (see Tables 4 and 5). The results show that speed (syllables per second) increased with age and, presumably, with language skill development. Overall, the results are somewhat similar to those reported above for the *Rapidly Presented Words* task. On the whole, the groups are very clearly separated (for L1 Finnish/Russian  $F = 230.66, p < 0.0001, df = 3$ ; and for L2 English/Finnish  $F = 210.08, p < 0.0001, df = 3$ . The effect sizes are very large: 0.474 for L1 and 0.452 for L2). Again, the primary school age students performed less well than the older students but this time the Russian-speaking students' L2 Finnish performance was not only better than their same-age Finnish peers' L2 English performance but was at the same level as the eighth graders (Table 5). On this task, the oldest group (gymnasium) outperformed all younger groups statistically significantly, both in the L1 and L2 tasks.

**Table 4.** Reading Aloud Word List in L1 (Finnish / Russian): syllables per second

Group	n	M	SD
4th grade	205	2.76	0.76
Russian primary	182	2.60	1.20
8th grade	195	4.13	0.97
Gymnasium	189	4.86	1.01

**Table 5.** Reading Aloud Word List in L2 (English / Finnish): syllables per second

Group	n	M	SD
4th grade	204	1.32	0.42
Russian primary	182	2.30	0.85
8th grade	194	2.36	0.66
Gymnasium	190	2.95	0.66

In the *Rapid Alternating Stimulus* (RAS) task, the performance speed also increased with age (see Tables 6 and 7). The results of the RAS tasks were almost identical to those reported above for the *Reading Aloud Word List*: the two younger groups (Finnish-speaking and Russian-speaking primary students) performed the same on the L1 RAS task but the Russian-speakers outperformed the Finns on the second/foreign language version of the same task, and reached the same average speed as the eighth-grade Finns. Again, the gymnasium students did these tasks significantly faster than any of the younger groups. (Details of the ANOVA results are as follows: for L1  $F = 82.99$ ,  $p < 0.0001$ ,  $df = 3$ ; and for L2  $F = 120.55$ ,  $p < 0.0001$ ,  $df = 3$ . The effect sizes are also quite large: 0.246 for L1 and 0.321 for L2).

**Table 6.** Rapid Alternating Stimulus in L1 (Finnish/Russian): average time in seconds spent on each item (50 items)

Group	n	M	SD
4th grade	204	0.87	0.18
Russian primary	177	0.83	0.31
8th grade	195	0.68	0.16
Gymnasium	190	0.59	0.13

**Table 7.** Rapid Alternating Stimulus in L2 (English/Finnish): average time in seconds spent on each item (30 items in primary school and 50 items in 8th grade and gymnasium)

Group	n	M	SD
4th grade	205	1.20	0.41
Russian primary	180	0.89	0.22
8th grade	195	0.86	0.25
Gymnasium	190	0.68	0.15

### 3.2. Regression analyses

Stepwise multiple regression analyses were used to test to what extent the time-pressure cognitive and linguistic tasks explain reading and writing skills. The dependent variables were reading and writing in L2 (ability measures derived from the Rasch analyses of the reading scores and ratings of the writing scripts) and the independent variables (IV) were the time-pressure linguistic and cognitive tasks (see Table 8).

Overall, the speed-related measures predicted writing in L2 better than reading in L2: The amount of explained variance was at least 10%, sometimes even 20%, higher for writing. For reading, about 20% of the variance could be predicted from performance on the speed measures included in the study. For writing, the explained variance rose to about 30–40%. Interestingly, the prediction was somewhat better for the L2 (Russian-speakers) learners than it was for the foreign language learners (Finnish-speakers), in both reading and writing.

For the Finnish-speaking learners of English, the best predictor for the primary school group was shown to be L2 the *Rapidly Presented Words* (RPW) either in L1 or L2, and *Rapid Alternating Stimulus* (RAS) in L2 for other grades. However, several other measures of speed, including some linguistic speed tasks, also predicted reading and writing performance significantly among the Finnish-speaking learners. In comparison, the set of predictors of Finnish as a L2 reading and writing among the Russian-speaking learners looks somewhat different. *Reading Aloud Word List* in L2 Finnish dominates as the best predictor for both reading and writing. *Rapidly Presented Words* in L2 and (for writing) also *Reading Aloud Word List* in L1 Russian appear as significant predictors of L2 macro skills.

**Table 8.** Speed in the time-pressure cognitive tasks as a predictor of reading and writing in a second/foreign language

Age groups	Dependent variable	Adjusted R Squared	% variance	1st IV	2nd IV	3rd IV	4th IV
4th grade Finnish L1	Writing L2	0.272	<b>27.2%</b>	RPW L1	RAS L2		
	Reading L2	0.188	<b>18.8%</b>	RPW L2	RAS L2		
Primary Russian L1	Writing L2	0.459	<b>45.9%</b>	Word List L2	Word List L1	RPW L2	
	Reading L2	0.300	<b>30.0%</b>	Word List L2	RPW L2		
8th grade Finnish L1	Writing L2	0.422	<b>42.2%</b>	RAS L2	Segmentation L1	RPW L2	Segmentation L2
	Reading L2	0.218	<b>21.8%</b>	RAS L2	RPW L2		
Gymnasium Finnish L1	Writing L2	0.322	<b>32.2%</b>	RAS L2	Spelling errors L1	Word List L1	
	Reading L2	0.242	<b>24.2%</b>	RAS L2	Segmentation L1	Word List L1	

## 4. Discussion

The study shows that speed in the fluency tasks increases with age and education, at least among the learner groups studied here. Three different age groups were distinguished in most tasks. In both primary school groups, in which second or foreign language skills are only starting to develop, performance on the L1 versions of the tasks was always better than in the L2 tasks. When language skills become more fluent, the difference starts to diminish. In this study the difference was still mostly visible also in the older groups, which is in accord with Jan Hulstijn, Amos Van Gelderen and Rob Schoonen (2009), for example. Where no difference between L1 and L2 could be found, in particular in the *Rapidly Presented Words* task, this was mainly due to the task construction. The English version of the RPW task was clearly too easy for the older students and the results were influenced by a ceiling effect. In the *Word List* task the Finnish words used were much longer with complex inflections while the English and Russian versions contained easier basic words. For the comparability of the results they were analysed as syllable/second. Interestingly, the Finnish version of the list was also the language version that did have the most predictive value with reading and writing skills in the regression analysis, as both L1 and L2 task. It was quite a challenging task even for older children as the word-length came up to a 22 letter-word (e.g. *prosessikirjoittaminen* 'process writing').

The analyses revealed some differences between the two language groups. The primary school Russian speakers were closest to the corresponding Finnish age group in L1 tasks, but in L2 tasks they were markedly better and even indistinguishable from the eighth grader Finns. This might be caused by several factors. For some learners labelled as Russian-speakers, Finnish was in fact their stronger language. Their Russian skills were more varied, as we can see from the standard deviations on their L1 tasks. For the Finnish students English is a foreign language learned at school; the youngest group had only studied English for slightly over a year. For the Russian speakers, Finnish is a second language learned in an authentic environment. There is also a slight age difference in favour of Russian speakers. Positive effects that are often associated with bilingualism can also have some effect on the results (e.g. Bialystok 2001).

The time-pressure tasks studied here were found to predict the learners' reading and writing performance in L2 rather well. The regression analyses indicated that the speed measures predicted writing in L2 better than reading in L2, and that the prediction was somewhat better for the second language group than it was for the foreign language groups. The best predictor of reading and writing measures among the Finnish-speaking primary school group was the *Rapidly Presented Words*, which was also a significant predictor for the Russian-speaking primary school learners. The current versions of the RPW task, which made use of simple, very frequent words were too easy for the older students whose word recognition skills could be expected to be more automatic. It is possible that a more difficult version of the RPW could be useful in predicting L2 macro skills such as reading and writing but this remains to be studied in the future. Also, in line with earlier studies, *Rapid Alternating Stimulus* explained a significant amount of variance in L2 reading and especially writing, at least for the Finnish speakers. It can be argued to be a good predictor of automaticity in language as it combines different skills that

are required for fluent reading. Overall, the results are in accordance with previous research in that the lower level processes can predict more holistic processes, such as reading and writing also in a foreign language. In addition to the L2 tasks, the L1 tasks also had predictive power in the current study which has not been the case in all studies (cf. Van Gelderen et al. 2007). It would be reasonable to assume that L1 had influence only on the lower grades, when the L2 skills are still at quite a low level. This study shows the L1 effect on the L2 skills also in the older groups. For them it was especially the L1 linguistic tasks (i.e. *Segmentation* and *Spelling errors*) that showed up in the regression analysis.

Hulstijn et al. (2009) conclude their analysis stating that processing speed may not be a very useful criteria for automaticity. Segalowitz (2010: 79) also proposes that for a skill to be regarded as automatized, speed cannot be the only criterion. This study has looked only at the speed component of fluency, and to examine the wider phenomenon of fluency, an analysis of accuracy of performance could bring a more balanced and informative view. From the data of this study it was evident that often the time to complete the task was only a part of the whole performance. The time-pressure in the tasks seemed to result in a trade-off of speed and accuracy, with quick performance sometimes revealing on a closer inspection to consist only of approximated responses. The combined effect of speed and accuracy should be therefore examined in the further analyses.

## 5. Conclusion

The main conclusion from the work presented here is that the speed in the time-pressure tasks in both L1 and L2 can be used in predicting L2 reading and especially writing skills. This applied also to the older students, even when some of the tasks used in this study were not very suitable for them. All in all, the tasks with the best predictive value differed somewhat for the different age groups and for the second or foreign language learners. For further study, an even more informative view could be attained by combining these speed of performance measures with accuracy measures from the same tasks.

### Abbreviations

L1	first language
L2	second or foreign language
M	mean
RAS	<i>Rapid Alternating Stimulus</i> test
RPW	<i>Rapidly Presented Words</i> test
SD	standard deviation

### References

- Bialystok, Ellen 2001. *Bilingualism in Development. Language, Literacy and Cognition.* Cambridge: Cambridge University Press.
- Bowey, Judith A.; Storey, Tiffany; Ferguson, Angela N. 2004. The association between continuous naming speed and word reading skill in fourth- to sixth-grade children. –

- Australian Journal of Psychology, 56 (3), 155–163. <http://dx.doi.org/10.1080/00049530412331283345>
- Cronin, Virginia; Carver, Paula 1998. Phonological sensitivity, rapid naming, and beginning reading. – *Applied Psycholinguistics*, 19 (3), 447–461. <http://dx.doi.org/10.1017/S0142716400010262>
- Di Filippo, Gloria; Brizzolaro, Daniela; Chilosi, Anna; De Luca, Maria; Judica, Anna; Pecini, Chiara; Spinelli, Donatella; Zoccolotti, Pierluigi 2005. Rapid naming, not cancellation speed or articulation rate, predicts reading in an orthographically regular language (Italian). – *Developmental Neuropsychology*, 30, 885–904.
- Dufva, Mia; Voeten, Marinus J. M. 1999. Native language literacy and phonological memory as prerequisites for learning English as a foreign language. – *Applied Psycholinguistics*, 20 (3), 329–348. <http://dx.doi.org/10.1017/S014271649900301X>
- Grabe, William 2010. Fluency in reading – thirty-five years later. – *Reading in a Foreign Language*, 22, 71–83.
- Hatch, Evelyn; Farhady, Hossein 1982. *Research Design and Statistics for Applied Linguistics*. Cambridge: Newbury House.
- Holopainen, Leena; Kairaluoma, Leila; Nevala, Jukka; Ahonen, Timo; Aro, Mikko (Eds.) 2004. *Lukivaikeuksien seulontamenetelmä nuorille ja aikuisille*. [Dyslexia Screening Test for Youth and Adults.] Jyväskylä: Niilo Mäki Instituutti.
- Hulstijn, Jan; Van Gelderen, Amos; Schoonen, Rob 2009. Automatization in second-language acquisition: What does the coefficient of variation tell us? – *Applied Psycholinguistics*, 30 (4), 555–582. <http://dx.doi.org/10.1017/S0142716409990014>
- Häyrynen, Taru; Serenius-Sirve, Silve; Korkman, Marit 1999. *Lukilasse. Lukemisen, kirjoittamisen ja laskemisen seulontatutkimus peruskoulun ala-asteen luokille 1–6*. [The Lukilasse Graded Achievement Package for Comprehensive School Age Children.] Helsinki: Psykologien Kustannus Oy.
- Kaakinen, Johanna K. 2004. Perspective effects on text comprehension: Evidence from recall, eyetracking, and think-alouds. *Annales Universitatis Turkuensis, Ser. B*, 275.
- Kirby, John R.; Parrila, Rauno K.; Pfeiffer, Shannon L. 2003. Naming speed and phonological awareness as predictors of reading development. – *Journal of Educational Psychology*, 95 (3), 453–464. <http://dx.doi.org/10.1037/0022-0663.95.3.453>
- Kroll, Judith F.; Sunderman, Gretchen 2003. Cognitive processes in L2 learners and bilinguals. – Catherine J. Doughty, Michael H. Long (Eds.). *Handbook of Second Language Acquisition*. Malden, MA: Blackwell Publishing.
- Landerl, Karin; Wimmer, Heinz 2008. Development of word reading fluency and orthographic spelling in a consistent orthography: An 8-year follow-up. – *Journal of Educational Psychology*, 100 (1), 150–161. <http://dx.doi.org/10.1037/0022-0663.100.1.150>
- Puolakanaho, Anne; Ahonen, Timo; Aro, Mikko; Eklund, Kenneth; Leppänen, Paavo; Poikkeus, Anna-Maija; Tolvanen, Asko; Torppa, Minna; Lyytinen, Heikki 2007. Very early phonological and language skills: estimating individual risk of reading disability. – *Journal of Child Psychology and Psychiatry*, 48 (9), 923–931. <http://dx.doi.org/10.1111/j.1469-7610.2007.01763.x>
- Torgesen, Joseph K.; Wagner, Richard K.; Rashotte, Carol A.; Burgess, Stephen; Hecht, Stephen 1997. Contributions of phonological awareness and rapid automatic naming ability to the growth of word-reading skills in second- to fifth-grade children. – *Scientific Studies of Reading*, 1 (2), 161–185. [http://dx.doi.org/10.1207/s1532799xssr0102\\_4](http://dx.doi.org/10.1207/s1532799xssr0102_4)
- Segalowitz, Norman 2003. Automaticity and second languages. – Catherine J. Doughty, Michael H. Long (Eds.). *Handbook of Second Language Acquisition*. Malden, MA: Blackwell Publishing, 382–408.
- Segalowitz, Norman 2010. *The Cognitive Bases of Second Language Fluency*. New York: Routledge.

- Stroop, John Ridley 1935. Studies of interference in serial verbal reaction. – Journal of Experimental Psychology, 18 (6), 643–662. <http://dx.doi.org/10.1037/h0054651>
- Van Gelderen, Amos; Schoonen, Rob; Stoel, R. D.; De Gloppe, Kees; Hulstijn, Jan 2007. Development of adolescent reading comprehension in language 1 and language 2: A longitudinal analysis of constituent components. – Journal of Educational Psychology, 99 (3), 477–491. <http://dx.doi.org/10.1037/0022-0663.99.3.477>
- Wolf, Maryanne; Bowers, Patricia Greig 1999. The double-deficit hypothesis for the developmental dyslexias. – Journal of Educational Psychology, 91 (3), 415–438. <http://dx.doi.org/10.1037/0022-0663.91.3.415>

## Online references

- DIALANG. <http://www.lancs.ac.uk/researchenterprise/dialang/about> (21.12.2012).
- DIALUKI = Diagnosing reading and writing in a second or foreign language. <https://www.jyu.fi/hum/laitokset/solki/tutkimus/projektit/dialuki> (21.12.2012).
- English frequency lists. [http://ucel.lancs.ac.uk/bncfreq/lists/1\\_2\\_all\\_freq.txt](http://ucel.lancs.ac.uk/bncfreq/lists/1_2_all_freq.txt) (15.10.2010).
- JLD = Jyväskylä Longitudinal Study of Dyslexia. [https://www.jyu.fi/ytk/laitokset/psykologia/huippututkimus/en/research/JLD\\_main/JLD](https://www.jyu.fi/ytk/laitokset/psykologia/huippututkimus/en/research/JLD_main/JLD) (21.12.2012).
- Russian frequency lists. <http://corpus.leeds.ac.uk/serge/frqlist/rnc-modern.num.html> (15.10.2010).
- Topling. Paths in Second Language Acquisition. <https://www.jyu.fi/hum/laitokset/kielet/topling/en> (21.12.2012).
- YKI test = National Certificate of Language Proficiency. [http://www.oph.fi/english/mobility/testing\\_language\\_skills](http://www.oph.fi/english/mobility/testing_language_skills) (21.12.2012).
- Winsteps program. <http://www.winsteps.com> (21.12.2012).

**Sanna Olkkonen** (University of Jyväskylä). Her academic research interests include cognitive processes in second language acquisition, especially development of fluency in second language. [sanna.olkkonen@jyu.fi](mailto:sanna.olkkonen@jyu.fi)

# KOGNITIIVSETE ÜLESANNETE LAHENDAMISE KIIRUS KUI VÕORKEELE / TEISE KEELE LUGEMIS- JA KIRJUTAMISOSKUSE NÄITAJA

Sanna Olkkonen

Jyväskylä Ülikool

Artikkel analüüsib võõrkeeleskuse automaatsuse arengut, mis on teise keele uurin-gutes veel väheksitletud valdkond. Käesolev läbilõikeuuring kasutas ajasurvega kognitiivseid ja keelelisi ülesandeid ning analüüsis, kuidas soorituskiirus vanusega ning õppimise käigus areneb. Andmed koguti aastatel 2010–2011 Soome koolide 823 õpilaselt. Soomekeelsed lapsed, kes õppisid põhikooli neljandas ja kahek-sandas klassis ning gümnaasiumi teises klassis, õppisid inglise keelt võõrkeelena. Venekeelsed lapsed olid algkooliõpilased ning õppisid soome keelt teise keelena. Hindamaks kiirust mõõtvate ülesannete võimet ennustada võõrkeele (teise keele) lugemis- ja kirjutamisoskust, kasutati regressioonanalüüsi.

Uuring näitas, et algkoolis, kui keeleskus alles hakkab arenema, oli sooritus emakeeles parem kui teises keeles või võõrkeeles. Kui keeleskus enam automa-tiseerus, hakkas see erinevus kaduma. Soorituskiirus üldiselt paranes vanusega, kuigi kohati ei olnud võimalik kahe vanema soomekeelse rühma tulemusi eristada. Venekeelsete algkooliõpilaste tulemused olid samas vanuses soomekeelse rühma omadest sageli paremad. Siiski ei olnud need enamasti vanemate soomekeelsete õpilaste tasemel. Parimateks võõrkeele lugemis- ja kirjutamisoskuste ennustaja-teks osutusid nooremate õpilaste puhul sõnade kiire esitamise ülesanne (*Rapidly Presented Words*) ja vanemate õpilaste puhul kiirelt vahelduv stiimul (*Rapid Alternating Stimulus*). Venekeelsete õpilaste puhul oli parimaks ennustajaks sõnaloendi valjult ettelugemise ülesanne (*Reading Aloud Word List*). Ka emakeelsed ülesan-ded suutsid teatud määral võõrkeele (teise keele) oskust ennustada. Kokkuvõttes ennustasid ajasurve ülesanded paremini kirjutamist kui lugemist.

**Võtmesõnad:** soorituskiirus, lugemine, kirjutamine, kognitiivne testimine, teine keel, soome keel, inglise keel, vene keel



# INTERVIEWER BEHAVIOUR DURING ORAL PROFICIENCY INTERVIEWS: A GENDER PERSPECTIVE

Edith Reemann, Ene Alas, Suliko Liiv

**Abstract.** The article investigates the behaviour of interviewers during the oral proficiency interview within the framework of the national examination in the English language in Estonia from the gender perspective. The research investigates whether noticeable differences can be traced in the behavioural patterns of male and female interviewers that could lead to placing the candidates in a disadvantaged position, not allowing them to display their language proficiency to the full. The study involves nine female and nine male interviewers whose interview language is subjected to content analysis. The results show an overall attempt to follow the scripted interview format envisaged for this examination, but alterations to the script prevail with both gender groups.\*

**Keywords:** interviewer, oral proficiency interview, validity, reliability, scripted interview, gender variation, standardisation

## 1. Introduction

Interviewer behaviour has been discussed in research literature from various angles: accommodation, i.e. how much interviewers adjust their speech to the person taking the test (Ross 1992, Ross, Berwick 1992, Lazarton 1996, 2002, Malvern, Richards 2002); interviewer proficiency, i.e. language proficiency combined with procedural proficiency (Morton et al. 1997, McNamara and Lumley 1997); variation in personality and communication style (Brown 2003, Luoma 2004); and rapport-establishing techniques with the candidate (Morton et al. 1997), all of which affect the assessment of candidates. For a detailed discussion of the above aspects, see Alas (2010: 25–28).

This article will concentrate on interviewer gender. O’Loughlin (2002) asserts that the female interview style is ‘collaborative, symmetrical, co-operative, and

---

\* The research was partly supported by the Estonian Science Foundation Grant No 9037.

supportive whereas its male equivalent is controlling, asymmetrical, uncooperative, and unsupportive' (2002: 170). Her research indicated that interviewers demonstrated a limited use of overlaps, a negligible use of interruptions and a widespread use of minimal responses in the interviews, but the use of these features did not follow any clear gender pattern. There was a high degree of variability in the use of overlaps and especially minimal responses within the different gender pairings. Most importantly, both female and male participants indicated their ability to make supportive contributions to the interviews through their use of positive overlaps and minimal responses in particular. O'Loughlin concludes that other interlocutor characteristics – interviewer training, professional orientation – probably affect the behaviour of the interviewer more than gender (2002: 190). For a detailed discussion of gender-related research see Alas (2010: 26).

Proceeding from the ambiguity of the findings of international research, and because relevant studies have not been conducted in Estonia, a research hypothesis was posed as follows: the conduct of interviewers in oral proficiency interviews (OPI) during the national examination in the English language in Estonia will display gender-related differences, which follow identifiable patterns.

## 2. Method

Interviewers involved in oral proficiency interviews for the English language national examinations in Estonia are practising teachers who are required to have undertaken standardised interviewer/assessor training. The interviewers must follow an interviewer script verbatim and adhere to the time limits for different parts of the interview. The candidates they interview are usually familiar to them.

Eighteen interviews were randomly selected from among the recordings of interviews conducted during the examination session of 2011. The principle of choosing the recordings was that the interview had to have taken place during that session, and there was to be an equal number of male and female interviewers. Although 7917 students took the English examination in the spring 2011 examination session (Põder 2011), the choice of the interviews for analysis was made from a limited pool because the recording is only made at the candidate's discretion. The interviews were transcribed and assigned letter-codes and numbers in a random order, M1 to M9 for male and F1 to F9 for female interviewers.

The speaking test of the national examination in the English language consists of an introduction (not rated), a monologue and follow-up questions (task 1), and a role-play (task 2). The overall time for a rateable sample from each candidate is estimated at between 14 and 16 minutes. The database of the eighteen recordings comprised 227 minutes and 49 seconds of interview time in total.

The data were processed using content analysis, which proceeded from first identifying the salient features of each interview section in the script (for a complete script, see REKK 2008), which then served as criteria for the evaluation of each interviewer. Once the features had been found, research proceeded to trace their presence or absence in each interview. This was followed by juxtaposition of the two gender groups to discover the group properties. Changes to the script also served as a subject to content analysis.

### 3. Content analysis and results

The section presents the findings of the current study along with some preliminary interpretation. As the number of participants is fairly small, the results are given in raw figures and not in percentage.

#### 3.1. Introduction

Analysing the representation of the salient features in the introduction (Table 1) we can see that in 17 interviews out of 18, the introductory stage is present.

**Table 1.** Interview introduction element representation in interviews

No.	Element	Present/missing in interviews	
		Men	Women
1.	Candidate number/code	7/2	9/0
2.	Greeting the candidate	4/5	5/4
3.	Introducing the interviewer	6/3	6/3
4.	Introducing the assessor	6/3	6/3
5.	Enquiring about well-being	9/0	7/2
6.	Asking warm-up questions	8/1	9/0

All women start with the warm-up stage, but with male interviewers, a discernable introductory stage is present in eight. In M6, the interviewer does not introduce himself or the assessor; or indeed ask any of the prescribed warm-up questions. Instead, he asks one question which seems to be based on prior knowledge of the candidate's habits.

Candidate's examination code is not skipped by any interviewer. There are, however, differences in the point at which the number is stated: M7 notes it at the beginning of task 1, M9 states it after the candidate has been asked to choose task 1 topic, i.e. half way through the explanation of task 1. Overall, female interviewers were more disciplined here.

Greeting the candidate divides the interviewers into two almost equal groups, those who greet the candidate at the start of the recorded interview (M1, M2, M4, M6, F2, F4, F6, F7, F9), and the rest do not. The behaviour of male and female interviewers does not differ in this aspect.

The interviewers must introduce themselves at the beginning of the exam. Six male interviewers (M1, M2, M5, M7, M8, M9) do so, M3, M4 and M6 do not. The number of female interviewers in this group is also six (F1, F4, F5, F6, F7, F9). The reason is probably because candidates and interviewers are already familiar with each other as teachers and students, thus introductions seem nugatory.

The interviewer has to ask the candidates how they are feeling, helping to relieve examination-related anxiety. All the men do so and respond appropriately; M2, M4, and M7 follow the script verbatim. M1 rephrases the question and M6 asks 'Are you nervous?', suggesting nervousness rather than deflating it, and follows it up with a seemingly inappropriate question ('Did you jog in the morning?'). Seven women

enquire about the candidate's well-being: F2, F3, F5, and F6 verbatim. The script expression is expanded in F1, F7 and F9. F4 and F8 skip this part and move straight to warm-up questions. In this section, men adhere to the script more than women.

The focal points of the interview introduction are four warm-up questions, which have the function of leading the candidate into the interview mode. The questions are easy, general interest questions that are set below the test's general B2 proficiency level. They are designed to relieve the candidate's anxiety and increase their confidence (cf. Alas, Liiv 2009). The questions were asked by all but one interviewer: M6 skips the warm-up completely. Fourteen interviewers confine themselves to the script without developing candidate responses, however, F1, F4, and F9 add questions here.

Different types of deviant behaviour are noted. Men are more likely to leave out questions, e.g. M7 only asks one of four warm-up questions. As a result, his introductory phase only lasts 46 seconds, which is over a minute less than the two minutes to which the candidate is entitled.

On one occasion (F1), the interviewer asks all the warm-up questions at once and lets the student decide which question to answer, then adds an additional question. F4 also asks additional questions and when the candidate does not answer in the expected manner, she guides the student in the appropriate direction by asking for clarification. F9 changes the questions to accommodate the student who seems reluctant to respond to the script questions.

The end of the introduction should be signalled by the discourse marker 'Thank you'. Five men (M1, M4, M5, M8, M9) end this stage properly. M2 ends with 'okay', M3 and M7 with 'alright' and in M6 the introductory stage is not present. The introduction is appropriately finished by six female interviewers (F1, F3, F4, F5, F6, F9). F2 and F7 ended the introduction with 'okay' and F8 does not use words at all, signalling the end with a back-channelling sound 'mhmh'.

The maximum length of the stage with men is 1 minute 22 seconds and the minimum is 28 seconds. The mean duration is 1 minute 1second, brought down by M6 skipping the warm-up phase. The maximum length of this stage with female interviewers is 1 minute 44 seconds and the minimum is 54 seconds. The mean time spent is 1 minute 14 seconds. The longer mean length can be attributed to female interviewers asking the warm-up questions, and occasionally adding supplementary questions (F1, F4, F9). It should be noted that the interviewers in general could have made more of the 2 minutes allotted to the introduction, as the mean as well as the individual introduction lengths fall considerably short of the given time. Clear gender differences are present in the minimum and maximum length of the interviews, as well as the mean length. All women spend more time with each candidate than men, they allow more speaking time for the candidate, appearing to be more accommodative conversation partners than men.

### **3.2. Task 1. Monologue and follow-up questions**

In task 1, the script has been developed to provide maximum support for the candidate and must be quoted verbatim to ensure uniform treatment of candidates. The criteria selected to analyse the transition to task 1, as well as their representation in the interviews are summarised in the Table 2 below.

**Table 2.** Element representation in task 1

No.	Element	Present/missing in interviews interviews	
		Men	Women
1.	Essence of task	8/1	8/1
2.	Expected monologue length	9/0	8/1
3.	Option to take notes	8/1	8/1
4.	Comprehension	4/4	4/4
5.	Providing pen and paper	7/2	8/1
6.	Requesting topic selection	9/0	9/0
7.	Asking for the topic number	9/0	8/1
8.	Signalling start of preparation time	7/2	9/0
9.	Signalling end of preparation time	8/1	9/0
10.	Allowing time for monologue	9/0	9/0
11.	All follow-up questions	8/1	9/0
12.	Signalling end of task 1	9/0	9/0

Previous research on the topic (Alas 2010) prompted the current research to consider also maintaining the order of the script elements; adhering to the wording of the lead-in information; supplementary information the interviewers include in the instructions to the candidate during the lead-in to task 1.

As can be seen, the elements feature in the interviews to varying degrees. Only three are present in all: topic selection, allowing time for delivering the monologue, signalling the end of the task.

### 3.2.1. Essence of the task

In general, the interviewers seem to recognise the need to explain the task. Sixteen interviewers explain it, although M6's lead-in is limited to: 'You have preparation time for the monologue', and the interviewer F9 is almost as laconic, adding only the length of the preparation time '[---] you have three minutes to prepare for the monologue [---]'. In those interviews, however, the candidate is not swayed by the minimum information provided and proceeds to prepare appropriately, which testifies to adequate examination preparation and familiarity with the procedural demands. It could prove to be a problem with candidates who are less prepared, i.e., do not know what the essence of the task is or how long they should be prepared to talk. If deprived of that information, they would be disadvantaged compared to others who get it. Hence, the validity is at stake. There are no discernable gender-related behavioural differences here, however.

### 3.2.2. The expected length of the monologue

Before the candidates start thinking about what to say, they must be reminded of the expected speaking time. All but one interviewer state the expected length: F9 gives the information only after the preparation time has ended. It is interesting

that she deems it more appropriate to shift it closer to the actual presentation of the monologue. However, as it is more important to know the speaking time when starting to plan the monologue, shifting the information may affect validity.

### 3.2.3. Length of the uninterrupted preparation time

For task 1, the candidate should have 3 minutes of uninterrupted preparation time. The candidates have the right to forgo this if they wish. Figure 1 below illustrates the amount of time men and women allowed for monologue preparation, the y-axis representing the time (minutes/seconds) and the x-axis the participants.

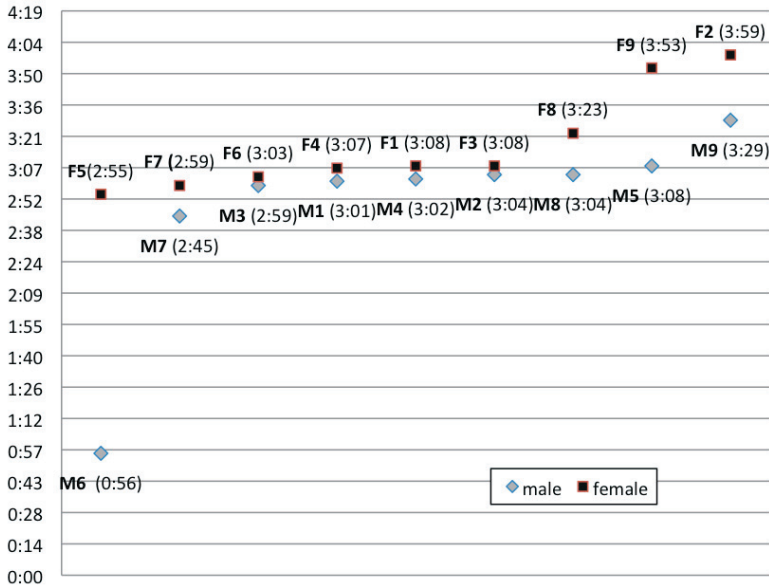


Figure 1. Monologue preparation time with male/female interviewers

In M6 and M7, the candidate wished to begin the monologue sooner and this was granted. Both of these cases were excluded from the calculation of the mean time in the male group. The mean is 3 minutes 6 seconds; the longest time allowed is 3 minutes 29 seconds, and the shortest 2 minutes 59 seconds. With the exception of M5 and M9, most male interviewers come close to the time required (+/- 4 seconds), varying marginally. A notable exception is M9 who allows nearly half a minute more than others, giving his student a clear advantage.

A cursory look at Figure 1 above indicates that women allow their candidates more preparation time. Also, considerably more fluctuation in the amount of time allowed is observed. The mean time is 3 minutes 17 seconds, i.e. most female interviewers (F1, F2, F3, F4, F6, F8, F9) allow the candidates more than 3 minutes, F5 and F7 fall marginally under the specified time. Compared with male interviewers, a difference is evident in the maximum length of preparation time: 3 minutes 59 seconds (F2). This is nearly a minute longer than envisaged. Almost the same can be observed with F9 (53 extra seconds). Both of these candidates have nearly 25%

more time to prepare for the task than their peers. There is a clear gender-related difference here: women are more lenient than men when keeping time. There is, however, great variation also within the groups.

### **3.2.4. Note-taking**

During preparation for task 1, students should have pen and paper at hand for taking notes. It is the interviewer's responsibility to make them aware of that.

With the exception of one female (F9) and one male interviewer (M7), all the interviewers conform to this requirement. The script is followed verbatim by four men (M1, M2, M3, M8) and four women (F1, F3, F5, F6). In M4, M5, F4, and F8, 'and' is inserted at the beginning. Other script deviations include paraphrasing, for instance, F7 tells the candidate: 'And yeah, feel free to take notes [---]', and M7 only implies that the candidate can take notes without telling the candidate clearly: 'Here's a pen ... here's some paper if you need it'. All in all, there is a marginal disadvantage to the candidates with male interviewers.

### **3.2.5. Enquiring about comprehension**

Asking if the candidate has understood the instructions should give the interviewer feedback about whether the candidate is ready to start monologue preparation. Compared with other elements this phase seems to be considered less important and is omitted by five male and five female interviewers: M1, M2, M3, M8, F1, F2, F5, F6 ask the question. It may be worthwhile to note that whenever the question is asked, the response is always positive. As the instructions are given in simple language and there might be accompanying body-language (nodding) on the part of the candidate, the interviewers might consider the query irrelevant.

### **3.2.6. Providing pen and paper**

Offering pen and paper for notes was less widespread among men than women. M1 uses the script verbatim. M2 seems unsure of what he needs to say: 'Alright, here's ... aaa ... you have a pencil and have some paper.' In M3, the word 'pencil' is replaced with 'pen'. M5 paraphrases the instruction: 'And you can use this paper as rough notes paper.' M6 tells the candidate he can take notes, but pen and paper is not mentioned. M7 says, 'Here's a pen ... here's some paper if you need it.' In M8, the interviewer's comment on pen and paper seems to indicate that the candidate is expected to move somewhere for the preparation stage: 'So, here's a pen and you'll find some paper there.' M9 mentions the option to take notes but does not state that there is pen and paper available.

Eight women provide pen and paper; however, only F3 and F6 use the script verbatim. F1, F2 and F5 add a connector. F2 and F9 make an insignificant word order change. In F4, the candidate is made aware of the possibility of taking notes but providing pen and paper is implied not stated: 'And you can make some notes if you wish'. F7 tells the candidate, 'There's pen and paper *over there*', which also

indicates that the candidate is expected to move to another seat. F8 seems to remember the need for paper and pen somewhat later, as she remarks: 'Oh yes, and a pen and some paper'.

Overall, the option of using pencil and paper is made known to the candidate, but the interviewers do not give the instruction in a consistent way. The gender distinction is slightly to the disadvantage of men, as they refrain from giving the option more often than women.

### 3.2.7. Requesting selection of a topic

Topic selection is one of the three elements present in all interviews, but the wording varies widely. Only two men (M1, M3) use the script verbatim, all others change the tone of the request. In the M2 interview, the candidate is told: 'And you can take one of the topics over there'. M4 softens the request by saying: '*And would you* please pick a topic'. M5 reduces the level of politeness by using the imperative: 'Choose a topic'. M6 utters, 'And please start with the topic', which is ambiguous and can be interpreted both as a request to choose as well as to start speaking on it. M7 seems unsure of what he needs to say, his remark, 'Alright, now ... I'd ... like ... you ... to pick ... a topic' may be a sign that he is not familiar with the script. Both M8 ('But before this would you please choose one of the cards in front of you'), and M9 ('I would like you to take a ticket to choose a topic') use an alternative wording. Here, synonyms have been introduced for the 'topic' – card and ticket – which is unnecessary and linguistically inappropriate. Also, the alternative wording suggested is unnecessarily long.

Women adhere more to the prescribed script – F1, F4, F5, F6, and F9 use the script verbatim. F2 asks the candidate to 'Please, pick *up* a topic', indicating a slight struggle with the language. F7 extends the script: 'Please, pick a topic *for yourself*'. Both F3 and F8 unnecessarily use the synonym 'card' for 'topic'.

### 3.2.8. Asking for the topic number

As it is important for the interviewer and the examiner to know which monologue topic the candidate will be speaking on, which set of follow-up questions to ask, and which role-play should ensue, asking for the topic number is critical.

All but one interviewer (F4) ask for the number although there are differences in wording. Slight changes involve either leaving out the word 'topic' (M2, M9) or replacing it by a pronoun (e.g. M4 'What's the number of *it*?'). M8 asks the candidate, 'What is the number of the card?', maintaining the synonym he had introduced earlier, and M3, by using 'Tell us the number of your topic', increases the level of directness of the interchange. The most notable change is present M6, who orders the candidate to 'Tell us the topic', completely changing the nature of the request: instead of the number of the topic the candidate is ordered to give the name of the topic, which is confusing as there are usually at least two sub-topics identified by the number under one general topic.

Three women (F1, F5, F6) quote the script, in others the wording is changed. F2 asks the candidate to 'Tell me the number'; F8 'Could you tell us the number?'



Similarly, F9 asks ‘What is the number?’; in F7, ‘Could you tell us the number as well?’; and in F3, the interviewer says ‘What is the number of the card?’. As can be seen, most interviewers include the element in the script, but communicate the message of the request rather than deliver it precisely. The number of women who did that exceeded that of men.

### 3.2.9. Signalling beginning of preparation time

The script phrase ‘*Now you will have 3 minutes.*’ serves both as a signal to the candidate to start topic preparation and the point for the interviewer for time-keeping. Three men quote the script exactly (M1, M2, M3). M4 and M5 add connectors to enhance cohesion; the latter also mentions the purpose of the given time similarly to M9 (*‘to make notes and to prepare for the task’*), M8 adds a comment about who ends the preparation time (*‘Okay, so I will tell you then the three minutes are up’*). M6 and M7 omit the phase altogether, which makes keeping time problematic.

Three women quote the script verbatim in (F1, F3, F5). F2 changes the tense, F4 alters the tone and the word order (*‘So, remember you have now 3 minutes’*). F6 uses a different connector (*‘So, you will have 3 minutes’*). F7, F8 and F9 clarify the purpose of the time given (*‘So, now you will have 3 minutes to prepare’*).

Although the amount of participants who follow the script verbatim in both groups is the same, and the changes introduced by the interviewers have very similar patterns, the data suggest that women try to follow the script more conscientiously, since none of them omit the phase as is the case with some male interviewers.

### 3.2.10. Transition from preparation to monologue delivery

The transition from the preparation phase to monologue delivery can be initiated by either the candidate or the interviewer. When the three minutes are up, the interviewer must instruct the candidate to start speaking, but if the candidates wish to forgo using all or some of the preparation time, they can. Regardless who stops the preparation, the interviewer should remind the candidate of the expected length of the monologue and signal the beginning of the speaking.

In the oral proficiency interviews analysed, two candidates ask to begin before the end of the allotted time (M6, M7). The two interviewers respond by diametrically different strategies: M7 follows the prescribed script verbatim, whereas M6 skips the reminder and allows the candidate to proceed by saying ‘Good, take a seat’. The comment is significant in that it seems to imply a procedural discrepancy noticed earlier whereby the candidate is required to move about in the examination room preparing at one desk and responding at another. This may be an attempt to provide space for the candidate during the preparation time, but with so little time for the interview, moving between tables seems to cause unnecessary confusion and waste of interview time.

The transition script is followed verbatim by M3 and M7. M1 adds a more pronounced ending signal, ‘Time is up’ (*‘Alright. Time is up. Remember, you have 2 minutes for speaking. I will tell you when the time is up.’*). M2 and M5 change the connector, starting with ‘right’ / ‘okay’ respectively. Information is left out on

three occasions: M4 leaves out the phrase 'Please start speaking now', making the beginning of the candidate's turn ambiguous, and M8 and M9 skip 'alright' – the signal to stop preparation.

Similar patterns occur with women. The script is followed verbatim in F1, F2 and F8. F3 and F9 use a longer phrase to stop preparation ('Alright, *the time is up*', 'Alright, *I have to interrupt you now*'). F3 changes the modality of the wording, implying that the candidate has to speak for two minutes, whereas there is no such obligation. Two minutes is just the time available to the candidate. ('And remember *you have to speak for 2 minutes*. I'll tell you when the time is up.'). Similarly to F5, there is no signal to start the monologue. In F4, 'alright' is left out at the beginning, so the interviewer resorts to no verbal signalling of the end of preparation time. In F6, the candidate is addressed by name; otherwise the instruction is intact. F7 shortens the signal to start speaking ('Alright. Remember you have 2 minutes for speaking, yeah. [---] So, *please start*.').

Once again, women adhere to the script more rigorously, but both groups need more awareness with regard to the functions of the different elements of the script.

### 3.2.11. Follow-up questions

After the monologue has been completed, four scripted follow-up questions should ensue. The interviewers should not alter the questions or leave any of them out unless the candidate has already responded to them in the monologue.

Seven men handle the questions appropriately. M1 asks a question although the candidate has already covered it in the monologue, whereas M6 leaves out a question although the student has not addressed it. Asking a question that has already been answered may confuse the candidates as they might think the answer was erroneous either from the point of view of content or language. Leaving questions out may deprive the students of the chance displaying the full spectrum of their language proficiency. It also makes it problematic to elicit a rateable sample from the candidate. Both behaviours undermine test validity.

Women appear more consistent than men, as they do not change the order of the questions or leave any of them out. Some interviewers, however, add information, paraphrase or ask additional questions (F3, F4). This seems to be done in an attempt to assist the candidates when they fail to respond, and can thus be perceived as an accommodating behaviour. F3, for example, offers additional information in the form of a clarifying comment to the candidate. In addition to the above, F4 resorted to evaluative comments (*mhmh* with a rising intonation, *good*), which, in addition to being a characteristic of teacher talk rather than ordinary interaction, may potentially mislead the candidate as to the quality of their performance. Both behaviours are detrimental to OPI validity. The above behaviours are not present in any of the interviews conducted by men.

### 3.2.12. Signalling the end of task 1

Most men (M1, M2, M3, M4, M8, M9) end this stage appropriately. M5 does not thank the candidate but informs her that they are moving to task 2, reducing the politeness level. M6 paraphrases the script, making the instructions more specific (*'Thanks! [---] Let's move to the role (.5) play'*). M7 adds a connector and extends the thank-you (*'Right, thank you very much. Let's move on to the next task.'*).

Female interviewers mostly follow the script; six use it verbatim (F1, F2, F5, F6, F8, F9). F3, F4 and F7 insert additional connectors or deictic expressions for the sake of cohesion (*'Alright, thank you. And let's move on to the next task'*, *'So, thank you. Now, let's now move to the next task'*, *'Thank you. Let's move on to the next task then.'*).

### 3.2.13. Monologue length

The designated monologue time is two minutes. If the monologue is shorter, the interviewer should prompt the candidate to continue to obtain a rateable sample. As can be seen in Figure 2 below there is a notable difference in the monologue length in the two gender groups.

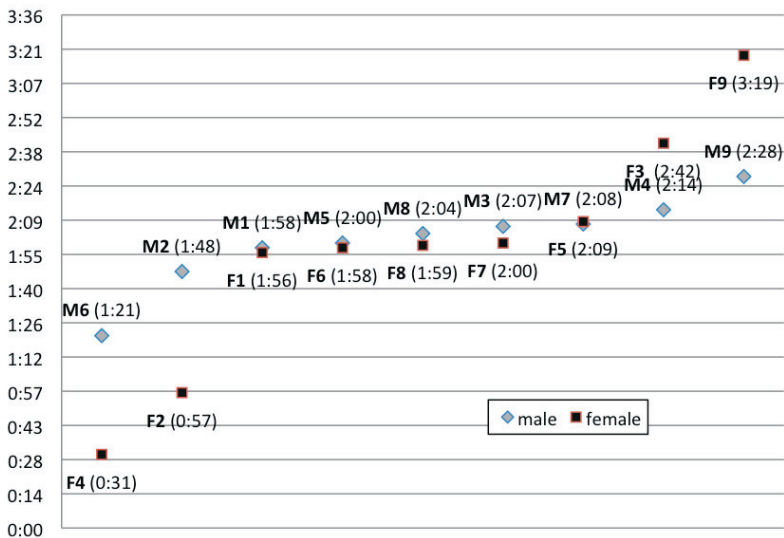


Figure 2. Monologue length with male/female interviewers

The monologue length is under two minutes in three male interviews (M1, M2, M6). The longest time is 2 minutes 28 seconds (M9) and the shortest 1 minute 21 seconds (M6). Two interviewers prompt the student to continue (M2 and M9). In M9, the interviewer does that unnecessarily as a sample of two minutes has already been elicited. M6, however, should have prompted the candidate to proceed, as the monologue here was more than 25% shorter than expected.

The candidate speaks for less than two minutes in F1, F2, F4, F6 and F8. The longest monologue is 3 minutes 19 seconds (F9) and the shortest 31 seconds (F4).

Three interviewers prompt students to continue (F2, F4, F9): in F2 and F4 this is justified, as the length falls considerably short of the expectation, whereas in F9 the monologue had already lasted over three minutes, so prompting to continue indicates difficulty in managing time.

Eliciting a rateable sample is achieved admirably by most interviewers but examples of deviant behaviour even in a small sample like ours seems to imply that there is a need to address the issue in interviewer training. The current dataset seems to indicate, however, that time-keeping practices do not depend on interviewer gender but rather individual interviewer idiosyncrasies.

### 3.3. Task 2. Role-play

The role-play poses the added challenge of performing a dual role for the interviewers: they need to adhere to the script’s procedural requirements as well as assume the role dictated by the role-play cue card and actively participate in communication. Both roles presuppose different behaviour: as the person managing the interview, their linguistic contribution is restricted by the script. As a role-play participant, they are expected to take the cue from the candidate and therefore modify their responses according to the candidate’s input.

The script criteria for the role-play and their representation in the interviews can be seen in Table 3 below.

**Table 3.** Element representation in task 2

No.	Element	Present/missing in interviews	
		Men	Women
1.	Appropriate lead-in to task 2	9/0	9/0
2.	Note preparation time	9/0	9/0
3.	Keeping time	9/0	9/0
4.	Prompt start of the role-play	8/1	8/1
5.	Participating in role-play	9/0	9/0
6.	Stating end of interview	9/0	8/1

#### 3.3.1. Lead-in to task 2

All men include the lead-in in their interviews: M1 and M8 follow the script verbatim. Others introduce minor additions: M2, M3 and M4 combine sentences with a conjunction; M5 makes the instructions more specific (*‘And you have mm one minute to think about it and I’ll tell you when the time is up, when you should start your role-play.’*); M6 additionally mentions the absence of the right to take notes at this stage. M7 increases the politeness level by adding, ‘would you’ at the beginning. M9 gives some information (time) twice and adds a comment about note-taking (*‘Please read it to yourself and you have one minute to think about it, I’ll tell you when the time is up and notes ... note taking is not allowed in this stage. You have one minute.’*).

Lead-in was also present in all female-led interviews. F4 and F6 quote the script exactly. F1, F2 and F5 add conjunctions *and* and *so* to connect different parts. F3 is struggling to locate necessary role-play materials. Once she finds them, she offers more specific information ('I must find a card for you. Is it A5? So, this is your role-play card. Please have a look at it. You have one minute for it. And when the time is up, I'll tell you.'). F7 engages all the parts of the lead-in but uses connectors to link the parts of the section. There are additional comments about note-taking and an apology to finish. ('So, here is your card. Please read it to yourself. Yeah, I'll take that one. And you have one minute to think about it and I will tell you when the time is up. You cannot take any notes. Sorry.'). The instructions are more specific and elaborate in F8, where the interviewer is paraphrasing most instructions ('The next task is role play. And I'll give you a card. You have on minute in order to familiarise yourself with the content of this card and I will tell you when the time is up and you should start.'). The same happens in F9, with an additional remark about the beginning of the role-play ('and then we start the role-play'). Overall, women are more accommodating, adding supplementary information in this stage.

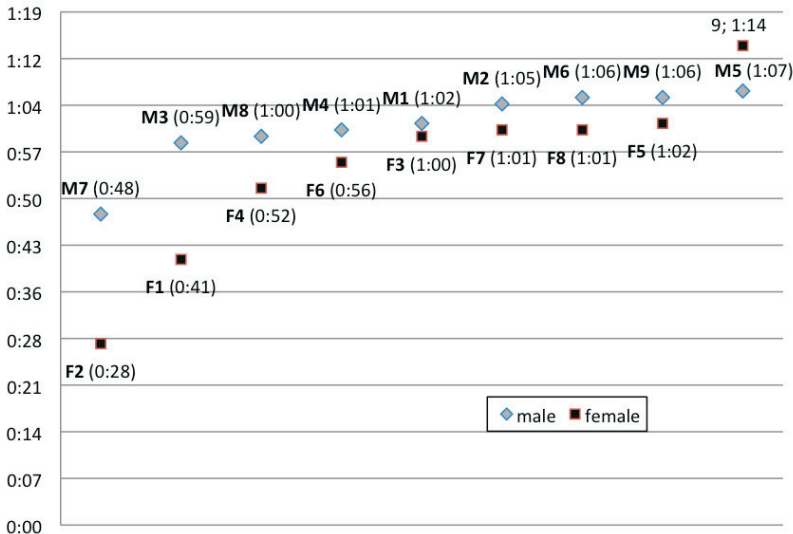
### **3.3.2. Noting the length of the preparation time**

The interviewer must remind the candidate of the one-minute time limit for role-play. Five men use the script verbatim (M1, M2, M4, M5, M8). M3 starts the instruction in a similar manner to that of the introduction to task 1, but then reverts to the appropriate script. M6 adds a comment about taking notes, which is not in script. M7 and M9 insert a connector. In six female-led interviews (F1, F2, F4, F5, F6, F7), the interviewer uses the script verbatim. F7 adds a connector to the beginning of the sentence, similarly with M7 and M9. F3 and F8 resort to paraphrase and by doing so, may mislead the candidate. F3 uses a prepositional phrase 'for it' instead of 'to think about it'. In doing so, she is misleading the candidate to think that the minute is meant for reading the task situation, whereas the minute is intended to think about task completion. F8 paraphrases the given instruction as above and inadvertently ends up giving the candidate false information similar to F3 ('You have one minute *in order to familiarise yourself with the content of this card.*'). Overall, the preparation time is mentioned by all interviewers and although there are more women who follow the script to the letter, the changes to instructions do not lead to ambiguity in interviews led by men, whereas in the female counterparts they do.

### **3.3.3. Length of role-play preparation time**

The time allowed by male and female interviewers for role-play preparation, presented in Figure 3 below, shows different patterns of time management in the respective gender groups.

Overall, men manage to adhere to the one-minute time-slot. The shortest role-play preparation time allowed is 48 seconds (M7), and the longest 1 minute 7 seconds (M5). More training should focus on precise time-keeping, ensuring similar conditions for all candidates.



**Figure 3.** Length of role-play preparation time with male/female interviewers

Women’s results show more variation. In F2 the time is noticeably shorter because the candidate initiates the role-play after 28 seconds of thinking. The most time allowed is 1 minute 14 seconds in F9, and the shortest time is 41 seconds (F1). Generally, women are more precise time-keepers (F3, F5, F6, F7, F8), their timing is closer to the one-minute limit than that of men.

### 3.3.4. Prompting start of role-play

Provided that the candidate has not initiated task 2, the interviewer must prompt the beginning of the role-play once one minute has elapsed. In the analysed interviews, the candidate starts the role-play earlier on two occasions (F2, M7). In male-conducted interviews, the script is followed exactly on five occasions (M2, M4, M6, M7, M8, M9). In interviews M1, M3, and M5 the instruction is modified: M1 adds an initial discourse marker ‘alright’ (*Alright, could you start now?*); M3 increases the politeness level by adding ‘please’ at the end. M5 supplies a connector at the beginning and drops the modifier ‘now’ at the end (*So, could you start the role play?*). Three women follow the script verbatim (F3, F5, F6). F1 shortens instruction (*Could you start now?*), and F7 adds ‘okay’ to the beginning. F4 grants the candidate permission to start speaking (*You may start*) thus changing modality. F8 uses the imperative (*I think you should start now*), and F9 used the phrase *‘Can we start?’* leaving it ambiguous who the initiator of the interaction should be. Thus, in the current script section, women are more direct, and at times less polite than male interviewers.

### 3.3.5. Role-play participation

Role-play is handled with mixed success. On the one hand, the interviewers actively participate in it and answer the candidate's queries according to the script. The problem is the interviewers' unwillingness to adapt their answers to the candidates' questions. Interviewers' cue cards for the role-play contain plausible answers to candidates' questions. As it is difficult to predict what form the candidate questions are going to take, the interviewer cue card answers are usually more detailed than necessary. The interviewers have been instructed to adapt cue card answers to particular questions. Interestingly, it transpires that the interviewers do not know if they can modify the lengthy answers or not. Thus unnecessarily long answers can be found in the recordings to questions requiring a short answer. Another reason for not modifying the response might be the interviewers' mistrust of their own language proficiency. It is safe to read out the long answer: some part of it will definitely answer the candidate's question and it is also linguistically correct.

The entire answer is read out in six role-plays by male interviewers (M1, M2, M3, M4, M8, M9), M5, M6 and M7 adapted the answer to the respective question. Women respond more appropriately. In fact the result is reversed compared to men: F1, F2 and F8 read out the entire answer, all others adapt their answers.

### 3.3.6. Signalling end of interview

The interviewer should say to the candidate clearly that the OPI has been completed, after which the recorder is switched off. Other procedural comments about signing the attendance form or learning about interview results should not be part of the recording.

Six men follow the script exactly (M1, M2, M3, M4, M5, M7). In the three remaining recordings, procedural comments are included too: M6 asks the candidate to sign his name and M8 and M9 inform the candidate about getting the results. Five women quote the script (F1, F2, F3, F4, F5). F6 switches off the recorder before announcing the end. F7 and F9 add a connector at the beginning and F8 supplements the closing remark by an explanation ('This is the end of the interview *and the end of the examination.*').

## 4. Discussion and implications

The hypothesis of the current study was that the interviewer conduct in oral proficiency interviews during the national examination in the English language in Estonia will display gender-related differences, and perhaps reveal identifiable patterns. Content analysis revealed that none of the interviewers completely followed the script. Deviations from the script ranged from leaving out entire sections (i.e. introduction to the interview) to introducing additional discourse markers (e.g. connectors; *and, but, ok*) to achieve cohesion. In some cases, both men and women added supplementary information to be more specific (e.g. substituting the general word 'task' with the specific 'role-play'), and gave the candidate more information than provided by the script (e.g. forbidding note-taking in task 2). We may argue

that while some of these changes are unimportant and should not affect the validity of the OPI (e.g. adding connectors) there are deviations, such as leaving out entire sections of instructions or changing the sequence of information, which alter the amount of support the student is getting during his/her performance and that may certainly influence the outcome.

Both men and women used accommodative strategies to help the candidate, e.g. they slowed down their speech when asked for repetition. However, women also asked additional questions to direct the candidates if they seemed to have difficulties with the task. This is not an acceptable accommodation technique and may affect validity, as some candidates have an advantage with more information/support. Women seemed to have more difficulty in separating their role as a compassionate language teacher from that of a friendly but neutral OPI interviewer.

Time-keeping during the OPI was, in general, a problematic area, as at least half of the interviewers of both genders were imprecise. There were some alarming cases where the interviewer allowed a significant amount of extra time for both task preparation and task delivery. As this occurred with both male and female interviewers, gender cannot be claimed to be a distinguishing feature. Overall, it seemed, however, that women were more lenient than men, and allowed for more speaking time. Here, again, it should be mentioned that this might have contaminated examination validity since all the candidates were not treated equally.

Women seemed to adhere more to the script demands than men, as the necessary interview elements were present more often. But both men and women changed the sequence of information, e.g. not all men stated the candidate's code number at the beginning of the interview and some women asked the candidate to pick a topic for the monologue before explaining the essence of the task. However, if we take into consideration the fact that female interviewers attempted to help the candidate to perform better by providing more direction and men seemed more likely to omit questions/phases, both gender groups are undermining the validity of the OPI in different ways. It is difficult to state which of the interviewers' behaviour is more detrimental. This could be a direction for further research.

With interviewer training, the differences in behaviour with regard to the characteristics pointed out in this research can and should be minimised. To ensure the validity of the OPI, more training seminars should be conducted, as it is evident that not all interviewers seem to acknowledge the necessity of adhering to the script requirements as a strategy for maintaining OPI validity. Recording of OPIs should be obligatory for all the candidates as an additional tool for the standardisation of interviewer behaviour. Recording the OPI would perhaps motivate the interviewers to be more consistent in their behaviour. Given, however, that the current study was conducted relying on recorded interviews, recording alone does not solve the standardisation problem.

## **5. Conclusion**

The current study is the first of its kind to attempt a comparison of the behaviour of male and female interviewers in the context of the oral proficiency interviews of the national examination in the English language in Estonia. The content analysis of the 18 interviews involved confirmed the hypothesis to some extent: there was



an overall attempt to follow the interviewer script in broad terms by both gender groups, but women tended to be more accommodative interview partners and more generous with time. Male interviewers seemed to be more direct in their interaction and also more likely to adhere to the general script demands rather than following it to the letter. Whether the differences in the behaviour can really be attributed to gender, or are the result of insufficient training or awareness of the interviewer role is still inconclusive at this point. More definitive conclusions about gender-related interviewer effects on the OPI would require the analysis of a larger number of interviews. While doing so, research could also make use of more advanced data mining and processing software for more substantiated conclusions.

## 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.
- Alas, Ene; Liiv, Suliko 2009. Constraints of measuring language proficiency in Estonia: The national examination in the English language. – Eesti Rakenduslingvistika Ühingu aastaraamat, 5 / Estonian Papers in Applied Linguistics, 5, 19–32. <http://dx.doi.org/10.5128/ERYa5.02>
- Brown, Annie 2003. Interviewer variation and the co-construction of speaking proficiency. – Language Testing, 20 (1), 1–25. <http://dx.doi.org/10.1191/0265532203lt2420a>
- Lazarton, Anne 1996. Interlocutor support in oral proficiency interviews: The case of CASE. – Language Testing, 13 (2), 151–172. <http://dx.doi.org/10.1177/026553229601300202>
- Lazarton, Anne 2002. A Qualitative Approach to the Validation of Oral Language Tests. Studies in Language Testing 14. Cambridge: Cambridge University Press.
- Luoma, Sari 2004. Assessing Speaking. Cambridge: Cambridge University Press.
- Malvern, David; Richards, Brian 2002. Investigating accommodation in language proficiency interviews using a new measure of lexical diversity. – Language Testing, 19 (1), 85–104. <http://dx.doi.org/10.1191/0265532202lt2210a>
- McNamara, Tim; Lumley, Tom 1997. The effect of interlocutor and assessment mode variables in overseas assessment of speaking skills in occupational settings. – Language Testing, 14 (2), 140–156. <http://dx.doi.org/10.1177/026553229701400202>
- Morton, Janne; Wigglesworth, Gillian; Williams, Donna 1997. Approaches to the evaluation of the interviewer performance in oral interaction tests. – G. Brindley, G. Wigglesworth (Eds.). Access: Issues in English Language Test Design and Delivery. Sidney: National Centre for English Language Teaching and Research, 175–196.
- O’Loughlin, Kieran 2002 The impact of gender in oral proficiency testing. – Language Testing, 19 (2), 169–192. <http://dx.doi.org/10.1191/0265532202lt2260a>
- Pöder, Irma 2011. 2011. aasta inglise keele riigieksami lühianalüüs. Riiklik Eksami- ja Kvalifikatsioonikeskus. [www.ekk.edu.ee/120917](http://www.ekk.edu.ee/120917) (04.03.2012).
- REKK 2008 = Inglise keele riigieksam 2008: Introductory stage (sample). Scripts (sample). Riiklik Eksami- ja Kvalifikatsioonikeskus. <http://www.ekk.edu.ee/valdkonnad/uldharidusvalishindamine/riigieksamite-materjalid-2008/inglise-keel> (04.03.2012).
- Ross, Steven 1992. Accommodative questions in oral proficiency interviews. – Language Testing, 9 (2), 173–185. <http://dx.doi.org/10.1177/026553229200900205>
- Ross, Steven; Berwick, Richard 1992. The discourse of accommodation in oral proficiency interviews. – Studies in Second Language Acquisition, 14 (2), 159–176. <http://dx.doi.org/10.1017/S0272263100010809>

# INTERVJUEERIJAJA KÄITUMISE ISEÄRASUSED KEELEPÄDEVUSTESTI LÄBIVIIMISEL: SOOLINE VARIATIIVSUS

**Edith Reemann, Ene Alas, Suliko Liiv**

Tallinna Ülikool

Artiklis analüüsitakse intervjuuerija käitumise soolisi iseärasusi inglise keele riigieksami suulise keelepädevustesti läbiviimisel. Autoreid huvitab, kas on olemas intervjuuerija käitumismalle, mis on iseloomulikud ühele või teisele sugupoolele. Samuti vaadeldakse artiklis, kas need käitumismallid võiksid takistada intervjuueeritavat oma keelepädevust intervjuu käigus täies mahus demonstreerimast. Uurimuses analüüsitakse üheksat nais- ja üheksat meesintervjuuerija poolt läbiviidud suulise keelepädevustesti intervjuud. Tulemused näitavad küll intervjuureeglite üldist järgimist mõlemas grupis, kuid mõlemas grupis ilmneb nii sarnaseid kui ka sooliselt determineeritud erinevusi.

**Võtmesõnad:** intervjuu, intervjuuerija, intervjueeritav, suuline keelepädevustest, valiidsus, sooline variatiivsus

**Edith Reemann** töötab gümnaasiumi inglise keele õpetajana ning on inglise keele riigieksami suulise keelepädevustesti intervjuuerija.

[edith.reemann@gmail.com](mailto:edith.reemann@gmail.com)

**Ene Alase** (Tallinna Ülikool) teadushuvid on keeletestimine, testide koostamine ja nende kvaliteedi hindamine, õpetajakoolitus, õppekirjanduse hindamine.

[ene.alas@tlu.ee](mailto:ene.alas@tlu.ee)

**Suliko Liivi** (Tallinna Ülikool) uurimisvaldkonnad on kontrastiivuuringud, kultuuridevaheline suhtluspädevus, keelepoliitika, võõrkeelte õpetamise meetodika.

[liiv@tlu.ee](mailto:liiv@tlu.ee)

# SISESTATUD TARINDID LEPINGUTE TEKSTIMOODUSTUSVÕTTENA

Riina Reinsalu

**Ülevaade.** Töövõtu- ja käsunduslepingute keeleanalüüs näitab, et lepingute lausestruktuur on võrdlemisi lihtsakoeline: tekst koosneb peamiselt lihtlausetest ja lihtsa ülesehitusega (enamasti esimese astme kõrvallausestega) liitlausetest. Ent ometigi on lepingute keelekasutus raskepärane. Kuna keerukaid lauseskeeme rakendatakse lepingutes harva, püütakse vajaminevat teavet edasi anda mitmesuguste sisetatud tarinditega. Infiniit- ja partitsiiparindite, nominalisatsioonide, adverbialisatsioonide ja predikaadita tarindite rohkus, kombineerituna kantseliitliku väljendusviisiga, muudab laused kohmakaks ja raskesti mõistetavaks. Teisalt aga täidavad need tarindid tekstimoodustuse seisukohalt lepingutes kindlaid funktsioone, milles peegeldub lepingu kui žanri olemus.

**Võtmesõnad:** žanr, žanrianalüüs, õiguskeel, eesti keel

## 1. Tekstide uurimisest žanrianalüüsi meetodil

Tekstide käsitlemisel kas igapäeva- ja tööelus või keeleteaduslikus uurimuses ei saa läbi liigitamise ja nimetamiseta. Me eristame ilukirjandus- ja tarbetekste ning tarbetekste alaliikidena näiteks reklaam-, teadus-, poliitika-, ajakirjandus- ja ametitekste. Meil on ettekujutus sellest, millised keele- ja struktuurivahendid kuuluvad näiteks ametitekste arsenalis, ning teeme seeläbi vahet protokollil ja aruandel, korraldusel ja otsusel, kiituskirjal ja diplomil, volikirjal ja lepingul. Ent žanrianalüüsiks üksnes sellisest kogemustele tuginevast liigitus- ja nimetusviisist ei piisa: vaja on sisu, struktuuri ja keelevahendite, aga ka situatsiooni- ja kultuuri-konteksti üksikasjalikku analüüsi.

Žanr liigituskategoriana on funktsionaalses keeleteaduses tugevasti juurdunud, kuid žanri olemust mõistetakse erinevalt. Üldjoontes ollakse üksmeelel selles, et žanr on abstraktne kategooria (vt nt Fairclough 2003). Samuti on levinud seisukoht, et žanr põhineb eesmärgipärasel tegevusel, kuigi puudub ühene arusaam sellest,

mida see eesmärk endast õigupoolest kujutab. Seetõttu otsivad žanrianalüüsi teoreetikud muu hulgas vastust küsimusele, kas tekstil on ainult üks eesmärk või sisaldab tekst peale üldtunnustatud eesmärgi ka varjatud eesmäärke, kas žanri uurimisel tuleks lähtekohaks seada seesama eesmärk (needsamad eesmärgid) või hoopis-tükkis teksti vormiline külg ning kuidas suhestuvad žanrid kultuurikontekstiga.

Kui 1990-ndate alguse funktsionaalses žanriteoorias juhinduti peamiselt põhimõttest, et žanri tuvastamiseks on vaja uurida teksti (kommunikatiivset) eesmärki (vt nt Swales 1990, Bhatia 1993), siis hiljem on ühe eesmärgi otsingutest loobutud. Ka John Swales on koos Inger Askehavega oma hilisemates žanrikäsitlustes sedastanud, et kuigi on tekste (nt retseptid), mille eesmärki on võrdlemisi lihtne tuvastada, leidub ka tekste, millel on hulk eesmäärke. Lisaks võivad teksti eesmärgid ajas muutuda, nt kui varem tutvustati ettevõtete brošüürides eeskätt tooteid, siis praegusajal püütakse brošüüride kaudu ka ettevõtet ennast reklaamida, luua temast arvestatava ja usaldusväärse partneri kuvand, tutvustada töötajatele ettevõtte põhiväärtusi ja tugevdada meeskonnavaimu. Niisamuti ei saa lähtekohaks seada üksnes teksti vormilist külge, sest leidub tekste, millel on küll sama vorm, kuid erinev sisu (eesmärk). Lisaks takistaks selline välistel tunnustel põhinev analüüs alażanrite eristamist, sest teatud tasandil ei tuvastata alażanreid mitte vormiliste, vaid sisuliste tunnuste alusel. (Askehave, Swales 2001: 195–204) Liiatigi leidub tekste, mille puhul on nimme tekitatud sisu ja vormi ebakõla, nt võib armastuskiri olla vormistatud lepinguna või ostunimekirjale mnemotehnilisel eesmärgil antud luuletuse vorm.

Ka kultuuriti võivad žanril olla erinevad eesmärgid. Näiteks selgub müügikirjade võrdlusest, et USA müügikirjad on enamjaolt sihitud lühiajaliste suhete loomisele ning toote või teenuse müümiseks kasutatakse sageli survemeetodit, seevastu Hiinas püütakse müügikirjade abil luua pikaajalisi töösuhteid, avaldamata selleks survet. Surve vältimine on märk harmooniast ja austusest. Lisaks hoidutakse hiinakeelsetes müügikirjades osutamast madalatele hindadele, sest madal hind seostub tihti halva kvaliteediga, mistõttu rõhutatakse erihindu. Kultuurierinevused tulevad esile ka kirjade formaalsuses: Hiina müügikirjad on kirjutatud ametlikus registris (keeleline distant väljendab austust lugeja vastu), kuid USA müügikirjadele on iseloomulik familiaarsem väljenduslaad (Zhu 2000: 52–61).

Niisiis võib žanrit pidada keelekasutusviisiks, mille eesmärk (eesmärgid) ja keelevahendid peegeldavad kultuuritavasid. Tekste žanrianalüüsi meetodil uurides tuleb arvesse võtta nii vormi kui ka sisu ning tõlgendada tulemusi konteksti taustal: vaadata, mis tekstis on ja kuidas see “miski” on esitatud ning miks on seda just sellisel viisil tehtud.

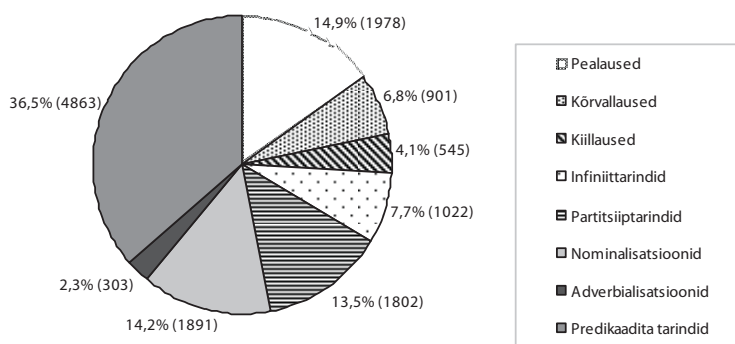
## 2. Lepingukeele lausestruktuuride analüüs

Olen seni käsitlenud lepingute üldstruktuuri: uurinud Otepää vallavalitsusega sõlmitud 39 töövõtu- ja käsunduslepingu<sup>1</sup> näitel, mis osadest lepingud koosnevad ning kuidas korduvad vormielemendid annavad lepingule just sellise kuju, mille järgi me tavatseme üht või teist teksti lepinguks pidada (vt Reinsalu 2011). Üldstruktuuri analüüsi rikastab lepingute lausestruktuuri analüüs, mis põhineb pind- ja

<sup>1</sup> <http://avalik.amphora.ee/otepaa/documents.aspx?org=66&unit=-1&fid=272>; <http://avalik.amphora.ee/otepaa/documents.aspx?org=66&unit=-1&fid=271> (25.01.2009).

süvastruktuuri lausete eristusel. Selle kohaselt koosnevad kõik pindstruktuuri laused (st laused sellisel kujul, nagu need lepingus esinevad) süvastruktuuri tasandil ühest või mitmest elementaarlausetest ehk propositsioonist. Olenevalt sellest, millise kuju on need propositsioonid pindstruktuuris võtnud, võib eristada pea-, kõrv- ja kiillauseid, mis on ka vormi poolest laused, ning infiniit-<sup>2</sup> ja partitsiiparindeid, nominalisatsioone, adverbialisatsioone ja predikaadita tarindeid<sup>3</sup>, mis kuuluvad sisestatud tarindite hulka. (Meetodi kirjeldust vt Reinsalu 2012, Kasik 2002.)

1405 pindstruktuuri lause (13 305 propositsiooni) analüüs näitab, et lepingukeele keerukuse seisukohalt on pea- ja kõrvallausete osatähtsus võrdlemisi tagasihoidlik: tekst on esitatud peamiselt lihtlausete ja elementaarsete liitlausete kujul, mistõttu puuduvad keerukad hierarhiasuhted pea- ja kõrvallausete vahel. Seevastu kiilud, mis lausetesse pikituna annavad mitmesugust taustteavet, kuid on ülejäänud lausega enamasti grammatiliselt seostamata, raskendavad lepinguteksti lugemist ja mõistmist tunduvalt. Lepingukeele keerukus seisneb aga sisestatud tarindite rohkuses ja funktsioonides.



**Joonis 1.** Analüüsimaterjali jagunemine eri lauselisuse tasandiga tarindite vahel

Jooniselt 1 nähtub, et predikaadita tarindid moodustavad selgelt kõige esindatuma lauselisuse tasandi. Pealauseid on küll hulga järgi teisel kohal, kuid neid on predikaadita tarinditest üle poole võrra vähem, lisaks jääb nende hulk samasse suurusjärku partitsiiparindite ja nominalisatsioonide omaga. Ka kõrvallausetel on lepingusüntaksis täita võrdlemisi marginaalne roll, sest nad edestavad oma hulga poolest vaid kiillauseid ja adverbialisatsioone (seega tasandeid, mis on mistahes tekstides niigi väiksema esindatusega) ning neid on ka infiniittarinditest vähem. Selles artiklis keskendun infiniit- ja partitsiiparinditele, nominalisatsioonidele, adverbialisatsioonidele ja predikaadita tarinditele.

### 3. Infiniittarindid

Infiniittarindeid on vaatlusaluse materjali hulgas 7,7% (1022 propositsiooni). Tarindiliikide alusel on jaotumus järgmine: supiinitarindid – 51,2% (524 propositsiooni), infiniitiivitarindid – 34,9% (356 propositsiooni) ja gerundiivitarindid – 13,9% (142 propositsiooni). Kui supiinitarindid hõlmavad nii *ma*-infinitiivi kui ka selle

<sup>2</sup> Infiniittarinditena käsitlen siin kõiki muid infiniitseid tarindeid peale partitsiiparindite.

<sup>3</sup> Predikaadita tarinditena käsitlen siin substantiivi- ja kvantorifraase, mis süvastruktuuris on enamasti *olema*-verbiga predikatiiviga lause ja omajalause kujul.

käändevorme (lepingutes *mast-* ja *mata-*vorm), siis infinitiivitarindite koguhulk väljendab üksnes *da*-infinitiivi hulka, sest *vat*-vormi lepingutes ei kasutata. Gerundiivitarinditest on lepingutes esindatud vaid *des*-gerundiiv.

### 3.1. Supiinitarindid

Infiniittarinditest moodustavad üle poole supiinitarindid, millest 85% esineb *ma*-infinitiivi (444 propositsiooni) ja 15% selle käändevormide kujul (80 propositsiooni). Vaatlusalused *ma*-infinitiivid laiendavad üksnes verbe (tabel 1).

**Tabel 1.** Verbid, mida *ma*-infinitiiv laiendab

Verb	Arv
<i>kohustuma</i>	286
<i>kohustama</i>	90
<i>pidama</i>	62
<i>asuma</i>	5
<i>volitama</i>	1
Kokku	444

Enim tarindeid on moodustatud deontilismodaalsete verbidega *kohustuma* (64,4%) ja *kohustama* (20,3%), mille suur esinemus on otseses seoses lepingu põhifunktsiooniga reguleerida lepingupoolte suhteid. Verb *kohustuma* on suunatud tegijale endale: lepingupool võtab endale kohustuse midagi teha (1). Verb *kohustama* aga eeldab, et on öeldud, kes kohustas keda mida tegema, kuid lepingutes moodustab *kohustama* üksnes passiivkonstruktsioone (2).

- (1) Töövõtja **kohustub** Tellijale **üle andma** KSH programmi eelnõu KSH programmi avalikustamiseks 25. jaanuariks 2009.a.
- (2) Tellija **on kohustatud andma** Töövõtja käsutusse Töö teostamiseks vajalikud vahendid ja informatsiooni.

Kohustuslikkustähendus on ka modaalverbil *pidama* (13,9%). Umbes pooltes *pidama*-tarindites esineb agent individualiseeritult (3), kuid konstruktsiooni *pidama + olema + -tud* sisaldavates lausetes on agent impersonaalne (4). Ka ajakirjandus- ja ilukirjandustekstide analüüsist ilmneb, et sageli välditakse *pidama*-tarindites agendi nimetamist (Penjam 2008: 181). Tihti kasutatakse modaalverbi *pidama* ka seisundilausetes, et märkida esemete vms kohustuslikke omadusi (5).

- (3) Töövõtja **peab** Tööde lõpetamisel **hankima** töö üleandmiseks vajalikud load ja kooskõlastused ning **koostama** vajaliku täitedokumentatsiooni.
- (4) Kõigi ülalmainitud infovahetuste puhul **peavad** lähetised **olema saadetud** allpool nimetatud või hiljem määratud teistele täievolilistele esindajatele.
- (5) Pooltevahelised käesoleva Lepinguga seotud teated **peavad olema** kirjaliikus vormis, välja arvatud juhtudel, kui sellised teated on informatsioonilise iseloomuga, mille edastamisel teisele poolele ei ole õiguslikke tagajärgi.

Deontilist modaalsust annab edasi ka verbi *volitama* ja supiini tarind (6), kuid volitamist kasutatakse lubatuse väljendamise vahendina vaid ühes lauses. Nii-samuti leidub vaid üksikuid näiteid alustamist väljendava verbi *asuma* ja supiini kooskasutuse kohta (7).

- (6) Otepää Vallavalitsus **volitab** advokaat Kalju Kutsarit **esindama** Otepää Vallavalitsust Tartu Halduskohtus haldusasjas nr 3-08-9 kohtulikul arutamisel kõigi seaduses sätestatud õigustega, välja arvatud õigus kaebus õigeks võtta ja/või sõlmida ilma täiendavate volitusteta kaebuse esitajaga kokkuleppeid asja kohtuväliseks arutamiseks.
- (7) Hankija võib teha kirjaliku hoiatuse järgmistel juhtudel:  
- **pole asutud** 48 tunni jooksul alates rikke avastamisest või sellekohase teate saamisest lampide läbipõlemise põhjusi **välja selgitama**, vigu **likvideerima** või lampe **vahetama**

Kohustuste seesugune esitusviis on kooskõlas nii registri (ametlik), valdkonna (lepinguõigus) kui ka kanaliga (kirjalik). Kuna puudub otsene adressaat, ei saa lepingutes rakendada käskivas kõneviisis väljendatud direktiivsust. Samuti pole õigustekstides tavaks kasutada kaudseid kohustamisstrateegiaid (uurida lepingu-poolle võimalust või soovi midagi teha).

Supiini käändevormidest on vaatlusaluses materjalis kasutatud peamiselt *mata*-vormi, mis esineb enamasti atribuudi (8) või sõltuvusmääruse (9), harvem seisundimääruse (10) või predikatiivi kujul (11). Verbidega *vaatama* ja *hoolima* moodustatud *mata*-tarindid (12) pakuvad aga näiteid sõnavormide kaassõnastumise kohta ning väljendavad määrdumustena mingeid tingimusi või piiranguid (vt Uuspõld 2001). *mast*-tarindit esineb tekstis vaid kahel korral, mõlemal juhul koos negatiivimplikatiivse verbiga *keelduma* (13). Supiini käändevormide vähesust ja liigivaesust võib pidada lepingute erijooneks.

- (8) Viivis **maksmata** summalt on 0,5 % päevas.
- (9) Juhul kui TELLIJAJ jätab projektiplaani ja pakkumised õigeks ajaks **esitamata**, lükkub TÄITJAJ esitatava töö tähtaeg viivitatud päevade võrra edasi.
- (10) Juhul, kui Tellija ei ole teostatud tööde kvaliteediga rahul või leiab, et töö jää **tegemata**, informeerib ta koheselt sellest Töövõtjalt (tel nr xxxx xxxx).
- (11) Tellijal on õigus nõuda Töövõtjalt viivist null koma viis (0,5) % tähtajaks esitamata Projekteerimistöode maksumusest iga kalendripäeva eest, kui Projekteerimistööd või projektdokumentatsioon on ettenähtud tähtajaks Tellijale kohaselt **üle andmata** või **koostamata**.
- (12) Tellija ei täida kirjalikust meeldetuletusest **hoolimata** käesoleva lepingu sätestatud maksekohustusi.
- (13) Kui Tellija keeldub tööde üleandmis-vastuvõtuaktile **alla kirjutamast**, kohustub ta selle kirjalikult põhjendama hiljemalt viie kalendripäeva jooksul, vastasel korral loetakse akt allakirjutatuks;

### 3.2. Infinitiivitarindid

Teise suure infiniititarindite rühma moodustavad 34,9%-ga infinitiivitarindid (356 propositsiooni), mis paistavad silma oma süntaktilise kasutusvaldkonna üheilbalisusega: 94% analüüsitud *da*-infinitiividest (333 propositsiooni) talitleb laiendina ja 4,5% (17 propositsiooni) otstarvet väljendava kõrvallause predikaatverbina, ülejäänud 1,5% (6 propositsiooni) täidab mingit muud funktsiooni (perifrastilise verbi koosseisus, käsklause predikaadina).

Kui *ma*-infinitiivid moodustavad tarindeid vaid verbidega, siis *da*-infinitiivid laiendavad nii substantiive (57%) kui ka verbe (37%), kuid ka siin avaldub sama seaduspära mis *ma*-infinitiivi puhul: välja on kujunenud kindlad sõnad, mille laiendina infinitiiv talitleb.

**Tabel 2.** Substantiivid, mida *da*-infinitiiv laiendab

Substantiiv	Arv
<i>õigus</i>	190
<i>kohustus</i>	6
<i>tahe</i>	3
<i>alus</i>	2
<i>volitus</i>	1
<i>valmisolek</i>	1
Kokku	203

**Tabel 3.** Verbid, mida *da*-infinitiiv laiendab

Verb	Arv
<i>võima</i>	77
<i>tulema</i>	15
<i>õnnestuma</i>	10
<i>saama</i>	6
<i>suutma</i>	5
<i>tohtima</i>	5
<i>võimaldama</i>	4
<i>soovima</i>	2
<i>vajalik olema</i>	2
<i>võimalik olema</i>	2
<i>aitama</i>	1
<i>püüdma</i>	1
Kokku	130

Tabelist 2 nähtub, et substantiivide hulgas domineerib 94%-ga selgelt sõna *õigus*, mis väljendab lubatust sooritada infinitiiviga tähistatud tegevust, ülejäänud 6% moodustavad sõnad *kohustus*, *tahe*, *alus*, *volitus* ja *valmisolek*. Tarindi koosseisus esinevate verbide rida on märksa pikem, kuid ka siin on kujunenud välja selge vastandus: modaalverb *võima* (59%) vs. muud verbid (41%) (tabel 3).



Suurima rühma moodustavad *da*-infinitiivi tarindites sõnad, mis esindavad deontilist modaalsust: lubatust väljendab sõna *õigus*, kuid ka *võima*, *tohtima* ja *volitus* ning kohuslikkust sõnad *tulema* ja *kohustus*.

Enamasti fikseeritakse substantiivi *õigus* ja *da*-infinitiivi tarindiga lepingu- poolte eriõigused (14), harvem väljendatakse sellega mõlema lepingupoole ühis- õigusi (15). Eitust kasutatakse substantiivi *õigus* ja *da*-infinitiivi tarindis harva (16), mis osutab sellele, et keelatus tähistatakse lepingutes muude vahenditega.

- (14) Töövõtjal on **õigus alustada** tootmist kohe peale käesoleva lepingu allakirjutamist.
- (15) Pooltel on **õigus** mõlemapoolsel kokkuleppel Leping **lõpetada** mistahes staadiumis.
- (16) Töövõtjal ja Tellijal **ei ole õigust anda** kolmandatele isikutele informat- siooni Lepingu sõlmimise käigus teatavaks saanud asjaolude kohta, välja arvatud seaduses sätestatud juhtudel.

Loatähendus on ka 60%-l *võima*-tarinditest, mis esinevad peamiselt agendiga tegevusprotsesse väljendavates pealausetes (17), harvem impersonaalse agendiga lausetes, kus agendi saab konteksti või lause põhjal hõlpsalt tuvastada. Näiteks pole lauses 18 öeldud, kes teadet edastab, kuid varasemate lepingupunktide ja üldise konteksti põhjal võib aimata, et teate edastaja on emb-kumb lepingupool. Deontilist tähendust kannab *võima* ka oleviku eitavas kõnes (19).

- (17) Tellija **võib nõuda** ebakvaliteetselt tehtud töö tasuta ümber tegemist.
- (18) Informatsioonilist teadet **võib edastada** ka suuliselt ja telefoni kaudu.
- (19) Lepingu tõlgendamise aluseks **ei või olla** ebaõige tähistus ega väljendus- viis, mida Pooled kasutasid eksimuse tõttu.

Peale *võima* moodustab *da*-infinitiiviga ahelverbe modaalverb *tohtima*, mis on lepingutes üksnes eitavas kõnes (20). Ka *da*-infinitiivi tarindite analüüs ajakir- jandus- ja ilukirjanduskeele põhjal kinnitab modaalverbi *tohtima* eituslembust võrreldes teiste modaalkonstruksioonidega (Penjam 2008: 135).

- (20) Täitja **ei tohi** käesolevast Lepingust tulenevaid õigusi ja kohustusi **üle anda** kolmandatele isikutele ilma Tellija kirjaliku nõusolekuta.

Tegevuse sooritamiseks antakse agendile luba ka substantiiviga *volitus*, kuid seda vaid ühel korral (21). Kuna *volitama*-tarindeid on samuti vaid üks, viitab see sellele, et volitama tegevusena pole lepingutele iseloomulik (õiguste andmist ei nimetata volitamiseks).

- (21) Lepingujärgne Töövõtja esindaja on P. K., Corrente Grupp OÜ projektijuht (tel nr xxxxxxx), kellel on **volitus kirjutada alla** Tehtud tööde aktile

Tegevuse kohustuslikkus on *da*-infinitiivi ühendites vormistatud peamiselt verbi *tulema* abil. Kõikidel juhtudel on agent jäetud nimetamata: see, kellele kohustus on suunatud, selgub eelnevatest ja järgnevatest lepingupunktidest. Verb *tulema* on *da*-infinitiivi tarindites alati jaatavas kõnes (22). Selline kasutus pole lepingu- spetsiifiline, sest samalaadseid tulemusi on andnud ka muude tekstide analüüs (vt nt Mandra 2007: 28, 39).

(22) Töö peatumise nõue **tuleb** Töövõtjale **teatada** kirjalikult.

*da*-infinitiivi tarindeid substantiiviga *kohustus* esineb vaatlusaluses materjalis kuus korda (23). Sellist väikest esinemust võib põhjendada sellega, et kohustuslikku tegevust ei märgita mitte substantiivi ja *da*-infinitiivi ühendi, vaid verbide *kohustama* ja *kohustuma* ning *ma*-infinitiivi abil või on kasutatud selleks mõnda muud kohustavat verbi (nt *pidama*).

(23) Lepingu ennetähtaegsest lõpetamisest on lepingupooltel **kohustus teatada** teisele osapoolle vähemalt kolm kuud ette.

Deontilise modaalsuse väljendusvahendite hulka kuuluvad ka mõned *da*-infinitiivi ja verbiühendi *vajalik olema* tarinid (24).

(24) Nimetatud tingimus ei laiene informatsioonile, mis on avalikkusele kättesaadav ning mida **on vajalik avaldada** Töö teostamiseks.

Teine osa infinitiivitarinditest väljendab dünaamilist võimalikkust ehk agendi võimalisust sooritada tegevus. Võimalikkustähendus on *da*-infinitiivi ühenditel verbidega *õnnestuma* ja *suutma*, modaalverbidega *saama* ja *võima* ning ühendiga *võimalik olema*.

Verbi *õnnestuma* ja *da*-infinitiivi tarinid esinevad üksnes (*juhul*) kui tingimused, seejuures on verb eitavas kõnes, väljendades tegevusi, mida pole võimalik teha (25). Viide agendile puudub, kuid võttes arvesse, et kõik *da*-infinitiiviga edasi antavad tegevused (vaidluste lahendamine, kokkuleppe sõlmimine, ühise tahte kindlakstegemine) nõuavad kaht osalist, siis on impersonaalseks agendiks nii teenuse tellija kui ka täitja. Sama moodi on üksnes eitavas kõnes ka verb *suutma*, mis kätkeb endas võimetust sooritada *da*-infinitiiviga märgitud tegevust (26). Verbi *suutma* sisaldavates tarindites on agent isikustatud, olles väljendatud kas kokkulepitud nimetuse või isikulise asesõnaga.

(25) Juhul, kui lahkkelisid **ei õnnestu lahendada** eelnimetatud korras, lahendatakse vaidlus kohtus.

(26) Garantiiperioodi jooksul kohustub töövõtja kõrvaldama omal kulul kõik defektid ja tegematajätmised, kui ta **ei suuda tõestada**, et seda on kohustatud tegema keegi kolmas.

Kuigi modaalverbi *saama* abil on võimalik väljendada nii võimalikkust, tõenäosust kui ka kohustuslikkust, on *saama* lepingutes üksnes võimalikkuse väljendamise teenistuses ja sedagi harva (27). Niisamuti leidub vaid paar näidet ühendit *võimalik olema* sisaldavate tarindite kohta (28). *da*-infinitiiviga väljendatud tegevuse võimalikkust annab edasi ka modaalverb *võima*, mis on kasutusel peamiselt kõrvalausetes, kus see on tavaliselt isikulises tegumoes ja jaatavas kõnes (29).

(27) Kui Töövõtja teade Tellijale on otstarbekohasel viisil edastatud (fax, e-mail, kiri), **saab** Töövõtja **tugineda** teatele, hoolimata selle hilinemisest, moondu misest või mitte kohale jõudmisest.

(28) Juhul kui üleskerkinud küsimust või probleemi **ei ole** põhjendatult või Tellijast olenematutel asjaoludel **võimalik** ülalmärgitud tähtaja jooksul **lahendada**, teatab Tellija sellest koheselt Käsundisaajale.

- (29) Lepingupool kohustub teist Poolt informeerima kõikidest olulistest asjaoludest, mis **võivad mõjutada** või **takistada** käesolevas Lepingus sätestatud kohustuste täitmist või õiguste realiseerimist.

Võimalikkustähendusele on lähedane võimaldamistähendus, mida väljendavad *da*-infinitiivide tarindid verbidega *võimaldama* (30) ja *aitama* (31). Võimalust väljendatakse üksikjuhtudel ka substantiividega *alus* ja *valmisolek*: esimene neist viitab välisele ressursile midagi teha (32), teine sisemisele (33).

- (30) (Tellija) tagab et Töövõtjale Ehitus- ja hooldustööde tegemiseks üleantavad pinnad on seisundis, milline **võimaldab** Töövõtjal **alustada** Ehitus- ja hooldustööde tegemist või nende korraldamist lepingus sätestatud tingimustel ja korras
- (31) Esitama Töövõtjale lisaks käesoleva lepingu dokumentides nimetatud lähteandmetele viivimatult projekteerimistööde teostamiseks vajalikud täiendavad lähteandmed mis **aitavad kiirendada** käesoleva lepingu täitmist.
- (32) Kui Tellijal on Töövõtja ees võlgnevusi käesoleva Lepingu alusel või muude Green Building Development OÜ ja Otepää Vallavalitsuse vahel sõlmitud lepingute alusel või kui Töövõtjal on pärast käesoleva Lepingu sõlmimist teatavaks saanud sellised asjaolud, mis annavad piisavalt **alust arvata**, et Tellija ei suuda Projekteerimistööde eest tasuta maksejõuetuse tõttu, on Töövõtjal õigus mitte anda valmis Projekteerimistöid üle enne kui Tellija on likvideerinud oma võlgnevused Green Building Development OÜ ees või tasunud Projekteerimistööde eest ette.
- (33) Pooled deklareerivad oma **tahet** ja **valmisolekut suunata** parimad jõupingutused Lepingust tulenevate kohustuste täitmisele, austades ja järgides teise poole seaduslikke ja Lepingust tulenevaid õigusi ja huve.

Agendi sisemine tahe avaldub ka soovimiskonstruktsioonide kaudu. *da*-infinitiiviga edasiantava tegevuse soovimist märgivad *soovima* (34), *püüdma* (35) ja *tahe* (33). Nende kohta leidub vaatlusaluses materjalis vaid üksikuid näiteid, seega väljendatakse lepingupoolte subjektiivseid tahteavaldusi *da*-infinitiivi abil harva.

- (34) Kui Tellija **soovib** Lepingu tähtaega **lühendada**, teeb ta vastava ettepaneku kirjalikult Töövõtjale.
- (35) Lepingust tulenevad vaidlused **püütakse lahendada** Poolte läbirääkimistel.

Peale selle, et *da*-infinitiiv moodustab tarindeid substantiivide ja verbidega, talitleb väike osa (4,5%) *da*-infinitiividest otstarvet väljendava kõrvallause predikaatverbina (36). 1,5% analüüsitud *da*-infinitiividest esineb perifrastilise verbi koosseisus (37), samuti on *da*-infinitiiv kasutusel käsklause predikaadina (38), *da*-infinitiivitarind alusena (39).

- (36) Teenus on käesolevas lepingus kirjeldatud kohustused, mida ehitusjärelvalve litsentsi omav isik täidab Tellija ülesandel, et **tagada** rekonstrueeritava ehitise vastavus ehitusprojekti dokumentidele, kehtivatele nõuetele ning heale ehitustavale.

- (37) Tellija annab objekti töövõtjale tööde teostamiseks üle pooltele teadaolevas seisundis, töö tegemise koht ja tingimused on töövõtjale eelnevalt **teada**.
- (38) Kokku **tasuda** summas 86 140.- (kaheksakümmend kuus tuhat üks sada nelikümmend).
- (39) Lepingu eesmärgiks on **reguleerida** Poolte vahel Lepingu alusel ja sellega seoses tekkivaid õigussuhteid.

### 3.3. Gerundiivitarindid

Peale *ma-* ja *da-*tegevusnime kuulub infiniititarindite hulka *des-*gerundiiv, mis on vaatlusaluses materjalis esindatud 13,9%-ga (142 propositsiooni). 65% gerundiividest (92 propositsiooni) on moodustatud verbidega *arvama*, *lähtuma*, *võrdlema*, *arvesse võtma*, *arvestama* ja *algama* (40) ning neid tuleks käsitleda kaassõnastunud vormidena (vt nt Uuspõld 2001). Ülejäänud 35% *des-*gerundiivi tarinditest (50 propositsiooni) on juhuslikumat laadi, andes edasi põhilause sündmusega samaaegset tegevust (41). Gerundiivide süntaktiline funktsioon on väljendada mingeid tingimusi.

- (40) Eelpool nimetatud asjaoludest peab Töövõtja Tellijat kirjalikult informeerima viie (5) päeva jooksul **alates** asjaolude ilmnemisest.
- (41) Pooled võivad omavahelisel kokkuleppel lepingutingimusi muuta, **vormistades** need vastavalt kehtivale korrale.

Tarindid, mille koosseisu kuuluvad kaassõnastunud gerundiivivormid *alates*, *arvates*, *lähtudes* ja *arvestades*, on suure esinemissagedusega, ülejäänute esinemus on väiksem. Ka selliste grammatiseerunud vormide nagu *arvesse võttes* ja *võrreldes* kohta leidub vaid üksikuid näiteid. Verbi *arvesse võtma* asemel kasutatakse sagedamini samatähenduslikku verbi *arvestama*, kuid sõna *võrdlema* väikest esinemust võib seletada lepingu üldise laadiga: lepingutele pole komparatiivi jm grammatiliste ja leksikaalsete võrdlusvahendite kasutamine omane.

## 4. Partitsiiptarindid

Partitsiiptarindid on vaatlusaluses materjalis esindatud 13,5%-ga (1802 propositsiooni). Nagu infiniititarindid ja nominalisatsioonid, nii ka partitsiiptarindid annavad edasi mitmesugust tegevust, muutes sündmuste edastamise mitmekihilisemaks. Lisaks täidavad partitsiiptarindid lepingutes eriülesandeid: kehtestavad piiranguid (nt *teostatud töö*, *kokkulepitud hind*), esitavad tekstisiseseid viiteid (nt *käesolev leping*, *antud olukord*) ja väljendavad kohustuslikkust (nt *tasumisele kuuluv summa*).

Lepingutes talitlevad partitsiiptarindid peamiselt atribuudina: 96% neist (1730 propositsiooni) väljendab substantiivist põhjaga tähistatud isiku, eseme, nähtuse vms mõnd tunnust. 1,6%-l partitsiiptarinditest (29 propositsiooni) on öeldistäite funktsioon (42). 2,3% (41 propositsiooni) esineb lauses lokatiivse (43) või latiivse seisundimääruse (44) või sõltuvusmääruse kujul (45), seejuures moodustab

*alljärgnev* ligi kolmandiku muus kui atribuudi positsioonis paiknevatest partitsiipidest, kuuludes 27 lepingu avalõiku. 0,1% partitsiipidest (2 propositsiooni) on sihitise koosseisus (46).

- (42) Käesoleva Lepingu täiendused ja muudatused **on kehtivad** ainult juhul, kui nad on koostatud kirjalikus vormis ning allkirjastatud mõlema Osa-poolt.
- (43) Käsundisaja kohustub hoidma tsiviilvastutuskindlustus **kehtivana** kogu käesoleva Lepingu kehtivuse perioodi jooksul.
- (44) Lepingu tingimuste tõlgendamisel eelistatakse tõlgendust, mis muudab Lepingu tingimuse seaduslikuks või **kehtivaks**.
- (45) Otepää Vallavalitsus, (edaspidi nimetatud Tellija), keda esindab ..... ja OÜ Vivocard Plus, (edaspidi nimetatud Töövõtja), keda esindab juhatuse liige A. K., sõlmisid käesoleva lepingu **alljärgnevas**:
- (46) Kui viivitused on tingitud Tellijast mitteolenevatel põhjustel, kohustub Töövõtja tegema kõik endast **sõltuva** selle takistuse kõrvaldamiseks.

Nagu partitsiipitarindite liigilisest jaotumusest ilmneb, on lepingutes kõige arvu-kamalt esindatud atribuutiivsed partitsiipitarindid. Nende hulgas on ülekaalus *tud-* (48%) ja *v-*partitsiibid (42,5%), *tav-* ja *nud-*partitsiipe on vähem (vastavalt 6,3% ja 3,2%). Partitsiibi kujul esitatud verbide loetelu on esinduslik, sageduse alusel võib tuua välja verbid *teostama* (nt *teostatud projekteerimistööd*), *tegema* (nt *tehtud tööd*), *esitama* (nt *esitatud aktid*) ja *nimetama* (nt *eelnimetatud summa*). Kahe-tine roll on verbil *kuuluma*: tavaliselt väljendatakse sellega füüsilist kuuluvussuhet (47), kuid lepingutes on see kasutusel pigem ülekantud tähenduses – midagi tuleb millegagi teha (48). Sellist väljendusviisi peetakse kantseliitlikuks. Niisamuti väl-jendab kaudset kuuluvussuhet verbist *omama* moodustatud *v-*partitsiip lepingutele omases lõpulauses (49).

- (47) Käesoleva Lepingu esemeks on Otepää vallale **kuuluva** Lipuväljak 13 haldushoone, avaliku WC ja külustuskeskuse (edaspidi – Objekt) hool-dustööde teostamine perioodil 01.01.2008-31.12.2008 (edaspidi - Töö) vastavalt käesolevas Lepingus toodud tingimustele.
- (48) Kui käesolev Leping lõpetatakse enne täiskuu möödumist, arvutatakse tasumisele **kuuluv** summa ümber proportsionaalselt tegelikult hooldatud ajale.
- (49) Leping on koostatud eesti keeles kahes võrdset juriidilist jõudu **omavas** eksemplar, millest kummalegi poolele jääb üks eksemplar.

Eraldi partitsiibirühma kuuluvad leksikaliseerunud partitsiibid. Need on oma lähteverbist eemaldunud nii tähenduslikult (nad saavad tähenduse üksnes kindlat tüüpi tarindites) kui ka süntaktiliselt (nad ei seo endaga verbi rektsioonilisi laiendeid) (Kasik 2008: 52). Lepingute tüüpnäiteks on partitsiip *käesolev*, mis esineb tavaliselt koos substantiiviga *leping*, muude substantiivide kohta leidub vaid üksikuid näiteid (*käesolev tehing*, *käesolev punkt*, *käesolev säte*). See, et *käesolev* moodustab umbes 15% kõikidest atribuutiivsetest partitsiipidest, näitab selgelt

lepingu koostaja soovi määrata või täpsustada, millest on juttu, lisaks võimendab seda soovi mõne substantiivi tarbetu esisuurtähestamine (50). Reili Argus osutab, et partitsiip *käesolev*, aga ka näiteks *antud*, *nimetatud*, *alljärgnev* ja *vastav* talitlevad eesti kirjakeeles määrava artiklina. Nende vohamist soosib traditsioon, lisaks kasutatakse lepingute jm dokumentide koostamisel olemasolevaid vorme, mis kinnistavad sõnastusstampe. (Argus 2005)

- (50) Ühe Poole poolt **käesoleva** sätte rikkumisel käsitletakse teise Poole saadetud teadet **käesolevas** Lepingus fikseeritud aadressile või faksile kohasel viisil saadetud teatena.

Partitsiipidega seonduvad ka *poolt*-tarindid. Võrdlus pea- ja kõrvallausete ning nominalisatsioonidega näitab, et just partitsiipitarindid on kõige altimad *poolt*-tarindeid moodustama: 61% *poolt*-tarinditest on seotud partitsiipidega (51), 31% nominalisatsioonidega (52) ning 8% pea- ja kõrvallausetega (53). Agendi paigutamine *poolt*-tarindisse on tingitud vajadusest väljendada subjekti ja objekti korraga.

- (51) Maksmine toimub **Tellija poolt kinnitatud** Tehtud tööde akti alusel koostatud arve järgi 20 tööpäeva jooksul.
- (52) Leping kuulub **täitmisele Poolte õigusjärglaste poolt** samadel tingimustel.
- (53) Käesolevat Lepingut võib muuta või täiendada ainult Poolte kirjalikul kokkuleppel, mis **allkirjastatakse** Poolte või Poolte valitud esindajate **poolt**.

## 5. Nominalisatsioonid

Pindstruktuuris on vormitud nominalisatsioonideks 14,2% süvastruktuuri elementaarlausetest (1891 propositsiooni), mille koguhulk on predikaadita tarindite ja pealausete järel suuruselt kolmas. Umbes 85% analüüsitavaatest nominalisatsioonidest (1607 propositsiooni) on *mine*-tuletiste kujul, teiste liidetega tuletised moodustavad heterogeense rühma (284 propositsiooni). Süntakiline funktsioon on neil aga üks: väljendada nimisõnatarindi abil tegevust (midagi muudetakse, täidetakse, tehakse, kasutatakse, korraldatakse jne).

Nominalisatsioonid on lepingute raskesti mõistetavuse üks põhjuseid. Kuna nende hulk jääb alla vaid predikaadita tarindite ja pealausete omale, kinnitab see nominalisatsioonide osatähtsust muude, vähem esindatud lauselisuse tasanditega võrreldes. Seejuures ei piirduta sageli vaid ühe nominaliseeritud protsessi siirdamisega lausesse, vaid laused võivad sisaldada mitut nominaalfraasi (54). Lisaks on nominaalfraasid suhteliselt pikad (55).

- (54) **Kooskõlastamine** käesoleva Lepingu mõistes tähendab **Tellijalt** kirjalikku **taasesitamist** võimaldavas vormis nõusoleku **saamist**.
- (55) Käsundisaajal on õigus nõuda ehitusettevõtjalt **kasutatud ehitustoodete ja paigaldatud seadmete vastavusdeklaratsioonide ja/ või -sertifikaatide (tunnistuste) ja muude vajalike dokumentide, nagu näiteks garantiikirjad, hooldus- või kasutusjuhendid, originaalide või koopiade esitamist**.

Nominaliseeritud protsessidega seoses pakub ka huvi, kas ja kuidas on agent tekstis väljendatud. Üht võimalust selleks pakub genitiivatribuut (56), kuid enamasti on genitiivatribuudiks siiski lähtelause objekt (57). Sellist tendentsi võib seletada transitiivsete verbide suure esinemusega, sest nende puhul on just objektiline genitiivtäiend keeleomane (Kasik 2006: 30). Näitelause 56 on nominaliseeritud verb alla kirjutama küll sihiline, kuid kuna objekt (leping) on tõstetud pealause subjekti positsioonile, võimaldab see nominaalfraasi koosseisus esitada ka lähtelause subjekti. Kui selline vangerdus ei ole võimalik, ja enamasti see ei ole, seotakse agent nominaalfraasiga sõna *poolt* abil.

(56) Leping jõustub mõlema **osapoole allakirjutamise** päevast.

(57) Leping loetakse lõpetatuks pärast tööde **vastuvõtuakti allkirjastamist** Tellija ja Töövõtja poolt.

Subjektilise genitiivtäiendi ja sõna *poolt* tarindis on nominaliseeritud protsessi agent protsessiga grammatiliselt seotud, kuid agent võib paikneda ka mujal lauses. Seda nimetatakse agendi taustalejätuks (Kasik 2006: 30). Agent võib sel juhul olla põhilause subjekti positsioonil (58), kuid nominaalfraas võib paikneda ka kõrvallauses, nominaliseeritud tegevuse agent aga pealause ja vastupidi (59). Lisaks võib agent olla põhilause ka muus kui subjekti vormis, nt ablatiivis habitiiivadverbiaalina (54).

(58) Lepingujärgsete kohustuste **täitmisel** juhinduvad **Pooled** alljärgnevatest Lepingu dokumentidest.

(59) **Töövõtjal** on õigus katkestada Leping ennetähtaegselt, kui Töö **lõpeta-**  
**mine** on võimatu Tellija süül.

Agent võib olla tekstist ka taandatud. Taandatus korral lause agendi ei sisalda, kuid kuna lepingutes on agendiks tavaliselt emb-kumb lepingupool või mõlemad korraga, saab konteksti põhjal agendi hõlpsalt tuvastada. Näiteks pole lauses (60) öeldud, kes alla kirjutavad, aga keegi teine peale lepingupoolte (või nende esindajate) seda teha ei saa, mistõttu puudub vajadus agendi eraldi nimetada. Niisamuti ilmneb kontekstist, kes peab töid tegema: kuna tellija ülesanne on maksta tasu ja töövõtja ülesanne on teha tööd, on tööde nõuetekohase tegemise kohustus just töövõtjal.

(60) Käesolev Leping jõustub **allakirjutamise** momendist ja kehtib kuni tööde nõuetekohase **teostamiseni**.

## 6. Adverbialisatsioonid

Adverbialisatsioonid on 2,3%-ga (303 propositsiooni) kõige tagasihoidlikumalt esindatud lauselisuse tasand. *lt*-adverbidenäidena väljendavad need tegevuse tingimusi. Umbes 90% neist on tuletatud adjektiividest, mille hulgas domineerivad *ne*- (nt *pahauskselt*, *visuaalselt*, *otsestelt*) ja *lik*-liitega adjektiivid (nt *tegelikult*, *ametlikult*, *tähelepanelikult*), kuid adverbide on tuletatud ka *line*- (nt *süüliselt*, *elektrooniliselt*, *faktiliselt*) ja *tu*-liitelistest adjektiividest (nt *viivitamatult*, *erapooletult*, *tähtajatult*), üksikuid näiteid on *kas*-adjektiivide kohta (nt *hoolikalt*).

Adjektiividest tuletatud adverbide puhul väärrib esiletõstmist kaks asjaolu. Esiteks, ligi neljandiku (27%) nende koguhulgast moodustab adverb *kirjalikult*.

Selle sagedast esinemust võib põhjendada lepingupoolte sooviga tagada, et kõik see, milles lepitakse kokku (nt lepingu muudatused, tingimuste kooskõlastused) ja/või mis võiks lepingu täitmist mõjutada (nt probleemid, pretensioonid), oleks kindluse ja ühese arusaadavuse tagamiseks vormistatud paberil. Teiseks, nagu ametitektide puhul on tavaks kujunenud, leidub ka lepingutes võrdlemisi palju ajamääruslikest sõnaühenditest saadud adjektiivide edasituletisi (nt *igapäevaselt*, *igakordselt*, *õigeaegselt*), samuti selliseid adverbe, mis on tuletatud substantiivi- ja abisõnatüve liitmisel saadud adjektiividest (nt *ennetähtaegselt*). Selline sõnakasutus on kantseliitlik ning soovitatav on eelistada sõnaühendeid (nt *iga päev*, *iga kord*, *õigel ajal*, *enne tähtaja lõppu*).

Teise rühma (10%) moodustavad partitsiipidest tuletatud adverbid. Enamasti on tuletusaluseks *v*-partitsiip (nt *püsivalt*, *piisavalt*, *eelnevalt*), harvem on adverbe tuletatud *tud*-partitsiipidest (nt *põhjendatult*,  *motiveeritult*). Adverbe saab moodustada ka tuletamata adjektiividest, mille näiteks on adverb *selgelt*, kuid selline moodustusviis on lepingutele ebatüüpiline.

Nagu ka eeltoodud näidetest nähtub, on adverbialisatsioonide hulgas ülekaalus ühesõnalised tarindid, laienditega adverbe leidub vaatlusaluses materjalis alla 5% (61). Ühesõnalisus on üks tunnusjooni, mis eristab adverbialisatsioone teistest sisestustest.

- (61) Tellija annab Järelvalvele kõik Ehitusprojekti Dokumente puudutavad muudatused, täiendused ja kokkulepped üle **samaaegselt nende üleandmisega** Ehitajale.

## 7. Predikaadita tarindid

Predikaadita tarindid moodustavad 36,5%-ga (4863 propositsiooni) vaatlusaluste lepingute suurima sisestatud tarindite hulga. Kui eeltoodud lauselisuse tasandite puhul ilmnesid sageli selged liigilised erijooned, nt partitsiipitarindite hulgas domineerivad atributiivsed partitsiibid ja nominalisatsioonide hulgas *mine*-tuletised, siis predikaadita tarindite liigiline pilt on kirjum: leidub genitiiv-, adjektiiv- ja adverbialatribuute ning apositsioone sisaldavaid substantiivfraase ja kvantorifraase.

Lepingutele omased predikaadita tarindid on substantiivifraasid, mis sisaldavad genitiiv- ja adjektiivatribuute. Kuna subjektiline (nt *poolte valmisolek*, *tellija nõudmine*) ja objektiline genitiivatribuut (nt *hooldustöö teostamine*, *akti allakirjutamine*) talitlevad tegevust väljendava põhja laiendina, olen neid käsitletud nominalisatsioonide koosseisus. Predikaadita tarinditeks liigituvates konstruktsioonides esinevad genitiivatribuutidena peaausjalikult adverbialised genitiivatribuudid (nt *töövõtja honorar*, *inseneri töötund*, *töövõtja kogemused*, *valla katlamaja*). Adjektiivatribuutide hulgas moodustavad suurima rühma adjektiivid, mis žanrile kohaselt paiknevad põhja suhtes neutraalses eesendis (nt *vahetu teostaja*, *lepingulised kohustused*, *õigeaegsed pretensioonid*, *materiaalne kahju*), kuid ka pronoomenitest adjektiivatribuudid on arvestataval hulgal ja liigirikkalt esindatud (nt *mõlemad osapooled*, *mis tahes asjaolu*, *selline olukord*, *muu normatiivakt*, *oma otsus*). Muude adjektiivatribuutide, st substantiivi käändevormide (nt *sularahata arveldamine*, *kvalifikatsiooniga töäjõud*), adverbide (nt *valmis töö*, *eraldi arve*) ja järgarvsõnade (nt *esimene osamakse*, *kolmas isik*) kohta leidub näiteid vähem.



Predikaadita tarindite hulka kuuluvad ka järgarvudega ühendid (nt *11. jaanuar, 2008. aasta*).

Peale genitiiv- ja adjektiivatribuutide laiendavad substantiive ka adverbiaalatribuudid, mille esindatus lepingutes on ootuspäraselt väike. Näidetena võib tuua sellised adverbiaaltribuute sisaldavad substantiivifraasid nagu *viivis päevas, muudatused lepingus, mõju lepingu hinnale ja tähtajale*.

Lisandit sisaldavate fraaside tüüpiline kasutusvaldkond on tähistada nimeliselt üksikisikuid, kes tuuakse välja esindajate ja kontaktisikutena. Fraasid on lühikesed, tavaliselt on piiratud üksnes ametinimetustega, nt *vallavanem A. P., juhatuse esimees A. M., advokaat K. K.* Ametinimetused on harilikult esitatud eeslisandina, kuid kontaktisikute jaotises on need paigutatud ka järellisandi kohale – nii on fookus isiku nimel (62). Harvem on nimeliselt väljendatud isiku tähistamiseks kasutatud nimetust *esindaja*, mis viitab lepingust tulenevale isiku funktsioonile, ettevõtlusvormi tähistavat nimetust *füüsilisest isikust ettevõtja* vm. Muid kui isikuid laiendavate lisandite esinemisjuhud on juhuslikumat laadi. Nii näiteks kuuluvad lisandite hulka ka ettevõtlusvormi tähistavad lühendid (*AS, MTÜ, OÜ, FIE*), mis paiknevad nime ees või järel.

- (62) Tellijat esindab Lepingust tulenevates ja sellega seotud õigustes ja kohustustes: K. R., **abivallavanem** tel xxxx xxxx

Kuna lepingutes fikseeritakse ka makstavad summad, makse- jm tähtajad, moodustavad ühe arvestatava predikaadita tarindite rühma kvantorifraasid. Kvantoriks on põhiarvsõnad (nt *kolm kalendripäeva, kaksikümmend kaheksa tuhat kaheksakümmend neli krooni*), seejuures on sageli arvsõnale vastav arv märgitud arvsõna järele sulgudesse ja vastupidi. Kvantiteedisubstantiivide ja -adverbide puudumine ning põhiarvsõnade domineerimine lepingutes osutavad vajadusele väljenduda võimalikult täpselt, ühtlasi annab selline jaotumus edasi ka tekstiliigilist eripära.

Nagu öeldud, puudub predikaadita tarindite seas mingi tarindiliigi märkimisväärt ülekaal, kuid predikaadita tarindite osatähtsus seisneb nende rohkuses. Genitiiv- ja atribuuttarindid kombineerituna partitsiipide ning subjekt- ja objektatribuutidega moodustavad pikki substantiivifraase, mis raskendavad tekstist arusaamist. Keerukust lisab fraaside ühendamine või mitme pika substantiivifraasi kasutamine ühes lauses. Näiteks jääb genitiivatribuutide rohkuse tõttu näitelausest (63) ebaselgeks, mis mida laiendab. Näites (64) paikneb aga substantiivifraasi põhi predikaadist kaugel, mistõttu on mõtet raske jälgida.

- (63) (Töövõtja kohustub:) tagama, et Ehitustööde teostamisega ei rikuta **Ehitise ning seoses Ehitustööde tegemisega Töövõtja valdusse antud, Ehitise asukoha maaüksuse ning seda ümbritseva maa-ala ja teede heakorda ja ohutust ümbruskonnale**
- (64) Töövõtja kohustub tegema **Tatra-Otepää-Sangaste mnt Otepää linna/vallas asuval lõigul Hurda tn teerist - Alajaama elamurajoon kergliiklustee rajamisel välisvalgustuse rekonstrueerimistööd.**

## 8. Kokkuvõte

Lepingute lausestruktuuri analüüsi kaudu paljastuvad lepingu kui tekstiliigi üld- ja erijooned. Analüüs näitab, et töövõtu- ja käsunduslepingute pindstruktuuri laused sisaldavad ohtralt mitmesuguseid sisestatud tarindeid. Puhtkvalitatiivselt ületab nende hulk pea-, kõrval- ja kiillausete oma ligi neli korda, kuid peale hulga tuleks žanritunnustena võtta arvesse ka nende tekstifunktsioone.

Lepingute olemuslik eesmärk on määrata kindlaks poolte õigused ja kohustused. Deontilist modaalsust – kohustatust ja lubatust – annavad kõige ilmekamalt edasi *ma-* ja *da-*infinitiivi tarindid mitmesuguste substantiivide ja verbidega, seejuures on sõnadel *kohustuma* ja *õigus* täita põhiroll. Niisamuti on lepingute lausestruktuuri žanrispetsiifiline omadus kõikvõimalike täpsustavate tingimuste esitamine, tegevuste ja võimaluste kitsendamine. Piiranguid võimaldavad seada, tingimusi täpsustada ja tekstisiseseid viiteid lisada partitsiipitarindid (peamiselt *tud-* ja *v-*partitsiibid), kuid ka adverbialisatsioonid ja predikaadita tarindid on lepingute täpsuse ja konkreetsuse teenistuses. Seevastu nominalisatsioonid täidavad lepingutes muud funktsiooni: nende abil nimetatakse tegevusi terminilaadselt (vahel liialdatult). Nominaalstiili toetab ka lepingule kui žanrile omane staatiline, st kirjeldav ja analüüsiv väljendusviis (nimetamine ja defineerimine, seostele osutamine).

Sisestatud tarindite rohkus, kuid ka nende funktsioonid muudavad lepingukeele tahes-tahtmata keerukaks, mis omakorda võib raskendada lepingu ühemõttelist tõlgendamist. Seetõttu tuleks suuremat tähelepanu pöörata lepingutekstide koostamisele ja ametnike koolitamisele.

### Viidatud kirjandus

- Argus, Reili 2005. Kui sõna ei tähenda, vaid ainult teeb. – Õpetajate Leht, 25. nov, nr 42.
- Askehave, Inger; Swales, John 2001. Genre identification and communicative purpose: A problem and possible solution. – Applied Linguistics, 22 (2), 195–212. <http://dx.doi.org/10.1093/applin/22.2.195>
- Bhatia, Vijai Kumar 1993. Analysing Genre: Language Use in Professional Settings. New York: Longman.
- Fairclough, Norman 2003. Analysing Discourse: Textual Analysis for Social Research. London, New York: Routledge.
- Kasik, Reet 2002. Uudiste süntaks. – Renate Pajusalu, Tiit Hennoste (toim.). Tähen-dusepüüdja. Pühendusteos professor Haldur Õimu 60. sünnipäevaks. Tartu: Tartu Ülikooli Kirjastus.
- Kasik, Reet 2006. Nominaliseeritud protsessi agent meediauudistes. – Emakeele Seltsi aastaraamat, 51 (2005), 21–37.
- Kasik, Reet 2008. Infojaotus uudistekstides: partitsiipitarindid. – Emakeele Seltsi aastaraamat, 53 (2007), 48–63.
- Mandra, Katrin 2007. Deontiline modaalsus eri tekstiliikides. Magistritöö. Tartu: Tartu Ülikool. <http://hdl.handle.net/10062/2866> (03.03.2013).
- Penjam, Pille 2008. Eesti kirjakeele *da-* ja *ma-*infinitiiviga konstruktsioonid. Dissertationes philologiae estonicae Universitatis Tartuensis 23. Tartu: Tartu Ülikooli Kirjastus.
- Reinsalu, Riina 2011. Leping tekstiliigina: žanristruktuur. – Eesti Rakenduslingvistika Ühingu aastaraamat, 7, 215–229. <http://dx.doi.org/10.5128/ERYa7.13>
- Reinsalu, Riina 2012. Lepingute lausestruktuur. – Emakeele Seltsi aastaraamat, 57 (2011), 218–234. <http://dx.doi.org/10.3176/esa57.10>

- Swales, John 1990. *Genre Analysis: English in Academic and Research Settings*. Cambridge: Cambridge University Press.
- Zhu, Yunxia 2000. Building knowledge structures in teaching cross-cultural sales genres. – *Business Communication Quarterly*, 63 (4), 49–68. <http://dx.doi.org/10.1177/108056990006300405>
- Uuspõld, Ellen 2001. *des-* ja *mata-*vormide kaassõnastumine ja eesti komareeglid. – Reet Kasik (Toim.). *Keele kannul*. Pühendusteos Mati Ereli 60. sünnipäevaks 12. märtsil 2001. Tartu Ülikooli eesti keele õppetooli toimetised 17. Tartu: 306–321.

## Võrgumaterjalid

Otepää vallavalitsuse dokumendiregister. <http://avalik.amphora.ee/OTEPAA/documents.aspx?org=66&unit=-1&fid=2> (24.02.2013).

**Riina Reinsalu** põhilised tegevus- ja uurimisvaldkonnad on terminoloogia, õigus- ja halduskeel, tekstianalüüs.

[riinareinsalu@gmail.com](mailto:riinareinsalu@gmail.com)

# INSERTED CONSTRUCTIONS AS THE TEXT FORMATION METHOD OF CONTRACTS

Riina Reinsalu

Concerning contracts as legal texts, besides accurate content, technical accuracy, which appears in the proper structure of the text, good syntax, precise word choice and neutral style, is also expected. However, the reality is somewhat different – lower-level legal and administrative texts that have not been edited tend to be syntactically more complicated than, for instance, legislative texts. Language analysis reveals that although the sentence structure of contracts is rather simple, the comprehension of a text is inhibited by several inserted constructions: non-finite clauses, participle constructions, nominalisations, adverbialisations and constructions without predicates.

Although the abundance of the inserted constructions is one of the textual characteristics of contracts, the functions that these types of constructions fulfil from the point of view of text formation have to be taken into account. The constructions of *ma-* and *da-*infinitive vividly pass on the essence of contracts – deontic modality – obligation and permission is expressed by different nouns and verbs, especially frequently by *kohustuma* and *õigus*. Moreover, genre-specific structural characteristics of contractual sentences set out various specifications of conditions, restriction of actions and options. With participle constructions limits are set, conditions specified and intratextual references added. Also, the adverbialisations and constructions without predicates are at the service of the accuracy and concreteness of the contracts. However, nominalisations perform a different kind of function in contracts – they pass the actions on in the same way as terms. This is supported by the genre-specific static language of a contract – that is to say, descriptive and analytical (naming and defining, referring to connections).

Both the abundance of inserted constructions and their functions make contractual language unavoidably complicated, which in turn may impede the unambiguous interpretation of a contract. Therefore, more attention should be paid to the compilation of contractual texts and the training of officials.

**Keywords:** genre, genre analysis, legal language, Estonian

# TÄIENDUSI HEINRICH GÖSEKENI SAKSA-LADINA-EESTI SÕNASTIKU (1660) JA SELLE UUSVÄLJAANDE (2010) KOOSTAMISE KOHTA

Kristel Ress

**Ülevaade.** Artikli eesmärk on kirjeldada ajaloolise leksikograafia põhimõtete rakendamist Heinrich Gösekeni saksa-ladina-eesti sõnastiku kui 17. sajandi mahukaima eesti keelt sisaldanud leksikoni näitel. Artikli alguses iseloomustatakse nii Gösekeni sõnastiku kui ka ülejäänud kahe samal sajandil ilmunud saksa-eesti sõnastiku sotsiokultuurilist tausta ja koostamise asjaolusid. Artikli põhiosas keskendutakse Gösekeni sõnastiku metodoloogiliselt uudsele analüüsile. Rõhk on 2010. aastal Tartu Ülikooli vana kirjakeele töörühmas valminud kogumiku “Heinrich Gösekeni grammatika ja sõnastik 350” sõnastikuosal, peamiselt on vaatluse all märksõnastamise üksikasjad ja sõnade tähenduse avamise protsess.\*

**Võtmesõnad:** ajalooline leksikograafia, mitmekeelne leksikograafia, leksikoloogia, sõnatähendus, kirjakeele ajalugu, eesti keel

## 1. Sissejuhatus

1660. aastal avaldas Kullamaa pastor Heinrich Göseken 547-leheküljelise teose “Manuductio ad Linguam Oesthonicam, Anführung Zur Öhstnischen Sprache / Bestehend nicht alleine in etlichen præceptis und observationibus, Sondern auch In Verdolmetschung vieler Teutschen Wörter” (Göseken 1660), mis koosnes eesti keele õpetusest ja saksa-ladina-eesti sõnastikust (seda sõnastikku on siinses artiklis nimetatud algsõnastikuks). Selle algsõnastiku esimene teadaolev kordustrukk ilmus faksiimileväljaandena 1977. aastal (Värri Haarmann 1977) ning 30 aastat hiljem lisati faksiimile Eesti Kirjandusmuuseumi arhiivraamatukogu haruldasi trükiseid sisaldavasse digitaalraamatukokku.<sup>1</sup>

\* Uurimisteema kuulub riikliku programmi “Eesti keel ja kultuurimälu” projekti “Eesti vana kirjakeele sõnastikud” (RFLEE09088). Artikli autor tänab Külli Prillopit abi eest sõnaraamatu ainese analüüsimisel.

<sup>1</sup> <http://www2.kirmus.ee/grafu> (03.03.2013).

Siinse artikli eesmärk on tutvustada artikli autori väljatöötatud meetodit Gösekeni sõnastiku ja teiste ajalooliste kakskeelsete leksikonide eesti ainese uurimiseks. Meetodi eripäraks on sõnastikus leiduva eesti keelematerjali tähelepanu keskmesse toomine ja selle analüüsimine nüüdisaja vaatepunktist. Selliselt läbitöötatud Gösekeni saksa-eesti sõnastiku uusväljaande (Kingisepp jt 2010) sõnaloend lähtub eesti sõnade esitähdest (s.t tegu on eesti-saksa-ladina sõnastikuga) ja märksõnadeks on keelendite tänapäevased kirjakujud. Meetodit järgides uuritakse ja kommenteeritakse eelkõige eesti ainek: seletatakse nüüdiskeeles läbipaistmatute märksõnade tähendust ning tänapäeva grammatikakirjelduse alusel lisatakse morfosüntaktilist infot (vt lähemalt artikli 4. osast). Nii on teistelgi huvilistel võimalik arhailisele ainesele hõlpsamalt ligi pääseda ja seda mõista ning kasutada nii sõnavara- ja grammatikauuringutes kui ka näiteks erialakeelte ajaloo käsitlemisel.

Artikkel koosneb seitsmest osast. Sissejuhatus järel vaadeldakse 17. sajandi misjonilingvistika erijooni Eesti alal ja tollaste sõnastike sotsiokultuurilist tausta. Kolmandas osas iseloomustatakse Gösekeni 1660. aasta sõnastikku ja neljas osa on pühendatud uurimismeetodi tutvustamisele. Viimasel osas sõnastatakse märksõnastamise ja tähenduse seletamise põhimõtted ning kirjeldatakse nende rakendamist Gösekeni sõnastiku uusväljaandes. Et see on 17. sajandi mahukaim ja keerukaim ülesehitusega eesti ainek sisaldav leksikon, saab selle varal tuua esile iseloomulikke juhtumeid valitud meetodi rakendamisel ning näidata, millised võivad olla komistuskivid. Artikli lõpetab kokkuvõte ja kirjanduse loend.

Kirjutise aluseks on autori kogemus Tartu Ülikooli (TÜ) vana kirjakeele töörühma kogumiku “Heinrich Gösekeni grammatika ja sõnastik 350” (Kingisepp jt 2010) sõnastikuosa koostamisega (sõnastikuosa nimetatakse artiklis uusväljaandeks). Siinne analüüs on mõeldud täienduseks nimetatud kogumiku saatetekstidele ning autori varasemale artiklile algsõnastikust tulenevate probleemide, tüvede rekonstrueerimise ja märksõnapesade üksiklemmadeks lahutamise kohta (Ress 2009).

## 2. Esimesed eesti keelt sisaldanud sõnastikud ja ajastu sotsiokultuuriline taust

TÜ vana kirjakeele töörühmas on elektroonilisel kujul olemas ja analüüsimisel kõik kolm 17. sajandi eesti keele sõnastikku: Heinrich Stahl<sup>2</sup> (1637, põhjaeesti) ja Johannes Gutslaff<sup>3</sup> (1648, lõunaeesti) saksa-eesti sõnastik ning artiklis vaatluse all olev Heinrich Gösekeni saksa-ladina-eesti sõnastik (1660, põhjaeesti).

### 2.1. Stahl, Gutslaffi ja Gösekeni sõnastiku ühisjooni

Neil kolmel leksikonil on hulk sarnaseid jooni. Esmalt jagavad nad sama aegruumi, sest ilmusid Eestis lühikese, 23-aastase ajavahemiku jooksul. Heli Laanekase (2004: 19) hinnangul on selle – eesti kirjakeele ajaloo teise – perioodi erijooniks tekstikorpuse suurenemine ja keele teadliku normimise algus. Teine periood kestab 17. sajandi algusest kuni 1686. aastani, mil avaldati Uue Testamendi tartukeelne

<sup>2</sup> Tollaste autorite nimekirjutusviis on sama kõikuv nagu muu ortograafia. Allikates esineb Heinrich Stahli eesnimi kujul *Henricus*, perekonnanime variantide hulgas on ülekaalus *Stahell* (Raag 2002: 186). Selles artiklis on kasutatud eesti keeleteaduses levinuimat nimekuju *Heinrich Stahl*.

<sup>3</sup> Eesnime leidub nii trüki- kui ka käsikirjalistes allikates muudelgi kujudel. Artiklis on eelistatud nime *Johannes Gutslaff*, mida on kasutanud Marju Lepajõe, lähtudes Arvo Teringi soovitusel (Lepajõe 1998: 298).

tõlge. Ka Arnold Kase varasema periodiseeringu järgi (1970: 21) on piiritähiseks 1686. aasta.

Ajaloolises kontekstis on nendest kolmest kujunenud esimesed trükis ilmunud eesti keele sõnastikud. Sõnastike koostajad olid Saksamaal hariduse saanud baltisaksa pastorid, kes olid omavahel tuttavad ja tegid koostöödki.

Leksikograafiliselt nimetatakse neid sõnastikke aktiivseteks, sest nad olid mõeldud emakeelest võõrkeelde tõlkimiseks (keelesuund saksa-eesti<sup>4</sup>). Kõik kolm olid koostatud õppematerjaliks Eesti- ja Liivimaa vaimulikele, aga ka ilmalike tegevusalade esindajaile, kes soovisid oma keeleoskust parandada.

Bibliograafiliselt on igäüks neist sõnastikest mahukama teose teine osa, sest esimese osa moodustas eesti keele õpetus. Keelekäsiraamatute või grammatikate nime all neid kolme kogumikku laiemalt tuntaksegi.

## 2.2. Misjonilingvistika 17. sajandi Eestis

Eelnimetatud asjaolusid hõlmavaks tunnusjooneks saab pidada misjonilingvistikat<sup>5</sup> kui Eesti 17. sajandi olusid vorminud kultuurinähtust. Misjonitööd tehti paljude ilma väljakujunenud kirjakeelela rahvaste juures üle maailma. Selle tegevuse eesmärgid ja vahendid olid kõikjal üldjoontes samad: õpiti ära võõras keel ning koostati selle jaoks õpikuid ja sõnastikke, et jõuda peamise eesmärgini tõlkida ära piibel. Piiblitõlke asjaolusid on viimastel aastatel põhjalikult käsitlenud Kai Tafenu (2011b) ja piibli tõlkekeele dünaamikat Annika Kilgi (2012).

Helmut Piirimäe osutab, et riigivõimu ja luterliku kiriku kõrval nõudis eesti keele oskust ka talurahvas ise, kaevates halva keeleoskusega pastorite peale (1992: 412). Eestimaa sinodi otsusega (1627) ei lubatud maakogudusse kirikuõpetajaks inimest, kes ei suuda pidada eestikeelset proovijutlust; keeleoskuse nõude kinnistas Karl XI 1684. aasta korraldus ja nõuet tugevdas tema 1685. aasta korraldus (Piirimäe 1992: 410).

Eesti ala kuulumist Rootsile, mitte katoliiklikule Poolale ega õigeusklikule Venemaale, on peetud võimalikest soodsaimaks ajal, mil oli käes keskaegselt kultuuritüübilt uusaegsele ülemineku põhietapp. Aleksander Loit (2006: 140) nimetab Rootsit “protestantismi kindluseks selles kontkõvas, luterlik-ortodoksses vormis”. Matti Klinge kirjeldab Rootsit kui fundamentalistliku luteri usuga riiki, mille aluseks oli Uppsala 1593. aasta kirikukogu. Klinge hinnangul olid tollase Rootsi riigikorra aluseks sõjaväelaslikkus ja absolutism ning võimu koondamise protsessis mängis suurt rolli just vaimulikkond, kelle ülesandeks tehti muu hulgas rahvahariduse andmine. (Klinge 2008: 90–91)

## 2.3. Stahl, Gutsloff ja Gösekeni sõnastiku erijooni

17. sajandi kolmel eesti keele sõnastikul on igaühel ka oma nägu. Nii on Stahl ja Gösekeni sõnastikus ülekaalus põhjaeestikeelsed väljendid, Gutsloffi omas aga lõunaeesti keel. Mahu poolest tõuseb vaieldamatult esile Gösekeni leksikon, mis

<sup>4</sup> Esimest trükitud eesti-saksa sõnaraamatut tuli oodata veel terve inimpõlv: selle avaldas alles 1732. aastal Anton Thor Helle, samuti koos eesti keele grammatikaga (Kilgi, Ross 2006: C25).

<sup>5</sup> Misjonilingvistika terminit on religioosse misjonitöö eesmärgil tehtava keeleuurimistöö kohta 17. sajandi Eesti ala kontekstis kasutanud mitu autorit, teiste seas Marju Lepajõe (1998), Kristiina Ross (nt 2003), Cornelius Hasselblatt (2006) ja Reet Kasik (2011).

sisaldab ligikaudu neli korda rohkem ainet kui sellele ajaliselt eelnenud Stahli teos. Gutsblaffi sõnastik on omakorda üle kahe korra väiksem kui Stahli oma.

Kui kaks viimati mainitud autorit avaldasid oma keeleõpetusega koos saksa-eesti sõnastiku, siis Göseken on esitanud ka hulga ladina vasteid (need on 38% märksõnadel), ehkki ta ei nimeta ladina keelt ei sõnastiku pealkirjas, alapealkirjades ega ka saatesõnades. Raamatu kaanel seisab “Juhatus eesti keele juurde, mis ei koosne mitte ainult mõningatest õpetustest ja tähelepanekutest, vaid ka paljude saksa sõnade tõlgetest” ning keeleõpetuse järel algava sõnastikuosa alapealkiri on “Saksa-eesti sõnastik tähestikulises järjekorras kirja pandud. See on Eesti keele sõnastik, kus tähtsamad saksa sõnad eesti keelde on tõlgitud ja tähestikulisse järjekorda pandud”<sup>6</sup> (Kingisepp jt 2010: 39, 279). Ladina ainese allikatest tuleb juttu järgmises jaotises.

## 2.4. Sotsiokultuurilise tahu arvestamine

Stahlist ja Gösekenist said oma aja autoriteetsemad kirjamehed Põhja-Eesti keelealal ning lõuna-eesti kontekstis võib sama väita Johannes Gutsblaffi kohta. Kui Heinrich Stahl pani oma sõnastikuga eesti leksikograafiale aluse, siis kahe järgmise sõnastiku autorid said tuge varem ilmunud eesti keelt sisaldanud teostest – neile viitavad nad ka oma teoste eessõnades. Kõigil kolmel autoril võis mõistagi olla kasutada selliseidki allikaid, millele nad ei viita. Kai Tafenau (2011a: 432 jj) on teinud näiteks kindlaks, et hulk Gösekeni leksikoni ladina sõnu pärineb Jan Amos Komenský temaatilisest ladina keele õpikust “*Janua linguarum reserata*” (esma-trükk 1631), mis tähendab, et Göseken käsitas sõnaraamatutööd avaramas võtmes, kui seda oleks eeldanud piibli tõlkimise siht. Tava võtta võõras aines oma teosesse üle ilma algautorit nimetamata on pärvinud hilisemal ajal kriitikat, ent 17. sajandil oli selline kompileerimine üldlevinud (Lotman 1995: 541).

Teiste seas on Euroopa tollast mõtteruumi ja grammatikakirjandust põhjalikumalt iseloomustanud Marju Lepajõe. Ta sedastab (1998: 291), et 20. sajandi lingvistid ei ole osanud 17. sajandi keeleteaduslike teoste väärtust hinnata, kuni Noam Chomsky neid selle eest 1966. aastal kritiseeris. Ka Stahli keeleõpetuse retseptisoonis oli võõrapärasuse ja ebahütlase kirjaviisi tõttu ülekaalus kriitika (vt nt Kask 1970: 40), suhtumisse tõi pöörde alles Huno Rätsep (1987). Viimastel aastakümnetel on 17. sajandi sõnastikesse suhtutud eelkõige hindavalt.

Veidi varem oli ka Harald Haarmann (1976: 9–22) rõhutanud tollaste grammatikate väljaandmise keelesotsioloogilist tahku. Et 17. sajandi teisel poolel ei olnud Eesti alal piisavalt ei kirikuid ega ka lugeda oskavaid talupoegi, andis kuningas välja resolutsioonid (1675), mis panid aluse koolide rajamisele. Need aktid olid sammuks pärisorjusliku talurahva õigusliku seisundi parandamise teel, nagu osutab kiriklikku rahvaharidust põhjalikumalt vaadelnud Aivar Põldvee (vt nt 2010: 17–24).

<sup>6</sup> Selles artiklis on kasutatud Kai Tafenau tõlkeid eesti keelde.



### 3. Heinrich Gösekeni algsõnastik

Gösekeni 1660. aasta teos algab saksakeelse dedikatsiooni ja eessõnaga, millele järgneb 84 lehekülge eesti keele õpetust. Keeleõpetuse ja põhisõnastiku vahele on paigutatud väike sõnastik, mille esimese, eraldiseisva osa moodustavad mõisteteliselt liigitatud sõnaloendid: lühendid, kontraheerunud ja apokoobiga verbivormid, Tartu murde erijooned ning kuude, nädalapäevade ja pühade nimetused koos selgitustega. Järgnevad 412 lehekülge hõlmab saksa-ladina-eesti sõnastik ning kogumiku lõpus asub 15 lehekülge lisa pealkirjaga "Supplementum", milles autor täiendab ja parandab nii keeleõpetust kui ka sõnastikuosa. Teose maht on 547 lehekülge.

Nii ulatusliku töö tegemiseks sai Göseken innustust Johannes Gutsclaffilt, kes oli teinud talle ettepaneku anda välja Uue Testamendi põhja- ja lõunaestikeelne paralleeltõlge. Et Gutsclaff 1657. aastal suri, tuli sellise tõlke avaldamise plaan kõrvale jätta. Gösekeni sõnastikku sattunud lõunaesteti väljendeid ongi selgitatud asjaoluga, et tal olid kasutada Gutsclaffi käsikirjalised piiblitõlkekatkendid. (Värri Haarmann 1977: 8–9)

Algul kavatses Göseken avaldada temaatilise sõnastiku koos sünonüümi- ja paronüümiloendiga, kuid see ei õnnestunud – eessõnas nimetab ta põhjusena kirjastaja kehvust. Nii koostas ta kiiresti alfabeetilise sõnastiku ja jättis ülejäänud materjali paremaid aegu ootama. Sõnastikus on saksa ja/või ladina märksõnadele antud 13 787 eesti vastet (sealhulgas korduvad). Lausalfabeetiline ülesehitus, mille puhul iga lemma algab uuel realt, on põimitud pesaalfabeetilise lähenemisega, kus koos on esitatud samast tüvest lähtuvad sõnad.

### 4. Uusväljaannete koostamise meetod

Siirdugem nüüd Gösekeni 1660. aasta sõnastiku juurest uusväljaande (2010) koostamise juurde.

Eesti keelt sisaldavate allikate uurimiseks on TÜ vana kirjakeele töörühmas viidud materjal elektroonilisele kujule. Sõnastiku aines tuleb seejuures süstematiseerida, sest selle sisestamine jooksva tekstina pole mõttekas. Seega algab analüüs juba sisestamisel. Kõigepealt püütakse algsõnastiku keelendid vormida üksühes-tekst vastepaarideks, tervikuna sõnaartikliteks. Alati pole see siiski võimalik, sest näiteks Gösekeni sõnaraamatus leidub 167 saksa märksõna, millel ei ole ei ladina ega eesti vastet, ning 142 saksa-ladina vastepaari on ilma eesti vasteta. Algsete märksõnapesade lahtiharutamine ning vastepaaride või -kolmikute tähestikuline järjestamine eesti keele alusel toob paberväljaandes tänapäeva lugeja jaoks kokku eesti sõnad ja väljendid, mis algsõnastikus asusid eri paikades ja olid erinevate saksa sõnade vasteks.

Lisaks raskesti loetavatele, kustunud või laialivalgunud grafeemidele seisab uurija ees hulk algsõnastiku mikro- ja makrostruktuuri ebahütlusest tulenevaid küsimusi. Osa neist on ajastuomased, seotud nii leksikoni käitsi koostamise kui ka selle laadiga, teine osa kaasneb kasutatava meetodiga – algsõnastiku läbimõeldud tervik võetakse osadeks lahti. Peale selle rekonstrueeritakse tarviduse korral puuduvad sõna(osa)d, selgitatakse sõnade tähendust ja lisatakse grammatilisi

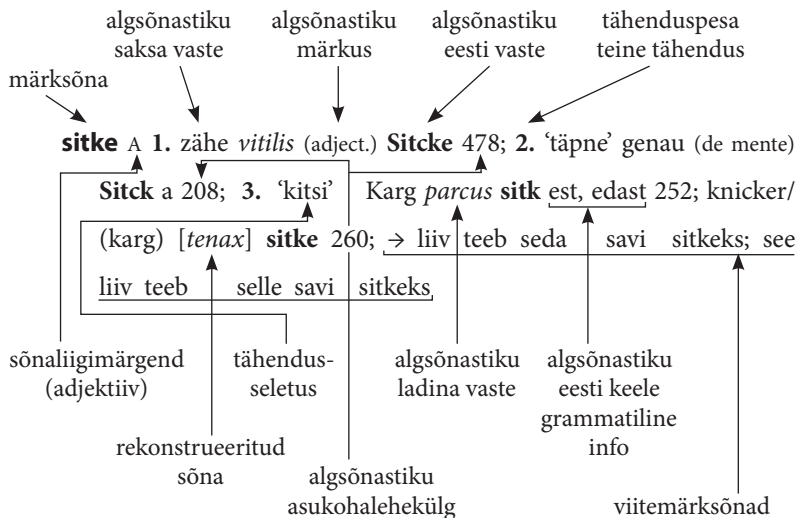
märgendeid. Et huviline saaks algsõnastikust tervikpildi, lisatakse ka viitemärksõnad (vt lähemalt artikli 5. osast).

Selliselt töödeldud keeleainest saab kasutada elektrooniliselt<sup>7</sup>, aga ka vormida paberil sõnaraamatuteks. Esimesel juhul pole keelesuuna küsimus peamine, kuid seda tasub materjali kasutades arvestada, sest keelesuund on mõjutanud algsõnastiku sisu. Olnuks Gösekenil lähtekeeleks eesti keel, valinuks ta tõenäoliselt mõnevõrra teistsugused märksõnad.

Kirjeldatud meetodit (vt ka Ress 2009) katsetati esimest korda Heinrich Stahli sõnastiku uusväljaande koostamisel (Kikas 2002), seejärel Johannes Gutsblaffi (käsikirjas, Viitkar 2005) ja Heinrich Gösekeni sõnastiku puhul (Kingisepp jt 2010).

## 5. Gösekeni algsõnastiku eesti ainese märksõnastamine ja tähenduse seletamine

Uusväljaandes<sup>8</sup> on töödeldud ja analüüsitud kujul avaldatud kogu Heinrich Gösekeni 1660. aasta keeleõpetuse ja sõnastiku materjal, sealhulgas faksiimileheküljed kuni sõnastikuni. Koostajate saatesõnadele järgneb algsõnastiku sisesejuhata-vate osade ja grammatika faksiimile koos kõrvutise eestindusega Kai Tafenault. Kogumiku teise poole moodustab algse saksa-ladina-eesti sõnastiku uusvariant, mida iseloomustab muudetud keelesuund ning märgendatud ja leksikoloogiliselt analüüsitud aines. Siinses kirjutises näiteks toodud sõnaartiklite ülesehitust aitab mõista joonisel 1 esitatud skeem. Sõnaartikli algusosas asub uusväljaande koostajate lisatud teave: nüüdisortograafias märksõna (nt *sitke* joonisel 1), sõnaliigimärgend (nt A ‘adjektiiv’) ja tähendusseletus (nt ‘kitsi’). Nii algsõnastiku eesti märksõnad (*Sitcke*, *Sitck*, *sitk*, *sitke*) kui ka uusväljaande koostajate lisatud tänapäeva eesti keele märksõna (*sitke*) on poolpaksus kirjas.



Joonis 1. Uusväljaande sõnaartikli struktuur (Kingisepp jt 2010: 34)

<sup>7</sup> Aines lisatakse eesti vana kirjakeele korpuse aadressil <http://www.murre.ut.ee/vakkur>.

<sup>8</sup> Sõnastikuosa koostajad Valve-Liivi Kingisepp ja Kristel Ress on tänulikud toimetajatele Kai Tafenaule, Külli Habichtile ja Külli Prillopile. Algsõnastiku materjali on arvutisse sisestanud Argo Mund ja Piret Sõmermaa.

Uusväljaandes on kirjakuju mõningase varieerumisega korduvad eesti märksõnad koondatud tänapäevase märksõnaga algavasse sõnaartiklisse. Kui see sisaldab mitut algsõnastiku sõnaartiklit (s.t sõna või fraas kordus Gösekeni algsõnastikus) ja eesti märksõna on mitmetähenduslik, on algsõnastiku sõnaartiklid rühmitatud eesti vaste tähenduse järgi. Suuremate erinevuste korral on abiks võetud araabia numbrid ja tähendusseletused, väiksemate puhul seletused. Tähendusrühmade sees on aines reastatud asukoha järgi algsõnastikus.

Polüsemidel esitatakse esimesena tänapäevane põhitähendus, millele üldjuhul tähendusseletust ei lisata (nt märksõna *kuulutama* joonisel 2), seejärel loetletakse muud tähendused koos seletusega. Sõnaartikli lõpetavad noolele järgnevad viitemärksõnad (kui neid on), mis osutavad seda sõna sisaldavatele mitmest sõnast koosnevalele põhimärksõnadele (nt viide märksõnadele *ette kuulutama*, *jutlust kuulutama*, *käsk, mis välja kuulutatud on*, *välja kuulutama* joonisel 2).

**kuulutama** v 1. ankündigen **kuhletama** 108; [Ausbringen] *effutire kuhlutama* 116; [verkündigen *annunciare*] **kuhlutama** 427; 2. 'jutustama' Erzählen **kuulutama** 180; 3. 'ette kuulutama' weissagen *vaticinari kuhlutama* 465; 4. 'avatuks kuulutama' eröffnen *promulgare kuulutama* 178; → ette kuulutama; jutlust kuulutama; käsk, mis välja kuulutatud on; välja kuulutama

**Joonis 2.** Uusväljaande sõnaartikkel *kuulutama* (Kingisepp jt 2010: 443)

Gösekeni sõnastikus on rohkesti mitmest sõnast koosnevaid märksõnu. Kui nende märksõnade koostises olevaid sõnu (peale esimese) sõnastikus iseseisvana ei leidu (nagu see on näiteks sõna *mina* puhul), on nende leidmiseks loodud viitemärksõnad. Need asuvad sõnastikus põhimärksõnadega samas loendis ja on sama tüpograafilise välimusega (vrd põhimärksõna *kuulutama* joonisel 2 ja viitemärksõna *mina* joonisel 3). Viidatakse seega fraasiliste märksõnade järelkomponentidele, et lugeja leiaks uusväljaande paberversioonist ka need kergema vaevaga. Kui sõnaartikli osaks on mitu viidet, on need järjestatud tähestikuliselt.

**mina** → ei ole minul sest süüd mitte; ei tule mu meelde; ei või mind näha mitte; et ei ma ütle; lase mind uinuda; mis sa heris pöörled end n[ing] kõheldad, ma tahan sulle rutu teha; oleksid sa minu kõrva midagi tõlkinud; seba mu asi; see kiitmine ei tee mulle head; see koda on minu päralt; see on minu osa; see saab minu peale tulema; see teeb mulle haiget; see tuleb minu osaks; säh sa minule see; tema mõistab minu poole; tema on üle minu; tema patub minu poole; ühtegi hakkab minu hõlmast kinni

**Joonis 3.** Uusväljaande viitemärksõna artikkel *mina* (Kingisepp jt 2010: 499)

## 5.1. Märksõnastamine

Eesti märksõnaks on uusväljaandes võimaluse korral tänapäevase kirjakuju sõna oma põhivormis, s.t käändsõna üldjuhul ainsuse nimetavas ja pöörd sõna *ma*-tegevusnime kujul. Mitmesaja-aastases sõnastikus on rohkesti sellist leksikat, näiteks

kunagisi laensõnu, mis ei ole nüüdiskirjakeelde jõudnud (vt joonis 4<sup>9</sup>). Ka nende keelendite puhul on märksõnaks tänapäevastatud ortograafias kuju.

**leeverstok** s 'leeskuptk' Liebstöckel **Lewerstock** 280

..  
**tolfter** N 'tosin' dutzet/ (dossin) **Tolffter** 163; **Zwelffter**/ (Dosin) **Tolffter** 496

**Joonis 4.** Uusväljaande sõnaartiklid *leeverstok* ja *tolfter* (Kingisepp jt 2010: 468, 666)

Endis- ja praegusaegsete sõnakujude vahekord on üks keerulisemaid küsimusi märksõnastaja ees. Tüvede ortograafiat on nüüdisajastatud sihiga pakkuda uurijatele lihtsasti mõistetavat andmestikku, mida kasutades ei ole tarvis pead murda, millise arhailise kirjakuju juurest mingit sõna leida. Märksõna kirjakuju tänapäevastamise põhimõtet on rakendatud ka TÜ vana kirjakeele tööruhma varasemates väljaannetes (Müller 2000, Rossihnius 2002).

Märksõnale sobiva kuju leidmisel ja/või tähenduse seletamisel on esmajoones toetunud järgmistele allikatele siin esitatud järjekorras: õigekeelsussõnaraamat (ÕS 2006) ja eesti keele seletav sõnaraamat (EKSS), Ferdinand Johann Wiedemanni eesti-saksa sõnaraamat (1973 [1893]) ja Andrus Saareste eesti keele mõisteline sõnaraamat (1958–1963). Kasuks on olnud ka näiteks Jacob ja Wilhelm Grimmi saksa keele sõnaraamatu (1838–1960) veebiversioon, Georg Tuksami (1939) saksa-eesti sõnaraamat ning ladina-eesti sõnaraamat (Kleis jt 2002).

Kui Gösekeni arhaismi ÕSis ei leidunud (vt joonis 5), saadi tihti abi Wiedemanni sõnaraamatust ja võeti märksõnaks sealne kuju, nagu näiteks märksõna *haevas on ees* 'ajab aevastama' koostises esineva sõna *haevas* puhul. Wiedemanni sõnaraamatust puuduvaid eesti sõnu otsiti Saareste sõnaraamatust, kust teiste seas sai oma kuju märksõna *sültkruut* 'hoidis'.

**haevas on ees** F 'ajab aevastama' da ist ein praust für/ (sagen Teutsche und Bauren) **hahwas on ees** 325

..  
**sültkruut** s 'hoidis' Eingemachte Sachen **Sültkruut** 168

**Joonis 5.** Uusväljaande sõnaartiklid *haevas on ees* ja *sültkruut* (Kingisepp jt 2010: 335, 645)

## 5.2. Verbilised märksõnad

Verbilisi eesti vasteid leidis algsõnastikus mitmel kujul. Sagedasimad olid a) *ma*-tegevusnimi (enamik, nt Gösekeni kirjakuju *kickertama*), b) *ma*-tegevusnimi koos mõne pöördvormiga, milleks peamiselt on ainsuse esimese pöörde oleviku ja lihtmineviku vorm (väga vähe, nt *tahhoma*, *tahhan*, *tahhosin*), c) *da*-tegevusnimi (väga vähe, nt *tehha*), mõnikord samuti koos pöördvormi(de)ga, ning d) ainult pöördvorm (väga vähe, nt *koitab*, *kommistab*, *ütteltaxe*).

Uusväljaandes on pöördvorm märksõnaks (vt joonis 6) eelkõige siis, kui ka algsõnastiku võõrvaste oli muus vormis kui infinitiivis või oli tegemist fraasiga. Nii on saadud sellised märksõnad nagu *komistab*, *koidab* ja *ütteldakse*. Kaks viimast esindavad nende verbide rühma, mida kasutatakse sageli just ühes kindlas vormis.

<sup>9</sup> Selle artikli joonistel, mis sisaldavad mitut sõnaartiklit, mis ei asu sõnaraamatus vahetult üksteise järel, on need järjestatud tähestikuliselt ja väljajätt on märgitud kaksikpunktiga. Sõnaartiklite asukohale sõnastikus osutavad samas järjekorras esitatud leheküljenumbriid joonise allkirjas.

**kikerdama** v 'püherdama (hobuste kohta)' weltzen/ (der Pferde) **kicker-**  
**tama** 466

..

**koidab** v es wird morgen **koitab** 299

..

**komistab** v [es] strauchelt **kommistab** 321

..

**kukkuma II** v '(käo häälightsuse kohta)' *cuculare* **Kuckub** 268

..

**tahuma** v <Behauen> **tahhoma** tahhan, tahhosin 128, 501S; [zimmern]  
**tahhoma** 485

..

**üteldakse** v man saget *ferunt üteltaxe* 288 ..

**Joonis 6.** Uusväljaande sõnaartiklid *kikerdama*, *koidab*, *komistab*, *kukkuma II*, *tahuma* ja *üteldakse* (Kingisepp jt 2010: 410, 421, 424, 433, 650, 759)

### 5.3. Substantiivsed märksõnad

Nimisõnalised eesti vasted olid algsõnastikus enamjaolt ainsuse nimetavas, üksikjuhtudel mõnes muus käändes. Peale selle on Göseken aga esitanud ainult mitmuse vormis selliseid substantiive, mis väärinuks siiski ainsust, sest nad ei tähista mitmuslikke nähtusi (nagu *prillid*, *talgud*). Sel juhul on uusväljaande koostajad pidanud algse autori tahtest lugu ja esitanud Gösekenil mitmuse vormis olnud nimisõnad samal kujul ka uusväljaandes, s.t arvu pole muudetud. Nii on mitmuse nimetavaline märksõna näiteks marjanimetustel (nt *kaarmad* 'murakad', *jõhvikad*), muud elusloodust tähistaval leksikal (nt *massakad* 'putukad', *kasikud* 'puuseened'), toidusõnavaral (nt *mämmid* 'küpsetised', *veisetoidud*), esemeid tähistavatel sõnadel (nt *tahkid* 'pussid, pistodad', *puuastjad*) jm.

Et tegu oli keeleõpet toetava sõnastikuga, lisas Göseken põhivormile sageli grammatilist teavet (vt joonis 7). Kõige tüüpilisemalt oli selleks tüvevokaal (nt Gösekeni kirjakuju *Keert* / *i*), seestütleva käände vorm või lõpp (*hack* / *ust*), harvem mitmuse nimetava käände vorm – näiteks juhul, kui sellega kaasnes tüvelõpumuutus (*lehexet*).

**hakk, -u** s 'võsa, noored kased' Meye (Pfungstmeyen) **hack** ust 294

..

**keert** s 'teemoon' [beysteur *collecta*] **keert** 138; Proviant (für Pferde) **Keert** i  
326

..

**leht** s blatt **lecht** pl. lehexet 141; Laub **lecht** lehhest, pl. lehexet 275; →  
laiade lehestega

**Joonis 7.** Uusväljaande sõnaartiklid *hakk*, *keert* ja *leht* (Kingisepp jt 2010: 336, 400, 468)

Nii mahukas sõnakogu nagu Gösekeni oma sisaldab paratamatult homonüüme. Uusväljaandes moodustavad need omaette märksõna ja homonüümide eristamiseks on kasutatud kaht viisi. Eelistatud on tüvevokaali lisamist (nt märksõnale *hakk* joonisel 8), ent kui sellest jäi väheks, on võetud abiks Rooma numbrid (vt märksõna

koon). Kui korduval sõnal erinevad nii sõnaliik, päritolu, muutmisviis kui ka tähendus, on need saanud oma sõnaartikli.

**hakk, -i** s 'lind)' Dohle **Hack** 159

**hakk, -u** s 'võsa, noored kased' Meye (Pfungstmeyen) **hack** ust 294

..

**koon I** s kinn *mentum kohn* 257; [kinn *mentum*] **koon** 257

**koon II** 'takukoonal' Deissel/ (am Wocken) **Koon** 157; Distel **Kohn** 189;

Anlage (Flachs Distel) *pensum koon* 499S; <Distel (Flachs) **Kohn**> 503S

**koon III** 'koonus' [Kegel *conus*] **koon** 254

**Joonis 8.** Uusväljaande homonüümsete märksõnade artiklid *hakk* ja *koon* (Kingisepp jt 2010: 336, 425)

#### 5.4. Tuletised ja liitsõnad

Üks osa Gösekeni eesti vasteid on eripärased tuletised ja tuletisetaolised sõnad, a) millel on teistsugune liide (*resp.* liitetaoline osis) kui tänapäeva sama tüve ja tähendusega sõnas (vrd nt joonisel 9 *k*-lõpuline *harak* ja *kas*-lõpuline *harakas* nüüdiskeeles; *line*-tuletis *pääskline* ja *ke(ne)*-tuletis *pääsukene* nüüdiskeeles), b) millel puudub liide, mis esineb sama tüve ja tähendusega nüüdisaja sõnas (nt joonisel 10 sõnas *haiguma* puudub tänapäevakeele sõnas *haigutama* olev *-ta-*) või c) mille puhul on tegu sõnaga, milles tänapäevakeeles ei ole enam liidet või liitetaolist osist (nt joonisel 10 *müüestelema*).

**harak** s 'harakas' elster *pica harogk* 172

..

**pääskline** s 'pääsuke' Schwalbe *hirundo pehscklinne* 371

**Joonis 9.** Uusväljaande sõnaartiklid *harak* ja *pääskline* (Kingisepp jt 2010: 338, 583)

Tuletusliiteid ei ole uusväljaande koostamisel märksõnadele lisatud, küll aga on tüvesid tänapäevastatud häälikuliselt. Edasise uurimise seisukohast on peetud vajalikuks säilitada liited, mis on nüüdseks sõnast kadunud, aga Gösekeni aineses veel näha.

**haiguma** v 'haigutama' Gähnen/ (Hojanen) **haickoma** 200

..

**müüestelema** v 'müüma' Feil haben oder tragen **mühhestellema** 185

**Joonis 10.** Uusväljaande sõnaartiklid *haiguma* ja *müüestelema* (Kingisepp jt 2010: 335, 514)

Liite poolest eripärastele märksõnadele on võimalik elektroonilises andmebaasis lisada originaalilähedased viitemärksõnad (joonisel 10 märksõna *haiguma* puhul *haikuma*). Praegu ei ole seda tehtud.

Liitsõnade puhul on märksõna nimetavalise ja omastavalise liitumise vahel valimisel lähtunud tänapäevakeele põhimõtetest: tulemuseks on näiteks *katelsepp* (Gösekeni kirjakuju *Katla Sepp*), *kihlvedu* (*kihlaweddo*), *purpurikarvaline* 'purpurne' (*purpur* (*karwalinne*)) ja *üksmeelelised* 'üksmeelsed' (*öhhe meelelisset*). Tänapäevastamise põhjenduseks on taas soov, et uurijail oleks hõlpsam sõnakuju

leida. Ka neile märksõnadele saaks korpuse lisada originaalilähedased viite-märksõnad, vastavalt *katlasepp*, *kihlavedu*, *purpurkarvaline* ja *ühemeelised*.

Kas märksõnasse kuuluvad sõnad kirjutada kokku või lahku, on esmajoones tänapäeva uurija otsus lähtuvalt praegusaja ortograafiareeglitest – algsõnastiku kokku-lahkukirjutus nagu ka muu ortograafia on ebaühtlane. Nii pakkus uusväljaande koostajatele mõneks ajaks mõtteainet eesti vaste *tühhi mahd* (esineb ka kujul *Tühha mahd*, *Tühhi mahd*, vt joonis 11).

das wüste im Leibe / tühhi mahd / Affeme.

weiche/ (zur munden)topp. weiche/ (des bauchs)  
Tühha mahd.

wüste im Leibe / Tühhi mahd / paick.

**Joonis 11.** Uusväljaande sõnaartikli *tühemed* aines algsõnastikus (Göseken 1660: 277, 462, 478)

Tõlgendus kaldus esialgu märksõna *tühi maa* poole, ent sellega ei haakunud saksa vastete tähendus. Märksõnaks sai Wiedemanni sõnaraamatu toel hoopis *tühemed* (vt joonis 12) ja sõnastiku analüüsijad said kasuliku õppetunni.

**tühemed** s 'lohk looma kerel puusa lähedal' das wüste im Leibe **tühhi mahd** 277; weiche/ (des bauchs) **Tühha mahd** 462; wüste im Leibe **Tühhi mahd** 478

**Joonis 12.** Sõnaartikkel *tühemed* uusväljaandes (Kingisepp jt 2010: 679)

## 5.5. Gösekeni märkused

Peale tõlkevastete ja grammatilise info leidub algsõnastikus metateksti, mida on uusväljaandes nimetatud Gösekeni märkusteks (vt nt joonis 13<sup>10</sup>).

Begleiten / einem das Geleit geben/  
in bonam partem johharan.  
in malam partem olletan.  
in utramq; partem sahtran.

**Joonis 13.** Märkused algsõnastiku saksa sõnaartikli *Begleiten* juures (Göseken 1660: 128)

Sõnastikumaterjali sisestamise käigus on püütud kõik 828 märkust siduda selle sõnaartikliga, mille juures Göseken nad oli esitanud. Märkused on sõnaartikli muust ainesest eristatud väiksema kirjaga ja paigutatud Gösekeni eesti kirjakuju ette (vt joonis 14). Ladina- ja saksakeelsetele märkustele ei ole uusväljaande koostajad tõlget lisanud. Lugejad on pidanud märkuste ühendamist vastava märksõnaga vajalikuks, sest nii saab infot algsõnastikus olnud terviku kohta (Laanekask 2012: 276).

<sup>10</sup> *in bonam partem* 'heas tähenduses', *in malam partem* 'halvas tähenduses', *in utramque partem* 'mõlemas tähenduses'.

**juhatama** v Begleiten/ einem das Geleit geben in bonam partem **johhatan** 128; [Führen/geleiten *ducere*] **johhatama** 196; geleiten **johhatama** 206; [leiten *ducere*] **johhatama** 279; → tee peale juhatama

..  
**saatma** v 1. [Begleiten/ einem das Geleit geben] in utramq; partem **sahtan** 128; Führen/geleiten *ducere* **saatma** 196; [geleiten] **saatma** 206; leiten *ducere* **saatma** 279; 2. 'tekutama' stiefften *machinari* **sahtma** 397; zeugen/ (erlangen) *comparare* **saatma** 483; → enesest ära saatma; hukka saatma; häbiduse sisse saatma; kes ostab ning saadab seda kõrvast rooga; kohtusse saatma; käest ära saatma; kätte saatma; mu naine on kui saatnud on; oma teed otsa saatma; otsa saatma; rahule saatma; sisse saatma; teotuse sisse saatma; und otsa saatma; vabaks saatma; viletsuse sisse saatma; ära saatma

..  
**õletama** v '(kedagi) saatma' [Begleiten/ einem das Geleit geben] in malam partem **olletan** 128

**Joonis 14.** Joonisel 13 näha olevad märkused uusväljaande sõnaartiklites *juhatama*, *saatma* ja *õletama* (Göseken 1660: 370, 607, 728)

Märkused on kirev rühm saksa-, ladina- ja eestikeelset materjali. Korduvad eelkõige ladinakeelsed, nt *activ.*, *adject.*, *adverb*, *nomen*, *subst.*, *est avis* 'on lind', *est herba* 'on taim', *est insectum* 'on putukas', *in bonam partem*, *in malam p.* Arvuka rühma moodustavad lühendiga *v.* ... või sõnaga *vide* ... 'vaata [vastavat sõna]' algavad ladina märkused – nendega juhata Göseken lugeja näiteks saksakeelse sünonüümi juurde. Sagedased on *vide suo loco* 'vaata oma kohta' ja *vide* koos tähestiku vastava tähega (nt *vide A*). Saksa märkused on harilikult ainukordsed, nt *werden auch gebraucht von denen vermüdeten die sich verpausten* märksõna *lõõtsuma* juures.

Märkusteks on tõlgendatud ka selline eestikeelne aines, kus sünonüümsele eesti vastele eelneb ladinakeelne lühend *it.* 'samuti' või märk &. Pelgalt selle lühendi või sümboli äratrükkimine märkusena poleks uusväljaande kasutaja huve teeninud, seetõttu on loetud kogu fraas (*it.* või & koos eesti vastega) märkuseks ja esitatud nii iseseisva sõnaartiklina kui ka esimese eesti vaste sõnaartiklis märkuse positsioonis. Näiteks algsõnastiku saksa sõnaartiklit *PalmSontag* (vt joonis 15) analüüsides on Gösekeni märkusteks määratud kõik, mis järgneb esimesele eesti vastele *urma pehw*, selle alusel on loodud kaks sõnaartiklit (vt joonis 16). Et algsõnastikus oli *urma pehw* eesti vastetest esikohal, siis on uusväljaandes märksõna *kanastepüha* kui teisena esitatud vaste juures saksa vaste ja Gösekeni eesti kirjakuju kirjas nurksulgudes. Nii näidatakse lugejale, et uusväljaande koostaja on need sõnad rekonstrueerinud.

**PalmSontagi urma pehw. it. kanaste pöha.**  
NB. urm die Knospe am Baum.

**Joonis 15.** Märkused algsõnastiku saksa sõnaartikli *PalmSontag* juures (Göseken 1660: 318)



**kanastepüha** s ‘palmipuudepüha’ [PalmSontag **kañaste pöha**] 318

..  
**urmapäev** s ‘palmipuudepüha’ PalmSontag NB. urm die Knospe am Baum/ it.  
kañaste pöha **urma pehw** 318

**Joonis 16.** Joonisel 15 näha olevad märkused uusväljaande sõnaartiklis *kanastepüha* eesti vastena ja sõnaartiklis *urmapäev* märkusena (Kingisepp jt 2010: 387, 685)

## 5.6. “Supplementumi” aines

Göseken kinnitab oma teose eessõnas, et on teadlik võimalikest trükivigadest, sest korrektuuri lugemiseks oli vähe aega ja inimesi. See selgitab, miks on kogumiku lõpus 12 leheküljel esitatud grammatika ja sõnastiku vigade parandus ning täienduste loend pealkirjaga (tõlkes) “Tähelepanekud eesti-saksa sõnaraamatu kohta ja täiendused sellele koos raskemate vigadega” (artiklis lühidalt “Supplementum”). Need leheküljed on pagineerimata – ainese asukohale osutamiseks on leheküljenumbri (499S–511S) lisanud uusväljaande koostajad, eristuseks on nende lõpus olev S-täht.

Neli viiendikku “Supplementumi” mahust moodustavad uued sõnaartiklid, ülejäänud on olemasolevate artiklite täiendused nii saksa, ladina kui ka eesti ainesega seoses. Uusväljaande koostajad on kõigi lisanduste-paranduste alusel algsõnastiku sõnaartikleid muutnud. Kui “Supplementumi” aines oli uus, on uusväljaande sõnaartikkel saanud S-tähga lõppeva leheküljenumbri. Kui Göseken soovis aga olemasolevat sõnaartiklit muuta või täiendada, lisati ainese põhiasukohale viitavale leheküljenumbrile (262) “Supplementumi” leheküljenumber (509S – mõlemad joonisel 17). “Supplementumi” põhjal muudetud-lisatud aineist ümbritsevad noolsulud.

**hobune on koodist vältinud** F ‘hobune on liigesest nikastunud’ das  
Pferd ist über die <Koht> geschossen <**Hobbone on kohtist weltinut**>  
262, 509S

**Joonis 17.** “Supplementumist” pärineva ainesega sõnaartikkel *hobune on koodist vältinud* uusväljaandes (Kingisepp jt 2010: 346)

## 5.7. Tähenduse seletamine

Märksõnastamisega käib tihedalt koos tähenduse seletamine, milles on uusväljaandes abiks sünonüüm või lühike kirjeldav seletus. Arhailise sõnastiku sõnade tähendust on keerulisem mõista kui tekstisõnu. Sõnaraamatus on toeks esmajoones võõrkeelsed vasted. Uusväljaande koostamisel on lähtutud seisukohast, et tänapäeva uurija jaoks oleks sageli eksitav esitada üksnes märksõna, ilma selle tähendusvarjundit selgitamata. Ainult üks kaalutlusi oli see, et saksa keele oskus on uuemal ajal kahanenud, nagu osutab ka Heli Laanekask (2012: 276).

Nii ei ole märksõna *kikerdama* eesti vastes hobusest küll juttu, ent saksakeel-sena on kirjas viide tähendusväljale (Gösekeni *der Pferde* ‘hobuste’). Ilma saksa vastet arvestamata poleks koostajad osanud õieti midagi arvata näiteks Gösekeni kirjakujust *tühhi kohhu*, mis on saanud märksõnaks *tühikohu* ‘enesekiitja’. Või kas kirjakuju *Kahpse Aidamees* esitades oli Göseken mõelnud *kapsaaidameest* või *kapsaaiameest*? Saksa *Kohl Gartner* alusel on valitud märksõnaks *kapsaaiamees*

'kapsakasvataja'. Samamoodi on eristatud polüsemsete sõnade tähendusnüansse, mille hulgast võib keerukamate juhtudena esile tõsta tegusõnad (nt *tasama*). Võõravastete abil tähenduse seletamisel on püütud vältida teise äärmusse kaldumist ning mitte lisada tähendusi, mille kohta kindlaid andmeid pole.

Loetletud märksõnade sõnaartiklid on esitatud joonisel 18.

**kapsaaiamees** s 'kapsakasvataja' Kohl Gartner **Kahpse Aidamees** 262

..

**kikerdama** v 'püherdama (hobuste kohta)' weltzen/ (der Pferde) **kicker-tama** 466

..

**tasama** v 1. 'tasuma' [entgelten/ (vergelten)] **tassama** 174; 2. 'tasaseks tegema' Eben machen **tassama** 164; 3. 'tasandama, lepitama' [beylegen/ (ein Sache)] **tassama** 138; → ära tasama

..

**tühikohu** s 'enesekiitja' der sich viel rühmet **tühhi kohho** 220

**Joonis 18.** Uusväljaande sõnaartiklid *kapsaaiamees*, *kikerdama*, *tasama* ja *tühikohu* (Kingisepp jt 2010: 390, 410, 655, 680)

Tähendus võib aja jooksul muutuda – mõne sõna praegu lisatähendusena käsitatav nüanss võis 17. sajandil olla sõna esmane või ainus tähendus. Tõlgendamisel jääb tihti õhku küsitavusi ja kaasneb subjektiivsus. Uusväljaandes on seletatud kõiki arhaisme<sup>11</sup> ja omapärase häälikkujuga laensõnu, erilise tähendusega märksõnu ja sõnaühendeid (vt joonis 19), mis ei ole meie keelde jäänud või on vähe tuntud. Ka neile eesti sõnadele, mis on küll õigekeelsussõnaraamatus olemas, aga esinevad Gösekenil muus tähenduses, on lisatud tähendusseletus.

**knuuflok** s 'küüslauk' [knoblauch] **knuufloki** 260

..

**neps** A 'niiske' Nas/ (Feucht) **neps** 306

..

**Saksamaa tang** F 1. 'hirss' hirse *milium saxa mah tango* 239; 2. 'riis' [reyß *oryza*] *Saxa mah tango* 336

**Joonis 19.** Uusväljaande sõnaartiklid *knuuflok*, *neps* ja *Saksamaa tang* (Kingisepp jt 2010: 418, 522, 609)

Vajaduse korral on tähendussisu avatud seletusega või osutatud tähendusvaldkonnale, kuhu sõna kuulub (vt joonis 20).

**sapp päeva peal** F 'päikesesapp, kõrval-, ebapäike (halonähtusena Kuu v Päikese kõrval paistev hele laik)' wenn ein sonne über der ander stehet/ saget der Baur **Sapp pehwa pehle** 384

..

**susi-ilv** s 'huntilves (jahimehed talvekarvas ilvese kohta)' Kalbluchs **sussi ilw** 284

**Joonis 20.** Uusväljaande sõnaartiklid *sapp päeva peal* ja *susi-ilv* (Kingisepp jt 2010: 611, 634)

Mitmeski küsimuses jagavad uusväljaande koostajad Heinrich Gösekeniga leksikograafide igikestvaid rõõme ja muresid. Nagu Gösekeni algsõnastikus on osa saksa märksõnu ilma eesti vasteta, nii tuli ka uusväljaande koostajail esialgu leppida sellega, et paarikümne märksõna tähendus jäi ähmaseks. Sel juhul on tähendusseletuse ees küsimärk, ühel juhul on seletuse asemel kirjas 'tähendus ebaselge'. Parema puudumisel on tänapäeva lugeja jaoks tõlgitud ära vähemalt Gösekeni saksa vaste, näiteks märksõna *hang* puhul (vt joonis 21).

**hang** s 'suured nupud puust ehitistel, esemetel (kirikul või reel)' die grossen knopffe oder köpffe an den Kirchen oder Schlitten **Hangk** 261

..  
**kopsumoos** s '(tähendus ebaselge)' <Lungen muß **kopse moes**> 284, 509S

**Joonis 21.** Uusväljaande sõnaartiklid *hang* ja *kopsumoos* (Kingisepp jt 2010: 338, 427).

Nüüdseks on jõutud märksõna *kopsumoos* (vt joonis 21) tähendusele jälile – selle sõnaga tähistas Göseken islandi samblikku, rahvasuus *kopsusamblikku*.

## 6. Kokkuvõte

17. sajandi Eestis anti välja kolm saksa-eesti sõnastikku. Ehkki neid on rohkesti kritiseeritud ladina ja saksa keele eeskujuks seadmise, ekslike vastete ning ebaühtlase ja võõrapärase kirjaviisi pärast, leidub neis lähemal vaatlusel palju väärtuslikku eesti kirjakeele varasema arengujärgu kohta. Sellise ainese tänapäeva keeleteaduse jaoks kergemini kasutatavaks tegemine ongi üks TÜ vana kirjakeele töörühma peamisi eesmärke. Ajaloolised leksikonid muudetakse elektrooniliseks ja neist avaldatakse järgemööda nii uusväljaandeid kui ka uurimistööd selgitavaid artikleid.

Uusväljaannete eripäraks on uudne meetod, mille puhul on kõige tähtsam sõnastiku eesti ainese esiletõstmine ning tänapäeva kasutaja jaoks analüüsimine, sealhulgas materjali ümberstruktureerimine ja selgituste lisamine. Meetodi rakendamisel kerkisid esile ajaloolise leksikograafia tüüpasjaolud, näiteks kohati katkendlik ja mitmeti mõistetav algaines, aga kindlasti ka see, et algsõnastikku ja uusväljaande sõnastikku lahutava kolme ja poole sajandi jooksul on keel ja selle kirjapanemise viis ulatuslikult muutunud. Milline kirjakuju sobib märksõnaks? Kuidas käituda tuletistega? Mil määral arvestada tähenduste seletamisel võõrvasteid, eriti kui need paistavad vahel eksitavad? Kuidas kujundada paberväljaandes sõnaartikli struktuur nii, et lugeja leiaks otsitava hõlpsasti? Nende ja paljude samalaadsete küsimustega puutuvad kokku ühe vana sõnastiku analüüsiijad.

Artiklis on avatud Heinrich Gösekeni 1660. aasta saksa-ladina-eesti sõnastiku uusväljaande koostamise tagamaid, kirjeldades tollaseid olusid nii Euroopas üldiselt kui ka Eestis, keskendudes siinsete kirikumeestest leksikograafide tegevusele. Seejärel on lähemalt selgitatud, millistel põhimõtetel on uusväljaandes eesti aines märksõnastatud ja kuidas on lähenetud tähenduse seletamise küsimusele.

Gösekeni koostatud 17. sajandi mahukaim eesti sõnastik oli oma keeruka struktuuri ja mitmekülgse ainesega hea materjal, millel rakendada varem Heinrich Stahli 1637. aasta saksa-eesti sõnastiku uusväljaandes katsetatud meetodit. Plaanis on jätkata tööd Gösekeni algsõnastiku sõnavaraallikate väljaselgitamisega ning

lisada TÜ vana kirjakeele korpusse uusväljaande saksa- ja ladinakeelsete sõnade indeks. Tulevane koondülevaade kolme esimese eesti keele sõnastiku ainese kohta võimaldab analüüsida neid nii kvantitatiivselt kui ka kvalitatiivselt.

## Viidatud kirjandus

- EKSS = Eesti keele seletav sõnaraamat I–VI. Toimetanud Margit Langemets, Mai Tiits, Tiia Valdre, Leidi Veskis, Ülle Viks, Piret Voll. Tallinn: Eesti Keele Sihtasutus, 2009.
- Gutsloff, Heinrich 1648. *Observationes grammaticę circa linguam Esthonicam. Judicia tentandi & experiendi causā in lucem editę a Johanne Gutsloff/ Pastore zu Vrbs. Dorpati Livonorum Excudebat Johannes Vogel/ Acad. Typogr. Anno M DC XLVIII.*
- Göseken, Heinrich 1660. *Manuductio ad Linguam Oesthonicam, Anführung Zur Öhstnischen Sprache / Bestehend nicht alleine in etlichen præceptis und observationibus, Sondern auch In Verdolmetschung vieler Teutschen Wörter. Der Öhstnischen Sprache Liebhabern mitgetheilet Von HENRICO GÖSEKENIO, Hannovera-Brunsvigo, Der Christlichen Gemeine zu Goldenberg in der Wysk Pastore, der umbliegenden Land Kirchen Præpositio, und des Königl. Consistorij zu Reval ordinario Assessore. Reval, Gedruckt und verleget von Adolph Simon, Gymnasij Buchdr. Anno 1660.*
- Haarmann, Harald 1976. *Die estnischen Grammatiken des 17. Jahrhunderts. I. Observationes Grammaticae circa linguam Esthonicam. Johann Gutsloff. Kommentiert und herausgegeben von Harald Haarmann (Fenno-Ugrica, Bd. 2). Hamburg: Helmut Buske Verlag.*
- Hasselblatt, Cornelius 2006. *Geschichte der estnischen Literatur: Von den Anfängen bis zur Gegenwart. Berlin: de Gruyter.*
- Kasik, Reet 2011. *Stahli mantlipärijad. Eesti keele uurimise lugu. Tartu: Tartu Ülikooli Kirjastus.*
- Kask, Arnold 1970. *Eesti kirjakeele ajaloost. Tartu.*
- Kikas, Kristel 2002. *Mida sisaldab Heinrich Stahli Vocabula? Toimetanud Valve-Liivi Kingisepp. Tartu Ülikooli eesti keele õppetooli toimetised 21. Tartu: Tartu Ülikool.*
- Kilgi, Annika; Ross, Kristiina 2006. *“Lühike sissejuhatus” ja eesti piiblitõlge. – Lühike sissejuhatus eesti keelde 1732. [Anton Thor Helle.] Saksa keelest tõlkinud ja järelsõna kirjutanud Annika Kilgi, Kristiina Ross. Eesti Keele Instituut. Tallinn: Eesti Keele Sihtasutus, C26–C53.*
- Kilgi, Annika 2012. *Tõlkekeele dünaamika piibli esmaeestinduse käigus: verbi morfosüntaksi areng ja lõplik toimetamisfaas. Tallinna Ülikooli humanitaarteaduste dissertatsioonid 27. Tallinn: Tallinna Ülikool.*
- Kingisepp, Valve-Liivi 2009. *Leide Heinrich Gösekeni sõnavaramust. – Keel ja Kirjandus, 8-9, 620–630.*
- Kingisepp, Valve-Liivi; Ress, Kristel; Tafenu, Kai 2010. *Heinrich Gösekeni grammatika ja sõnastik 350. Tartu: Tartu Ülikool.*
- Kleis, Richard; Torpats, Ülo; Gross, Lalla; Freymann, Heinrich 2002. *Ladina-eesti sõnaraamat. Teine, täiendatud ja parandatud väljaanne. Tallinn: Valgus.*
- Klinge, Matti 2008. *Läänemere maailm. Tallinn: Ilo.*
- Laanekask, Heli 2004. *Eesti kirjakeele kujunemine ja kujundamine 16.–19. sajandil. Dissertationes philologiae Estonicae Universitatis Tartuensis 14. Tartu: Tartu Ülikooli Kirjastus.*
- Laanekask, Heli 2012. *Sajandid käeulatuses: vanad grammatikad eesti keeles. – Emakeele Seltsi aastaraamat, 57 (2011), 271–281.*
- Lepajõe, Marju 1998. *Gutsloff, Johannes 1648. Observationes grammaticae circa linguam Esthonicam ... Dorpat: Johannes Vogel. [Faksiimile ja tõlge: Observationes grammaticae circa linguam Esthonicam. Grammatilisi vaatlusi eesti keelest. Johannes*

- Gutslaff. Koostanud ja tõlkinud Marju Lepajõe. Tartu Ülikooli eesti keele õppetooli toimetised 10. Tartu: Tartu Ülikool, 1998.]
- Loit, Aleksander 2006. Ühiskondlike vastuolude peegeldus Tartu-Pärnu rootsiaegse ülikooli tegevuses (1632–1710). – Kai Tafenau (Toim.). Läänemere provintside arenguperspektiivid Rootsi suurriigis 16/17. sajandil II. Eesti Ajalooarhiivi toimetised 12 (19). Tartu: Eesti Ajalooarhiiv, 123–150.
- Lotman, Piret 1995. Veel kord Heinrich Stahli Ingerimaal kirjutatud katekismusest. – Keel ja Kirjandus, 8, 541–547.
- Müller 2000 = Habicht, Külli; Kingisepp, Valve-Liivi; Pirso, Urve; Prillop, Külli 2000. Georg Mülleri jutluste sõnastik. Toimetanud Jaak Peebo. Tartu Ülikooli eesti keele õppetooli toimetised 12. Tartu: Tartu Ülikool.
- Piirimäe, Helmut 1992. Kuidas Rootsi ajal kontrolliti kirikuõpetajate eesti keele oskust. – Keel ja Kirjandus, 7, 409–412.
- Põldvee, Aivar 2010. Bengt Gottfried Forselius ja rahvahariduse lätted Eesti- ja Liivimaal. *Dissertationes historiae Universitatis Tartuensis* 20. Tartu: Tartu Ülikooli Kirjastus.
- Raag, Raimo 2002. Mis oli öieti esimese eesti keele grammatika autori nimi? – Keel ja Kirjandus, 2, 183–192.
- Ress, Kristel 2009. XVII sajandi saksa-eesti sõnastike töötlemise probleemidest. – Keel ja Kirjandus 8-9, 631–641.
- Ross, Kristiina 2003. Uue ajastu misjonilingvist Eduard Ahrens 200. Tallinn: Eesti Keele Sihtasutus.
- Rosshinius 2002 = Kingisepp, Valve-Liivi; Habicht, Külli; Prillop, Külli 2002. Joachim Rosshiniuse kirikumanaalide leksika. Tartu Ülikooli eesti keele õppetooli toimetised 22. Toimetanud Jaak Peebo. Tartu: Tartu Ülikool.
- Rätsep, Huno 1987. Heinrich Stahli keeleõpetus oma aja peeglis. – Keel ja Kirjandus, 12, 709–715.
- Saareste, Andrus 1958–1963. Eesti keele mõisteline sõnaraamat I–IV (indeks: Uppsala, 1979). Stockholm: Vaba Eesti.
- Stahl, Heinrich 1637. *Anführung zu der Esthnischen Sprach / auff Volgemeinten Rath / vnd Bittliches Ersuchen / publiciret von M. Henrico Stahlen. Revall / Druckts Chr. Reusner der älter / in Verlegung des Authoris.* [Faksiimileväljaanne: Heinrich Stahl, *Anführung zu der Esthnischen Sprach.* Brampton-Tartu: Maarjamaa, 2000.]
- Tafenau, Kai 2011a. Heinrich Gösekeni sõnaraamatu seni märkamata eeskuju. – Keel ja Kirjandus, 6, 425–439.
- Tafenau, Kai 2011b. Uue Testamendi tõlkimisest Rootsi ajal: käsikirjad, tõlkijad ja eesti kirjakeel. *Dissertationes philologiae Estonicae Universitatis Tartuensis* 27. Tartu: Tartu Ülikooli Kirjastus.
- Tuksam, Georg 1939. Saksa-eesti sõnaraamat. Käsikirja läbi töötanud ja redigeerinud E. Muuk. Tartu: Kool.
- Viitkar, Urve 2005. Johannes Gutslaffi “*Observationes grammaticae circa linguam Esthonicam*” (1648). Bakalaureusetöö Tartu Ülikooli eesti keele osakonnas. Tartu.
- Värrü Haarmann, Anna-Lisa 1977. Die estnischen Grammatiken des 17. Jahrhunderts. II. Heinrich Göseken. *Manuductio ad Linguam Oesthonicam.* *Anführung zur Öhstnischen Sprache.* Reval 1660. Herausgegeben und mit einer Einleitung versehen von A.-L. Värrü Haarmann (Fenno-Ugrica, Bd. 3). Hamburg: Helmut Buske Verlag.
- Wiedemann, Ferdinand Johann 1973 [1893]. Eesti-saksa sõnaraamat. Neljas, muutmata trükk teisest, Jakob Hurda redigeeritud väljaandest. *Estnisch-Deutsches Wörterbuch.* Vierter, unveränderter Druck nach der von Jakob Hurt redigierten Auflage. Tallinn: Valgus.
- ÕS 2006 = Eesti õigekeelsussõnaraamat ÕS 2006. Toimetanud Tiiu Erelt, koostanud Tiiu Erelt, Tiina Leemets, Sirje Mäearu, Maire Raadik. Tallinn: Eesti Keele Sihtasutus.

## Võrgumaterjalid

Grimm, Jacob; Grimm, Wilhelm 1838–1960. Das deutsche Wörterbuch von Jacob und Wilhelm Grimm auf CD-Rom und im Internet. Ein Projekt des Kompetenzzentrums für elektronische Erschließungs- und Publikationsverfahren in den Geisteswissenschaften an der Universität Trier in Verbindung mit der Berlin-Brandenburgischen Akademie der Wissenschaften Berlin. <http://germazope.uni-trier.de:8080/Projekte/DWB> (15.05.2012).

Manuductio ad Linguam Oesthonicam..., 1660. Heinrich Göseken. Manuductio ad Linguam Oesthonicam, Anführung Zur Öhstnischen Sprache, Bestehend nicht alleine in etlichen praeceptis und observationibus, Sondern auch In Verdolmetschung vieler Teutschen Wörter. Der Öhstnischen Sprache Liebhabern mitgetheilet Von HENRICO GÖSEKENIO, Hannovera-Brunsvigo, Der Christlichen Gemeine zu Goldenberg in der Wyck Pastore, der umbliegenden Land Kirchen Praeposito, und des Königl. Consistorij zu Reval ordinario Assessore. – REVAL, Gedruckt und verlegt von Adolph Simon, Gymnasij Buchdr. Anno 1660. <http://www2.kirmus.ee/grafo/index.php?gid=1> (23.08.2012).

Vana kirjakeele korpus. <http://www.murre.ut.ee/vakkur> (20.08.2012).

**Kristel Ress** (Tartu Ülikool) on uurinud 17. sajandi saksa-eesti sõnaraamatutes leiduvat eesti leksikat. [kristel.ress@ut.ee](mailto:kristel.ress@ut.ee)

# **SUPPLEMENT TO THE COMPILATION OF THE NEW EDITION (2010) OF HEINRICH GÖSEKEN'S GERMAN-LATIN-ESTONIAN GLOSSARY (1660)**

**Kristel Ress**

University of Tartu

In 1660 a 547-page book written by the Baltic-German pastor Heinrich Göseken working in Estonia, titled *Manuductio ad Linguam Oesthonicam* was published. The book consisted of a part concerning teaching the Estonian language and a German-Latin-Estonian glossary. This article gives an overview of the method developed by the author for studying the material concerning Estonian in Göseken's glossary and in other bilingual glossaries. The method is characterised by a shift in focus to the material concerning Estonian and the analysis thereof from a contemporary point of view. The list of entry words of the new edition of the German-Estonian glossary, *Heinrich Göseken's Grammar and Glossary 350* (2010), analysed in the manner described, proceeds from the alphabetical order of Estonian words. The contemporary spellings of speech forms are used as entry words. The study of Estonian material is of primary importance: the meaning of opaque entry words is explained and morphosyntactic information is added based on a contemporary description of grammar. Hence those interested will have easier access to archaic material and they can use it both to study Estonian lexis and grammar as well as, for example, to discuss the history of particular languages.

The article consists of seven sections. The introduction is followed by an overview of the characteristics of the seventeenth-century missionary linguistics on the territory of Estonia and the socio-cultural background of the dictionaries of the time. The third section describes Göseken's glossary of 1660. The fourth section is dedicated to an overview of the method and section five describes the principles underlying the selection of entry words and the explanation of their meaning and the application of these principles in the 2010 edition of Göseken's glossary. The article concludes with a summary and the list of references.

The article is based on the author's experience acquired in the drafting of the glossary section of the 2010 edition. The analysis herein is meant to be a supplement to the accompanying texts of the collection in question and the author's earlier article on the problems arising from the original glossary, the reconstruction of roots and the separation of entry word nests into single lemmas.

**Keywords:** historical lexicography, multilingual dictionaries, lexicology, lexical semantics, old literary Estonian, Estonian





# MEETODEID TEKSTIDE LEKSIKAALSETE JA GRAMMATILISTE ERINEVUSTE TUVASTAMISEKS MEDITSIINILISTE TARBETEKSTIDE NÄITEL

Raul Sirel

**Ülevaade.** Artiklis käsitletakse seni uurimata eestikeelset ressursi: ravimipakendites sisalduvaid infolehti ja arstidele suunatud ravimeid tutvustavaid kokkuvõtteid. Nimetatud ainestiku analüüsimiseks kasutatakse mõnda läbipaistvat statistilist meetodit, mis võimaldavad kerge vaevaga tuvastada tekstide tüübist ja žanrist tulenevaid leksikaalseid ja grammatilisi erinevusi. Taolise analüüsi eesmärgiks on ühest küljest katsetada nimetatud meetodite efektiivsust tekste eristavate karakteristikute leidmisel, kuid ka koguda andmestikust lähtuvaid taustateadmisi keeletehnoloogiliste rakenduste efektiivsemaks loomiseks.\*

**Võtmesõnad:** korpuslingvistika, tekstilingvistika, tekstikorpused, žanrianalüüs, keeletehnoloogia, eesti keel

## 1. Sissejuhatus

Infotehnoloogia võidukäiguga seoses on kasvanud kõikisugu dokumentide ja tekstide hulk ning kättesaadavus, mis omakorda on loonud nõudluse tehnoloogiate ja algoritmide järele, mis võimaldaksid seda ressursi efektiivsemalt töödelda ja hallata. Nii näiteks on näha jätkuvalt kasvavat huvi erinevatest tekstidest informatsiooni kogumise (ingl *information extraction*), nende automaatse liigitamise (*document classification*), automaatse kokkuvõtmise (*automatic summarisation*) jms vastu.

Käsikäes saadaval olevate tekstide hulga on suurenenud ka nende varieeruvus, kuna tekste luuakse väga erinevatel eesmärkidel ning erinevatele sihtgrupsidele. Keeletehnoloogias on paraku üsna tavapärane olukord, kus ühe keelekogu alusel loodud mudel või meetod toimib selle kogu raames suurepäraselt, kuid teisele korpusele rakendades annab meetod märksa tagasihoidlikemaid tulemusi (Kilgarriff 2001). Põhjus ongi enamasti erinevate tekstikogude keeikasutuse suures varieeruvuses, mistõttu on tarvis uurida funktsioonilt ja keelelistelt tunnustelt

\* Artikli valmimist on toetanud Euroopa Regionaalarengu Fond Eesti Arvutiteaduse Tippkeskuse ja Tarkvara Tehnoloogia Arenduskeskuse kaudu, Eesti Teadusfond (grandiprojekt 9124) ning Haridus- ja Teadusministeerium (projektid SF0180078s08 ja EKT11005). Uurimistöök vajalikud andmed on taganud Ravimiamet. Suured tänud ka anonüümsetele retsensentidele sisulise tagasiside eest.

täiesti erinevaid tekstitüüpe ja -žanre, et luua universaalsemaid või vastupidi tekstispetsiifilisemaid mudeleid ja meetodeid.

Kuigi igapäevaelus on tekstižanrid tihtipeale määratletavad juba intuitiivselt või selle vormi analüüsid, siis Fairclough'i (2003: 66) järgi võib nõnda jõuda väärjaärduseni, kuna teksti vorm ja sisu ei pea tingimata kokku langema. Seetõttu on funktsionaalses žanrianalüüsis teksti klassifitseerimisel tarvilik lähtuda eelkõige teksti sisust (Reinsalu 2011). Sisu funktsionaalne analüüs on paraku aeganõudev tegevus, mis nõuab lisaks ajale valdkondlikke teadmisi. Sel põhjusel käsitleb käesolev artikkel statistilisi meetodeid, mis võimaldavad üsna väikese vaevaga leida leksikaalsed ja grammatilised tunnused, mille suhtelised kasutussagedused on tekstiklassiti märkimisväärselt erinevad. Taoliste tunnuste analüüsimine võimaldab ühest küljest efektiivsemalt luua tekstispetsiifilisemaid mudeleid nende tekstide töötlemiseks, kuid samuti on need kasutatavad erinevate tekstitüüpide ja -žanride keelekasutuse profileerimiseks.

Tekstides sisalduvat keelekasutust on varasemalt uuritud näiteks nende formaalsuse ja kontekstuaalsuse alusel, milleks on analüüsitud tekstides esinevaid lausestruktuure (Pajupuu, Kerge 2010). Leksikaalsete ilmingute poolelt on eesti aja- ja ilukirjanduskeele erinevusi kirjeldanud Tiit Hennoste ja Kadri Muischnek (2000), kes on välja toonud märkimisväärse vahe (eriti isikuliste) pronoomenite kasutamises (ilukirjanduses selgelt enam; põhjuseks peetakse ilukirjanduse suuremat dialoogilisust).

Artiklis käsitletakse kahte seni uurimata eestikeelset ressursi, milleks on ravimipakendites sisalduvad infolehed ning arstidele suunatud ravimeid tutvustavad kokkuvõtted. Nimetatud andmetest moodustati kaks tekstikorpust, analüüsiti neid leksikaalsest ja grammatilisest aspektist ning võrreldi tulemusi Tasakaalus korpusega<sup>1</sup>, mis sisaldab võrdsel määral ilukirjanduse, ajakirjanduse ja teaduse keelt. Maailmas on ravimipakendi infolehti keeletehnoloogiliste ülesannete lahendamisel kohati rakendatud: nii näiteks on infolehtedest tekstikaeve meetoditega ekstraheeritud patsienti hoiatavaid fraase (Nabeta jt 2012) ning samuti on infolehtedest üritatud leida (Kuhn jt 2010) ning ennustada (Atias, Sharan 2011) ravimite kõrvaltoimeid. Eestikeelseid ravimipakendi infolehti ja ravimi omaduste kokkuvõtteid seni uuritud ei ole.

Artikli põhiosa on jagatud neljaks peatükiks, millest esimeses antakse ülevaade tekstide liigitamisega seotud problemaatikast, teises tutvustatakse uuritavat aines-tikku ning kolmandas käsitletakse korpuste analüüsimiseks kasutatud meetodeid. Neljas peatükk on analüüsitulemuste tutvustamiseks.

## 2. Tekstide liigitamine teorias ja praktikas

Tekstide liigitamine on mõneti keerulisem ülesanne, kui võiks tausta põhjalikumalt tundmata arvata. Seda eeskätt seetõttu, et puudub üheselt aktsepteeritav tüpologia tekstide klassifitseerimiseks ning liigitamine toimub väga erinevatel alustel: nii näiteks on võimalik tekste klassifitseerida kommunikatsioonikanali (nt suuline või kirjalik), auditooriumi ehk sihtgrupi, kasutusvaldkonna või ka suhtluseesmärgi vms järgi.

Üheks tuntumaks ning aktsepteeritumaks klassifikatsiooniks peetakse Werlich'i tüpoloogiat, mis jagab tekstid viide tüüpi: deskriptiivsed, narratiivsed, ekspositsioonilised, argumenteerivad ning instruktiivsed (Chilton, Schäffner 2002: 19). Taoline geneeriline jaotus põhineb Reet Kasiku (2007: 29) järgi inimese vajadusel kasutada keelt erinevatel eesmärkidel ning sellest lähtuvalt on tekstitüübid ka universaalsed – kultuurist ja harjumustest sõltumatud. Mõneti sarnaselt Werlich'ile on Kasik ise jaganud tekstid nende tüübi alusel kolme klassi: deskriptiivsed, narratiivsed ja argumenteerivad.

Lisaks tekstitüüpidele kasutatakse tekstide klassifitseerimisel tekstiliigi ehk žanri mõistet. Žanr on kultuurisidus keelekasutusviis, mis on aja jooksul välja kujunenud ning erinevalt tekstitüüpidest ei ole neis mingit universaalset alust, vaid need on pigem seotud harjumuste, kultuuri ning tavadega. Keelekasutusvaldkonna järgi eristatakse argikeelt, ilukirjanduskeelt ja tarbekeelt. Viimase alaliikidena räägitakse nt ajakirjanduskeelest, ametikeelest ja teaduskeelest, millest igäihel on veel hulgaliselt alaliike (Kasik 2007: 35).

Mõneti keeruliseks teeb tekstitüübi ja -žanri eristamise aga asjaolu, et ingliskeelses kirjanduses kasutatakse termineid *genre* ja *type* tihtipeale samatähenduslikuna, viidates enamasti tekstitüübile (Vardi 2000, Askehave, Swales 2001). Nii näiteks on James Robert Martin (1985) tekstitüübi (kasutades terminit *genre*) defineerinud kui keele sihipärase kasutamise mingi konkreetse eesmärgi saavutamiseks. Antud artiklis käsitletakse termineid *tekstitüüp* ja *tekstžanr* siiski sarnaselt Kasikule erinevate mõistetena.

Artiklis käsitletavat tekstid (ravimi infolehed ja omaduste kokkuvõtted) kuuluvad oma keelekasutusvaldkonna järgi tarbetekstide hulka: need on mõeldud suurele auditoriumile ning nende eesmärk on anda lugejale edasi mingi asja sisu ning teha seda võimalikult selgelt ja üheselt mõistetavalt (Kasik 2007: 43–44). Veel täpsemalt võib nimetatud tekste liigitada meditsiinivaldkonda kuuluvaks. Kuigi tekstid on käsitletavat tarbetekstidena ning kuuluvad samasse keelekasutusvaldkonda, siis tegelikkuses esineb neis ka märkimisväärseid erinevusi.

Tekstide erinevused on ühest küljest põhjendatavad erinevusega auditoriumis: neil on üsnagi erinevad sihtgrupid, kuna ravimipakendi infolehed on suunatud eelkõige ravimi tarvitajale, ravimi omaduste kokkuvõtted seevastu aga arstile. Teisest küljest ei saa kindlasti jätta tähelepanuta mõningaid erinevusi neid näiteks Werlich'i tüpoloogiasse paigutades. Nimelt kuuluvad ravimi omaduste kokkuvõtted kahtlemata deskriptiivsete tekstide hulka (neis kirjeldatakse ravimi (kõrval)toimeid, näidustusi, manustamisrežiime jne). Sama võiks esmapilgul arvata ka ravimipakendi infolehtede kohta (need sisaldavad samuti kirjeldusi kõrvalnähtude, näidustuste jms kohta), ent neis on ka teatavaid instruktiivseid jooni (juhendatakse ravimit õigesti tarbima ning probleemide ilmnemisel korrektselt käituma).

Seetõttu võiks eeldada, et ravimipakendi infolehed sisaldavad enam käskivat kõneviisi ning personaalpronoomeneid (tekstides pöördatakse ravimi kasutaja poole, kellel puudub igasugune valdkondlik kogemus). Võrreldes Tasakaalus korpus sisalduvate tekstidega võib prognoosida mõnevõrra väiksemal hulgal tingivat kõneviisi kasutamist ravimitekstides. Seda peamiselt seetõttu, et tingivat kõneviisi kasutatakse väljendamaks viisakat käsku, kõneleja (täitumatut) soovi või hinnangut

tegevusele (EKK)<sup>2</sup>. Leksikaalsest küljest võiks aga eeldada, et ravimitekstide ja Tasakaalus korpuse leksikaalne ühisosa (vähemasti substantiivide lõikes) on üsna tagasihoidlik, kuna ravimitekstid sisaldavad suurel hulgal valdkonnaspetsiifilist terminoloogiat. Sarnaselt võib ka oletada, et ravimipakendi infolehtede ja ravimi omaduste kokkuvõtete leksikaalne ühisosa on märksa suurem kui ravimitekstidel ja Tasakaalus korpusel.

Käesoleval artiklil on kaks peamist eesmärki: katsetada eestikeelsete meditsiiniliste tarbetekstide peal statistilisi meetodeid, mis võimaldaksid automaatselt tuvastada tunnuseid, mille poolest kaks tekstitüüpi või -žanri erinevad, ning nimeetatud meetodeid kasutades eelnevas lõigus esitatud hüpoteese kontrollida.

### 3. Andmestik

Uuritavaks ainekseks on koostatud kaks tekstikorpust, millest esimene sisaldab 3977 Eestis registreeritud müügiloaga ravimipakendi infolehte ning teine samade ravimite omaduste kokkuvõtteid. Olemuslikult on esimeses informatsioon ravimi tarvitajale ning teises informatsioon arstile.

Käsitlev ainekse on saadud Ravimiametist ning on kättesaadav PDF-failidena Eesti ravimiregistri kodulehelt<sup>3</sup>. Korpuste koostamisel on järgitud põhimõtet, et korpustesse kuuluksid ainult selliste ravimite tekstid, millel on kättesaadavad nii pakendi infolehed kui ravimi omaduste kokkuvõtted. PDF-failidest ekstraheeritud andmed on analüüsitavates korpustes muutmata kujul, mis tähendab, et alles on jäetud näiteks peatükkide pealkirjad, tabelites sisaldunud arvandmed jms. Samuti ei ole korpustest eemaldatud suurel hulgal esinevaid korduvaid tekstiüksuseid, kuna selline korduvus on antud uurimistöö kontekstis tõlgendatav tekste iseloomustava parameetrina.

Selleks, et leida kvantitatiivsete meetoditega kaht tekstitüüpi või -žanri eristavad parameetrid, on tarvis võrrelda neid tekste sisaldavaid korpuseid. Võttes aga näiteks kaks meelevaldset tekstikorpust, võib juba enne nende analüüsimist täie kindlusega ütelda, et need on mingil määral erinevad. Nii võib näiteks aja-lookirjandus tunduda majanduskirjandusest märksa narratiivsem, kuid võrreldes ilukirjandusega on mõlemad siiski äärmiselt mittenarratiivsed. Selleks, et näidata, kas erinevused on relevantid või mitte, kasutatakse üldjuhul multifaktoriaalset analüüsi (Biber jt 1998: 169).

Artiklis on loodud tekstikorpuseid perspektiivi saavutamiseks võrreldud ka Tasakaalus korpusega, kuna see on ammendavalt kirjeldatud ning seda võib põhjendatult pidada kirjaliku eesti keele neutraalseks baastasemeks. Tasakaalus korpus sisaldab kirjaliku keelekasutuse kolme tähtsamat tekstiklassi: ilukirjanduse, aja-lookirjanduse ja teaduse keelt. Kasutatud tekstikorpuste mahud on näidatud tabelis 1.

**Tabel 1.** Analüüsitavate tekstikorpuste maht

	Ravimipakendi infolehed	Ravimi omaduste kokkuvõtted	Tasakaalus korpus
Sõnade arv	6 562 114	11 515 652	14 823 078

<sup>2</sup> Vt <http://www.eki.ee/books/ekk09/index.php?id=198> (28.12.2012).

<sup>3</sup> Vt [http://193.40.10.165/register/register.php?keel=est&inim\\_vet=inim](http://193.40.10.165/register/register.php?keel=est&inim_vet=inim) (28.12.2012).

## 4. Analüüsimeetodid

### 4.1. Morfoloogiline märgendamine

Võrreldavad korpused on morfoloogiliselt märgendatud, kasutades eesti keele morfoloogilist analüsaatorit ja ühestajat. Morfoloogilise ühestamise käigus määratakse tekstisõna lokaalset konteksti (selle vahetus ümbruses asuvaid sõnu) analüüsides selle lemma ehk algvorm, sõnaliik ja sõnavormis kodeeritud grammatilised kategooriad (Kaalep 1997, Kaalep, Vaino 1998).

Morfoloogilise ühestaja väljundist on enne tekstikorpuste võrdlemist eemaldatud lausemärgi (nt –, /, ? jne) märgendi saanud sõned. Uurimistöös on morfoloogilist ühestajat kasutatud järgmiste grammatiliste tunnuste määramiseks:

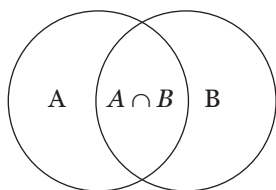
- sõna algvorm ehk lemma,
- sõnaliik,
- substantiivide kääne,
- verbide kõneviis,
- verbide aeg (olevik ja lihtminevik).

Suurest hulgast verbikategooriatest on valitud kõneviis ja aeg eeskätt sel põhjusel, et need avalduvad ühes sõnavormis (erinevalt näiteks verbi liitaegadest, mis avalduvad kahe sõnavormi ühendina), mistõttu ei ole tarvis rakendada automaatset süntaksianalüüsi.

### 4.2. Venni diagrammid leksikoni visualiseerimiseks

Venni diagrammiks nimetatakse diagrammi, millel hulgad esitatakse mingile pinnale joonistatud piirkondadena (Liikane, Kesa 2006). Tegemist on meetodiga hulkadevaheliste seoste tasapinnaliseks illustreerimiseks.

Käesolevas uurimistöös on Venni diagramme kasutatud võrreldavate korpuste leksikaalsete erinevuste illustreerimiseks. Täpsemalt on üritatud näidata, kui suur on erinevate korpuste leksikaalne ühisosa (kui leksikaalselt sarnased võrreldavad korpused üksteisele on) erinevate sõnaliikide lõikes. Korpuste leksikaalsete erinevuste visualiseerimisel ei ole kasutatud kõiki korpustes sisalduvaid sõnu, vaid ühes katses kümnet tuhandet kõige sagedamat lemmat ning teises katses kolme sõnaliigi (adjektiivid, substantiivid ja verbid) tuhandet kõige sagedamat lemmat.



**Joonis 1.** Näide Venni diagrammist, kus  $A \cap B$  on hulkade A ja B ühisosa

Diagrammide loomisel on kasutatud veebipõhist tarkvara Area-Proportional Venn Diagram Plotter and Editor<sup>4</sup>.

<sup>4</sup> Vt <http://bioinform.com/free/bxarrays/venndiagram.php> (28.12.2012).

### 4.3. Sagedusprofileerimine

Sõnaliikide ja käänete kasutussagedusi on analüüsitud sagedusprofileerimist rakendades. Nimetatud meetod annab võimaluse leida kahe tekstikorpuse suurimad erinevused, analüüsides sõnavormide või märgendite sagedusi (Rayson, Garside 2000). Nimetatud meetodit on käesolevas uurimistöös kasutatud morfoloogilise ühestaja määratud sõnaliikide ja käänete märgendite esinemissageduste analüüsimiseks.

Sagedusprofili koostamiseks on tarvis arvutada eeldatavad väärtused  $E_1$  ja  $E_2$  järgmiste valemite järgi, kus  $a$  on märgendi absoluutne sagedus esimeses korpuses,  $b$  on sama märgendi absoluutne sagedus teises korpuses,  $c$  on märgendite koguarv esimeses korpuses ning  $d$  on märgendite koguarv teises korpuses:

$$E_1 = \frac{c(a+b)}{c+d} \quad E_2 = \frac{d(a+b)}{c+d}$$

Iga märgend saab logaritmilise tõepära (inglise keeles *log-likelihood*) väärtuse, asendades arvutatud  $E_1$  ja  $E_2$  väärtused järgmisesse valemisse:

$$\text{Logaritmiline tõepära} = 2 \left( (a \times \log \frac{a}{E_1}) + b \times \log \frac{b}{E_2} \right)$$

Logaritmiline tõepära näitab antud märgendi sageduse erinevust kahes korpuses (mida suurem on logaritmiline tõepära, seda suurem on kahe tekstikorpuse vaheline erinevus antud parameetri suhtes). Logaritmilise tõepära alusel saab hinnata, kuivõrd sarnased või erinevad on kaks vaadeldavat teksti. Seda teadmist saab omakorda kasutada prognoosimaks, kui edukalt on võimalik muu tekstikogu jaoks väljatöötatud algoritmi või meetodit analüüsitava tekstile rakendada (Kilgarriff 2001) ning milliseid muudatusi tuleks algoritmi või meetodi edukamaks kasutamiseks sisse viia.

### 4.4. Suhteline sagedus

Kõneviiside ja sõnavara kasutussageduse analüüsimiseks on kasutatud suhtelist sagedust, mis annab võimaluse analüüsida mingi omaduse esinemissagedust korpuse suurust arvestades. Suhtelist sagedust on kasutatud näiteks kõneviiside ja muude verbikategoriate ning käänete ja sõnaliikide kasutuse kvantitatiivseks iseloomustamiseks vaadeldavate tekstikorpuste lõikes. Suhtelist sagedust arvutatakse järgmist valemit kasutades:

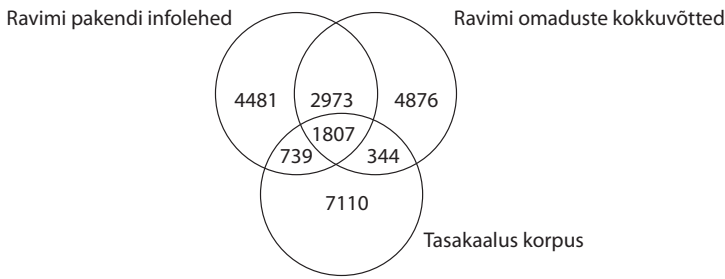
$$\text{Suhteline sagedus} = \frac{\text{sagedus}}{\text{corpuse suurus}} \times 100$$

## 5. Analüüsitulemused

### 5.1. Leksikaalsed erinevused

Võrreldud on ravimipakendite infolehtede korpust ravimi omaduste kokkuvõtete korpusega ning mõlemat eelnevat Tasakaalus korpusega. Joonisel 2 on kujutatud Venni diagrammina kõigist kolmest korpusest leitud kümne tuhande kõige sagedama lemma kattuvus kõigi sõnaliikide lõikes.

Kõigi kolme korpuse kümne tuhande sagedama lemma ühisosaks osutus 1807 lemmat ehk ligikaudu 18% iga korpuse kümnest tuhandest kõige sagedamast lemmast, kusjuures ravimipakendi infolehtede ja ravimi omaduste kokkuvõtete korpuste sagedama leksikoni ühisosa oli 47,8%. Ravimipakendi infolehtede ja ravimi omaduste kokkuvõtete leksikoni Tasakaalus korpusega võrreldes selgus, et esimese ühisosa Tasakaalus korpusega oli 25,5% ning teise oma 21,5%, millest võib järeldada, et patsiendile suunatud tekstid on leksikaalselt mõneti tavakeelele sarnasemad kui arstile suunatud tekstid.



Joonis 2. Venni diagramm 10000 kõige sagedama lemmaga kolmes korpuses

Tabelis 2 on esitatud kõigis kolmes korpuses leiduvate unikaalsete lemmade arv kolme sõnaliigi lõikes, kusjuures lemmade arv on antud nii absoluutarvu kui suhtelise sagedusena. Lemmade koguarvu teades saab teha üldistusi, kui võrd leksikaalselt piiratud analüüsitava korpust, mida omakorda saab ära kasutada algoritmide ja meetodite kohandamiseks analüüsitava tekstiühikule. Nii näiteks võib eeldada, et oluliselt piiratumale leksikoniga tekstides kasutatakse piiratumat terminoloogiat, mis tähendab, et sünonüümide hulk on piiratud.

Tabel 2. Korpustes sisalduvate unikaalsete lemmade hulk kolme sõnaliigi lõikes

Sõnaliik	Ravimipakendi infolehed	Ravimi omaduste kokkuvõtted	Tasakaalus korpus
Adjektiivid	6 460 0,098%	10 339 0,090%	46 969 0,317%
Substantiivid	28 311 0,43%	42 770 0,371%	311 678 2,103%
Verbid	1 734 0,026%	1 920 0,017%	11 178 0,075%

Tabelis 3 on näidatud unikaalsete lemmade hulk, kusjuures välja on jätud lemmad, mis on vaadeldavates korpustes esinenud kõigest üks kord. On selgelt näha, et Tasakaalus korpuses esines märksa enam lemmasid, mida oli kasutatud vaid

üks kord. See tähendab, et selles korpuses sisalduv keel on leksikaalselt märksa mitmekesisem kui ravimipakendite infolehtede ja ravimi omaduste kokkuvõtete korpustes sisalduvad. Samuti peegeldub tulemustes tõsiasi, et loodud korpused sisaldavad oma formaadi tõttu märkimisväärsel hulgal korduvaid tekstiüksuseid.

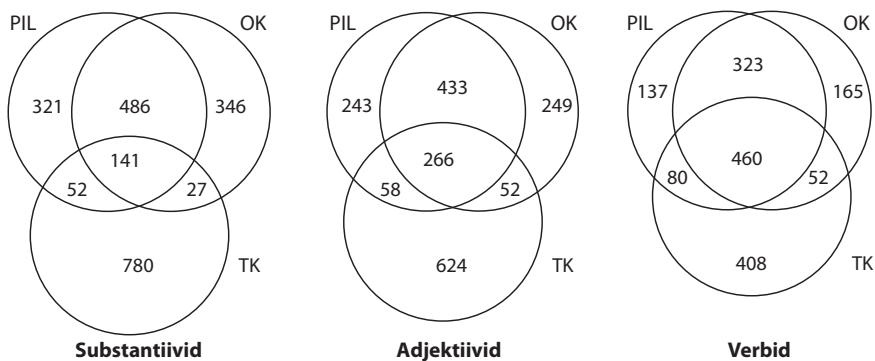
**Tabel 3.** Korpustes enam kui kord sisaldunud unikaalsete lemmade arv kolme sõnaliigi lõikes

Sõnaliik	Ravimipakendi infolehed		Ravimi omaduste kokkuvõtted		Tasakaalus korpus	
Adjektiivid	4 547	0,069%	7 670	0,067%	20 558	0,139%
Substantiivid	19 251	0,293%	30 407	0,264%	120 017	0,810%
Verbid	1 324	0,020%	1 539	0,013%	5 805	0,039%

Tabelites 2 ja 3 näidatud andmete suurusjärke hinnates on selgesti näha, et Tasakaalus korpuses sisalduv keel on oluliselt rikkalikum kui teistes vaadeldavates korpustes. Selle ning joonise 3 alusel saab järeldada, et ravimipakendi infolehtede ja ravimi omaduste kokkuvõtete korpustes sisalduvad tekstid on kõik küllaltki sarnase leksikoni ning ühtlase terminoloogiaga. Keeletehnoloogia seisukohalt on see teadmine oluline, kuna hõlbustab oluliselt näiteks informatsiooni ekstraheerimist tekstidest.

Loomulikult ei saa tabelite 2 ja 3 puhul teha lõplikke järeldusi lemmade täpse hulga kohta korpustes, kuna morfoloogilise märgendamise käigus on kahtlemata sisse sattunud vigu, mille ulatust on raske prognoosida.

Leksikoni sarnasuse hindamiseks sõnaliikide lõikes on joonisel 3 esitatud 1000 kõige sagedama lemmaga adjektiivide, substantiivide ja verbide Venni diagrammid. Kolme sõnaliigi lõikes osutus kolme korpuse tuhande kõige sagedama lemma ühisosa suurimaks verbidel (46%) ja adjektiividel (26,6%), väikseimaks aga substantiividel (14,1%). Kolme korpuse substantiivide küllaltki väike ühisosa on ilmselt põhjendatav küllaltki ravimispetsiifilise terminoloogiaga (toimeained, kõrvaltoimed, kaebused jms). Sellega on põhjendatav ka pakendi infolehtede ja ravimi omaduste kokkuvõtete üsna suur ühisosa substantiivide (62,7%) ja adjektiivide (69,9%) lõikes. Pakendi infolehtede ja ravimi omaduste kokkuvõtete tuhande kõige sagedama lemma ühisosa on 78,3%.



**Joonis 3.** Venni diagrammid 1000 kõige sagedama lemmaga ravimipakendi infolehtede korpuses (PIL), ravimi omaduste kokkuvõtete korpuses (OK) ja Tasakaalus korpuses (TK) kolme sõnaliigi lõikes



Kõigi kolme sõnaliigi puhul on sarnaselt kümne tuhande lemma katsega näha, et pakendi infolehed on leksikaalselt tavakeelele mõneti sarnasemad kui ravimi omaduste kokkuvõtted. Tekstikorpuste analüüsitulemused näitavad, et kuigi substantiivide ja adjektiivide ühisosa kolmes korpuses on üsna väike, siis märkimisväärselt suurem ühisosa pakendi infolehtede ja ravimi omaduste kokkuvõtete korpuses ning lemmade väike hulk nimetatud korpustes annavad üsna head eeldused edukaks tekstikaevaks.

## 5.2. Sõnaliigi- ja käändekasutuse sagedusprofiilid

Käände- ja sõnaliigikasutuse kirjeldamiseks on kasutatud peatükis 2.3 kirjeldatud sagedusprofileerimist. Joonistel 4 ja 5 kujutatud diagrammid on tõlgendatavad järgmiselt: vertikaalteljel on kujutatud logaritmiline tõepära, mis näitab kahe korpuse erinevust horisontaalteljel olevate atribuutide suhtes (mida suurem on logaritmiline tõepära, seda suurem on erinevus atribuudi esinemissageduste vahel kahes korpuses). Kuna diagrammidelt ei ole nähtav, kummas võrreldavas korpuses atribuut rohkem esindatud on, kasutatakse mõnel juhul erinevuse kirjeldamiseks suhtelist sagedust.

Joonisel 5 kujutatud käänete kasutussageduse erinevusest näeb selgesti, et enamiku atribuutide puhul on ravimipakendi infolehed ja Tasakaalus korpus üksteisele küllaltki sarnased (vastavad tulbad on diagrammil madalad). Joonisel 4 tulevad suuremad erinevused esile vaid kardinaalnumeraalide ja lühendite osas, kuna pakendi infolehed sisaldavad tarvitajale informatsiooni ravimite tarvitamisrežiimi kohta, mis on tihti peale esitatud numbreid ja lühendeid kasutades (näited 1, 2, 3).

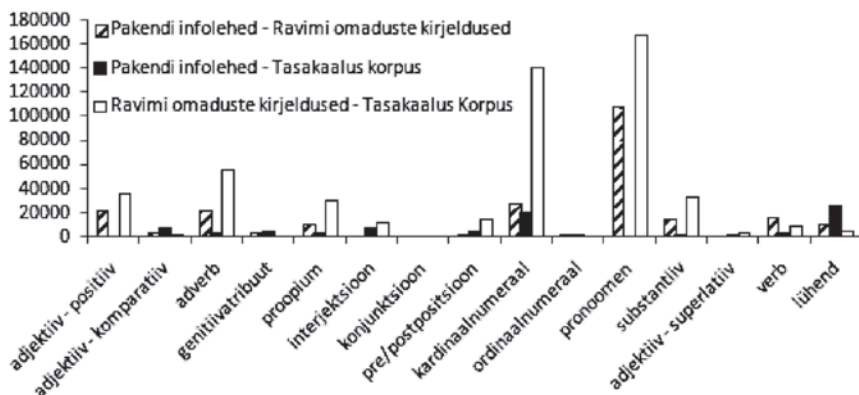
- (1) 1 tbl. päevas.
- (2) Kate: makrogool 400, makrogool 6000.
- (3) Rytmonorm, 300 mg õhukese polümeerikattega tablettidel on märgistus "300".

Oluline on ka täheldada suurt erinevust pronoomenite kasutuses, kusjuures ravimipakendi infolehtedel on väga suurel hulgal personaalpronoomeneid, kuna tekstides pöörduetakse ravimi tarvitaja poole (näited 4 ja 5):

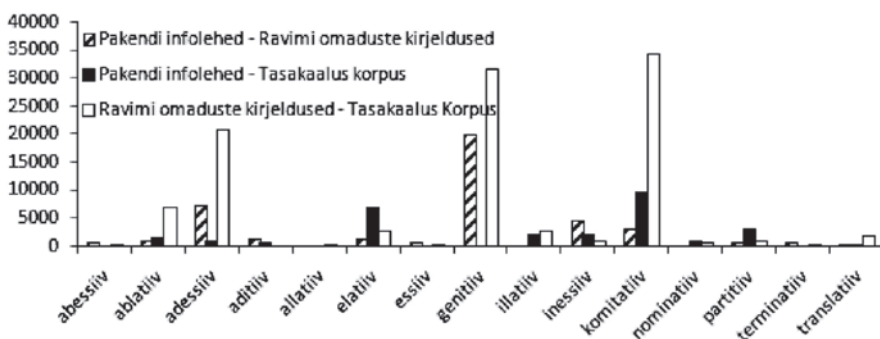
- (4) .. kui teie veres on madal kaaliumisisaldus.
- (5) Kui arst teile öelnud, et teil on teatud suhkrute talumatus ..

Võrreldes Tasakaalus korpuse ja ravimipakendite infolehtedega on suur pronoomenite kasutamise erinevus ka ravimi omaduste kokkuvõtete korpuses, kus need on tugevasti alaesindatud. Ravimi omaduste kokkuvõtete korpuses on seevastu suuresti ülesindatud kardinaalnumeraalid, kuna need tekstid sisaldavad hulganisti numbrilist informatsiooni ravimite toime ja manustamise kohta. Samuti on nii pakendi infolehtede kui ravimi omaduste kokkuvõtete korpuses tugevasti ülesindatud pärisnimed, kuna mõlemad sisaldavad suurel määral viiteid ravimi nimele, millest parasjagu räägitakse (näited 6 ja 7).

- (6) Äрге võtke ravimit Rispolept ..
- (7) AVAXIM 160 U kasutatakse ETTEVAATUSEGA.



**Joonis 4.** Korpuste sagedusprofiilide võrdlus (vertikaalteljel logaritmiline tõepära, horisontaalteljel sõnaliigid)



**Joonis 5.** Korpuste sagedusprofiilide võrdlus (vertikaalteljel logaritmiline tõepära, horisontaalteljel käänded)

Erinevused on ka näiteks interjektsiooni kasutuses, kuna pakendi infolehed ja ravimi omaduste kokkuvõtted neid ei sisalda. Oluline kvantitatiivne erinevus tuleb eri tekstide lõikes sisse ka adjectiivide kasutuses, kus Tasakaalus korpuses ja pakendi infolehtede korpuses olid suhtelised sagedused vastavalt 8,75% ja 8,73%, kuid ravimi omaduste kokkuvõtetes oli suhteline sagedus oluliselt väiksem: 3,31%.

Sõnaliigikasutuse sagedusanalüüs kinnitab leksikoni sagedusanalüüsist selgunut, et ravimipakendi infolehtedel sisalduv keel on Tasakaalus korpuses esindatud keelele oluliselt sarnasem kui ravimi omaduste kokkuvõtetes sisalduv. Samuti on huvitav tõdeda, et arstile ja patsiendile suunatud ravimitekstid on seniste tulemuste alusel selgesti statistiliselt eristatavad (näiteks on suured erinevused pronomnenite, pärisnimede, adverbide ja adjectiivide hulgas).

Käändekasutust kolmes võrreldavas tekstikorpuses on graafiliselt kujutatud joonisel 5, kust on näha, et suurimad erinevused tulevad esile taas Tasakaalus korpuse ja ravimi omaduste kokkuvõtetete võrdluses, kus kõige enam on ülesindatud komitatiivi kasutus (suhteline sagedus 2,81%). Võrdluseks on Tasakaalus korpuses komitatiivi suhteline sagedus kõigest 1,25% ning sama väärtus pakendi infolehtede korpuses 2,09%. Komitatiivi on ravimi omaduste kokkuvõtetes kasutatud tihtipeale struktuurides nagu näidetes (8), (9) ja (10).

- (8) Madopar 200 mg/50 mg ristuva poolitusvaoga tablett.
- (9) Lisainformatsiooni saamiseks pidage nõu oma apteekriga.
- (10) .. leevendab kudede pinget ja sellega kaasnevat valu.

Suuremad erinevused tulevad esile ka genitiivi ja adessiivi kasutuses, kusjuures suurimad lahknevused on täheldatavad taas ravimi omaduste kokkuvõtete ja Tasakaalus korpuse võrdluses ning mõlema parameetri järgi on pakendi infolehed oluliselt sarnasemad Tasakaalus korpusega kui ravimi omaduste kokkuvõteted.

Sarnaselt leksikoni ja sõnaliikide sagedusanalüüsile on ka käänete kasutussagedustest näha, et pakendi infolehed sarnanevad Tasakaalus korpuses sisalduvale keelele enam kui ravimi omaduste kokkuvõtetes sisalduv keel. Huvitavateks parameetriteks osutusid komitatiivi üleesindatus ravimi omaduste kokkuvõtetes ning elatiivi mõnetine alaesindatus pakendi infolehtede korpuses.

### 5.3. Verbikategooriate kasutus

Mõnede verbikategooriate suhtelist kasutussagedust kolmes tekstikorpuses iseloomustab tabel 4, kus esimeses veerus on kategooria, teises kuni kaks näidet antud kategooriast, kolmandas kategooria suhteline esinemissagedus pakendi infolehtede korpuses, neljandas kategooria esinemine ravimi omaduste kokkuvõtete korpuses ning viiendas kategooria esinemine Tasakaalus korpuses.

**Tabel 4.** Lihtaegade, käskiva ja tingiva kõneviisi esinemine analüüsitavates korpustes

Verbikategooria	Näide	Pakendi infolehed	Ravimi omaduste kokkuvõteted	Tasakaalus korpused
Olevik	<i>loeb, loeme</i>	10,99	8,70	9,49
Lihtminevik	<i>luges, lugesime</i>	1,03	2,31	5,48
Käskiva kõneviisi oleviku aktiivi jaatav kõne	<i>lugege</i>	1,73	0,04	0,03
Käskiva kõneviisi oleviku 2. isiku mitmuse aktiivi eitav kõne	<i>ärge</i>	0,350	0,004	0,003
Tingiva kõneviisi oleviku aktiivi jaatav kõne	<i>loeks, loeksid</i>	0,13	0,16	0,34

Tabelis 4 antud tulemustest võib olulise erinevusena välja tuua käskiva kõneviisi mitmuse aktiivi jaatava ning eitava kõne suurt üleesindatust pakendi infolehtede korpuses võrreldes teiste korpustega. See on põhjendatav pakendi infolehtedes sisalduvate pöördumistega ravimi tarvitaja poole, kus soovitatakse või keelatakse mõne toimingu läbiviimist (näited 11, 12 ja 13).

- (11) Enne ravimi võtmist lugege hoolikalt infolehte.
- (12) Ärge kasutage Madopar'i raseduse ajal.
- (13) Kõrvaltoimete ilmnemisel konsulteerige arstiga.

Seevastu esineb ravimipakendi infolehtede ja ravimi omaduste kokkuvõtete korpustes märkimisväärselt vähem tingiva kõneviisi aktiivi jaatavat kõnet (vastavalt

0,13% ja 0,16%) kui Tasakaalus korpuses (0,344%). Aegade kasutusest võib välja tuua lihtmineviku vähese esindatuse pakendi infolehtede ja ravimi omaduste kokkuvõtete korpustes (vastavalt 1,03% ja 2,31%) võrreldes Tasakaalus korpusega (5,48%). Pakendi infolehtede puhul on see põhjendatav käskiva kõneviisi ulatusliku kasutusega.

Verbikategooriate suurima erinevusena saabki ilmselt välja tuua n-ö käskimise (või soovitamise) ja keelamise küllusliku esinemise pakendi infolehtedes, mis paistabki olevat selle tekstiliigi üheks olulisimaks karakteristikuks. Selline teadmine on oluline näiteks tekstidest kokkuvõtete tegemisel, kuna käsud-keelud on infolehtedes oluline informatsioon ja sisaldab juhtnööre ravimi doseerimiseks, tüsistuste vältimiseks jms.

## 6. Kokkuvõte

Artiklis on käsitletud mõnda läbinähtavat statistilist meetodit leksikaalsete ja grammatiliste erinevuste lihtsamaks tuvastamiseks tekstides sisalduvat keelekasutust analüüsid. Meetodeid on rakendatud ravimipakendi infolehtedest ja ravimi omaduste kokkuvõtetest loodud korpuste statistiliseks analüüsimiseks ja võrdlemiseks.

Loodud korpuseid analüüsid selgus, et ravimipakendi infolehed ja ravimi omaduste kokkuvõtted on võrreldes Tasakaalus korpuses sisalduvate tekstidega küllaltki piiratud leksikoniga, kusjuures pakendi infolehed sarnanevad Tasakaalus korpuses sisalduvale keelele mõneti enam kui ravimi omaduste kokkuvõtteid. Samalaadne sarnasus tuli ilmsiks ka käände- ja sõnaliigikasutust analüüsid. Seevastu verbikategooriate kasutuses osutus oluliseks käskiva kõneviisi ning personaalpronoomenite laialdane esindatus ravimipakendi infolehtedes, mis tähendab, et patsientidele suunatakse oluliselt enam soovitusi-manitsusi kui arstidele.

Kõneviisi- ja sõnaliigikasutuse analüüs kinnitab eelnevalt püstitatud hüpoteese seoses teatavate erinevustega ravimipakendi infolehtede ja ravimi omaduste kokkuvõtete keelekasutuses. See tähendab, et kuigi mõlemad tekstid on žanrilt tarbetekstid ja kuuluvad samasse keelekasutusvaldkonda, siis tüübilt on tegemist siiski mõneti erinevate tekstidega, kuna nende auditorium ja funktsioon on erinevad.

Kahtlemata on käsitletud tekstide näol tegemist andmestikuga, mis väärrib ka edaspidist uurimist ning rakendamist. Viimast eriti tekstikaeve vallas, mis võimaldaks näiteks luua meditsiinitöötajatele kasutamiseks andmebaase ravimitest, nende kõrvaltoimetest, manustamisrežiimidest jms. Sellistest andmebaasidest oleks kliiniliste andmete kättesaamine oluliselt lihtsam ja efektiivsem kui infolehti ja omaduste kokkuvõtteid lugedes.

## Viidatud kirjandus

- Atias, Nir; Sharan, Roded 2011. An algorithmic framework for predicting side-effects of drugs. – *Journal of Computational Biology*, 18, 207–218.
- Askehave, Inger; Swales, John M. 2001. Genre identification and communicative purpose: A problem and a possible solution. – *Applied Linguistics*, 22 (2), 195–212.
- Biber, Douglas; Conrad, Susan; Reppen, Randi 1998. *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge: Cambridge University Press.

- Chilton, Paul; Schäffner, Christina 2002. *Politics as Text and Talk: Analytic Approaches to Political Discourse*. Amsterdam: John Benjamins Publishing Co.
- EKK = Erelt, Mati; Erelt, Tiit; Ross, Kristiina 2007. *Eesti keele käsiraamat*. 3., täiendatud trükk. Tallinn: Eesti Keele Sihtasutus.
- Fairclough, Norman 2003. *Analysing Discourse: Textual Analysis for Social Research*. London, New York: Routledge.
- Hennoste, Tiit 2002. Suulise kõne uurimine ja sõnaliigi probleemid. – Renate Pajusalu, Ilona Trigel, Tiit Hennoste, Haldur Õim (Toim.). *Teoreetiline keeleteadus Eestis*. Tartu Ülikooli üldkeeleteaduse õppetooli toimetised 4. Tartu: Tartu Ülikooli Kirjastus, 56–73.
- Hennoste, Tiit; Muischnek, Kadri 2000. Eesti kirjakeele korpuse tekstide valiku ja märgendamise põhimõtted ning kahe allkeele võrdluse katse. – Tiit Hennoste (Toim.). *Arvutuslingvistikalt inimesele*. Tartu Ülikooli üldkeeleteaduse õppetooli toimetised 1. Tartu: Tartu Ülikooli Kirjastus, 183–317.
- 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://www.cl.ut.ee/yllitised/chum1997.pdf> (28.12.2012). <http://dx.doi.org/10.1023/A:1000668108369>
- Kaalep, Heiki-Jaan; Vaino, Tarmo 1998. Vale meetodiga õiged tulemused? Eesti keele morfoloogiline ühestamine statistika abil. [http://www.cl.ut.ee/yllitised/kk\\_yhest\\_1998.pdf](http://www.cl.ut.ee/yllitised/kk_yhest_1998.pdf) (28.12.2012).
- Kasik, Reet 2007. *Sissejuhatus tekstiõpetusse*. Tartu: Tartu Ülikooli Kirjastus.
- Kilgariff, Adam 2001. Comparing Corpora. – *International Journal of Corpus Linguistics*, 6 (1), 1–37.
- Kuhn, Michael; Campillos, Monica; Letunic, Ivica; Jensen, Lars J.; Bork, Peer 2010. A side effect resource to capture phenotypic effects of drugs. – *Molecular Systems Biology*, 6, artikkel nr 343. <http://dx.doi.org/10.1038/msb.2009.98>
- Liikane, Lauri; Kesa, Marilin 2006. *Arvutisõnastik*. Elektrooniline versioon. [www.keeleeveeb.ee](http://www.keeleeveeb.ee) (28.12.2012).
- Martin, James R. 1985. Process and Text: Two aspects of human semiosis. – J. D. Benson, W. S. Greaves (Eds.). *Systemic Perspectives on Discourse*, Vol 1: Selected Theoretical Papers from the 9th Int. Systemic Workshop. Norwood, N.J.: Ablex, 248–274.
- Nabeta, Keita; Kimura, Masaomi; Ohkura, Michiko; Tsuchiya, Fumito 2012. Analysis on descriptions of precautionary statements in package inserts of medicines. – V. G. Duffy (Ed.). *Advances in Human Factors and Ergonomics in Healthcare*. USA: CRC Press, 257–266.
- Pajupuu, Hille; Kerge, Krista 2010. Characteristics and assessment of educated L1 and L2 dialogue. – Rosario Caballero Rodrigues, Jesus Pinar Sanz (Eds.). *Modos y formas de la comunicacion humana / Ways and Modes of Human Communication*. Cuenca: Universidad Castilla-La Mancha, 339–348.
- Rayson, Paul; Garside, Roger 2000. Comparing corpora using frequency profiling. – WCC '00 Proceedings of the workshop on Comparing corpora, 9, 1–6. <http://dx.doi.org/10.3115/1117729.1117730>
- Reinsalu, Riina 2011. Leping tekstiliigina: žanrstruktuur. – *Eesti Rakenduslingvistika Ühingu aastaraamat*, 215–229.
- Vardi, Iris 2000. Developing critical writers at the undergraduate level: some insights from critical thinking pedagogy and linguistics. – *Cornerstones of higher education*. Selected papers from the 1999 HERDSA Annual International Conference, Melbourne, Australia, 12-15 July 1999.

**Raul Sirel** (Tartu Ülikool) tegeleb peamiselt meditsiinilise tekstikaave, korpuslingvistika ning dialoogsüsteemidega.  
[rsirel@ut.ee](mailto:rsirel@ut.ee)

# METHODS FOR IDENTIFYING LEXICAL AND GRAMMATICAL DIFFERENCES IN MEDICAL APPLIED TEXTS

**Raul Sirel**

University of Tartu

This paper introduces some transparent statistical methods for identifying characteristics distinctive for patient information and specification leaflets for human medicines. Though the patient information leaflets and specifications for human medicines have been published by the Estonian State Agency of Medicines and been digitally available for some time, they have not been linguistically analysed nor used in the development of language technology applications.

It has been generally accepted that improving the quality of language technology applications often requires genre-specific approaches, for it is common that a model trained on one genre does not produce equally good results when applied to some other genre.

It is the aim of the present paper to identify the linguistic features that differentiate the patient information leaflets and specifications for human medicines from each other and from language represented in the Balanced Corpus of Estonian. In order to achieve that, two text corpora containing the texts from 3977 patient information leaflets and 3977 specifications for human medicines have been created and statistically compared with each other and the Balanced Corpus of Estonian.

The comparison of the corpora revealed that patient information leaflets and specifications for human medicines contain relatively limited lexicon compared with the Balanced Corpus. This knowledge is relevant, because confined lexicons tend to facilitate the tasks of information mining, automatic summarisation, etc. Furthermore, it appeared that the language in patient information leaflets was somewhat similar (compared to the language in specification leaflets) to the language represented in the Balanced Corpus.

Indubitably the collected corpora of patient information leaflets and specifications for human medicines are valuable resources and should be subjects for further research.

**Keywords:** corpus linguistics, text linguistics, text corpora, genre analysis, language technology

# EESTI JA SOOME SÕNALUGEMISTESTIDE VÕRDLU

Piret Soodla, Maigi Vija, Renate Pajusalu

**Ülevaade.** Artiklis analüüsitakse kahte lugemistesti – sõnaloendi lugemise testi ja sõnade äratundmise testi – võrdlevalt eesti ja soome keeles. Uurimusel on kaks eesmärki: hinnata nimetatud testide eesti- ja soomekeelsete versioonide omavahelist vastavust ja arendada lugemistestide võrdlemise meetodeid. Testide võrdlemisel rakendatakse paralleelselt testide keelelist analüüsi, täiskasvanud lugejate testitulemuste ning laste testisoorituste ja intervjuude analüüsi. Tulemused osutavad nimetatud testide suurele sarnasusele. Teste ei saa siiski pidada identseteks eesti ja soome keeles, seda osalt mõningaste erinevuste tõttu testides endas ning osalt keeleliste iseärasuste tõttu.\*

**Võtmesõnad:** testimine, hindamine, lugemisoskus, eesti keel, soome keel

## 1. Sissejuhatus

Lugemaõppimine on üks olulisematest ülesannetest esimestel kooliaastatel. Lugemise peamine eesmärk on loetu mõistmine, mis võimaldab saada teavet, kogeda teisi (kujuteldavaid) maailmu, edukalt suhelda ning õppida. Lugemisraskused on üheks sagedamaks õpiraskuste tekkepõhjuseks – peaaegu kõik õppeained koolis nõuavad teatud lugemisoskuse taset. Kui lapsel esineb eakohase teksti mõistmisel raskusi, võivad selle põhjused olla erinevad. Sobiva õpiabi pakkumiseks tuleb esmalt välja selgitada mõistmisraskuse põhjus. Laste lugemisoskuse jälgimisel ja hindamisel on olulised õpetaja argitähelepanekud, kuid lisaks soovitatakse kasutada eri tüüpi lugemisteste. Praegusel ajal napib Eestis usaldusväärseid vahendeid õpilaste vaimsete võimete, akadeemiliste ja sotsiaalsete oskuste hindamiseks – kasutusel on neid küll rohkesti, kuid valdavalt on tegemist üliõpilastööde raames koostatud või teistest keeltest tõlgitud testidega, mille sobivust ei ole kontrollitud esinduslike valimitega (Kikas 2006). Sarnane on Eesti hetkeolukord ka lugemisoskuse hindamise vahendite vallas, samal ajal kui näiteks Soomes on laste lugemisoskuse hindamiseks

\* Artikkel on valminud Euroopa Sotsiaalfondi programmi Eduko alamprojekti nr. 30.2-10.2/1247, Eesti Teadusfondi grantide nr 9101 ja sihtfinantseeritava teema SF 0180056s08 toel.

kasutusel mitmeid standardiseeritud testipakette (vt Lerkkanen 2007: 150–155). Laste oskuste hindamisvahendite vajadusele on viidanud ka Eesti teadusuuringud (Soodla 2012, Soodla, Kikas 2010).

Viimastel aastatel on mitme uurimisprojekti<sup>1</sup> raames koostatud ja kasutatud eesti algklassiõpilaste lugemisoskuse hindamise vahendeid, millest mõne puhul on aluseks võetud Soomes kasutatavad materjalid. Siinses artiklis analüüsitakse kahe lugemistesti – sõnaloendi lugemise testi ja sõnade äratundmise testi – eestikeelset ja soomekeelset varianti. Testide võimalikult suur sisuline ja keeleline sarnasus võimaldaks neid kasutada analoogsete hindamisvahenditena keeltevahelistes uuringutes, mille osatähtsus on viimastel aastatel haridusvaldkonnas oluliselt tõusnud. Ideaalis peaksid võrdlusuuringutes kasutatavad hindamisvahendid olema statistiliselt identsed keskmise tulemuse, standardhälbe ning vigade struktuuri poolest. Praktikast on testide statistilist identsust saavutada siiski ülimalt keerukas, kui mitte võimatu (Arffman 2010).

Võrreldavad andmed eesti ja soome laste lugemisoskuse arengu kohta võimaldavad teha järeldusi lugemaõppimise ja -õpetamise kohta suhteliselt reeglipärase ortograafiaga keeltes. Siinse uurimuse kitsam eesmärk ongi hinnata sõnaloendi lugemise testi ja sõnade äratundmise testi eesti- ja soomekeelsete versioonide omavahelist vastavust, rakendades meetoditriangulatsiooni<sup>2</sup>: testide keelelist analüüsi, täiskasvanute testitulemuste ning laste testisoorituste ja intervjuude analüüsi. Laiem eesmärk on aga aidata kaasa üldise lugemistestide pädeva võrdlemise meetodi väljatöötamisele.

## 2. Lugemaõppimine eri keeltes

Kirjaliku teksti mõistmine on keerukas kognitiivne toiming, mis hõlmab palju erinevaid osaoskusi ja protsesse. Loetu mõistmine on kahe oskuste rühma – sõnade lugemise ja kuuldu mõistmise – tulemus (Hoover, Gough 1990). Sõna lugemine hõlmab dekodeerimist ehk tähtede ja nende kombinatsioonide “tõlkimist” häälduseks ning sõna äratundmist ehk kirjutatud sõna kokkuviiimist selle tähendusega. Kuuldu mõistmise all peetakse silmas protsesse, mis tagavad sõnumi mõistmise lause- ja tekstitasandil. Siinses uurimuses analüüsime sõnade lugemisoskuse hindamisvahendeid.

Sõnade lugemise oskus – lugemise õigsus ja kiirus – on loetu mõistmisega oluliselt seotud: korrelatsioonid eri uuringutes on mõõdukad kuni tugevad. Lugemistehniliste oskuste paranedes väheneb dekodeerimisoskuse mõju teksti mõistmisele ning üha olulisemaks muutuvad sõnavara ja nn kõrgema tasandi oskused – järeldamine, integreerimine, metalingvistilised oskused ja teksti struktuuri tundmine; vabanevad ka ressursid, mida kasutatakse kõrgema tasandi mõistmisprotsesside rakendamiseks. (Perfetti 1994: 878, Oakhill, Cain 2007, Priya, Wagner 2009). Siiski näitab ka mõõdukas seos sõnade lugemise ja loetu mõistmise vahel, et suurel osal tehniliste lugemiskustega lugejatest on probleeme ka kirjaliku teksti mõistmisega.

Sõnade lugemisele avaldab otsest mõju lugemismaterjali keerukus – sõna sagedus keeles ja sõna pikkus. Keeles harva esinevate sõnade lugemisel peatub pilk

<sup>1</sup> Eesti Teadusfondi grantid 7388 ja 9101, Eesti Sotsiaalfondi programmi Eduko projekt nr 30.2-10.2/1247.

<sup>2</sup> Sotsiaalteadustes määratletakse triangulatsiooni kui “kahe või enama meetodi kasutamist inimkäitumist puudutavate andmete kogumiseks” (Cohen jt 2001: 112). Ühe nähtuse uurimisel lähenetakse sellele erinevatelt teoreetilistelt seisukohtadelt, kasutades kvalitatiivseid ja/või kvantitatiivseid meetodeid. See võimaldab hõlmata nähtust mitmekülgsemalt ja täielikumalt ning tagab ka suurema valliidsuse.



sõnaosadel kauem ja rohkem kordi ning ka lugemisvigade (sõnaosade või sõnade asendamine, moonutamine, lugemata jätmise, lisamine) hulk on oluliselt suurem kui tuttavate sõnade lugemisel. Sarnaseid tulemusi on saadud tähendusega sõnade ja tähenduseta sõnade lugemise uurimisel: viimaste lugemine on vigaderohkem ja aeglasem. Lugemise tulemust mõjutab ka sõna pikkus – mida pikem on sõna, seda rohkem lugeja pilk selle eri osadel peatub ning seda kauem võtab sõna äratundmine aega. (Hyönä, Olson 1995, De Luca jt 2002, Hutzler, Wimmer 2004, Dürrwächter jt 2010). Sõnade sageduse ja pikkuse mõju lugemise õigsusele ja kiirusele on täheldatud eri vanuses lugejate ning nii eakohase kui ka puuduliku lugemisoskuse puhul, olles eriti tugev just viimasel juhul.

Lugemaõppimise ja -õpetamise juures on tähtis roll ka vastava keele kirjaviisil. Alfabeetilise kirjasüsteemiga keelte võrdlemisel on leitud, et lugemisoskuse kujunemist ja omandamise kiirust mõjutab oluliselt keele ortograafia reeglipärasus ehk tähe-hääliku vastavus. Reeglipärase ortograafiaga keeltes (nt soome, saksa) toimub lugemaõppimine oluliselt kiiremini kui ebareeglipärase kirjasüsteemiga keeltes (nt inglise, taani, prantsuse) (Aro, Wimmer 2003, Seymour jt 2003). Reeglipärase ortograafiaga keeltes on lugemisvigade periood suhteliselt lühike ning lugemiskeskused seostuvad eelkõige lugemise kiiruse, mitte niivõrd õigsusega (Aro 2004: 21, Landerl, Wimmer 2008). Teise võimaliku keelespetsiifilise mõjutegurina on Philip Seymour jt (2003) välja toonud keele fonoloogia, täpsemalt silbistruktuuri keerukuse: keeltes, kus on palju lahtisi silpe (nt itaalia, hispaania), on lapsed (mitte-sõnade) lugemises edukamad kui keeltes, mille sõnadele on iseloomulik kinniste silpide ja konsonantühendite rohkus (nt saksa, taani, inglise).

### 3. Eesti ja soome keele ortograafia ning silbistruktuuri põhijooned

Eesti ja soome keel on lähisugulaskeeled ja nende ortograafia on põhimõtteliselt sarnane: mõlema keele kiri on alfabeetiline ja järgib võrreldes paljude teiste keeltega suhteliselt hästi printsiipi üks häälik – üks täht. Eesti kirjakeeles on 23 tähte ja 24 foneemi, soome keeles 21 tähte ja 21 foneemi<sup>3</sup> (eesti keele kohta vt EKK 2007: 49–54, Hint 2004: 74, soome keele kohta vt Aro 2004: 14, Hakulinen jt 2004: 37–42). Mõlemas keeles on siiski ka kõrvalekaldumisi printsiibist üks häälik – üks täht. Näiteks ei erista kumbki keel kirja pildis häälikuid /n/ ja /ŋ/. Nii eesti kui soome keeles võib võõrsõnades häälikute pikkus häälduses erineda kirja pildi järgi oletatust (sm *arkeologia* [arkkeologia], ee *apelsin* [apelsiin]). Soome keeles ei märgita kirjas piirigeminatsiooni (*sinnekään* [sinnekkään]), mida eesti keele hääldamises esineb palju vähem.

Eesti keeles on mittevastavust fonoloogilise ja ortograafilise süsteemi vahel siiski rohkem kui soome keeles. Kui soome keele ortograafia kohta on öeldud, et see on lugemaõppimise seisukohalt optimaalne (Aro 2004: 16), siis ei saa sama öelda eesti keele kohta, kus kirjutatud kujul pole võimalik vahet teha pikal ja ülipikal häälikul (v.a sulghäälikud), konsonantühendil ega diftongil. Teine oluline mittevastavus eesti ortograafilise ja häälikusüsteemi vahel on palatalisatsiooni mittemärkimine (nt kirja pilt *palk* võib olla hääldatud [palk] ja [pal´k]). Ortograafia ei kajasta kummaski keeles sõnarõhku, mis soome keeles on alati esimesel silbil, kuid eesti võõrsõnade

<sup>3</sup> Võõrsõnatähti ja vastavaid foneeme arvestamata.

lugemisel võib raskendavaks asjaoluks olla see, et mõnikord on rõhk ka järgsilbil. Selle põhjal võiks oletada, et eesti keelt on raskem lugema õppida kui soome keelt.

Eesti ja soome silbistruktuur on põhijoontes sarnane (soome k kohta vt Aro 2004: 14). Kummalegi ei ole algupäraselt omased konsonantühendid silbi alguses, kuid eesti keeles on neid siiski ka vanemates laensõnades, samas kui soome keeles ainult päris uutes võõrsõnades (ee *klaas*, *traav*, sm *lasi*, *ravi*, aga ee *traditsioon*, sm *traditio*). Siinse artikli seisukohalt on oluline lahtiste ja kinniste silpide vahetõrge keeles, kuna viimaseid arvatakse olevat raskem lugeda (Seymour jt 2003).<sup>4</sup> Eesti keeles on soome keelega võrreldes kinniseid silpe ajaloolise arengu tõttu tõenäoliselt rohkem, neid on tekitanud lõpukadu (ee *madal*, sm *matala*) ja sisekadu (ee *kandnud*, sm *kantanut*), teiselt poolt on sõnalõpulise *n*-i kadu eesti keeles tekitanud mõnedesse vormidesse lahtiseid silpe just juurde (ee *jökke*, sm *jokeen*). (Ajaloaliste muutuste mõjust võrdlevalt eesti ja soome keelele vt nt Remes 2009: 50–56.) Teadaolevalt ei ole võrdlevat statistilist uurimust kinniste ja lahtiste silpide sageduse kohta eesti ja soome keeles tehtud,<sup>5</sup> kuid võib siiski oletada, et soome (kirja)keeles on lahtisi silpe rohkem.

Nii eesti kui soome keeles on lugemaõppimist raskendavaks faktoriks keerukas morfoloogia: sõnadel on mitmeid eri tüvesid, vormid on tihti väga pikad (soome keeles pikemadki kui eesti keeles) ja sõna äratundmine seetõttu raskendatud.

## 4. Lugemistestid – mida ja kuidas hinnata?

Lugemisega seotud oskuste hindamine on oluliseks lapse arengu toetamise vahendiks nii enne kooli kui ka koolieas. Lugemisoskust hinnatakse erinevatel eesmärkidel: lapse individuaalse arengu jälgimiseks, võimaliku lugemisraskuse, selle sügavuse ja põhjuse väljaselgitamiseks, tagasisideks õpetuse efektiivsuse kohta ning õpetuse planeerimiseks. Kuna lugemisoskus hõlmab erinevaid kognitiivseid oskusi, on vajalik määratleda, mida täpsemalt hinnata soovitakse: sõnade lugemise eeloskusi, lugemistehnilisi oskusi, funktsionaalset lugemisoskust vm. Artiklis anname ülevaate sõnade lugemise oskuse hindamise võimalustest.

Sõnade lugemisoskuse hindamise eesmärkideks on välja selgitada dekodeerimise õigsus (st võimalike lugemisvigade tüübid ja sagedus), juhtiv lugemisviis, lugemise kiirus ja/või sõnade äratundmise oskus. Kui ebareeglipärase ortograafiaga keeltes peetakse oskuse hindamisel oluliseks õigsust, siis reeglipärase kirjasüsteemiga keeltes rõhutatakse sõnade lugemise kiiruse hindamist (Cain 2010: 201).

Dekodeerimisoskuse mõõtmiseks kasutatakse sageli teksti(lõigu), sõna- või mittedõnade loendite lugemist (nt Landerl, Wimmer 2008). Kõikide nende meetodite puhul on vajalik individuaalne testimine, lugemine toimub häälega ning hinnata on võimalik lugemise õigsust ja kiirust. Sõnade äratundmise oskuse mõõtmiseks võib kasutada sõnaketi meetodit (nt Tong jt 2011), kus lugejal tuleb mingi aja jooksul võimalikult kiiresti eraldada sõnad kokkukirjutatud sõnade jadast, või sõna ja pildi ühendamise ülesannet (nt Torppa jt 2007), mis nõuab valikuks antud sõnade hulgast pildile vastava sõna leidmist. Sõnade äratundmise teste saab läbi viia rühmas või individuaalselt, hääletult või häälega lugedes.

<sup>4</sup> Omaette seni uurimata probleem on seegi, kas eesti keeles kirja pildis lahtine, kuid häälduses kinnine silp (lühikese klusiilgeminaadiga sõnades nt esisilbid sõnades *lõpu*, *soki*, *koti*) on lugemaõppimise seisukohalt pigem kergem või raskem variant. Kirjapilt on sel juhul nagu lihtsate lahtiste silpide puhul ja väga võimalik, et häälduslik gemineerimine ei avalda tegelikult mingit mõju.

<sup>5</sup> Eesti ilukirjanduskeele silbistruktuurielementide fonostatistilise ülevaate on koostanud Mati Hint (1988), kesknendes silbi järjekorranumbri ning silbi koostisosade (silbialgus, -tuum ja -lõpp) seoste.

Viimasel aastakümnel on oluliselt tõusnud rahvusvaheliste uurimuste osakaal haridusvaldkonnas. Taoliste uurimuste juures on väga oluline, et õpilaste oskuste hindamise vahendid oleksid sisuliselt ja keeleliselt võimalikult sarnased. Siinses artiklis kirjeldame kahte eesti- ja soomekeelset sõnalugemistesti ning tutvustame testide võrdlemise meetodit.

## 5. Meetod

### 5.1. Materjal: sõnaloendi lugemise test ja sõnade äratundmise test

Sõnaloendi lugemise testi soomekeelne variant kuulub Soome standardiseeritud testipaketti *Lukilasse* (Häyrinen jt 1999), eestikeelse testi on koostanud Piret Soodla ja Inna Marats. Test koosneb kolme veergu paigutatud 90 sõnast, mis on järjestatud sõna häälik- ja silbistruktuuri keerukusastme tõusu põhimõttel: testi esimesed sõnad on lihtsa struktuuriga ühe- ja kahesilbilised sõnad (nt ee *ema, lai, nüüd*, sm *eli, soi, jäät*), järgnevad pikemad ja keerulisema silbistruktuuriga sõnad ja võõrsõnad (nt ee *märkavad, vabaajarõivastus, mikroskoobid*, sm *markkinat, kotitalouskoneet, mikroskooppi*). Test viiakse läbi individuaalselt. Ülesanne on sõnu võimalikult kiiresti häälega lugeda 45 sekundi jooksul. Testi sooritamise ajal märgib testija üles õigesti ja valesti loetud ning vahele jäetud sõnad. Sõnaloendi lugemise test võimaldab hinnata dekodeerimise õigsust ja kiirust.

Sõnade äratundmise testi soomekeelne variant kuulub Soome testipaketti *Ala-asteen lukutesti* (Lindeman 1998), eestikeelse versiooni koostas Piret Soodla. Test koosneb 80 pildist ja 320 sõnast. Iga pildi juures on neli sarnase häälikkoostisega sõna, millest üks vastab pildile (nt pilt 'tigu' – sõnad *tegu, tibu, tigu, tina*). Testi sõnad on järjestatud häälik- ja silbistruktuuri keerukusastme tõusu põhimõttel: testi alguses on sõnad lihtsamad (nt ee *kuu, tibu*, sm *kuu, pupu*), testi edenedes muutuvad sõnad pikemaks ja struktuurilt keerukamaks (nt ee *võtmed, kummirõngas, dinosaur*, sm *farkut, kumipatja, dinosaur*). Testi saab läbi viia nii rühmas kui ka individuaalselt, sõnu loetakse hääletult. Ülesanne on kahe (soome originaalvariandis viie) minuti jooksul ühendada joonega võimalikult palju pilte sobiva sõnaga. Testi sooritust näitab õigesti ja valesti ühendatud ning ühendamata jäetud sõnade hulk. Testi abil saab hinnata sõnade äratundmise õigsust ja lugemiskiirust.

### 5.2. Testide võrdlemise protseduur

Sõnalugemistestide raskusaste võib tuleneda mitmest asjaolust, mida tuleb arvestada nii sama- kui erikeelsete testide võrdlemisel. Esmatähtis on see, mis oskust või oskuste rühma testiga hinnatakse: kas hinnatakse dekodeerimise õigsust, kiirust ja/või sõnade äratundmist. Teiseks on oluline sõnade tuttavus, sest keeles sagedamad sõnad tuntakse lugemisel kiiremini ära. Laste testide puhul oleks relevantne võrrelda sõnade sagedust nendes tekstides, mida lapsed tegelikult loevad (lasteraamatud, õpikud, veebileheküljed, filmide subtiitrid), kuid selline sagedusloend meil paraku puudub ja sõnade sageduse hindamiseks tuleb kasutada olemasolevaid korpusi. Kolmas faktor on loetavate sõnade struktuur: kas tegemist on liht- või liit-sõnadega, milline on sõnade häälik- ja silbistruktuur jms. Samuti võib oluline olla,

millisel alusel eristuvad sõnavaliku testis valed variandid (segajad), sest ilmselt on raskem valida häälikuliselt sarnaste kui täiesti erinevate sõnade vahel. Lõpuks ei ole vähetähtis ka testi vormistuslik külg: kas loetavad sõnad on esitatud trüki- või kirjatähtedega, kas on esitatud silbipiirid jms, piltidega testis on oluline ka piltide kvaliteet. Lisaks testide endi võrdlemisele on võimalik võrrelda eri vanuserühmade lugemissooritusi.

Kahe kirjeldatud eesti- ja soomekeelse lugemistesti omavahelise vastavuse analüüsil rakendasime mitut meetodit. Esiteks võrdlesime testisõnade koostist, et selgitada, kui võrd sarnased on eesti ja soome testid keeleliselt. Teiseks kontrollisime testide võrreldavust eesti ja soome üliõpilaste testitulemuste põhjal, et selgitada välja, kuidas õnnestub testi tegemine neil, kellel eeldatavasti ei tohiks enam lugemiskursi olla. Kolmandaks viisime sõnade äratundmise testi läbi eesti ja soome lastega intervjuuna, milles küsisime, millised testisõnad ja -pildid võivad lastele raskusi valmistada ja mis võib olla selle põhjuseks.

### 5.2.1. Testisõnade keeleline analüüs

Selleks, et võrrelda eesti- ja soomekeelsete sõnalugemistestide keelelist külge, analüüsisime mõlema testi sõnade esinemissagedust keeles, sõnade pikkust, silbi- ja häälikstruktuuri. Esinemissageduse määramiseks tegime kindlaks, mitmendal kohal on vastav sõna sagedussõnastikus lemmade sagedusloendis.<sup>6</sup> Testisõnade pikkuse määrasime tähtede, silpide ja liitsõnaosade arvu põhjal. Testisõnade silbistruktuuri iseloomustamiseks silbitasime sõnad kirja- ja pildist lähtuvalt ning arvutasime kokku lahtised häälikuühendita silbid (V, CV, CV<sub>1</sub>V<sub>1</sub>), kinnised häälikuühendita silbid (VC, CVC, CV<sub>1</sub>V<sub>1</sub>C) ja häälikuühendit (konsonantühendit ja diftongi) sisaldavad silbid mõlemas testis. Sõnade häälikstruktuuri analüüsisides lugesime kokku häälikuühendid sõnas ning tegime kindlaks, kas häälikupikkus varieerub pildile vastava sõnarühma sees (sõnade äratundmise testis). Nimelt lugesime kokku sõnarühmad, mille lugemisel on õige sõna valiku kriteeriumiks häälikupikkus (nt eesti testis pilt 'kassid' – sõnad *kassad*, *kaasad*, *kasid*, *kassid*; soome testis pilt 'vaip' (sm *matto*) – sõnad *mato*, *matto*, *mutta*, *musta*).<sup>7</sup>

### 5.2.2. Üliõpilaste testimine

Sõnaloendi lugemise testi sooritasid 60 eesti keelt ja soome keelt emakeelena kõnelevat üliõpilast. Eesti üliõpilased ( $n = 30$ , 15 meest ja 15 naist, keskmine vanus 23,5 a) õppisid testimise perioodil erinevatel erialadel eri Eesti ülikoolides. Soome üliõpilased ( $n = 30$ , kümme meest ja 20 naist, keskmine vanus 26,1 a) õppisid erinevatel erialadel Soome ja Eesti ülikoolides. Eesti üliõpilastega viidi läbi eesti-keelne, soome üliõpilastega soomekeelne test. Testimise protseduur oli mõlemas keelerühmas samasugune: testimisele eelnes juhised, test viidi läbi individuaalselt, ajalimiit oli 45 sekundit. Sooritused salvestati diktofonile ning hiljem kodeeriti testitulemused (loetud sõnade koguhulk, õigesti ja valesti loetud sõnade arv).

Sõnade äratundmise test viidi läbi 58 üliõpilasega. Eesti üliõpilased olid samad, kes sooritasid sõnaloendi lugemise testi. Soome üliõpilased ( $n = 28$ , viis meest ja

<sup>6</sup> Ainus võrreldav materjal sõnade esinemissageduse võrdlemiseks eesti ja soome keeles on ajakirjanduskorpuse põhjal tehtud sagedussõnastikud, vt Eesti kirjakeele sagedussõnastik: <http://www.cl.ut.ee/ressursid/sagedused1/> (20.09.2012) ja Suomen sanomalehtikielen taajuussanasto: <http://www.csc.fi/tutkimus/alat/kielitiede/taajuussanasto-B9996/view> (20.09.2012).

<sup>7</sup> Andmete statistilisel analüüsil on olulisuse tase 0,05.

23 naist, keskmine vanus 24,0 a) õppisid erinevatel erialadel Eesti ja Soome ülikoolides. Testimisele eelnes mõlemas keeles ühesugune juhised, testi sooritamiseks oli kaks minutit, eesti üliõpilastega viidi test läbi individuaalselt, soome üliõpilastega rühmas. Peale testi sooritamist kodeeriti tulemused (loetud sõnade koguhulk, õigesti ja valesti loetud sõnad ning ühendamata pildi-sõna paarid).

### 5.2.3. Laste testimine ja intervjuerimine

Sõnade äratundmise test intervjuuna viidi läbi eesti ja soome lastega. Eesti lapsi oli kokku viis: Peep<sup>8</sup> (1. kl, vanus 8 a), Iti (1. kl, 8 a), Ats (2. kl, 8 a), Hans (3. kl, 9 a) ja Kati (5. kl, 11 a). Soome lapsi oli neli: Jussi (3. kl, 9 a), Henna (3. kl, 9 a), Marja (5. kl, 11 a), Oskari (6. kl, 11 a). Testi läbiviimise käik erines eelpool kirjeldatust (vt 5.1) – lastel paluti sõnad valjusti ette lugeda ning iga pildi-sõna paari ühendamise järel küsis testija, miks laps just selle sõna valis ning kas mõni teine valikuks antud sõna võiks veel pildiga kokku sobida. Lastel lubati testi teha neile sobivas tempos, ajapiirangut ei rakendatud. Intervjuud lindistati diktofonile ning transkribeeriti.

## 6. Tulemused

### 6.1. Sõnaloendi lugemise test

#### 6.1.1. Testisõnade keeleline analüüs

Eesti- ja soomekeelse testi 90 sõnast oli sagedussõnastikes esindatud vastavalt 43 ja 46 sõna. Testisõnade sagedusi keeltes võrdlesime kahe meetodiga. Esmalt võrdlesime eesti ja soome testi sõnu nende keskmise järjekorranumbri alusel sagedussõnastikes (t-testi abil). Analüüs näitas sõnade keskmise sageduse sarnasust eesti ( $M = 3240,8$ ;  $SD = 2920,2$ ) ja soome testis ( $M = 3613,1$ ;  $SD = 3159,9$ ),  $t(87) = 0,58$ ,  $p = 0,57$ . Järgmisena jaotasime testisõnad sageduse alusel viide kategooriasse: sõnad, mille järjekorranumber sõnastikus oli vahemikus 1) 1–100, 2) 101–500, 3) 501–1000, 4) 1001–5000, 5) üle 5000 (vt tabel 1, kategooriad K1–K5). Testisõnade jaotumist vastavatesse kategooriatesse analüüsisime Pearsoni hii-ruut-testiga, mis näitas, et sõnad jaotusid kategooriatesse sarnaselt,  $\chi^2(4) = 2,04$ ;  $p = 0,73$ .

**Tabel 1.** Eesti- ja soomekeelsete sõnalugemistestide sõnade jaotumine kategooriatesse sageduse alusel

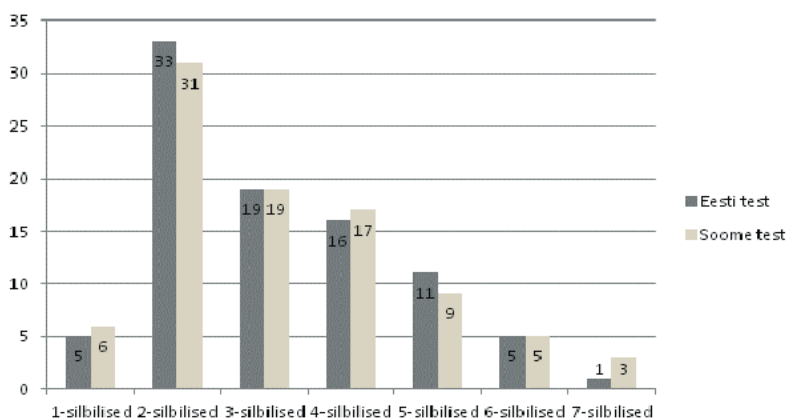
Test	Testi keel	Sõnade arv ja osakaal protsentides					Kokku
		K1	K2	K3	K4	K5	
Sõnaloendi lugemise test	eesti	2 (5%)	6 (14%)	3 (7%)	21 (49%)	11 (26%)	43 (100%)
	soome	3 (7%)	5 (11%)	5 (11%)	17 (37%)	16 (35%)	46 (100%)
Sõnade ära-tundmise test	eesti	–	7 (11%)	2 (3%)	26 (43%)	26 (43%)	61 (100%)
	soome	–	2 (4%)	1 (2%)	28 (60%)	16 (34%)	47 (100%)

**Tähised:** K1 = sõna järjekorranumber sagedussõnastikus vahemikus 1–100, K2 = sõna järjekorranumber sagedussõnastikus vahemikus 101–500, K3 = sõna järjekorranumber sagedussõnastikus vahemikus 501–1000, K4 = sõna järjekorranumber sagedussõnastikus vahemikus 1001–5000, K5 = sõna järjekorranumber sagedussõnastikus > 5000.

<sup>8</sup> Kõigi laste nimed on artiklis muudetud.

Testisõnadest, mida sagedussõnastikus ei leidunud, on üks osa selliseid, mis on laste jaoks kindlasti tuttavad (nt ee *puru*, *kilkab*, *vannitab*, *kõrvitsad*; sm *vaari*, *puuro*, *sääri*, *lakritsi*), teine osa sõnadest on keeles suhteliselt harva esinevad (nt ee *taimetoitus*, *padrikus*, *vingerdus*; sm *vingahdus*, *kettingit*, *graniittilouhos*). Lisaks leidus testisõnade hulgas pärisnimesid (ee testis *Anu*, *Kalle*, *Aafrika*, *Austraalia*; sm testis *Esa*, *Kalle*, *Afrikka*, *Australia*), mis võivad laste jaoks tuttavad olla, kuid mida eesti sagedussõnastik ei sisalda, soome sagedussõnastik aga sisaldab.

Testisõnade pikkust võrdlesime tähtede, silpide ja sõnaosade arvu alusel. Keskmise tähtede arv eesti ja soome sõnades oli väga sarnane: eesti testis 8,6 ( $SD = 3,7$ ) ja soome testis 8,6 ( $SD = 3,8$ ),  $t(178) = 0,06$ ;  $p = 0,95$ . Silpide arv sõnades varieerus mõlema testi puhul vahemikus 1–7 ning ka sõnade jaotumine silpide arvu alusel oli testides sarnane (vt joonis 1),  $\chi^2(6) = 1,38$ ;  $p = 0,97$ .



**Joonis 1.** Eesti- ja soomekeelse sõnaloendi lugemise testi sõnade jaotumine silpide arvu alusel

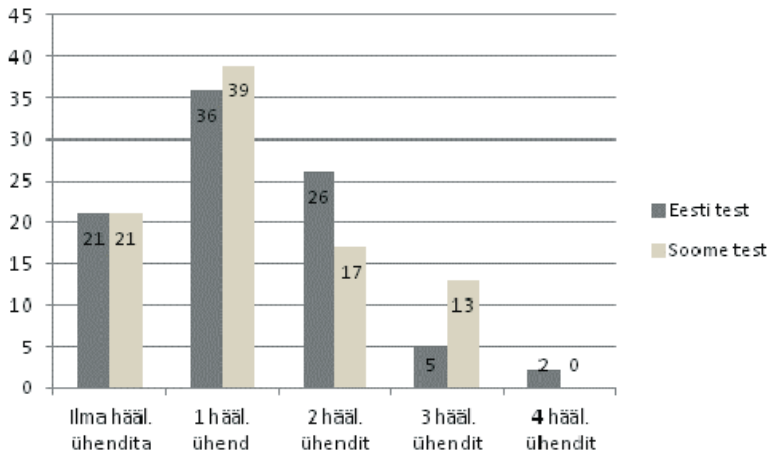
Eesti ja soome testisõnad olid sarnased ka sõnaosade arvu poolest: mõlemas testis olid valdavalt lihtsõnad (vastavalt 70 ja 73 sõna), kuid esines ka kahest (18 ja 16 sõna) ning kolmest osast (2 ja 1 sõna) koosnevaid liitsõnu,  $\chi^2(2) = 0,51$ ;  $p = 0,77$ . Seega, eesti ja soome sõnaloendi lugemise testi sõnad olid pikkuselt sarnased nii tähtede, silpide kui ka sõnaosade arvu poolest.

Sõnade silbistruktuuri võrdlemiseks silbitasime kõik testisõnad ja jaotasime silbid lahtisteks, kinnisteks ning häälikuühendiga silpideks. Iga sõna puhul summeerisime kõik sama tüüpi silbid (nt *broilerid*: üks häälikuühendiga, üks lahtine, üks kinnine silp). Ülevaadet kõikide testisõnade silbilise koostise kohta (iga silbitüübi osas eraldi) vt tabel 2. Eesti ja soome testisõnade jaotumist vastavat tüüpi silpide arvu alusel võrdlesime hii-ruut-testiga iga silbitüübi osas. Kõigi kolme silbitüübi osas oli sõnade jaotuvus sarnane (lahtised silbid:  $\chi^2(6) = 5,99$ ,  $p = 0,42$ ; kinnised silbid:  $\chi^2(4) = 4,97$ ,  $p = 0,29$ ; häälikuühendiga silbid:  $\chi^2(2) = 0,96$ ,  $p = 0,62$ ).

Lisaks võrdlesime testisõnade häälikkoostist, st kas sõna sisaldab häälikuühendit või mitte, sest silbistruktuuri analüüs ei võimalda määrata kõikide häälikuühendite arvu, kuna sageli jaotub konsonantühend mitme silbi vahel (nt ee *tüh/jen/da/tud*, sm *kor/tis/tot*). Analüüsil selgus, et eesti ja soome testisõnad jaotusid häälikuühendite arvu alusel sarnaselt,  $\chi^2(4) = 7,56$ ;  $p = 0,11$  (vt joonis 2).

**Tabel 2.** Eesti- ja soomekeelse sõnaloendi lugemise testi sõnade jaotumine lahtiste, kinniste ja häälikühendiga silpide arvu alusel

Silbi tüüp	Testi keel	Sõnade arv							Kokku
		0 silpi	1 silp	2 silpi	3 silpi	4 silpi	5 silpi	6 silpi	
Lahtine	eesti	25	29	22	7	7	–	–	90
	soome	17	35	23	8	4	2	1	90
Kinnine	eesti	23	35	18	13	1	–	–	90
	soome	35	23	18	13	1	–	–	90
Häälikuühendiga	eesti	51	30	9	–	–	–	–	90
	soome	47	36	7	–	–	–	–	90



**Joonis 2.** Eesti- ja soomekeelse sõnaloendi lugemise testi sõnade jaotumine häälikuühendite arvu alusel

### 6.1.2. Üliõpilaste testitulemused

Soome ja eesti üliõpilaste testitulemused (vt tabel 3) olid loetud sõnade koguhulga ja õigesti loetud sõnade osas sarnased: t-testi tulemused olid vastavalt  $t(58) = 1,20$ ,  $p = 0,24$ , ja  $t(58) = 1,49$ ,  $p = 0,14$ . Sõnade lugemisel tegid eesti üliõpilased soome tudengitega võrreldes siiski rohkem vigu ( $t(58) = 4,06$ ,  $p < 0,001$ ). Testi esimeste sõnade lugemisel tehti mõlemas keeles vigu harva, erinevus ilmnis keerukamate sõnade puhul, kus eesti tudengid tegid vigu oluliselt sagedamini (nt sõnade *harutaksin* (2'), *padrikus* (2'), *professorid* (2'), *kandikutelt* (2'), *rannikuil* (2') lugemisel) kui soome tudengid. Viimased eksisid vaid sõnade *vaari* (1'), *lakritsi* (1') ja *vingahdus* (1') lugemisel. Eestikeelses testis tegi lugemisvigu kokku 16 üliõpilast, soomekeelses testis eksis vaid kolm testitavat.

**Tabel 3.** Eesti ja soome üliõpilaste sõnalugemistestide tulemused: aritmeetilised keskmised ja standardhälbed (sulgudes)

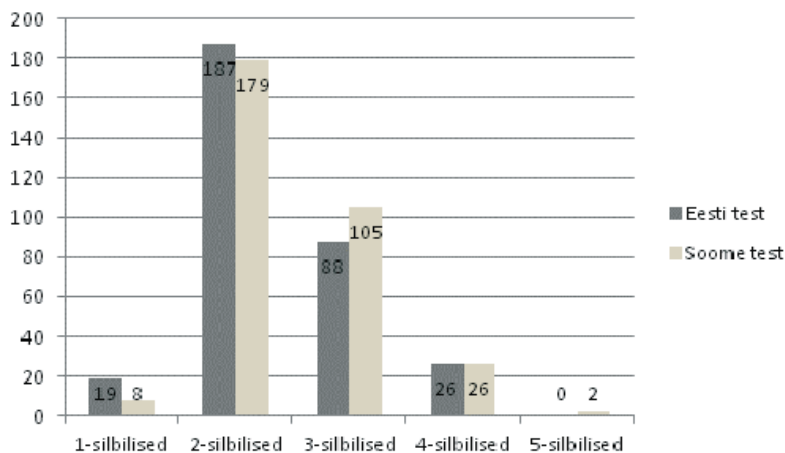
Test	Keel	Sõnade koguhulk	Õigesti loetud sõnad	Valesti loetud sõnad	Vahele jäetud sõnad
Sõnaloendi lugemise test	eesti (n = 30)	68,13 (10,07)	67,20 (10,12)	0,93 (1,08)	–
	soome (n = 30)	71,43 (11,27)	71,33 (11,29)	0,10 (0,31)	–
Sõnade äratundmise test	eesti (n = 30)	60,80 (10,63)	60,17 (10,63)	0,23 (0,50)	0,40 (1,07)
	soome (n = 28)	63,93 (9,22)	62,93 (8,78)	0,53 (0,84)	0,46 (1,37)

## 6.2. Sõnade äratundmise test

### 6.2.1. Testisõnade keeleline analüüs

Eesti ja soome testisõnade sagedusanalüüsi viisime mõlema testi puhul läbi nende 80 sõna osas, mis vastasid pildile. Sagedussõnastikes oli esindatud eesti testisõnadest 61 ja soome testisõnadest 47 sõna. Sõnade keskmised järjekorranumbrid sagedusloendites olid eesti ja soome testis vastavalt 5055,9 ( $SD = 4165,6$ ) ja 4331,6 ( $SD = 2719,3$ ) ning erinevus testide vahel ei olnud statistiliselt oluline,  $t(106) = 1,03$ ,  $p = 0,30$ . Sarnaselt sõnaloendi lugemise testiga jaotasime testisõnad sageduse alusel viide kategooriasse (vt tabel 1). Hii-ruut-testi tulemused näitavad, et eesti ja soome testi sõnad jaotusid vastavatesse kategooriatesse sarnaselt,  $\chi^2(3) = 3,82$ ,  $p = 0,28$ . Sagedussõnastikest ei leia laste jaoks ilmselgelt tuttavaid testisõnu (ee *silmad*, *lilled*, *konnad*; sm *pupu*, *muki*, *farkut*). Lisaks ei kajastunud sagedussõnastikes mitmed eesti ja soome keeles sagedased liitsõnad (ee *vanaema*, *tuletikud*, *piparkoogid*; sm *kumipatja*, *tulitikku*, *piparkakku*).

Testisõnade pikkust võrdlesime kõikide sõnade osas (mõlemas testis 320 sõna) tähtede, silpide ning sõnaosade arvu alusel. Keskmise sõna pikkus tähtedes oli testides väga sarnane: eesti testis 6,1 ( $SD = 1,6$ ) ja soome testis 6,1 ( $SD = 1,8$ ),  $t(638) = 0,40$ ,  $p = 0,69$ . Sarnaselt jaotusid sõnad ka silpide arvu alusel (vt joonis 3),  $\chi^2(4) = 8,15$ ,  $p = 0,09$ .



**Joonis 3.** Eesti- ja soomekeelse sõnade äratundmise testi sõnade jaotumine silpide arvu alusel



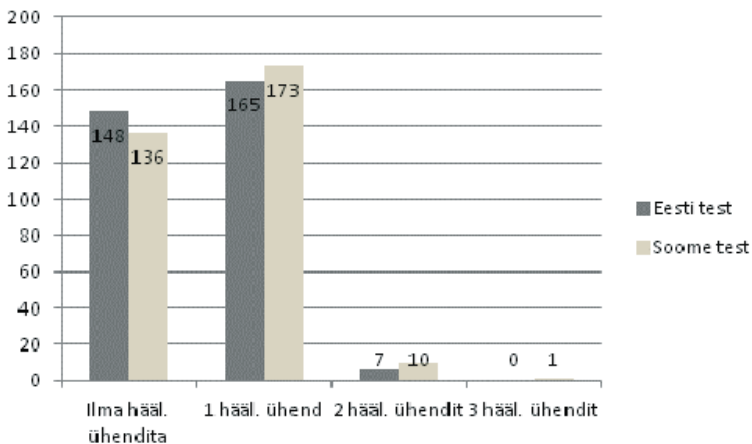
Eesti ja soome testisõnad olid sarnased ka sõnaosade arvu poolest: mõlemas testis oli enim lihtsõnu (vastavalt 299 ja 292 sõna), vähem esines ka kahest osast koosnevaid lihtsõnu (21 ja 28 sõna). Sõnade pikkus sõnaosade arvu alusel testide vahel ei erinenud,  $\chi^2(1) = 1,08, p = 0,30$ . Sõnade pikkuse osas olid eesti ja soome sõnade äratundmise testid seega sarnased.

Testisõnade silbistruktuuri analüüs, mille käigus võrdlesime lahtiste, kinniste ja häälikuühendiga silpide arvu sõnades, näitas erinevusi eesti ja soome testi vahel (vt tabel 4). Eesti testisõnad sisaldasid oluliselt vähem lahtisi silpe kui soome testi sõnad ( $\chi^2(4) = 75,74, p < 0,001$ ), seevastu oli eesti testis enam kinniseid silpe sisaldavaid sõnu ( $\chi^2(3) = 49,35, p < 0,001$ ). Häälikuühendit sisaldavate silpidega sõnu oli testides sarnaselt ( $\chi^2(1) = 2,60, p = 0,11$ ).

**Tabel 4.** Eesti- ja soomekeelse sõnade äratundmise testi sõnade jaotumine lahtiste, kinniste ja häälikuühendiga silpide arvu alusel

Silbi tüüp	Testi keel	Sõnade arv					Kokku
		0 silpi	1 silp	2 silpi	3 silpi	4 silpi	
Lahtine	eesti	112	101	73	26	8	320
	soome	24	141	113	38	4	320
Kinnine	eesti	90	137	88	5	–	320
	soome	144	149	26	1	–	320
Häälikuühendiga	eesti	244	76	–	–	–	320
	soome	226	94	–	–	–	320

Sõnade häälikkoostise analüüsimisel võrdlesime kõigepealt lihthäälikutest koosnevate ja häälikuühendeid sisaldavate sõnade sagedust eesti ja soome testis (vt joonis 4). Tulemused näitasid, et eesti ja soome testisõnad jaotusid häälikuühendite arvu alusel sarnaselt,  $\chi^2(3) = 2,23, p = 0,53$ . Teiseks analüüsisime, mitmes sõnarihmas on õige sõna valikukriteeriumiks häälikupikkus, st mitme sõnarihma puhul tuleb testi sooritajal sõna valikul toetuda vaid sõna lugemisele õige häälikupikkusega. Varieeruva häälikupikkusega sõnade rühmade arv oli testides sarnane (eesti testis 9 ja soome testis 7),  $\chi^2(1) = 0,28, p = 0,60$ .



**Joonis 4.** Eesti- ja soomekeelse sõnade äratundmise testi sõnade jaotumine häälikuühendite arvu alusel

## 6.2.2. Üliõpilaste testitulemused

Üliõpilaste tulemused sõnade äratundmise testis olid eesti ja soome rühmas küllaltki sarnased (vt tabel 3). Soome testitulemused olid küll mõnevõrra kõrgemad sõnade koguhulga ja õigesti loetud sõnade hulga osas, kuid mitte statistiliselt olulisel määral: t-testi tulemused olid vastavalt  $t(56) = 1,19$ ,  $p = 0,24$ , ja  $t(56) = 1,08$ ,  $p = 0,29$ . Lugemisel jäeti ühendamata sarnane hulk pildi-sõna paare ( $t(56) = 0,20$ ,  $p = 0,84$ ). Piltide ja sõnade ühendamisel tegid soomlased eestlastega võrreldes mõnevõrra enam vigu, kuid rühmadevaheline erinevus ei olnud statistiliselt oluline ( $t(56) = 1,67$ ,  $p = 0,10$ ).

Tulemuste detailsem analüüs näitas, et eesti testis tegi lugemisvigu kokku kuus üliõpilast, seejuures eksiti viie sõna äratundmisel: *sillid* pro *siilid* (2'), *killid* pro *kiilid* (2'), *ripakil* pro *rippuma* (1'), *viski* pro *veski* (1'), *varakas* pro *vaarikas* (1'). Soomekeelses testis eksis kokku 10 üliõpilast ning vead jaotusid järgmiselt: *pamppu* 'kumminui' pro *pumppu* 'pump' (7'), *tahti* 'takt; tempo' pro *tähti* 'täht' (2'), *silli* 'heeringas' pro *siili* 'siil' (2'), *nauha* 'pael' pro *naula* 'nael' (1'), *hirveä* 'kole' pro *hirvi* 'põder' (1'), *apina* 'ahv' pro *apila* 'ristikhein' (1'), *mustekala* 'kaheksajalg' pro *mustikka* 'mustikas' (1'). Enamik nimetatud lugemisvigadest peegeldab kiirel lugemisel tehtud dekodeerimisviga, v.a soome testisõnade *pamppu*–*pumppu* osas, kus sage eksimine viitab pildil kujutatud objekti ebaselgusele.

## 6.2.3. Laste intervjuude tulemused

Kuna testide sooritamise ajal ei ole võimalik registreerida laste kõhklusid ja kahtlusi – näha on vaid lõplikud valikud –, viisime läbi testid ja intervjuud üheksa eesti ja soome lapsega (vt eespool 5.2), kellel paluti lisaks sõnade ettelugemisele kommenteerida oma kõhklusid. Laste testisooritused ja intervjuud andsid olulist lisateavet võimalike sõnade lugemisel ja piltide tajumisel ettetulevate probleemide kohta.

Eesti ja soome laste testisooritused ja intervjuud näitasid, et valdav osa pildi-sõna paare ühendati suuremate raskusteta. Nooremad lapsed tegid lugemisel siiski rohkem vigu, mis kohati raskendas pildi juurde õige sõna valimist, vt näited (1)–(3).

- (1) (ee) Peep: ratas kujutatavale pildile sobivat sõna (*radas*, *rattas*, ***ratas***, *ruttas*)<sup>9</sup> valides loeb *ratas* pro *radas*, mistõttu valib esialgu sõna *radas*. Suunamise peale valib õige sõna.
- (2) (ee) Iti: kahte kassi kujutatavale pildile sobiva sõna (*kassad*, *kaasad*, *kasid*, ***kassid***) valimisel loeb *kassid* pro *kasid* ja *kassid* (III v) pro *kassid* (II v) ning kahtleb, kumb sõna on õige.
- (3) (ee) Ats: pildi juurde sobivat sõna (*kutsikad*, *kalurid*, *keeravad*, ***kartulid***) valides loeb *kalurid* pro *kartulid*. Kui testija kordab lapse valitud sõna, märkab Ats oma viga. Testija küsimusele, miks ta valis sõna *kalurid*, vastab, et ei lugenud sõna lõpuni.

Näidetes (4)–(6) teevad lapsed lugemisel vea, kuid see ei takista õige sõna ühendamist pildiga.

<sup>9</sup> Näidetes on pildi juures esitatud sõnade loendis alla joonitud õige sõna.

- (4) (sm) Henna: loeb *katkera* pro *katketa*, kuid see ei sega õige sõna ühendamist pildiga (*katketa*, *kihara*, ***kitara***, *kätkeä*).
- (5) (ee) Hans: õiget pilti (*randa*, ***panda***, *pandi*, *anda*) valides küsib, mis see *randa* (II v) on?
- (6) (ee) Hans: enne sõna valimist (*huugad*, ***huuled***, *huikad*, *huilgad*) küsib, mis sõna see *huugad* (II v) on?

Näited (5)–(6) osutavad sellele, et häälikupikkuse varieerumine eesti keeles teeb lugemise raskemaks seetõttu, et sõnarühma sees mõjutab sõna lugemist ka eelnevate ja järgnevate sõnade häälikute pikkus. Kuna antud sõnades (*randa*, *huugad*) kirjapilt hääldust täpselt ei fikseeri, siis võib see aeglustada lugemise protsessi.

Lisaks kirjeldatud juhtudele, kus sõna valikut raskendasid dekodeerimisprobleemid, esines testides ka pilte, millel kujutatud objekti lapsed ei tundnud (7) või oli mõni pildi kõrval pakutud sõnadest lapse jaoks võõras, nagu näites (8) *Iti* jaoks *sõrnik* ja *sõrmik*. Näites (9) tunnistab Kati, et ta ei tea sõna *kallur* tähendust. Henna valis küll pakutud variantidest õige sõna pildi juurde (10), kuid ütles, et ei tea sõna *pamppu* tähendust.

- (7) (sm) Marja: enne sõna valimist (***höylä***, *höyry*, *nöyrä*, *hölmö*) küsib hõövlit kujutava pildi kohta, mis see on.
- (8) (ee) *Iti*: sõrmikut kujutava pildi juurde sõna (*sõrnik*, ***sõrmik***, *sõrmus*, *sõõrik*) valimisel loeb sõnad õigesti, kuid ei oska valida, kas õige sõna on *sõrnik* või *sõrmik*. Testija suunava küsimuse peale, kuidas nimetatakse kinnast, millel on sõrmed küljes, valib õige sõna.
- (9) (ee) Kati: kalluri pildi juurde sobiva sõna (*kalur*, ***kallur***, *kallis*, *kaalus*) valikul ütleb, et õige sõna on *kallur*, aga ta ei tea, mida see sõna tähendab.
- (10) (sm) Henna: valis pildi kõrvalt (*pamppu*, ***pumppu***, *limppu*, *lamppu*) õige sõna, kuid tunnistas intervjuerijale, et ei tea, mis see *pamppu* on.

Laste kommentaarid osutavad sellele, et mõnel pildil kujutatu (soome testis tomat, kurk, pump, eesti testis kohviuba) ei olnud piisavalt selgelt ja ühemõtteliselt joonistatud:

- (11) (sm) Jussi leiab pildiga sobiva sõna, kuid kommenteerib seejärel tomati pilti, et see ei ole tomati moodi, see on pigem kõrvitsa moodi.
- (12) (sm) Jussi kommenteerib kurgi pilti: kurk on suvikõrvitsa moodi.
- (13) (sm) Oskari kommentaar (jalgratta)pumba pildile on, et see oleks nagu pipett.
- (14) (ee) Peep ütleb (*kohvik*, ***kohviuba***, *koolimaja*, *kohvitama*) õiget sõna valides, et sellest pildist küll aru ei saa, tundub nagu hamburger või kohv.

Laste intervjuude põhjal selgus, et nii eesti kui soome sõnavaliku testid on koostatud lapsele arusaadavalt, kuid mõlemas testis on mõned problemaatilised sõna ja pildi ühendamise juhud.

## 7. Kokkuvõte

Analüüsisime võrdlevalt kahte eesti- ja soomekeelset sõnalugemistesti – sõnaloendi lugemise testi ja sõnade äratundmise testi. Võrreldes teste meetoditriangulatsiooni (testide keeleline analüüs, täiskasvanud lugejate testisoorituste ja laste intervjuude analüüs) abil, püüdsime hinnata lugemistestide vastavust ja arendada sellega lugemistestide võrdlusmeetodeid. Mitme meetodi paralleelne kasutamine võimaldab nii eri- kui ka samakeelsete lugemistestide ekvivalentsust hinnata mitmekülgsemalt kui vaid ühte meetodit rakendades.

Sõnaloendi lugemise testide keeleline analüüs näitas, et eesti- ja soomekeelne test on sarnased nii sõnade sageduse, pikkuse kui ka hääliku- ja silbistruktuuri poolest. Üliõpilaste testisooritused viitasid aga eesti testi sõnade mõnevõrra suuremale keerukusele. Nimelt lugesid eesti üliõpilased antud ajalimiidi jooksul küll ligikaudu sama palju sõnu kui soome tudengid ja ka õigesti loetud sõnade keskmine hulk oli sarnane, kuid nad tegid rohkem lugemisvigu. Erinevus vigade arvus viitab eesti testi mõnevõrra kõrgemale raskusastmele, mille põhjused testide keelelisel analüüsil ei selgunud. On võimalik, et eesti test sisaldas siiski rohkem keeles harva esinevaid sõnu – nimelt kajastasid kasutatud sõnade sagedusloendid vähem kui poolt testisõnadest. Teine põhjus võib olla ka mõni siinses artiklis käsitlemata aspekt (näiteks sõnade morfoloogiline koostis), mis võis mõjutada sõnade lugemise õigsust. Kolmas võimalik põhjus võib seisneda keelelistes iseärasustes: kuna eesti keeles on sõnad üldiselt lühemad kui soome keeles, aga testisõnad olid pikkuselt sarnased, siis võib oletada, et eesti lugeja jaoks valmistab väga pikkade sõnade lugemine mõnevõrra suuremaid raskusi kui sama pikkusega sõnade lugemine soome keeles. Eesti testi sooritamise võib muuta raskemaks ka asjaolu, et paljude sõnade puhul tuleb lähtuvalt sõna tähendusest valida lugemisel õige sõnavälde (nt *kuula, purgid*), mida soomekeelse testi sooritajal teha ei tule. Nimetatud keeleline iseärasus ei pruugi oluliselt mõjutada kogenud lugejate testisooritust, küll aga raskendab see dekodeerimise protsessi lugemaõppimise alguses.

Sõnade äratundmise teste võrdlesime testisõnu keelelisel analüüsis, üliõpilaste testitulemuste ja laste intervjuude põhjal. Testide keeleline analüüs näitas, et eesti- ja soomekeelne test olid väga sarnased sõnade sageduse ja pikkuse poolest. Sõnade silbistruktuuri analüüsil selgus, et eesti testi sõnades oli soome testi sõnadega võrreldes mõnevõrra vähem lahtisi ja rohkem kinniseid silpe. Häälikuühendeid oli testisõnades siiski sarnasel hulgal. Kuna lahtisi silpe peetakse lugemise seisukohalt lihtsamaks kui kinniseid ja häälikuühendeid sisaldavaid silpe (Seymour jt 2003), siis võib oletada, et eesti testi sooritamine võib valmistada mõnevõrra suuremaid raskusi lugemisoskuse arengu algetapil kui soome testi sooritamine samas vanuses.

Üliõpilaste sooritused sõnade äratundmise testis osutasid testide suurele sarnasusele nii loetud sõnade koguhulga, õigesti loetud sõnade kui ka valesti loetud sõnade osas. Mõnevõrra rohkem eksisid küll soome üliõpilased, tulemuste täpsem analüüs osutas ühele problemaatilisele pildile (rattapump) ja raskustele vastava sõna valikul. Laste testisooritused ja intervjuud näitasid, et valdav osa pildi-sõna paare ühendati suuremate raskusteta, nooremad lapsed eksisid sõnade lugemisel enam ning kõhklesid seetõttu sagedamini ka sõnale vastava pildi valikul. Mõlemas testis oli siiski ka sõnu, mille tähendust lapsed ei teadnud ja pilte, millel kujutatut

ei tuntud ära või selles kaheldi. Laste intervjuude põhjal võib siiski väita, et nii eesti kui soome sõnade äratundmise testi sõnad ja pildid on lastele mõistetavad ning suuri erinevusi kahe testiversiooni vahel ei esine.

Lõpetuseks võib öelda, et artiklis analüüsitud eesti- ja soomekeelset sõnaloendi lugemise testi ja sõnade äratundmise testi võib kasutada võrreldavate hindamisvahenditena keeltevahelistes uuringutes. Siiski ei saa neid pidada identseteks testideks eesti ja soome keeles, seda osalt mõningaste erinevuste tõttu testides endas ning osalt keeleliste iseärasuste tõttu. Testide tõlkimisel ja kohandamisel ühest keelest teise on küll oluline, et testid oleksid võimalikult sarnased, kuid erikeelsete aja- ja piiranguga sõnalugemistestide identsus on ülimalt keerukas, kui mitte võimatu saavutada. Nimelt on sõnade lugemist mõjutavad keelelised faktorid lugemaõppimise eri faasides keeleti erinevad ning testid, mida saab pidada identseteks ühe lugejarühma (näiteks algajate lugejate) oskuste võrdlemiseks, ei pruugi seda olla teise lugejarühma (näiteks kogenumate lugejate) puhul.

## Viidatud kirjandus

- Arffman, Inga 2010. Equivalence of translations in international reading literacy studies. – *Scandinavian Journal of Educational Research*, 54, 37–59.
- Aro, Mikko 2004. Learning to Read: The Effect of Orthography. <https://jyx.jyu.fi/dspace/bitstream/handle/123456789/13355/9513917223.pdf?sequence=1> (14.09.2012).
- Aro, Mikko; Wimmer, Heinz 2003. Learning to read: English in comparison to six more regular orthographies. – *Applied Psycholinguistics*, 24 (4), 621–635. <http://dx.doi.org/10.1017/S0142716403000316>
- Cain, Kate 2010. *Reading Development and Difficulties*. West Sussex: BPS Blackwell.
- Cohen, Louis; Manion, Lawrence; Morrison, Keith 2000. *Research Methods in Education*. 5th ed. London: Routledge Falmer.
- De Luca, Maria; Borrelli, Marta; Judica, Anna; Spinelli, Donatella; Zoccolotti, Pierluigi 2002. Reading words and pseudowords: An eye movement study of developmental dyslexia. – *Brain and Language*, 80 (3), 617–626. <http://dx.doi.org/10.1006/brln.2001.2637>
- Dürrwächter, Ute; Sokolov, Alexander N.; Reinhard, Jens; Klosinski, Gunther; Trauzettel-Klosinski, Susanne 2010. Word length and word frequency affect eye movements in dyslexic children reading in a regular (German) orthography. – *Annals of Dyslexia*, 60 (1), 86–101. <http://dx.doi.org/10.1007/s11881-010-0034-9>
- EKK 2007 = Erelt, Mati; Erelt, Tiit; Ross, Kristiina 2007. *Eesti keele käsiraamat*. Kolmas, täiendatud trükk. Tallinn: Eesti Keele Sihtasutus.
- Hakulinen, Auli; Vilkkuna, Maria; Korhonen, Riitta; Koivisto, Vesa; Heinonen, Tarja-Riitta; Aho, Irja 2004. *Iso suomen kielioppi*. SKS:n toimituksia 950. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Hint, Mati 2004. *Eesti keele foneetika ja morfoloogia*. Viies, ümbertöötatud trükk. Tallinn: Avita.
- Hint, Mati 1988. *Eesti ilukirjanduskeele statistiline fonotaktika*. Silbistruktuuri elemendid. Tallinn: E. Vilde nim Tallinna Pedagoogiline Instituut.
- Hoover, Wesley A.; Gough, Philip B. 1990. The simple view of reading. – *Reading and Writing*, 2 (2), 127–160. <http://dx.doi.org/10.1007/BF00401799>
- Hutzler, Florian; Wimmer, Heinz 2004. Eye movements of dyslexic children when reading in a regular orthography. – *Brain and Language*, 89 (1), 235–242. [http://dx.doi.org/10.1016/S0093-934X\(03\)00401-2](http://dx.doi.org/10.1016/S0093-934X(03)00401-2)
- Häyrynen, Taru; Serenius-Sirve, Silve; Korkman, Marit 1999. *Lukilasse*. Helsinki: Psykologien kustannus.

- Hyönä, Jukka; Olson Richard K. 1995. Eye fixation patterns among dyslexic and normal readers: Effects of word length and word frequency. – *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21 (6), 1430–1440. <http://dx.doi.org/10.1037/0278-7393.21.6.1430>
- Kikas, Eve 2006. School psychology in Estonia. – Shane R. Jimerson, Thomas Oakland, Peter T. Farrell (Eds.). *The Handbook of International School Psychology*. California: Sage Publications, 91–102.
- Landerl, Karin; Wimmer, Heinz 2008. Development of word reading fluency and spelling in a consistent orthography: An 8-year follow-up. – *Journal of Educational Psychology*, 100 (1), 150–161. <http://dx.doi.org/10.1037/0022-0663.100.1.150>
- Lerkkanen, Marja-Kristiina 2007. *Lugema õppimine ja õpetamine alus- ja algõpetuses*. Tartu: Tartu Ülikooli Kirjastus.
- Lindeman, Johanna 1998. *ALLU–Ala-asteen lukutesti*. Åbo: Åbo Akademis förlag.
- Oakhill, Jane; Cain, Kate 2007. Introduction to comprehension development. – Kate Cain, Jane Oakhill (Eds.). *Children's Comprehension Problems in Oral and Written Language*. New York: The Guilford Press, 1–40.
- Perfetti, Charles A. 1994. Psycholinguistics and reading ability. – Morton Ann Gernsbacher (Ed.). *Handbook of Psycholinguistics*. San Diego: Academic Press, 849–894.
- Priya, Kanu; Wagner, Richard K. 2009. The roles of fluent decoding and vocabulary in the development of reading comprehension. – Richard K. Wagner, Christopher Schatschneider, Caroline Phytian-Sence (Eds.). *Beyond Decoding: The Behavioral and Biological Foundations of Reading Comprehension*. New York: The Guilford Press, 124–139.
- Remes, Hannu 2009. Muodot kontrastissa: Suomen ja Viron vertailevaa taivutusmorfologiaa. *Acta Universitatis Ouluensis B. Humaniora* 90. Oulu: Oulun yliopisto.
- Seymour, Philip H. K.; Aro, Mikko; Erskine, Jane M. 2003. Foundation literacy acquisition in European orthographies. – *British Journal of Psychology*, 94 (2), 143–174. <http://dx.doi.org/10.1348/000712603321661859>
- Soodla, Piret 2012. Teachers' judgments of monolingual and bilingual students' reading skills: Child-, class-, and teacher-related factors. – Aaro Toomela, Eve Kikas (Eds.). *Children Studying in a Wrong Language: Russian Speaking Children in Estonian School. Twenty Years after the Collapse of the Soviet Union*. Frankfurt am Main: Peter Lang, 123–151.
- Soodla, Piret; Kikas, Eve 2010. Teachers' judgment of students' reading difficulties and factors related to its accuracy. – Aaro Toomela (Ed.). *Systemic Person-Oriented Study of Child Development in Early Primary School*. Frankfurt am Main: Peter Lang, 73–94.
- Tong, Xiuli; Deacon, S. Hélène; Kirby, John R.; Cain, Kate; Parrila, Rauno 2011. Morphological awareness: A key to understanding poor reading comprehension in English. – *Journal of Educational Psychology*, 103, 523–534. <http://dx.doi.org/10.1037/a0023495>
- Torppa, Minna; Tolvanen, Askko; Poikkeus, Anna-Maija; Eklund, Kenneth; Lerkkanen, Marja-Kristiina; Leskinen, Esko; Lyytinen, Heikki 2007. Reading development subtypes and their early characteristics. – *Annals of Dyslexia*, 57 (1), 3–37. <http://dx.doi.org/10.1007/s11881-007-0003-0>

## Võrgumaterjalid

- Eesti kirjakeele sagedussõnastik. <http://www.cl.ut.ee/ressursid/sagedused1/> (20.09.2012).
- Suomen sanomalehtikielen taajuussanasto. <http://www.csc.fi/tutkimus/alat/kielitiede/taajuussanasto-B9996/view> (20.09.2012).

**Piret Soodla** (Tallinna Ülikool) uurimistö valdkonnad on suulise ja kirjaliku kõne areng, kõnearengu ja lugemisoskuse hindamine.

[piret.soodla@tlu.ee](mailto:piret.soodla@tlu.ee)

**Maigi Vija** (Tartu Ülikool) on uurinud esimese keele omandamist, sh morfoloogiat ja pronoomenikasutust.

[maigi.vija@ut.ee](mailto:maigi.vija@ut.ee)

**Renate Pajusalu** (Tartu Ülikool) uurimistö valdkonnad on semantika ja pragmaatika, eriti referentsiaalsusega seotu. Uurinud on ka esimese ja teise keele omandamist.

[renate.pajusalu@ut.ee](mailto:renate.pajusalu@ut.ee)

# COMPARISON OF ESTONIAN AND FINNISH WORD READING TESTS

Piret Soodla<sup>1</sup>, Maigi Vija<sup>2</sup>, Renate Pajusalu<sup>2</sup>

<sup>1</sup>Tallinn University, <sup>2</sup>University of Tartu

In the present study two word-reading tests in Estonian and Finnish are compared. The word- list reading test assesses decoding speed and accuracy. The word-recognition test assesses fluency and accuracy of word recognition in the reading process. The study had two aims. The first was to evaluate the comparability of the reading tests in Estonian and Finnish. The second aim was to contribute an appropriate method for comparing word reading tests. The comparative analysis of the two tests involved multiple methods: linguistic analysis, analyses of adult readers' and children's test performances and children's interviews. The results point mainly to similarities but also to some differences between the Estonian and Finnish tests.

**Keywords:** testing, assessment, reading skills, Estonian, Finnish



# MULTILINGUALISM OF A SOUTHERN ESTONIAN – COMPARISON OF ESTONIAN, VÕRU AND FINNISH NARRATIVES

Liina Tammekänd

**Abstract.** The main subjects of the present case study are structural analysis of multilingual autobiographical narratives, and multilingualism and emotions. The respondent told an emotional story from her childhood in Estonian, in the Võru dialect and in Finnish. The author recorded, transcribed and analysed the three narratives keeping in mind four research questions: Do the narratives differ from each other structurally and thematically? Is the respondent's L1 emotionally closer to her? Does the respondent present different selves and emotional repertoire in her different languages? The narratives are structurally similar but thematically different. Because of the emotional vocabulary and themes in the Võru narrative, the Võru dialect seems emotionally closer to the respondent than the other two languages. No extensive conclusions could be made regarding the different selves. The speaker uses a different emotional repertoire in Finnish compared with the other languages.\*

**Keywords:** multilingualism, oral narrative, emotion repertoire, Estonian, Võru dialect, Finnish

## 1. Introduction

The present article presents a case study on individual multilingualism in a Southern Estonian person. The author compared multilingual narratives on the same topic narrated by the respondent in Estonian, in the Võru dialect and in Finnish. The following research questions were posed:

1. Do multilingual narratives differ structurally and thematically from each other?
2. Is the respondent's L1 emotionally closer to her than her other languages? (Cf. Bond, Lai 1986, Javier et al. 1993)

---

\* This research was partly supported by ETF grant 9098 *The Development of Dialogue Interpreting in Estonia: Reflections of a Transition Society*.

3. Does the respondent have different selves in her different languages? (Cf. Koven 1998)
4. Does the respondent have different emotional repertoires in her different languages? (Cf. Pavlenko 2002, Marian, Kaushanskaya 2008)

The article has five parts. After the introduction, the research method and the subject matter are introduced, followed by an overview of the respondent's language biography to give a context to the research. The main part of the article consists of the analysis and interpretation of the results, followed by the discussion and conclusion.

## 2. Method and subject matter

### 2.1. Concepts and terms

Researchers (Ervin-Tripp 1964, Javier et al. 1993, Anooshian, Hertel 1994, Pavlenko, Dewaele 2002, Dewaele 2004, Wierzbicka 2004) have described the emotional closeness of L1 thus: a multilingual might feel that her communication in L1 is emotionally more intense than her communication in L2. L2, however, could be used to distance oneself from emotions (Bond, Lai 1986, Dewaele 2004).

Jeanette Altarriba (2003) maintains that the seeming emotional closeness of L1 relates to the emotional vocabulary of L1 being stored at the deeper representational level of memory as it has been used in multiple contexts. Michael Bond and Tat Ming Lai (1986) think that the emotional closeness of L1 is connected to the language learning context. L1 is learned mainly at home which is a more emotional environment where certain words could be connected with anxiety during socialisation. L2 is learned at school which is an emotionally neutral environment where the vocabulary items might not form emotional connotations. However, not all studies give reason to claim that L1 is the emotional language of a person and L2 is not. For instance, there are cases where L1 cannot be determined because people have reached the level of a native speaker later in their life and attrition of L1 might have occurred (Piller 2002).

A distinction should be drawn between objectively measurable identity and subjectively describable self. In the case of multilinguals, in Estonian the term *mitmikidentiteet* (multiple identity) (Ehala 2004) is used, but in my view the term is more related to the cultural and socio-political side of society (e.g. an Estonian Russian, a speaker of the Võru dialect). Koven (1998) and Aneta Pavlenko (2006) use the term *self* to describe how a person feels when expressing herself in L2.

Some researchers (Koven 1998, Dewaele, Pavlenko 2001–2003, Marian, Kaushanskaya 2004, Pavlenko 2006) try to describe the selves of multilinguals claiming that multilinguals feel themselves as different persons when speaking different languages.

According to Pavlenko (2007, 2008), elicited fictional narratives and autobiographical narratives (Javier et al. 1993) are among of the most used ways of collecting data to study multilingualism. The autobiographical narrative can be divided into two sub-groups: the personal experience narrative and the language biography. Researchers have used language learners' diaries and language memoirs to study multilingualism (Verschik 2010).

## 2.2. Relevant research

Susan Ervin-Tripp was one of the first researchers to use multilingual narratives to study multilingualism. Her research (1964) treated bilinguals who offered explanations to Thematic Apperception Test (TAT) pictures in their two languages. Ervin-Tripp found the L1 narratives more emotional and the narrative in L2 more abstract. The themes in the L1 and L2 narratives were different as well.

Rafael Javier, Felix Barrosi and Michele Muños (1993) gave Spanish-English bilinguals five minutes to tell an emotional story from their past. The respondents chose the first session's language themselves. In the second session, the respondents told the same story in their other language. Between two sessions, there was a 20-minute interval. The researchers concentrated on the number of and elaboration on ideas expressed in the different languages. They found that the L1 narrative was richer in detail than the one in L2, which seemed to be more of a summary of the L1 narrative.

Michèle Koven (1998) studied personal experience narratives of two Portuguese-French bilinguals. There was no interval between the sessions. She analysed the narratives linguistically and had the respondents and their peers assess the narration. The linguistic analysis showed that the respondents used different lexical and morphosyntactic means and registers when speaking in L1 and L2. Both the respondents and their peers found that the respondents sounded like different persons in their different languages. Therefore, Koven concluded that multilingual people use different selves when speaking in their different languages.

In their research on emotional repertoires of multilinguals, Pavlenko and Viktoriia Driagina (2007) found that multilinguals used different lexical and morphosyntactic devices for emotions in their different languages. Similar studies by Alexia Panayiotou (2004) and Viorica Marian and Margarita Kaushanskaya (2008) showed that a language might activate the socio-cultural framework connected to it so that a multilingual uses an appropriate emotional reaction in an appropriate context.

## 2.3. The present case study

In the present case study, the respondent told one emotional story about her childhood in three languages. There were long intervals (at least two weeks) between the sessions. All narratives were recorded and transcribed. The Võru, Estonian and Finnish narratives were 716, 610 and 434 words respectively. The narratives were analysed thematically and linguistically. The linguistic analysis concentrated on the use of deixis, grammatical tenses and emotion repertoires.

The use of deixis indicates how the narrator perceives her character and other characters in the narrative, how she positions the characters in relation to each other and how she conceives the narrated events. This implies emotional closeness of L1 and existence of different selves for different languages. The use of grammatical tenses indicates where the narrator places herself and the narratee and whether she uses emphatic narrative techniques. This implies again emotional closeness of L1. Differences in the use of emotion vocabulary and affective deixis signal the existence of different emotional repertoires that are at disposal of the narrator. From all the above thematic and structural differences arise in the three narratives.

The respondent's language biography was also recorded to give the case study a more specific context. Language autobiographies are described by Pavlenko (2007) as life histories that 'focus on the languages of the speaker and discuss how and why these languages were acquired, used, or abandoned'. The respondent describes all learned languages, the time of study, the context and frequency of use and determines her passive and active skills in each language.

### **3. Respondent's language biography**

The respondent, a 31-year-old woman with higher education, comes from Misso in Võrumaa. She is from a bilingual family where the parents spoke both standard Estonian and the Võru dialect to their children. The respondent acquired Estonian and the dialect simultaneously. She cannot remember whether her parents favoured standard Estonian or the dialect for different topics (e.g., family matters in the Võru dialect and school problems in Estonian). The respondent communicated with her grandparents and other older people in the Võru dialect. Children in the village and at school spoke standard Estonian.

For the respondent, standard Estonian is the language of education and profession. She speaks the Võru dialect to these peers from her village who are not embarrassed by the fact that they know the dialect. She says that despite the fact that her father tries to speak standard Estonian with her when visiting, she feels a strong connection between her father and the dialect, which leads her to speak the dialect to her father. The respondent does not have a similar reaction to her mother. She uses a mix of standard Estonian and the dialect when speaking to her as the mother has been living with the respondent for six years in town.

In her second year at school (7–8 years), she started studying Russian. She watched soap operas on Russian TV and asked a classmate of Russian origin to speak Russian to her. The respondent cannot remember when she reached fluency in Russian, but she uses the language both in everyday and professional communication and in writing.

In the sixth year at school (12 years), a second foreign language was added to the curriculum. The respondent learned German according to the general curriculum until the last term of the last year at school. In the last term, she changed school and joined a class that learned German according to a special curriculum, which meant more German classes in a week. She had to work hard to catch up with her classmates who had been studying according to the German special curriculum for some years already. However, she scored well in the national examination of German and was accepted at university where she continued her German studies. During her BA studies she received a month-long grant to study German in Germany. The respondent uses German professionally and in writing.

The respondent also communicates in Finnish. She achieved fluency during actual communication in the professional context. She says that the Võru dialect and Finnish are very similar for her and reaching fluency at the communicational level posed no problems. She recalls that at first she understood only 30% of what was spoken in Finnish and had to guess the reminder from the context. She describes that she learned 'like a child' by mimicking and imitating. She has asked the meanings

of words from native speakers and has sometimes written the vocabulary items down. Finns assess her speaking skills as very good.

In the first years at university, the respondent took an elementary English course (60 academic hours). She has not studied English in a formal context since and has instead practised the language through actual communication. She uses the language both in everyday and professional communication and in writing.

The respondent is able to understand the gist of a conversation in Italian and Spanish and she can express the most important ideas related to her work in these languages. She has also had some contacts with French, but she says that her skills in this language are insufficient. It can be presumed that her passive knowledge of Romance languages arises from the fact the respondent studied Latin intensively for two terms at university, where she studied also two terms of classical Greek and biblical Hebrew, some Modern Greek and Modern Hebrew, none of which she uses either actively or passively.

## **4. Analysis and interpretation**

### **4.1. The general description of the three narratives – thematic blocks, utterances and words**

Both the personal experience narrative and the language biography were used to collect data. The respondent was asked to tell an emotional story from her childhood. At first, the respondent told the story in Estonian, then, after two weeks, in the Võru dialect and finally, after four months, in Finnish. All three narratives were recorded and transcribed. The respondent recounts the events of one summer day when she herded cattle with her friend. To fight boredom, she decides to learn how to ride a cow. She is successful, but her friend gets jealous and drives the cows home from the pasture. The narratives have three main characters; the narrator, her girlfriend and the cow.

At first, the narratives were analysed thematically. The thematic blocks of all narratives were determined, named provisionally and juxtaposed to find unique details from each narrative. Later, the words and utterances in each thematic block of every narrative were counted to calculate the average length of an utterance in each language (see Table 1). The comparison of the number of unique details, words and utterances in the Estonian, Võru and Finnish narratives showed which narrative has more information and of what nature this information is. Also the use of tenses and Sg1 verb form was determined, helping to understand how the respondent perceives herself in the three languages.

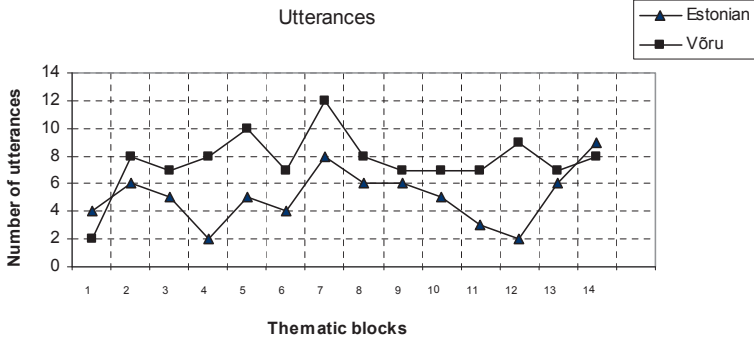
The Estonian and Võru narratives are structurally very similar. Both narratives have fourteen thematic blocks with almost identical content and succession. The blocks, however, differ thematically. The Finnish narrative has only eleven thematic blocks, but its structure is similar to the other two narratives. There are also thematic similarities to the other two narratives but with noticeable differences as the narrator combines three successive thematic blocks and loses two blocks present in the other narratives. She introduces two new thematic blocks. The structure of all narratives is presented in Table 1.

**Table 1.** Overview of the narratives

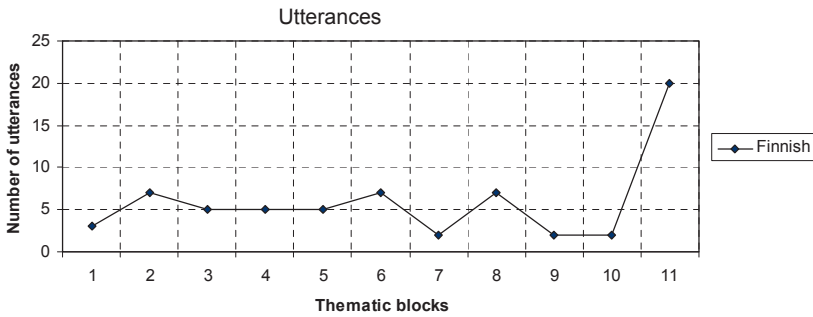
Provisional names given to the thematic blocks	Number of utterances (U) and words (W); notes		
	Estonian narrative	Võru narrative	Finnish narrative
1. Introduction	4U, 22W unique detail	2U, 55W	3U, 135W
2. Working as a milk maid	5U, 38W	7U, 42W	7U, 54W three thematic blocks are combined
3. Hard work of a milk maid	3U, 27W	4U, 54W	
4. Easy work of a shepherd	3U, 28W	5U, 47W two unique details	
5. Tediousness	3U, 33W	7U, 60W two unique details	5U, 39W 5U, 32W two blocks with different subject matter (Dream and The Friend); two unique details
6. Description of the cow	6U, 32W unique detail	1U, 8W	
7. Teaching the cow	11U, 74W	10U, 78W	5U, 27W
8. Riding the cow	5U, 41W	5U, 49W	7U, 49W
9. Friend is jealous	7U, 71W	7U, 60W	2U, 17W
10. Explanation-interlude	7U, 73W	6LU, 57W unique detail	block is absent
11. Cows go home	6U, 42W	5U, 66W	7U, 46W
12. A quick ride	2U, 56W	11U, 87W unique detail	2U, 20W
13. Unexpected twist	6U, 31W	6LU, 46W unique detail	2U, 14W
14. Conclusion	5U, 42w	7U, 59W unique detail	20U, 116W block with different subject matter; unique detail

There are 106, 71 and 67 utterances in the Võru, Estonian and Finnish narrative respectively (see Figures 1 and 2). The average length of the Estonian utterance is 8.6 words and that of the Võru utterance 6.8 words, which makes the average Võru utterance 21% shorter than the average Estonian utterance. The average length of the Finnish utterance (6.5 words) is rather similar to that of the average Võru utterance.

The Võru narrative has 35 utterances more than the Estonian narrative, which shows that the narrator makes more pauses in the Võru narrative and thus apparently structures the narrative better. The Võru introduction is laconic, but later the number of utterances increases and remains high through the whole Võru narrative. The Estonian narrative has more utterances than the Võru narrative only in the first and the last thematic block. Regarding the number of utterances, the Võru narrative has four peaks and the Estonian narrative has three peaks. Two of these peaks coincide (see Figure 1). The Finnish narrative has the peak in its last thematic block that corresponds to the last block if the Estonian and Võru narrative (see Figure 2).

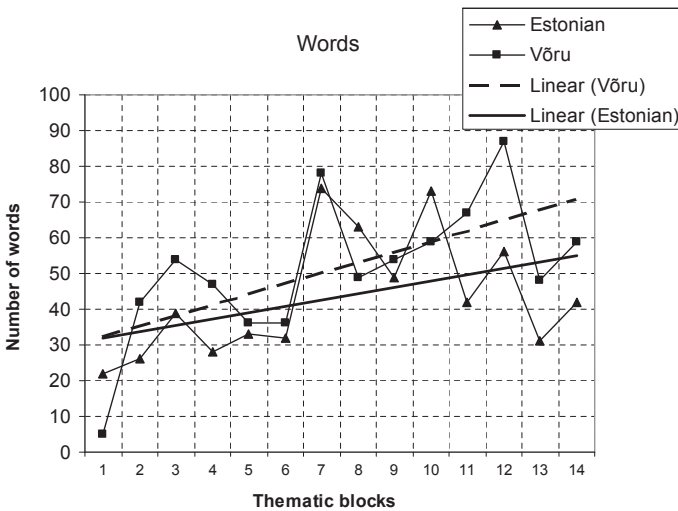


**Figure 1.** Number of utterances in thematic blocks in the Estonian and Võru narrative

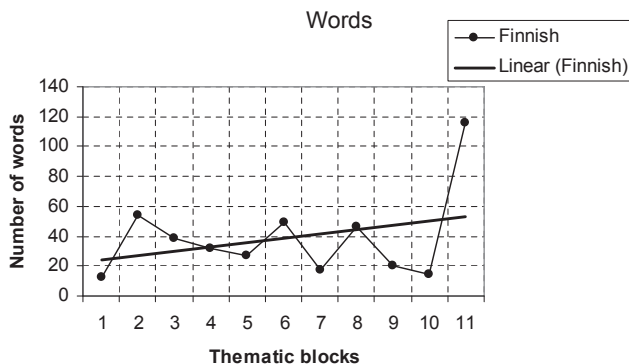


**Figure 2.** Number of utterances in thematic blocks in the Finnish narrative

The Võru narrative with its 716 words is the longest narrative. The Estonian narrative has 610 words and is 15% shorter than the Võru narrative. The Finnish narrative has 434 words and is 40% and 29% shorter than the Võru and Estonian narrative respectively (see Figures 3 and 4).



**Figure 3.** Number of words in thematic blocks in the Estonian and Võru narratives



**Figure 4.** Number of words in thematic blocks in the Finnish narrative

Regarding the number of words, the Võru narrative has three peaks and the Estonian narrative has four peaks. One of these peaks coincides (see Figure 3). Some thematic blocks in the Finnish narrative have more words than others because the narrator integrates and adds thematic blocks. In the last block, she has introduced a new theme as well (see Figure 4).

All three narratives are structurally rather similar. However, in the Finnish narrative the narrator integrates, loses and adds thematic blocks, which makes the Finnish narrative three thematic blocks shorter than the other two narratives.

There is a 21% difference in the length of the average utterance in the Võru and Estonian narratives. Both in utterances and in words, the Võru narrative is the longest and the Finnish narrative is the shortest. It can be assumed that the longest narrative, the Võru narrative, has the highest number of details and the shortest narrative, the Finnish narrative, gives a clipped overview of the event. However, the shortness of the Finnish narrative might be because the narrator uses the language in her professional communication and she does not possess the necessary vocabulary.

#### 4.2. Use of tenses in the narratives

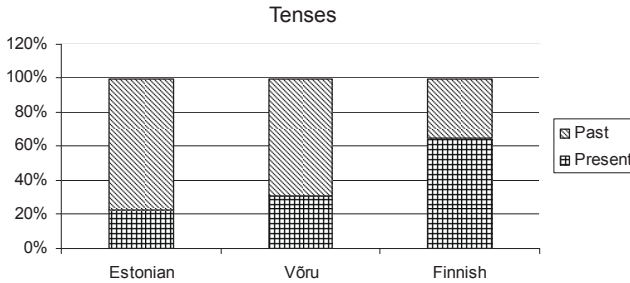
Monika Fludernik (2003) mentions five tenses used in narratives: narrative past, epic preterite, narrative present, historical present and epic present. In the context of the current study, three tenses take on a specific importance. According to Fludernik, narrative past is the default tense in past narratives, specifying the time and place of the events of the narrative. Narrative present refers to the deictic centre of the narrator and the narratee and is the tense in which the narrator addresses the narratee and conveys universal truths. Historical present is used to highlight important places in the narrative. Other authors (Wolfson 1979, Schriffrin 1981) describe historical past also as a tense (1) that adds to the dramatic effect of the narrative so that the narratee feels like she herself has witnessed the events being narrated, or (2) as a tense that leaves the impression of the narrator being absorbed in the narrative that she seems to relive the events being narrated (discourse time and story time appear to coincide).

As the case study deals with a past narrative, the narrator should use mainly past tenses. This is true for the Estonian narrative as there are 72% more past tense verbs



than present tense verbs (74 verbs in past; 21 verbs in present). Compared with the Estonian narrative, the narrator, however, uses present tense forms approximately 50% more in the Võru narrative (95 verbs in past; 41 verbs in present). Additionally, the narrator uses also perfect and past perfect in the Võru narrative, but she uses them both only once.

The narrator uses many present tense forms in the Finnish narrative: 63% of the verbs are in the present tense (22 verbs in past; 40 verbs in present). additionally, the narrator uses perfect tense once.



**Figure 5.** Use of tenses in the three narratives

In Estonian, the narrator uses predominantly the past tense, as if looking back to the event. It might be assumed that with such a choice, she distances herself from the event and acts as a bystander. In the Võru narrative, the narrator uses the present tense approximately 50% more than in the Estonian narrative. The use of the past tenses is proportionally higher in the Võru narrative, as the narrative itself is longer, but it can still be presumed that by the use of the present (historic present, in this case) the narrator has set herself in the middle of the event. She is not a neutral bystander, but rather a participant. The narrator has used also perfect and past perfect in the Võru narrative, which diversifies the use of tenses on a small scale but at the same time introduces a temporal dimension to the narrative. The narrator does not switch between past and present like she does in the Estonian narrative. however, any further conclusion about the use of perfect and past perfect cannot be drawn as these appear in the Võru narrative only once.

As the differences in the use of tenses in the Estonian and Võru narrative arise probably from the different thematical stresses (see subdivision 4.5), it can be presumed that the high use of the present in the Finnish narrative is rather the result of the narrator's low functionality in the language. It seems that the narrator finds it easier to form verbs in present. There is a noticeable influence of the Võru dialect on forming tenses in Finnish (for example *kado* vs. *katson* 'I look', *hellitä* vs. *hellin* 'I caress', *es ole* vs. *ei ollut* 'was not').

### 4.3. Use of Sg1 verb form with and without a pronoun

Lindström et al. (2009) conducted a study the aim of which was to determine whether the use of the pronoun 'I' is influenced by the verb ending in the Võru dialect (the indicative affirmative verb in the first person singular is either marked or not). The study concludes that the use of Sg1 pronoun does not depend on the verb

marking. It is also noted that the Sg1verb form is mostly used with the pronoun in the Võru dialect. The table below, however, demonstrates that in the present case study the narrator uses the pronoun especially extensively.

**Table 2.** Use of Sg1 verb form with and without a pronoun

Study	Sg1 with a pronoun	In all:	Sg1 without a pronoun	In all:
Lindström et al. (2009)	63.1%	195	36.9%	48
Here (in the Võru dialect)	89.5%	43	10.5%	5

Similarly, the narrator prefers to use Sg1 verb form with a pronoun than without it in the Estonian and Finnish narratives also. At first glance it seems that the narrative is very ‘self-centred’ in all languages. Below, the structure of the narrative and themes in different languages are described and it can be seen that the narrator actually shifts focalisation away from herself from time to time in all languages. At the moment, it has to be presumed that the extensive use of a pronoun with Sg1 verb form in the Võru dialect is a characteristic trait of the narrator or the sign of the language change.

To sum up, the respondent uses Sg1 pronouns extensively in all three narratives.

#### 4.4. Deictic relations

The narrator uses almost the same number of different deictics in the Estonian and Võru narratives (59 and 57 deictics in the Estonian and Võru narrative respectively). In the Finnish narrative, the narrator uses 32 different deictics. Table 3 gives an overview of the use of deictics in the three languages.

**Table 3.** Overview of deictics in the three narratives

Category of deixis	Estonian	Võru dialect	Finnish	
Place (demonstrative)	32	46	18	
Person	1Sg	48	25	
	2Sg	2 (addressing the listener)	2 (generic “you”)	4 (addressing the listener)
	3Sg	17	3	7
	1Pl	1	–	1
	2Pl	–	–	–
	3Pl	1	–	–
Time	8	4	2	

The narrator uses the primary role deixis (part of Estonian personal deixis, expressed by Sg1 and Sg2) and the demonstrative deixis most in all narratives. Compared with the Estonian narrative, the narrator uses 36% more of demonstrative deixis in the Võru narrative (32 and 46 demonstrative deictics in the Estonian and in the Võru narrative respectively). In both the Estonian and Võru narratives, the narrator uses the demonstrative deictic *see/too* ‘this’/‘that’ most (18 and 40 times in the Estonian and Võru narrative respectively). The demonstrative deictic *seal* ‘there’ is used more in the Estonian narrative (10 and 4 times in the Estonian and Võru

narrative respectively). Different forms of the pronoun *tää* ‘this’ are used in the Finnish narrative. The pronoun *tu* ‘that’ is used once.

The narrator uses deictics approximately 50% more in the Estonian narrative than in the Võru narrative (six and four times respectively). In both the Estonian and the Võru narrative, the most frequent time deictic is *siis* ‘then’ marking the succession of events. The narrator uses deictics just twice in the Finnish narrative *nüüd* ‘now’ and *sis* ‘then’ (the pronunciation by the respondent).

In all three narratives, the narrator uses the primary role deixis the most (48, 51 and 25 times in Estonian, the Võru dialect and Finnish respectively). Compared with the other narratives, the narrator uses much Sg3 deixis in the Estonian narrative. It seems that in the Võru narrative Sg3 deixis is substituted with the demonstrative pronoun *tuu* ‘that’ as Sg3 deictics are used only three times. Sg3 deictics appear seven times in the Finnish narrative. In the Estonian narrative, the narrator addresses the listener using Sg2 and uses both an impersonal and a passive verb once. In the Võru narrative she uses generic ‘you’ three times.

Pajusalu (1999) mentions that applying the demonstrative deictic *see* ‘this’ to a person could have a derogatory effect and in this way, the narrator might distance herself and the listener from the person she is speaking about (demonstrative deixis takes the role of affective deixis). Potts and Schwarts (2010) note about English *this* that the narrator might wish to engage the listener emotionally and the referent of ‘this’ is often judged by the narrator. In the Võru narrative, the use of the demonstrative deictic *tuu* ‘that’ in the way Pajusalu described can be discerned:

- (1) aga mul tuu sõbranna **tuu** X (.) **tuu** brigadiri tütar (.) **tuule** tuu plaan es meeldü  
‘but my that girlfriend, **that** X (.) **that** daughter of the foreman (.) **that** (friend) didn’t like that plan’

According to Potts and Schwarts, the narrator can be seen as trying to create an emotional connection in the Estonian narrative:

- (2) lõpuks ma hüppasin selga sõitsin **selle** lehmaga seal mingi kümme minutit [---] ja ta tahtis mind **selle** lehma seljast maha hirmutada  
‘finally, I got up to rode on **this** cow for some ten minutes [---] and she wanted to scare me off **this** cow’

In the Võru narrative, the distancing *tuu* ‘that’ and in the Estonian narrative emotionally connecting *see* ‘this’ appear in the same thematic block. It is not possible to find any affective deixis in the Finnish narrative.

In all three narratives, the narrator uses the primary role deixis the most. The second most frequently used type of deixis is demonstrative deixis. The narrator favours demonstrative deictics *see/too*, *tuu* and *tää* ‘this’/ ‘that’. Although the event took place in the past, the narrator does not use many time deictics. In the Estonian narrative, she uses the time deictic *siis* ‘then’ the most. The narrator uses demonstrative deictics *see* ‘this’ in Estonian and *too* ‘that’ in the Võru dialect as affective deixis.

#### 4.5. Emotional repertoire

In the first thematic block of the Estonian narrative, mostly positive emotion words are used.

- (3a) *kihvt karjapäev* ‘a memorable day in the pasture’
- (3b) *hea lugu* ‘a good story’
- (3c) *suur lehmastõber* ‘a great cow fancier’
- (3d) *väga rahul* ‘very happy’
- (3e) *meeldima* ‘to like’
- (3f) *lemmikud* ‘favourites’

The Võru and Finnish narratives do not hold such emotions at all.

Parallels can be observed in the emotion vocabulary in all three narratives in the third thematic block.

- (4a) *kergem töö* (Estonian) ‘an easier job’
- (4b) *hää elo, ull töö* (Võru) ‘a good life, a crazy job’
- (4c) *tosi hüvä tüö paha tüö* (Finnish) ‘a very good job, a bad job’

After the third thematic block, the Estonian and Võru narrative concentrate on boredom, but the Finnish narrative tells about a big dream.

Later in the narrative, the Estonian narrative conveys irritation, jealousy and shouting.

- (5a) *ta oli ärritunud* ‘she was irritated’
- (5b) *ta ei vaadanud selle peale küll positiivselt* ‘she did not have any positive feelings towards it’
- (5c) *teda hakkas see ärritama* ‘this situation irritated her’
- (5d) *ta oli ilmselgelt kade* ‘she was obviously jealous’
- (5e) *ta hakkas karjuma* ‘she started shouting’
- (5f) *ja karjub seal* ‘and she is shouting there’

The Võru narrative describes mainly dislike and fright in the same thematic blocks or the adjacent thematic blocks.

- (6a) *tuule tuu plaan es meeldü* ‘that (girl) didn’t like that plan’
- (6b) *nakasi lehmä hirmutamás* ‘she started frightening the cow’
- (6c) *nakse plaksutama õkvalt tämä niimoodi hirmutama* ‘she started clapping in order to frighten’
- (6d) *hirmutas takka* ‘she scared [the cow] even more’

In the Finnish narrative, the narrator uses only words conveying positive emotions even when she wants to describe a negative emotion.

- (7a) *ma nii onnelik* ‘I’m so happy’
- (7b) *onnellinen ihminen* ‘a happy person’
- (7c) *hän es rakasta* ‘she doesn’t like’

Then, in Estonian the narrator describes falling into manure as simply an unpleasant event.

- (8) *see ei olnud kõige parem, see oli ebameeldiv* ‘it wasn’t the best thing, it was unpleasant’

In the Võru dialect she uses much more vivid words for describing the same event.

- (9) *ullumaja tuu tuu tunne ja tuu hais* ‘it was bedlam there that that feeling and that smell’

In the Finnish narrative, the narrator only gives a summary of the event and does not use emotion words at all.

At the end of the narrative, the narrator describes her anger and her friend’s gloating in Estonian:

- (10) *ja ma olin väga vihane X peale aga X-l oli väga hea tunne väga hea meel* ‘and I was very angry at X but X had a very good feeling very happy’

In the Võru dialect she describes her guilt and her friend’s gloating:

- (11) *ja X-l oll hää müil, ja süüdi jäi iks ma* ‘and X was happy, and I was the guilty one [in letting the cows go home earlier]’

In the Finnish narrative, the narrator describes also the friend’s gloating, but positive emotions dominate:

- (12) *X-l nii hüvä miel, hän naara viela, hänen naara viela* ‘X [is] so happy, she is laughing, she is laughing’

The only negative emotion word *vihaista* ‘angry’ in the Finnish narrative is at the end:

- (13) *kaiki oli viela vihaista* ‘all were angry [at us]’

The choice of emotion words in the Estonian and Võru narrative is slightly different (Estonian ‘irritate’ vs. Võru ‘scare’; Estonian ‘I was angry’ vs. Võru ‘I was the guilty one’). In the Finnish narrative, the narrator uses positive emotion words even when she wants to express negative emotions. As in tense formation in Finnish, where the influence of the Võru dialect was noticeable, the Finnish narrative presents many cases of quasi-Finnish creations, similar to the situation described in Verschik’s study (2012) on receptive multilingualism in communication between Estonian frontline employees and Finnish customers.

## 5. Interpretation

The narrator’s perception of time and her place in it seems to be different in the Võru and Estonian narratives. In the Võru dialect, she speaks more slowly, the average utterance is shorter than that of the Estonian and there are more pauses between the utterances. On the one hand, time adverbials give a time frame to the narrative; on the other hand, they create an impression that time runs somehow slower for the narrator. The narrator uses time deictics approximately 50% less than in the Estonian narrative, which also gives an impression of time running slowly. Many present tense forms are used. This makes the narrative emphatic and places the narrator in the middle of the events.

She speaks more rapidly in Estonian, her average utterance is longer than that of the Võru utterance and there are few pauses between the utterances. Time adverbials show that the narrative lacks a clear time frame. In Estonian, the adverb *siis* 'then' is used as a time deictic to signal the quick succession of events. The narrator uses many past tense forms and so places herself outside the events, becoming a neutral observer. Compared with the Võru narrative, the Estonian narrative is less structured and the variety of tenses used is small.

The average length of the Finnish narrative is comparable with that of the Võru narrative, but here the pauses demonstrate the narrator looking for appropriate words. Also the use of tenses does not show whether the narrative is emphatic. The narrative is divided into three clear-cut parts by the use of tenses: the introduction is in the past, the description of the events in the present and the conclusion again in the past. The historical present of the Võru narrative is absent from the Finnish narrative and the present is used illogically from time to time. It can be assumed that the narrator finds it easier to form present tenses than past tenses. In the whole Finnish narrative, the narrator uses only two time deictics *nüt* 'now' and *sis* 'then'. Of the three narratives, the Finnish narrative seems to be the most rigid.

All narratives are rather 'self-centred' as the narrative is about a personal experience. However, in the narratives the narrator shifts focalisation away from herself using different means. In the Estonian narrative, she uses Sg3 deixis, the passive voice, impersonal verbs, the direct speech and addressing the listener directly. In the Võru narrative, she uses the affective deictic *tuu* 'that', the direct speech, generative 'you' and the secondary role deixis. Shifting the focalisation in the Finnish narrative is not clearly noticeable although the narrator uses the Sg3 deictic *hän* 'she'.

Both in the Võru and Estonian narratives, demonstrative deixis is used as affective deixis in the same thematic blocks. In the Estonian narrative, the narrator uses the affective deictic *see* 'this' (see Potts and Schwarts 2010) to engage the listener emotionally. In the same block in the Võru narrative, the narrator uses the affective deictic *tuu* 'that' (Pajusalu 1999) to distance herself and the listener from the person referred to. By doing so, she also judges the person. In the Finnish narrative, demonstrative deixis (mainly different forms of the pronoun *tää* 'this') is not used as affective deixis.

Emotion words used in the three narratives presumably demonstrate the differences in the emotional repertoires that are available to the narrator in her different languages. In the Estonian narrative the narrator expresses her positive feelings towards her work. This positive emotion is absent from the other narratives. The common denominator of the Estonian and Võru narratives is the friend's gloating, but in Estonian the narrator expresses her anger towards her friend, whereas in the Võru narrative the narrator communicates the injustice she feels.

The narrator describes falling into manure neutrally in Estonian, but with emotion in the Võru dialect.

In the Finnish narrative, the narrator expresses both positive and negative emotions by positive emotion words. The only negative emotion word in the Finnish narrative is 'to get angry'. This can be explained by the fact that as the narrator uses Finnish mainly in oral professional communication (service industries), she does not have the vocabulary to express negative emotions.

As the Võru narrative is the longest, it can be presumed to be most detailed. The narrative has eight unique details missing from the other narratives. These details are mostly connected with village life, the cattle and specific names and numbers. The Estonian narrative has two unique details connected with positive emotions towards work. There is also a detailed description of the cow in Estonian that is absent from other narratives. The Finnish narrative has three unique details that describe people as the narrative has two new thematic blocks absent from the other narratives.

It is interesting how the narrator describes the other main character – the friend, “this daughter of a foreman”. In the Võru dialect, the friend appears at very the beginning of the narrative. In Estonian, the friend is presented only in the second half of the narrative before the thematic block where the friend is one of the central characters. In the Võru narrative, the narrator stresses three times that the friend is a foreman’s daughter; in the Estonian narrative she mentions this only once. So it can be said that in the Estonian narrative the main characters are the narrator and the cow, and the friend is just a secondary character whose task is to complicate the events. The Võru narrative, however, has three characters – the narrator, the cow and the friend. Stressing that the friend is a foreman’s daughter in the Võru narrative could refer to the hierarchy of *kolkhozy* and *sovkhozy* of the Soviet era where everyone knew their place as department heads, head accountants, foremen, milkmaids etc. The narrator does not stress the hierarchy in Estonian and expresses her pride in being a milkmaid’s daughter. However, the word choice in the Võru narrative suggests that the narrator is ashamed rather than proud of being a milkmaid’s daughter.

The Finnish narrative does not stress the *sovkhos* hierarchy either. Most likely, it is because the narrator lacks the appropriate vocabulary. Obviously, it was difficult for the narrator to tell this story in Finnish which lacks the thematic vocabulary. For example, in Finnish the foreman’s daughter is described as ‘the big boss’s daughter’.

## 6. Discussion

The language biography shows that the respondent learned Estonian and the Võru dialect simultaneously. Therefore, it is difficult to determine whether the respondent’s L1 is Estonian or the Võru dialect. The respondent used the Võru dialect in informal contexts (with the family and older people) and Estonian in formal contexts (at school). The situation has not changed: the respondent uses the Võru dialect within the family and Estonian in the professional context. Therefore, it can be said that the present case study somewhat supports the hypotheses by Ervin-Tripp (1964), Bond and Lai (1986) and Javier et al. (1993). As in the research by Ervin-Tripp, the Estonian narrative seems to be more abstract and the themes are different from those of the Võru narrative. Estonian might sound more abstract as the respondent used it mostly in the education system, which deals with more abstract topics (cf. Bond, Lai 1986). The Võru narrative is the longest of the three narratives, presenting details that were missing in the other two narratives. This agrees to some extent with the conclusion by Javier et al. (1993), who found that the L2 narrative is rather a summary of the L1 narrative. However, the present case

study does not allow us to claim that any of the three narratives is a mere summary of the events. In the research by Javier et al., the impression of a summary could be because of too short an interval between the two sessions. During the second session the respondents were probably able to remember the exact content of the narrative and thus might have preferred to summarise their previous narrative. The researchers also allowed the respondents to choose the language of the first session. It is likely that the respondents chose their L1 or dominant language for the first session and were therefore able to narrate in a more lively and detailed way. In the present case study, it was impossible to determine whether Estonian or the Võru dialect is the narrator's L1. The case study therefore demonstrates that L1 being the dominant language is only one possible multilingual situation.

Also, it should be taken into account at which level the narrative is recycled and whether it has been told before. Some respondents may not consider the second session as important since they have already told their story. Some respondents regard the second session as a good opportunity to elaborate on the previously told narrative. It is rather likely that in the present case study the multilingual narrative is a deeply recycled one, which is the reason why all three narratives have a rather similar structure.

Like in Koven (1998), this respondent was asked her subjective opinion about whether she feels differently when speaking her other languages. The respondent denied it. The linguistic, thematic and narratological analysis could allow the finding that the respondent has different selves for her different languages. However, the case study is too limited to give enough information to claim this. I believe that the respondent could take on different roles in her different languages, or she could develop different narratological identities. Studying the development of narratological identities in different languages and the adequate analysis requires longer narratives, though. This, however, is not within the scope of the present study.

Like in Pavelnko (2002), Pavlenko and Driagina (2007), Panayiotou (2004), and Marian and Kaushanskaya (2008), it appeared that the respondent of the present case study has indeed different emotional repertoires in her different languages. As the respondent does not use Finnish for narrating or higher level monologues but only to communicate professionally, presumably she does not have a wide emotional repertoire in Finnish. In the Finnish narrative, she uses positive emotion words for expressing negative emotions which could show that she has not found any need for negative emotion words in her professional communication and therefore has not acquired them. This confirms the research of Pavlenko (2002) and Pavlenko and Driagina (2007). For the respondent, Estonian and the Võru dialect seem to activate different socio-cultural frameworks as well. Although the structures of the Estonian and Võru narrative are identical, the themes of these narratives are different. In Estonian, the respondent distances herself from the events, expresses pride in her work ethics and shows herself as independent in every way. In the Võru dialect, however, the narrator describes the hierarchy, expresses shame over her place in this hierarchy and the feeling of injustice because of the events. There are no clear themes in the Finnish narrative probably because of the respondent's insufficient language skills.



## 7. Conclusion

The multilingual narratives of the present case study are not very different structurally. There are differences in thematic stresses, however. The structure of the Finnish narrative has been changed by adding and losing thematic blocks. The Estonian and Võru narratives are identical by structure but different by themes. The Finnish narrative does not have a distinctive thematic stress.

The Võru dialect seems to have a greater emotional intensity for the respondent, although she learned Estonian and the Võru dialect simultaneously and the Võru dialect cannot be considered as her L1. However, the Võru narrative is the longest of the three narratives and contains the highest number of unique details. The respondent also uses affective deixis in the Võru dialect. Finnish seems to be the most emotionally distant of the three languages for the respondent.

The claim that the respondent has different identities in her different languages seems arbitrary at the moment, although the thematic differences in the Estonian and Võru narratives could allow us to infer that. It should be determined whether, in the case of the Estonian and Võru narratives, the respondent could have deployed different roles or whether the reasons for thematic differences in the narratives could be because of the separate narratological identities connected to different languages.

The emotional repertoires used in the Estonian and Võru narratives are rather similar to each other. The differences in the emotional repertoires, and also in the themes, may arise most likely from the fact that the Estonian and the Võru dialects trigger different socio-cultural frameworks in the respondent. Estonian, the language the respondent uses in education and professionally, helps her to distance herself from the events. Using the Võru dialect, the respondent describes the Soviet hierarchical *sovkhos* environment. However, the Finnish emotional repertoire of the respondent is rather one-sided because of the limited language means.

The present case study is part of a more extensive study that includes narratives by ten respondents and is conducted in a similar way to the present case study.

### Abbreviations

1, 2, 3	1st, 2nd, 3rd person
L1	first language
L2	second language
Pl	plural
Sg	singular

### References

- Altarriba, Jeanette 2003. Does carino equal “liking”? A theoretical approach to conceptual nonequivalence between languages. – *International Journal of Bilingualism*, 7 (3), 305–322. <http://dx.doi.org/10.1177/13670069030070030501>
- Anooshian, Linda; Hertel, Paula 1994. Emotionality in free recall: Language specificity in bilingual memory. – *Cognition & Emotion*, 8 (6), 503–514. <http://dx.doi.org/10.1080/02699939408408956>
- Bond, Michael; Lai, Tat Ming 1986. Embarrassment and code-switching into a second language. – *Journal of Social Psychology*, 126 (2), 179–186.

- Dewaele, Jean-Marc; Pavlenko, Aneta 2001–2003. Web questionnaire “Bilingualism and “Emotions”. University of London.
- Dewaele, Jean-Marc 2004. The emotional force of swearwords and taboo words in the speech of multilinguals. – *Journal of Multilingual and Multicultural Development*, 25 (2-3), 204–222. <http://dx.doi.org/10.1080/01434630408666529>
- Ehala, Martin 2004. Keelehoold ja mitmikidentiteet. – Karl Pajusalu, Jan Rahman (Toim.). *Kiil ja hindätiidmine / Keel ja identiteet*. Võro Instituudi Toimõndusõq 16. Võru: Võro Instituut.
- Ervin-Tripp, Susan 1964. Language and TAT content in bilinguals. – *Journal of Abnormal and Social Psychology*, 68, 500–507. <http://dx.doi.org/10.1037/h0044803>
- Fludernik, Monika 2003. Chronology, time, tense and experientality in narrative. – *Language and Literature*, 12 (2), 117–134. <http://dx.doi.org/10.1177/0963947003012002295>
- Javier, Rafael; Barroso, Felix; Muñoz, Michele 1993. Autobiographical memory in bilinguals. – *Journal of Psycholinguistic Research*, 22 (3), 319–338. <http://dx.doi.org/10.1007/BF01068015>
- Koven, Michèle 1998. Two languages in the self / the self in two languages: French-Portuguese bilinguals’ verbal enactments and experiences of self in narrative discourse. – *Ethos*, 26 (4), 410–455. <http://dx.doi.org/10.1525/eth.1998.26.4.410>
- Lindström, Liina; Kalmus, Mervi; Klaus, Anneliis; Bakhoff, Liisi; Pajusalu, Karl 2009. Ainususe 1. isikule viitamise eesti murretes. – *Emakeele Seltsi aastaraamat*, 54 (2008), 159–185.
- Marian, Viorica; Kaushanskaya, Margarita 2004. Self-construal and emotion in bicultural bilinguals. – *Journal of Memory and Language*, 51 (2), 190–201. <http://dx.doi.org/10.1016/j.jml.2004.04.003>
- Marian, Viorica; Kaushanskaya, Margarita 2008. Words, feelings, and bilingualism: Cross-linguistic differences in emotionality of autobiographical memories. – *The Mental Lexicon* 3 (1), 72–90. <http://dx.doi.org/10.1075/ml.3.1.06mar>
- Pajusalu, Renate 1999. Deiktikud eesti keeles. <http://www.folklore.ee/seminar/deiktikud> (22.09.2012).
- Panayiotou, Alexia 2004. Switching codes, switching code: Bilinguals’ emotional responses in English and Greek. – *Journal of Multilingual and Multicultural Development*, 25 (2-3), 124–139. <http://dx.doi.org/10.1080/01434630408666525>
- Pavlenko, Aneta 2002. Bilingualism and emotions. – *Multilingua*, 21 (1), 45–78. <http://dx.doi.org/10.1515/mult.2002.004>
- Pavlenko, Aneta 2006. Bilingual minds: emotional experience, expression, and representation. – Aneta Pavlenko (Ed.). *Bilingual Selves*. Clevedon: Multilingual Matters, 1–33.
- Pavlenko, Aneta 2007. Autobiographic narratives as data in applied linguistics. – *Applied Linguistics*, 28 (2), 163–188. <http://dx.doi.org/10.1093/applin/amm008>
- Pavlenko, Aneta 2008. Narrative analysis. – L. Wei, G. M. Moyer (Eds.). *The Blackwells Guide to Research Methods in Bilingualism and Multilingualism*. Blackwell Publishing Ltd, 311–325.
- Pavlenko, Aneta; Dewaele, Jean-Marc 2002. Emotion vocabulary in interlanguage. – *Language Learning*, 52 (2), 263–322. <http://dx.doi.org/10.1111/0023-8333.00185>
- Pavlenko, Aneta; Driagina, Viktoria 2007. Russian emotion vocabulary in American learners’ narratives. – *The Modern Language Journal*, 91 (2), 213–234. <http://dx.doi.org/10.1111/j.1540-4781.2007.00541.x>
- Piller, Ingrid 2002. Passing for a native speaker: Identity and success in second language learning. – *Journal of Sociolinguistics*, 6 (2), 179–208. <http://dx.doi.org/10.1111/j.1540-4781.2007.00541.x>
- Potts, Christopher; Schwarts, Florian 2010. Affective ‘this’. – *Linguistic Issues in Language Technology*, 3, 1–30.

- Schiffrin, Deborah 1981. Tense variation in narrative. – *Language* 57 (1), 45–62. <http://dx.doi.org/10.2307/414286>
- Verschik, Anna 2010. Keeleliste elulugude uurimisvõimalusi: Dagmar Normeti mitmekeelne lapsepõlv Eestis. – *Methis. Studia humaniora Estonica*, 5/6, 175–188.
- Verschik, Anna 2012. Practicing receptive multilingualism: Estonian–Finnish communication in Tallinn. – *International Journal of Bilingualism*, 16 (3), 265–286. <http://dx.doi.org/10.1177/1367006911426465>
- Wierzbicka, Anna 2004. Bilingual Lives, Bilingual Experience. – *Journal Of Multilingual and Multicultural Development*, 25 (2-3), 94–104. <http://dx.doi.org/10.1080/01434630408666523>
- Wolfson, Nessa 1979. The conversational historical present alternation. – *Language* 55 (1), 168–182. <http://dx.doi.org/10.2307/412521>

**Liina Tammekänd** (Tartu Ülikool) on erialalt tõlkija, uurimisvaldkonnaks on mitmekeelne narratiiv. [liina.tammekand@ut.ee](mailto:liina.tammekand@ut.ee)

# LÕUNA-EESTLASE MITMEKEELSUS – EESTI, VÕRU JA SOOME NARRATIIVIDE VÕRDLUS

**Liina Tammekänd**

Tartu Ülikool

Artikkel käsitleb juhtumiuuringut, mille aluseks on 31-aastase Võrumaalt pärit naise esitatud kolm samateemalist isikliku kogemuse narratiivi võru murdekeeles, eesti ja soome keeles ning tema keelelulugu.

Autor salvestas, transkribeeris ja analüüsis narratiive pidades silmas nelja uurimisküsimust: kas narratiivid erinevad üksteisest struktuuriliselt ja temaatiliselt, kas keelejuhi esimesena omandatud keel on talle emotsionaalselt lähemal, kas keelejuht demonstreerib erinevates keeltes osidentiteedi erinevaid tahke ning kas keelejuht kasutab erinevates keeltes erinevaid emotsioonirepertuaare.

Eesti ja võru narratiivid on struktuurilt identsed, kuid temaatiliselt erinevad. Soome narratiivi struktuur sarnaneb üldiselt teiste narratiivide struktuuriga, kuid temaatika nii selgelt välja ei joonistu. Temaatika ning valitud emotsioonisõnade tõttu tundub, et võru murdekeel on keelejuhile võrreldes teiste keeltega emotsionaalselt lähemal. Lingvistiline ja temaatiline analüüs ei luba hetkel oletada, et keelejuht demonstreerib erinevates keeltes osidentiteedi eri tahke. Erinevused eesti ja võru emotsioonisõnavarades tulenevad peamiselt narratiivide erinevast temaatikast. Soome narratiivi ebaselge temaatika ning ühekiuline emotsioonisõnavara tulenevad ilmselt sellest, et keelejuht on soome keele omandanud töökontekstis.

**Võtmesõnad:** mitmekeelsus, suuline narratiiv, emotsioonisõnavara, eesti keel, võru murre, soome keel

# ÕIGUSLIKUD VÄLJAKUTSED JA VÕIMALIKUD LAHENDUSED KEELERESSURSSIDE ARENDAMISEL

Arvi Tavast, Heiki Pisuke, Aleksei Kelli

**Ülevaade.** Keeleressursside arendamisel ei osata praktikas sageli arvestada lähtematerjali puudutavate õiguslike piirangutega, mis tekitab tarbetuid õigusriske ja võib takistada loodud ressursside sihipärasest kasutamist ja levitamist. Artikli eesmärk on kaardistada levinumad probleemid ja pakkuda kehtivale õigusele tuginevaid lahendeid, viimaste puudumisel aga teoreetilisi ja praktilisi tõlgendusi.

Autoriõiguslikult kaitstavate teoste kasutamiseks keeleressurssides esitatakse lisaks litsentsipõhisele meetodile kaks teoste vaba kasutamise viisi, tsiteerimine ja nn teadustöö erand, koos vastavate täpsemate juhistega. Lisaks kirjeldatakse täiendavaid tingimusi, mis puudutavad suulist ja eriti spontaanses olukorras salvestatud kõnet, varem avaldamata teoseid, tuletatud teoste loomist ja keeleressursside avalikustamist.\*

**Võtmesõnad:** intellektuaalne omand, autoriõigus, litsentsid, isikuandmete kaitse, teoste vaba kasutamine, teadustöö erand, tsiteerimine

## 1. Sissejuhatus

Eesti Keeleressursside Keskuse (EKRK) praktilises tegevuses on päevakorda kerkinud mitmed õiguslikud küsimused: millist keelematerjali tohib keele uurimiseks ja keeletehnoloogiliste lahenduste loomiseks kasutada, millistel tingimustel, kuidas need tingimused mõjutavad uute keeleressursside hilisemat levitamist jt. Võimalikke vastuseid nendele küsimustele tuleb otsida intellektuaalse omandi<sup>1</sup> jts õigusaktidest, teooriast ja praktikast. Just nimelt otsida, sest puuduvad valmislahendused ja kuni viimase ajani oli see valdkond jäänud välja nii filoloogide kui ka juristide uurimisalast. Autoriõigus on keeleressursside arendamisel ja kasutamisel

\* Artikkel on valminud Euroopa Sotsiaalfondi kaasabil "Teadus- ja innovatsioonipoliitika seire" programmi raames. Kõik selles artiklis avaldatud seisukohad kuuluvad autoritele ning ei väljenda mingil viisil Euroopa Liidu Nõukogu või muude Euroopa Liidu institutsioonide ametlikke seisukohti.

<sup>1</sup> Intellektuaalse omani legaalmääratlus on antud Ülemaailmse Intellektuaalse Omandi Organisatsiooni (WIPO) asutamise konventsiooni artiklis 2, mis sätestab, et intellektuaalne omand sisaldab õigusi, mis tulenevad intellektuaalsest tegevusest tööstuse, teaduse, kirjanduse ja kunsti alal (RT II 1993, 25, 55).

keskse tähendusega (autoriõiguse kohta üldiselt vt Pisuke 2006). Kas selguse huvi-des peaks autoriõigus nägema ette eraldi keeleressursside loomise regulatsiooni ja vastava teoste kasutamise erandi või võimaldab ka kehtiv autoriõiguse süsteem keeleressursside loovat kasutamist ja arendamist, see on väljakutse, millele vastuse otsimise on allkirjutanud endale võtnud (vt ka Kelli jt 2012).

Keeleressursside koostamine, eriti eesti keele puhul, ei pea reeglina silmas ärilisi eesmärgi. Ka õiguste omajad – autorid ja kirjastajad – on seni seda protsessi toetanud, andes lube oma teoste kasutamiseks. Õiguslikud probleemid tõusetuvad eelkõige seoses nende lubade puuduliku juriidilise vormistamisega seadusele vastavate litsentsidena. Praktikast tekivad probleemid samuti seoses loa taotlemata jätmisega, isikuandmete kaitsega ja kõne salvestamisega spontaanses olukorras.

Käesoleva artikli esimene osa liigitab keeleressurssid õiguslike probleemide alusel, mis võivad nende puhul tekkida ja raskendada nende kättesaadavaks tegemist EKRC ja CLARIN ERICu<sup>2</sup> kaudu. Teises osas analüüsivad autorid, milliste õiguslike aspektidega tuleb keeleressursside koostamisel arvestada, keskendudes keeletehnoloogias seni raskusi valmistanud teemadele. Kolmandas osas antakse autorite soovitusel keeleressursside edasise kogumise jaoks, millega arvestamine võimaldaks vältida õiguslike probleemide tekkimist kogu protsessi jooksul.

Käesoleva artikli kontekstis määratletakse keeleressurssi andmebaasina, mis sisaldab mitmeid kirjalikke ja suulisi tekste. Tekstide valimine on loominguiline tegevus, mistõttu on keeleressurss kui andmebaas autoriõiguslikult kaitstav teosena.<sup>3</sup>

## 2. Keeleressursside liigitus õiguslike väljakutsete seisukohast

### 2.1. Leksikaalsed andmebaasid

Leksikaalsete andmebaaside hulka kuuluvad sõnastikud, tesaurused, leksikonid, terminibaasid ja muud inim- ja/või masinloetavad andmebaasid, mis lihtsamal juhul sisaldavad üksiksõnu ja levinud väljendeid ühes või mitmes keeles. Lisaks võivad leksikaalsed andmebaasid sisaldada näitelauseid, mille puhul võib olla oluline nende autentsus, st pärinemine mõne teise autori teosest vastandina leksikograafi omaloomingule. Samuti võivad need baasid sisaldada enam või vähem formaliseeritud infot keelendite päritolu, kasutusviisi (muuttüüp, reksioon vms), tähenduse (täenduseseosed, mõisteseosed) või muude sõna või mõiste parameetrite kohta ning tähenduse verbaalseid kirjeldusi (seletused, määratlused).

Siinses artiklis käsitletavad õiguslikud väljakutsed leksikaalsete andmebaaside juures piirduvad autentsete näitelauseite kasutamise teiste autorite teostest

<sup>2</sup> CLARIN ERIC [www.clarin.eu](http://www.clarin.eu) (22.09.2012).

<sup>3</sup> AutÕS § 4 lg 3 p 22 sätestab, et teosed, millele tekib autoriõigus, on mh "teoste kogumikud ja informatsiooni kogumikud (sealhulgas andmebaasid). Andmebaas käesoleva seaduse tähenduses on süstemaatiliselt või metoodiliselt korraldatud iseseisvate teoste, andmete või muu materjali kogu, mis on individuaalselt kättesaadav elektrooniliste või muude vahendite abil. [...] Käesoleva seaduse alusel kaitstakse autoriõigusega andmebaasi, mis oma sisu valiku ja korralduse tõttu on autori enda intellektuaalse loominguga tulemus, ning ei kohaldata mingit muud kriteeriumi."

Andmebaas võib olla kaitstav ka AutÕS 8<sup>1</sup> peatüki alusel kui *sui generis* andmebaas. See tähendab, et õiguste saamiseks andmebaasile piisab investeringu tegemisest andmebaasi loomisesse. Autoriõiguse seadus määrab *sui generis* andmebaasi tegijana isikut, kes on "teinud kas laadilt, väärtuselt või suuruselt olulise investeringu selle andmebaasi sisu olemate andmete kogumiseks, omandamiseks, kontrollimiseks, süstematiseerimiseks või kättesaadavaks tegemiseks" (AutÕS § 75<sup>3</sup> lg 1). Iseenesest on keeleressurssid kaitstavad *sui generis* andmebaasina, kuid sellel puudub oluline tähendus, sest teosena kaitstava andmebaasi õiguskaitses on ulatuslikum. Põhiline eelis *sui generis* kaitsesega seoses on see, et investeringute jätkumisel pikeneb *sui generis* 15-aastane kaitsetähtaeg *ad infinitum* (vt AutÕS § 75<sup>7</sup>).

(vt 3.1). Lisaks on valdkonnas üldlevinud praktika koostada uusi leksikaalseid andmebaase varasemate põhjal, kuid selle kohta on hea tava suhteliselt hästi välja kujunenud ja võimalikud lahkarvamused püsivad valdkonna ekspertide siseriingis, mistõttu seda teemat siin ei käsitleta.

## **2.2. Korpused märgendatuse järgi**

Tekstikorpused kui suured struktureeritud ja eesmärgipärasel viisil koostatud tekstikogumid koosnevad tervenisti teiste autorite teostest või nende osadest. Lihtsamal juhul ehk märgendamata korpuse juures on koostaja töö piirdunud tekstide valimise ja süstematiseerimisega. Selliseid korpuseid on võimalik (ja keeletehnoloogiliste rakenduste seisukohalt vajalik) koostada poolautomaatselt ja väga suurtes mahtudes, mis tõstab päevakorda sisendlitsentside hankimise töömahukuse ja tekitab vajaduse otsida alternatiivseid viise ressursside seaduslikuks kogumiseks (vt 4.2–4.3, vt ka Vider jt 2013 siinses kogumikus).

Teine variant ehk märgendatud korpus (sh puudepank kui teatud viisil süntaktiliselt märgendatud korpus) sisaldab lisaks teiste autorite tekstidele ka lingvisti ulatuslikku panust nende märgendamisel ehk eksplitsiitse info lisamisel tekstis esinevate keelendite sõnaliigi, häälduse, lauses täidetava funktsiooni, tähenduse vms kohta. Teatud piirini on märgendada võimalik poolautomaatselt, kuid tõeliselt väärtuslikud on just käsitsi märgendatud korpused, sest need annavad keeletehnoloogiale juurde sellist infot keele kohta, mida automaatselt tekitada ei ole tehnika selle hetke arengutaseme juures võimalik. Märgendatud korpuste täiendav õiguslik väljakutse seisnebki teiste autorite teoste muutmises märgenduse lisamise teel ehk tuletatud teose loomises. Harjumuspärane viis seda vormistada on kirjutada sõna kohta käiv metainfo iga märgendatud sõna juurde, näiteks selle järele mingite koodide kujul, kuid tehniliselt peaks olema teostatav ka märgenduse hoidmine eraldi failis, et teise autori teksti sõnade vahele midagi kirjutama ei peaks.

## **2.3. Korpused keelematerjali liigi järgi**

Ajalooliselt esimesed ja praegu levinuimad tekstikorpused koosnevad kirjalikest õiguspäraselt avaldatud tekstidest, mille puhul väljakutsed piirduvad intellektuaalomandi temaatikaga.

Mõned korpusele lisatavad tekstid võivad olla ka avaldamata, näiteks koolitööd. Sellise materjali kasutusvõimalused on piiratumad kui avaldatud tekstidel, sest tsiteerimise ja teaduskasutuse luba laieneb vaid avaldatud tekstidele (vt 3.1).

Lisaks kirjutatud tekstidele läheb tarvis ka suulise kõne korpuseid, mis koosnevad helisalvestusest, videosalvestusest ja/või nende transkriptsioonidest, millega lisaks intellektuaalomandile kaasneb ka isikuandmete kaitse ja salvestamise õiguspärasuse temaatika (vt 3.2). Salvestamiseks tuleb küsida luba, kuid ehtsa suulise keelekasutuse uurimiseks vajalikku tõeliselt loomulikku kõnet saab ainult nii, et uuritav ei tea, et tema kõnet salvestatakse. Kui aga salvestatav kõneleja esitab kellegi teise teksti, siis tuleb võtta arvesse ka tolle teksti autoriõigustest tulenevaid piiranguid.

## 2.4. Keelemudelid, grammatikad

Seni nimetatud ressursid ise ei ole vajalikud millekski muuks kui keeleteadusliku või -tehnoloogilise uurimis- ja arendustöö jätkamiseks järgmistel astmetel. Korpuste ja sõnastike põhjal ning nende kasutaja loominguulise panuse abil saab hankida infot selle kohta, milline uuritav keel on või kuidas seda kasutatakse ning luua keelemudeleid, grammatikaid ja muid keele formaalseid kirjeldusi. Mõned neist ei sisalda üldse nende aluseks olevaid korpusetekste sõnast või kinnisväljendist pikeimate fragmentidena, näiteks kui lingvist (või tema projekteeritud arvutisüsteem) vaatab läbi suure koguse autentseid tekste ja jõuab järeldusele, et “dokumente” on eesti keeles oluliselt levinum mitmuse osastava vorm kui “dokumentisid”. Sellisel juhul on tuletuslik seos uuritud tekstide ja saadud järelduse vahel üsna vaieldav. Mõnel teisel juhul võivad tekstide äratuntavad osad keelekirjelduses siiski sisalduda, näiteks kui statistiline trigramm-mudel loetleb kõik uuritavates tekstides leiduvad kolmesõnalised järjendid ja nende tõenäosused, siis needsamad kolmesõnalised järjendid on mudelist ka leitavad.

## 2.5. Tarkvara

Keeleressursside hulka võidakse mõnikord lugeda ka keeletehnoloogiline tarkvara: analüsaatorid, süntesaatorid, sõnastikusüsteemid, otsingu- ja päringusüsteemid, masintõlkesüsteemid, dialoogsüsteemid, märgendusvahendid jms. Nende toodete endiga on küll seotud õiguslikke küsimusi, nt vajavad keeletehnoloogia riiklikust programmist finantseeritud tarkvara litsentsitingimused mingi kompromissi saavutamist finantseerija nõudmiste (et kõik riikliku programmi tulemused oleks kõigile vabalt kasutatavad) ja programmi autorite tööandjate soovide (et tööülesannete täitmise käigus loodud teoste varalised õigused kuuluvad tööandjale ja need tuleks rakendada tööandjale tulu genereerimise teenistusse) vahel. Kuna aga tarkvara eraldi vaadelduna ei sisalda keelematerjali, vaid üksnes kasutab seda, siis tarkvaraga seotud küsimused jäävad siinse artikli käsitusalaast välja.<sup>4</sup>

# 3. Õiguslikke probleeme seoses keeleressurssidega

## 3.1. Kasutatavate teoste autoriõiguslik kaitse

Keeleressursside loomine tugineb suuliste ja kirjalike tekstide kasutamisele, millest suur osa on autoriõiguslikult kaitstavad teosed. Autoriõiguse seadus määratleb teost kui “mis tahes originaalset tulemust kirjanduse, kunsti või teaduse valdkonnas, mis on väljendatud mingisuguses objektiivses vormis ja on selle vormi kaudu tajutav ning reprodutseeritav kas vahetult või mingi tehnilise vahendi abil”. Teos on originaalne, kui see on “autori enda intellektuaalse loomingu tulemus” (AutÕS § 4 lg 2). Eeltoodud regulatsiooni tuleb mõista selliselt, et kui isik on midagi loonud kirjanduse, kunsti või teaduse valdkonnas ning see loomingu on väljendatud mingisuguses tajumist võimaldavas objektiivses vormis, siis on see autoriõiguslikult kaitstav teos. Teose objektiivseks vormiks võib olla näiteks salvestatud suuline

<sup>4</sup> Ka autoriõiguse seadusest tuleneb, et andmebaasi mõiste ei hõlma selle tegemiseks ega käivitamiseks vajaminevat arvutiprogrammi (AutÕS § 4 lg 3 p 22).



kõne, käsikirjaline tekst, trükitud artikkel või raamat, arvutifail jms. Keskseks teose tunnuseks on seejuures originaalsus. Autoriõigusega kaitstavalt teoselt ei nõuta midagi ennenägematut või seniolematut, nagu see on patendiõiguses. Nõutakse lihtsalt autori enda isiklikku panust, mis ei tohi üks-üheselt kopeerida juba olemasolevaid teoseid. Originaalsus ei eelda seega uudsust, vaid on piisav, kui autoril on vähemalt minimaalne enda loomingiline panus. Autoriõigus põhineb kahel eeldusel ehk presumptsioonil: 1) teose originaalsust ja seetõttu ka kaitstust autoriõigusega eeldatakse ning vastupidise tõendamise kohustus lasub vaidlustajal (AutÕS § 4 lg 6) ja 2) isiku, kes avaldab teose oma nime all või oma üldtuntud pseudonüümi või autorimärgi all, autorsust eeldatakse, kuni pole tõendatud vastupidist. Tõendamiskohustus lasub autorsuse vaidlustajal (AutÕS § 29 lg 1). Kõik seniõeldu kehtib nii ainuisikuliselt kui ka kollektiivselt (AutÕS § 30) loodud teoste puhul.

Samas eksisteerib teatud hulk loomingut, mille autoriõiguslik kaitse on *expressis verbis* välistatud. Nii välistab autoriõiguse seadus autoriõigusliku kaitse alt ideed, mõisted, teooriad ja meetodid, rahvaloominguteosed, õigusaktid, haldusdokumendid, kohtulahendid ja nende ametlikud tõlked, faktid ja andmed ning muud seaduses otseselt nimetatud inimloomingu tulemused (AutÕS § 5).

Väga oluline on veel pidada silmas kahte printsiipi: 1) autoriõigusliku kaitse tähtajalisust ja 2) territoriaalsust.

- 1) Autoriõiguse üldine täheaeg on autori kogu elu ja 70 aastat pärast tema surma (70 aastat *post mortem auctoris* ehk *pma*). Seda tähtaega hakatakse lugema autori surma-aastale järgneva aasta 1. jaanuarist (AutÕS § 38 lg 1 ja § 43). On mõningad erisätted, mis kehtestavad kas lühema või isegi pikema tähtaja<sup>5</sup>, kuid kõige sagedamini esinev üldreegel on 70 a *pma*. Pärast teose kaitsetähtaja lõppu võivad kõik isikud seda vabalt kasutada ehk teos on nn ühises kasutuses (ingl *public domain*). Kasutaja põhiliseks kohustuseks on üksnes viitamine teose autorile, sest autoriõiguse seaduse kohaselt “[i] siku konkreetse teose autoriks olemist (teose autorsust), autori nime ning autori au ja väarikust kaitstakse tähtajatult” (AutÕS § 44 lg 1).

Vanade tekstide kasutamisel, mille puhul on möödunud 70 aastat autori surmast, on EL liikmesmaades kehtestatud üks erand juhuks, kui autori teos jäi tema eluajal või hiljem avaldamata.<sup>6</sup> Autoriõiguse seaduse kohaselt saab isik, kes pärast autoriõigusliku kaitse lõppu avaldab varem avaldamata teose, sellele teosele autori varalised õigused 25 aastaks (AutÕS § 74<sup>1</sup> lg 1). Regulaatsiooni mõte on motiveerida varem avaldamata teoste uurimist ja avaldamist ja kompenseerida sellega kaasnevad võimalikud kulud. (Täiendavaks analüüsiks vt Walter ja Lewinski 2010: 568–578.)

- 2) Territoriaalsuse printsiip tähendab seda, et igal maal kehtib oma autoriõiguse normistik ja põhimõtteliselt ei pea kaitsma välismaiste autorite välismaal avaldatud teoseid. Autoriõigusliku kaitse territoriaalsuse printsiipi laiendatakse rahvusvaheliste lepingute sõlmimise ja nendega ühinemise teel. Rahvusvaheliste konventsioonidega ühinemine toob kaasa kohustused

<sup>5</sup> Näiteks anonüümselt või pseudonüümi all avaldatud teoseid, kollektiivseid teoseid (AutÕS § 31) või töökohustuste täitmise korras loodud teoseid (§ 32) kaitstakse 70 aastat pärast selle teose õiguspärasest avalikustamisest, kuid see tähtaeg võib teatud tingimustel pikeneda ka üldtähtajani (§ 40 ja 41). Samas autoriõigus teosele, mille on loonud kaks või enam isikut oma ühise loomingulise tegevusega, kehtib teised autorid üleelanud ühise autori eluajal ja 70 aastat pärast tema surma (§ 39).

<sup>6</sup> Avaldamiseks loetakse teose kättesaadavaks tegemist määramata isikute ringile ehk seaduse sõnastuses: “Teos loetakse avaldatuks, kui teos või teose mis tahes vormis reprodutseeritud koopiad on autori nõusolekul antud üldsusele kasutamiseks koguses, mis võimaldab üldsusel sellega tutvuda või seda omandada. Teose avaldamiseks loetakse muu hulgas teose trüki väljaandmist, teose eksemplaride panemist müügile, jaotamist, laenutamist, rentimist ja muul viisil tasuta või tasu eest kasutada andmist” (AutÕS § 9).

ühtlustada teatud konventsioonis sisalduvad üldprintsüübid (näiteks mida tuleb kaitsta ja kui pika tähtaja jooksul) ja kaitsta oma territooriumil kõigi konventsiooniga ühinenud liikmesriikide autorite teoseid<sup>7</sup>. EL liikmesmaade puhul tulenevad teatud täiendava ühtlustamise ja vastastikuse kaitse kohustused EL direktiividest. Konventsiooniga mitteühinenud riikide autorite teoseid (näiteks nendes riikides väljaantud sõnaraamatuid või muid keeleressursse) Eestis kaitsma ei pea. Selliseid riike on aga võrreldes konventsioonidega ühinenud riikidega tühine vähemus.

Keeleressursside loomise seisukohalt on eelneval väga oluline tähendus. Nimelt, enamus tekste, mida keeleressursside loomisel kasutatakse, on autoriõiguslikult kaitstavad. Sisuliselt saab nende kasutamine tugineda kas õiguste omaja nõusolekule või siis teose vaba kasutamise eranditele.

Keeleressursside kasutamisel omab teatud juhtudel tähendust autoriõiguse eristamine ühest teisest intellektuaalse omandi instituudist – autoriõigusega kaasnevatest õigustest. Näiteks teose esitaja, fonogrammitootja ja ringhäälinguorganisatsiooni (ehk seaduse sõnastuses televisiooni- ja raadioteenuse osutaja) õigused on iseseisvad õigused ja autoriõiguse sätted nende suhtes ei kehti (AutÕS 8. peatükk).

### 3.2. Suulise ja spontaanses olukorras salvestatud kõne kasutamine

Suulise ja spontaanses olukorras salvestatud kõne kasutamine keeleressurssides on juriidiliselt komplitseeritud valdkond. Üldjoontes võib praktikas sagedamini esinevad probleemid jagada järgmistesse valdkondadesse kuuluvaiks: üldsusele avaldamata tekstide kasutamine, autori ja esitaja õigused ning isiku privaatsuse riive.

Autoriõiguslik kaitse laieneb ka suulistele teostele, mis tähendab, et kaitse saamiseks sellistele teostele ei ole vaja teksti eelnevalt kirja panna. Teksti suulist esitamist loetakse teosele objektiivse vormi andmiseks, kuid see peab olema kuulajate poolt tajutatav ja reprodutseeritav (näiteks salvestamise või üleskirjutamise teel). Kaitse alguspunktiks on teose loomine ehk selle väljendamine suulises vormis.<sup>8</sup> Suuliste teoste puhul kehtivad kõik autoriõiguse üldprintsüübid, sealhulgas ei nõuta loodud teose deponeerimist või registreerimist (AutÕS § 7). Kaitstavad on seejuures nii avalikustatud kui ka avalikustamata teosed. Avalikustamine tähendab teose avaldamist või üldsusele suunamist (AutÕS § 8). Seega pole tähtsust, kas suuline tekst loodi intervjuu käigus ühele ajakirjanikule või rahvusvahelise konverentsi ettekande esitamise ajal.

Keeleressursside kontekstis kasutatav suuline kõne on tihti avalikustamata. Avalikustamata teoste puhul on aga seadusega kehtestatud üks väga oluline lisapiirang. Teose vaba kasutamine ilma autori nõusolekuta võib toimuda ainult avalikustatud ehk autori enda selgel nõusolekul üldsuse ette viidud teoste puhul. Teose tsiteerimine ning kasutamine teaduslikel ja õppe-eesmärkidel eeldab, et see teos on avaldatud, ja lisatingimusena – avaldatud õiguspäraselt, seega autori selgelt väljendatud soovil ja tahtel. Asjaolu, et teos on avaldamata, ei tähenda, et seda ei võiks üldse kasutada. Kui teose autor ehk isik, kelle suulist kõnet salvestatakse või

<sup>7</sup> Tähtsaim autoriõiguse rahvusvaheline leping, millega ka Eesti ühines, on Berni kirjandus- ja kunstiteoste kaitse konventsioon (Berni konventsioon; RT II, 1994, 16, 49). Teatud kohustused Eesti jaoks tulenevad ka Maailma Kaubandusorganisatsiooni asutamislepingu lisast 1C ehk Intellektuaalomandi õiguste kaubandusaspektide lepingust (TRIPS lepingust; RT II, 1999, 22, 123) ning Maailma Intellektuaalse Omandi Organisatsiooni autoriõiguse lepingust (RT II, 2006, 14, 39).

<sup>8</sup> AutÕS § 4 lg 3 p 4 on autoriõigusega kaitstavate teoste näidisloetelus fikseeritud "kõned, loengud, ettekanded, jutlused jt teosed, mis koosnevad sõnadest ja on väljendatud suuliselt (suulised teosed)".

üles kirjutatakse, annab sellise salvestise või kirjalikult fikseeritud teksti kasutamiseks oma nõusoleku, siis on kasutamine õiguspärane.

Suulise kõne autoriõiguslikud küsimused võivad olla seotud mitme isiku loome-tegevusega. Näiteks kui esitatava teksti autoriks on üks isik ja see esitatakse teise isiku poolt, siis kehtivad teksti autori suhtes autoriõiguse normid, esitaja<sup>9</sup> suhtes aga autoriõigusega kaasnevate õiguste kohta käivad normid. Teksti kui autoriõiguslikult kaitstava teose kasutamisel (esitamisel, salvestamisel või muul reprodutseerimisel) saab olla kaks õiguslikku alust: 1) õigustatud isiku ehk õiguste omaja<sup>10</sup> nõusolek, mis on selgelt väljendatud loa (ehk litsentsi) või õiguste loovutamise vormis, või 2) seaduses ammendavalt sätestatud teose vaba kasutamise sätted.

Suulise ja spontaanses olukorras salvestatud kõne lülitamine keeleressursside hulka omab lisaks autoriõigusele tihedat kokkupuudet isikuandmete kaitsega, kuna kõneleja võib olla tuvastatav ning tema kõne kasutamine tähendab ka isikuandmete töötlust. Isikuandmete kaitse seaduse kohaselt on isikuandmed “mis tahes andmed tuvastatud või tuvastatava füüsilise isiku kohta, sõltumata sellest, millisel kujul või millises vormis need andmed on” (IKS § 4 lg 1). Vastavalt seadusele on isikuandmete töötlemine “iga isikuandmetega tehtav toiming, sealhulgas isikuandmete kogumine, salvestamine, korrastamine, säilitamine, muutmine ja avalikustamine, juurdepääsu võimaldamine isikuandmetele, päringute teostamine ja väljavõtete tegemine, isikuandmete kasutamine, edastamine, ristkasutamine, ühendamine, sulgemine, kustutamine või hävitamine, või mitu eelnimetatud toimingut, sõltumata toimingute teostamise viisist ja kasutatavatest vahenditest” (IKS § 5). Isikuandmete töötlemine on lubatud üksnes puudutatud isiku nõusolekul või seaduse alusel (IKS § 10).

Keeleressursside seisukohalt taanduvad isikuandmete kaitse küsimused eelkõige sellele, et lisaks autoriõigusest tulenevale nõusolekule teose kasutamiseks on vajalik ka nõusolek isikuandmete töötlemiseks. Nõusolekuta isikuandmete töötlemine on karistatav (IKS § 42 ja 43).

Spontaanses olukorras salvestatud kõne kogumisega seonduvad probleemid on juriidiliselt kõige komplitseeritumad. Spontaanse olukorraga on tegemist siis, kui rääkija ei tea, et tema kõnet salvestatakse, mis tagab kõnelemisel täieliku vabaduse. Salaja lindistamisega võidakse samas riivata isikute privaatsusõigust. Jälitustegevuse seadus loeb andmete varjatud kogumist jälitustoiminguks, mida tohib läbi viia jälitusasutus (JTS § 12 lg 1 p 1) ning karistusseadustik näeb koguni ette kuni kolmeaastase vangistuse, juhul kui jälitustegevuseks seadusliku õigusega isik jälgib teist isikut tema kohta andmete kogumise eesmärgil (KarS § 137). Karistusseadustiku kommentaarides on rõhutatud, et nimetatud koosseisuga kaitstakse nii elulisi kui ka tööalaseid andmeid, kusjuures ei ole oluline nende kvaliteet. Samuti on isiku vabadus piiratud, kui ta peab pidevalt kartma, et tema tegevust salaja talletatakse (Sootak, Pikamäe 2009: 405).

<sup>9</sup> Autoriõiguse seadus loeb teose esitajaks “näitlejat, lauljat, muusikut, tantsijat, muud isikut või kollektiivi, kes näitleb, laulab, deklameerib, mängib muusikariistal või muul viisil esitab kirjandus-, kunsti- või rahvaloominguteoseid või juhendab teisi isikuid teoste esitamisel, aga samuti isikut, kes esitab estraadi-, tsirkuse-, nukuteatri- jms numbreyid” (AutÕS § 64).

<sup>10</sup> Õiguste omajaks ei pruugi olla tingimata autor, vaid selleks võib olla ka näiteks kirjastaja, ülikool, uurimisinstituut või muu tööandja.

## 4. Võimalikud lahendused

### 4.1. Korrektsed litsentsid

Kui keeleressursside arendamine toimub ärilistel eesmärkidel, siis on vajalik õiguste omajaga kirjaliku litsentsilepingu sõlmimine. See on nn litsentsipõhine keeleressursside kasutamine. Litsentsileping on autorilepingu liik ja selles peavad kajastuma kõik seaduses nõutavad elemendid.<sup>11</sup> Eriti tuleks tähelepanu pöörata selle fikseerimisele, millisel eesmärgil, ulatuses ja vormis litsentseeritavat teost kavatsetakse kasutada. Oluline on kohe välja selgitada, kellele kuuluvad autoriõigused. Keeleressursside kasutamisel tuleb arvestada asjaoluga, et autor ei pruugi olla tingimata õiguste omaja ning suure tõenäosusega võivad autoriõigused kuuluda teose kirjastajale või tööandjale, mistõttu tuleb just nendega sõlmida litsentsileping. Kirjastajaga litsentsilepingu sõlmimise eeliseks keeleressursside looja seisukohalt on asjaolu, et kirjastajale kuuluvad reeglina ka paljude teiste teoste autoriõigused ning läbi-rääkimise objektiks saab olla korruga suurema hulga teoste kasutamine. Üldjuhul ei saa eeldada, et taoline litsentsilepingu sõlmimine riivaks kuidagi kirjastaja ärihuvisid, sest teoste kasutus keeleressursside kontekstis on erinev kui see, kuidas kirjastaja neid kommertsialiseerib (nt müüb raamatuna, e-raamatuna, annab filmi või teleõiguseid jne). Keeleressursside arendamisel kasutatakse teoseid eelkõige “toomaterjalina” ning teoste kirjanduslik või majanduslik väärtus ei ole oluline.

Litsentsil põhineva keeleressursside loomise suurim puudus ilmneb eelkõige massikasutuse tingimustes. Kuna keeletehnoloogia kasutab üha enam statistilisi meetodeid, siis läheb keeleressurssse vaja väga suures mahus. See omakorda muudab lepingute sõlmimise kõigi autoriõiguste omajatega väga töö- ja ressursimahukaks. Seetõttu, kui teoste kasutamine plaanitud viisil ei ole ilma autori loata lubatud, ja autori loa hankimine on liiga keeruline, tasuks uurida, millisel viisil tohib teost kasutada ilma autori nõusolekuta. Keeleressursside puhul tuleb niisugustest viisidest praegu kehtiva autoriõiguse süsteemi raames kõne alla peamiselt kaks: tsiteerimine ja nn teadustöö erand.

### 4.2. Tsiteerimine

Teadustöös on tsiteerimine väga levinud teiste autorite teoste kasutamise viis. Ka keeleressurssid võivad sisaldada tsitaate teiste autorite teostest. Tsiteerimiseks on seadusega kehtestatud väga konkreetsed reeglid<sup>12</sup>. Tsiteerida tohib autori nõusolekuta ja autoritasu maksmata (AutÕS § 19 p 1), kuid ainult avaldatud teoseid. Vajalik on autorile viitamine, eeldusel, et autor on teosele märgitud.

Tsiteerimise tingimused on seaduse üldnormi alusel samad kõigile teose liikidele, järelikult ka nii keeleressursside või näiteks teadusartiklite puhul. Konspektiivselt on need järgmised.

<sup>11</sup> Vastavalt AutÕS § 48<sup>1</sup> “Autorilepingu sisu” tuleb autorilepingus fikseerida: 1) teose kirjeldus (vorm, maht, nimetus jms); 2) üleantavad õigused, õigused, mille osas antakse luba, litsentsilepingu liik (liht- või ainulitsentsileping) ja all-litsentsi andmise õigus; 3) teose kasutamise viis ja territoorium; 4) autorilepingu kehtivuse tähtaeg ja teose kasutamise algustähtaeg. Samuti määratakse poolte kokkuleppel, kas leping on tasuta või tasuline. Autoritasu maksmise viis (protsent teose müügihinnast, konkreetselt kindlaksmääratud summa, protsent teose kasutamisel saadavast kasumist vms), tasu suurus, selle väljamaksmise tähtaeg ja kord määratakse autorilepingu poolte kokkuleppel.

<sup>12</sup> AutÕS § 19 sätestab: “Autori nõusolekuta ja autoritasu maksmiseta, kuid kasutatud teose autori nime, kui see on teosel näidatud, teose nimetuse ning avaldamisallika kohustusliku äranäitamisega on lubatud:

1) õiguspäraselt avaldatud teose tsiteerimine ja refereerimine motiveeritud mahus, järgides refereeritava või tsiteeritava teose kui terviku mõtte õige edasiandmise kohustust.”

- 1) Tsiteeritav teos peab olema õiguspäraselt avaldatud.
- 2) Iga tsitaadi juures peab reeglina olema viide autori nimele, teose pealkirjale ja avaldamisallikale, juhul kui need rekvisiidid on teosel näidatud. Samas tuleb arvestada ka sellega, et konkreetse valdkonnas on oma viitamise spetsiifika. Viitamise kohustus võib olla täidetud ka siis, kui allikate loetelu antakse lõpus ning iga tsitaadi juures viidet ei tehta. Näiteks on avaldatud õpikuid, kus kasutatud teosed viidatakse iga peatüki ees ning teksti sees puuduvad igasugused viited.
- 3) Tsitaat on teksti reprodutseerimine täpselt samal kujul (samas sõnastuses) nagu see on algse teoses; reeglina kasutatakse jutumärke.
- 4) Tsitaadiga ei tohi algse teose mõtet valesi edasi anda (ei tohi moonutada algse teose mõtet, sest tegemist on autori au ja väärikuse kaitsega). Keele-ressursside kontekstis on selle nõude vastu eksimine suhteliselt vähe tõenäoline.
- 5) Tsiteerida tohib motiveeritud mahus.

Seaduses ei ole sätestatud, mis on motiveeritud maht. Seadusega polegi seda võimalik teha, sest erinevate teose liikide puhul on motiveeritud maht erinev. Tsiteerida saab kõiki teose liike – kirjandusteoseid, muusikateoseid, arhitektuuriteoseid jne. Tsiteerimine põhineb väljakujunenud heal taval, mis on kohati väljendatud ka näiteks teadusajakirjade, ajakirjandus- ja ringhäälinguorganisatsioonide või autorite ühingu poolt väljatöötatud reeglites. Sellised tsiteerimise hea tava reeglid võiksid olla kehtestatud ka keeleressursside kasutamisel. Nähtavasti tuleks need välja töötada EKRR ja kirjastajate organisatsioonide poolt ühiselt ja kehtestada vastava hea tava kokkuleppena.

Ühtse lähenemise tagamiseks ja üldise mahuorientiiri andmiseks teeme siinkohal ettepaneku lugeda keeleressursside puhul motiveeritud mahuks järgmist:

- tsiteerida tohib kuni 5–10% tsiteeritava teose mahust, kuid mitte rohkem kui 10 lk (18 000 täheruumi),
- üks järjestikune tsitaat ei tohi ületada 1 lk (1800 täheruumi) ega olla teraviklik alamteos (nt üks terve novell novellikogust).

Keeleressursside loomisel tsiteerimise motiveeritud mahu väljatöötamisel tuleks arvestada nende ressursside loomise eesmärke. Kuna korpusi luuakse eelkõige rahvuskultuuri säilitamiseks, mitte esmaeesmärgiga koguda materjali ärilisi huvisid ja kommertsprojekte silmas pidades, siis võivad tsiteerimise motiveeritud mahud olla proportsionaalselt suuremad kui muude kasutusliikide puhul (näiteks ajakirjanduses või ringhäälingus). Samas peaksid nad jääma ühtsesse proportsiooni teadustöös heaks tavaks peetavate mahtudega<sup>13</sup>.

Teos, mis sisaldab teiste autorite teoste õiguspäraseid tsitaate, on uus iseseisev teos. Selle loomiseks ja edasiseks kasutamiseks ei ole tarvis kelleltki luba küsida ega autoritasu maksta. Uut teost võib selle autor oma äranägemisel reprodutseerida, levitada ja litsentsida, seda ka ärilistel eesmärkidel. Ka avalikustada tohib autor sellist uut teost oma äranägemisel.

<sup>13</sup> Järgnevas alaosas käsitletakse nn kolmeastmelist testi, mis seab teatud piirid motiveeritud mahu määramiseks teoste tsiteerimisel.

### 4.3. Teadustöö erand

Keeleressursid on oma olemuselt autoriõiguslikult teosena kaitstavad andmebaasid, mis sisaldavad elektroonilisi tekste. Seega eeldab keeleressurssidega tegelemine tekstide reprodutseerimist ehk kopeerimist. Kuna Eestis puudub spetsiifiline autoriõiguslik erand keeleressursside loomiseks<sup>14</sup>, siis on mõeldav tuginemine niinimetatud üldisele teadustöö erandile. Autoriõiguse seadus sätestab selle järgnevalt: “[a]utori nõusolekuta ja autoritasu maksmiseta, kuid kasutatud teose autori nime, kui see on teosel näidatud, teose nimetuse ning avaldamisallika kohustusliku äranäitamiseega on lubatud: 1) õiguspäraselt avaldatud teose reprodutseerimine õppe- ja teaduslikel eesmärkidel motiveeritud mahus haridus- ja teadusasutustes, mille tegevus ei taotle ärilisi eesmärke” (AutÕS § 19 p 3).

Täiendavalt peab teose vaba kasutamise, sealhulgas teadustöö erandile tuginemine vastama nn kolmeastmelisele testile (ingl *three step test*). See tähendab, et teose vaba kasutamine ei tohi olla vastuolus teose tavapärase kasutamisega ega kahjustada põhjendamatult autori seaduslikke huve ning võib toimuda ainult seaduses otseselt ette nähtud juhtudel (AutÕS § 17)<sup>15</sup>. Kolmeastmeline test on garantiiks autorile. Selle funktsiooniks on anda õiguste omajale võimalus vaidlustada teose vaba kasutamine juhul, kui see on küll formaalselt seadusega lubatud, kuid õiguste omaja arvates kahjustab oluliselt tema huvisid; sellised huvid võivad olla nii majanduslikud kui ka isiklikud. Seega on iga teose vaba kasutamine ka teatud kasutaja poolt võetud risk, et õiguste omaja vaidlustab sellise vaba kasutamise, sealhulgas tsiteerimise ja õppe- ning teaduslikel eesmärkidel kasutamise. Seetõttu ongi oluline teose vaba kasutamise õiguspärasuse hindamisel lähtuda nii vormilistest kui ka sisulistest aspektidest.

Niinimetatud teadustöö erandile tuginemine peab vastama järgmistele tingimustele.

- 1) Teos on õiguspäraselt avaldatud ning kasutamisel näidatakse ära autor, teose nimetus ja avaldamisallikas. Teos loetakse avaldatuks, kui see on autori nõusolekul antud üldsusele ehk määramata isikute ringile kasutamiseks (AutÕS § 8 ja 9). Seega tuleb arvestada, et teadustöö erand ei võimalda kasutada avaldamata materjali ilma õiguste omaja (autori, kirjastaja vms) nõusolekuta.
- 2) Teose reprodutseerimine toimub haridus- ja teadusasutuses. Selle tingimuse juures võib tõusetuda formaaljuriidiline küsimus, kas teadustöö erandile toetuja peab olema samal ajal haridus- ja teadusasutus või piisab, kui tegemist on haridus- või teadusasutusega. Näiteks Eesti Keele Instituut (EKI) tegeleb küll keeleressursside loomisega, aga ei ole haridusasutus: oma põhimääruse järgi on EKI Haridus- ja Teadusministeeriumi hallatav riigi teadus- ja arendusasutus. Käesoleva artikli autorite arvates võimaldab vaadeldav seaduse säte laiendavat tõlgendamist ning antud erand laieneb ka EKI-le.

<sup>14</sup> Ühendkuningriigis on kaalumisel eraldi teksti ja andmete analüüsiks erandi (ingl *exception for copying works for use by text and data analytics / text and data mining for research*) kehtestamine, mis täpsustaks üldist teadustöö erandit (Consultation on copyright 2012).

<sup>15</sup> Selline nõue tuleneb Berni konventsioonist ja teistest rahvusvahelistest instrumentidest. Berni konventsiooni art 9 (2), TRIPS leping art 13, Euroopa Parlamendi ja Nõukogu direktiiv 2001/29/EÜ, 22. mai 2001, autoriõiguse ja sellega kaasnevate õiguste teatavate aspektide ühtlustamise kohta infoühiskonnas (Infoühiskonna direktiiv art 5 (5)); Maailma Intellektuaalse Omandi Organisatsiooni autoriõiguse leping. – RT II 2006, 14, 39 (WIPO autoriõiguse leping), art 10.

- 3) Teose reprodutseerimine toimub õppe- ja teaduslikel eesmärkidel, seega äriülistel eesmärkidel on reprodutseerimine keelatud. Ka siin tekib küsimus, kas õppe- ja teaduslikud eesmärgid peavad esinema samaaegselt. Analoogiliselt eelmisele punktile võib seega eeldada, et õppe ja teaduslikud eesmärgid võivad esineda ka eraldi. Eesti keele säilitamine ja arendamine on eeldatavasti vaadeldav teadustööna.

Vaadeldava sätte tõlgendamisel omab tähendust veel see, kas nimetatud teadusasutus taotleb äriülist eesmärki (nagu see on näiteks erakõrgkoolide või tasuliste koolitusfirmade puhul) või mitte. EKI puhul võib väita, et selle tegevuse eesmärgiks ei ole äriülist eesmärki teenimine.

Kuna keeleressursid ei ole üldiselt eesmärk iseeneses, vaid neid kogutakse keele uurimise ja kirjeldamise eesmärgil, siis võib püstitada küsimuse, kas kopeeritud tekstide põhjal loodud statistiline mudel, grammatika, puudepank või muu keelekirjeldus on lahutatav nendest tekstidest ning kasutatav äriülistel eesmärkidel. Antud küsimusele vastamisel on oluline, kuivõrd reprodutseeritud teosed sisalduvad nende põhjal loodud kirjelduses. Kui teosed ei sisaldu keelekirjelduses, siis ei peaks olema takistatud ka selle äriülistel eesmärkidel kasutamine.

- 4) Teose reprodutseerimine motiveeritud mahus. “Motiveeritud mahu” nõue tähendab, et ei reprodutseerita rohkem kui on minimaalselt vaja algse teose autori originaalse idee, seisukoha, stiili, kujundi vms edasiandmiseks uues loodavas teoses. Võib tekkida küsimus, mille suhtes peab maht olema motiveeritud. Infoühiskonna direktiivi art 5 (3) a aitab nimetatud küsimust selgitada: “Liikmesriigid võivad ... näha ette erandeid või piiranguid ... tingimusel, et märgitakse ka allikas, sh autori nimi, kui see ei ole võimatu, ning *ulatuses, mis on õigustatult vajalik taotletava mitteäriülist eesmärgi saavutamiseks*”. Seega on võimalik tõlgendus, et motiveeritud mahu määrab kasutuseesmärk. Siin peab uue teose autor olema valmis vajadusel selgitama, miks reprodutseeriti just selles mahus, nagu seda tehti. Argumendid saavad tulla selle konkreetse valdkonna teadlastelt. Isegi kui maht on motiveeritud, võib kasutus olla vastuolus kolmeastmelise testi tingimustega, st olla vastuolus teose tavapärase kasutamisega ja riivata põhjendamatult autori huve.
- 5) Reprodutseerimine ei ole vastuolus teose tavapärase kasutamisega. Eksperdid on soovitanud tavapärase kasutamise hindamisel lähtuda küsimusest, kas konkreetne kasutus on selline, millega õiguste omaja loodab tulu teenida. Arvestada tuleb nii juba tuntud kui ka veel kasutamise ajal tundmatute teose kasutusviisidega. Kui teose vaba kasutus ei konkureeri majanduslikult õiguste omaja huvidega, siis on see lubatud (Ricketson, Ginsburg 2006: 769–770). Erialakirjanduse järgi “tavapärase kasutamine” eeldab, et hinnatakse nii võimalikke tulevasi kui ka praeguseid kasutusviise, mis võimaldavad kasu saamist teosest” (samas: 769).

Antud küsimust on käsitletud ka intellektuaalomandi õiguste kaubandus- aspektide lepingu (TRIPS) kontekstis<sup>16</sup> Maailma Kaubandusorganisatsiooni (WTO) aruandes ning selgitatud, et piirang on vastuolus teose tavapärase kasutamisega juhul, kui piiranguga lubatud kasutus hakkab konkureerima

<sup>16</sup> TRIPS lepingus nähakse samuti ette kolmeastmeline test ning argumendid selle sisustamisel võivad olla abiks ka Eesti autoriõiguse seaduse tõlgendamisel.

sellega, kuidas õiguste omaja harilikult oma teost ekspuuteerib, jättes see läbi õiguste omaja ilma sissetulekuta. WTO aruande kohaselt omab tähtsust ulatus, milles autoriõiguste omaja oma õigusi ekspuuteerib. Sõltuvalt sellest saab määrata, mis on teose tavakasutus vastaval turul. Teatud tingimustes ei ole olemasolev teose litsentseerimispraktika siiski küllaldaseks juhiseks hindamaks vaba kasutuse konkreetset mõju teose tavakasutusele (näiteks juhul, kui puudub tõhus õiguskaitse). Seega lisaks sellele, kuidas õiguste omaja kasutab hetkel oma teost tulu saamiseks, tuleb hinnata ka selliseid teose kasutamise viise, mis teatud tõenäosusega hakkavad tulu looma tulevikus. (WTO 2000: 48, 50; Correa 2007: 134–155)

- 6) Reprodutseerimine ei kahjusta põhjendamatult autori seaduslikke huve. Õiguskirjanduses on juhitud tähelepanu asjaolule, et sõnastus “ega kahjusta põhjendamatult autori seaduslikke huve”<sup>17</sup> viitab asjaolule, et piirang võib kahjustada autori seaduslikke huve tingimusel, et see vastab kahele eelnevale tingimusele (vaba kasutus on otseselt seaduses ette nähtud ja see ei ole vastuolus teose tavapärase kasutamisega) ning see on proportsionaalne (Ricketson, Ginsburg 2006: 776). Huvide kaalumisel tuleb võtta arvesse, millistest eesmärkidest on kantud teadustöö erandile tuginemine. Keeleressursside kontekstis on selleks eesti keele kaitse ja arendamine, mis on oluliseks argumendiks erandi kasutamise seaduspärasuse üle otsustamisel.

#### **4.4. Isikuandmete töötlemine teadusuuringu eesmärgil ja salvestamise nõusolek**

Isikuandmete töötlemisel keeleressursside kontekstis on risk kõige enam maandatud, kui puudutatud isik on andnud teadva nõusoleku. Samas näeb isikuandmete kaitse seadus ette ka võimaluse isikuandmete töötlemiseks ilma nõusolekuta. Teadusuuringute eesmärgil võib andmeid ilma puudutatud isiku nõusolekuta töödelda eelkõige kodeeritud kujul. Kui isik on tuvastatav, siis tema isikuandmete töötlemine teadustöö eesmärgil peab põhinema ülekaalulaval avalikul huvil. Andmete töötleja peab lisaks võtma organisatsioonilisi, füüsilisi ja infotehnoloogilisi meetmeid isikuandmete kaitseks (IKS § 16).

Kuigi salajane lindistamine on üldreeglina keelatud, on siiski mitmeid argumente, mis seda teatud kontekstis lubavad. Kui igasugune isikute eelnevalt teavitamata audio- või videosalvestamine oleks keelatud, siis ei saaks ka teha “Varjatud kaamera” laadseid huumorisaateid ja ka dokumentaalfilme. Antud temaatika on käsitlust leidnud Riigikogu istungil karistusõiguse kontekstis, kus M. Soosaar juhib tähelepanu asjaolule, et dokumentalisti jaoks on varjatud kaameraga töötamine möödapääsmatu. R. Lepiksoni vastus tõstatatud probleemile oli järgmine: “[p]alun väga, filmi terviseks, aga kasutada saad seda filmitut siis, kui selleks on antud nõusolek. Küsimus on selles, et neid toiminguid ei tehtaks salaja, vaid teavitataks neid osapooli, keda fikseeritakse”. (Riigikogu 2006) Tegemist on seega komplitseeritud õiguslike küsimustega, mis kuuluvad nii eraõiguse (autoriõigus, isiku kujutise kaitse) ja avaliku õiguse (põhiseaduslikud õigused, isikuandmete kaitse, jälitustegevus) valdkondadesse.



Minemata selle temaatikaga käesolevas artiklis sügavuti, võib keeleteadlastele teha järgmised soovitused. Esiteks, alati kui on vähegi võimalik, tuleb isikut, kelle kõnet salvestatakse, eelnevalt teavitada ning küsida temalt teadev nõusolek. See tähendab, et isik saab aru, millistel eesmärkidel ja kuidas tema kõne salvestist kasutatakse. Teiseks, kui ikkagi on vajadus täiesti loomuliku kõne lindistamiseks, siis tuleb alati pärast salvestuse tegemist teavitada puudutatud isikut ja küsida temalt nõusolekut lindistuse kasutamiseks keeleressursside kontekstis. Kui isik ei ole nõus, siis tuleb kogu materjal kohe kustutada. Kolmandaks, tuleb vältida lindistamist olukorras, kus juba eelnevalt on väga üheselt selge, et isik ei taha, et tema juttu lindistatakse ning, neljandaks, peab lindistust läbi viima lindistaja enda juuresolekul. Äärmiselt probleemne on olukord, kui keeleteadlane varustab erinevaid ruume, kohti ja telefone pealtkuulamiseadmetega eesmärgiga teha hiljem keelelist analüüsi. Riigikohtu kriminaalkolleegium on mõnevõrra teises kontekstis rõhutanud, et “[e]nda telefonivestluse kuulamine ja lindistamine ei vasta ühelegi seaduses kirjeldatud jälitustoimingule, sest enda telefonikõnet ei saa isik ise salaja ega varjatult pealt kuulata ega sellel eesmärgil lindistada. Neil põhjustel ei teki eraelulise vestluse salvestise kui tõendi lubatavuse otsustamisel küsimust võimaliku eraviisilise jälitustegevuse kohta KarS § 137 tähenduses”. (Riigikohus 2009: punkt 10) Sellest otsusest võib kaudselt järeldada, et Riigikohus ei loe kuriteoks olukorda, kui lindistatakse vestlust, milles ka lindistaja ise osaleb.

#### **4.5. Keeleressursside avalikustamise vajadus**

Keeletehnoloogia riikliku programmi finantseerimistingimustes on nõutud, et saadud tulemused oleksid kõigile vabalt kättesaadavad. See on läbimõeldud nõue, eesmärgiga võimaldada kõigile potentsiaalsetele uurimis- ja arendustöö jätkajatele juurdepääs seniloodud ressurssidele, edendada seeläbi eesti keele tehnoloogilist tuge ja lõpuks suurendada keele ellujäämise tõenäosust, mis ju ongi riigi sellealase finantseerimistegevuse eesmärk. Samas on pea kõik keeleressursid nii nõudlikud oma kasutamiseks vajalike eelteadmiste ja oskuste suhtes, et “inimesel tänavalt” ei ole nendega mitte midagi peale hakata ja ressursside täiesti piiranguteta avalikustamine ei ole tegelikult hädavajalik nimetatud eesmärkide saavutamiseks. Kuna aga avalikustamine sisaldab autoriõiguste omajate huvide kahjustamise riski, tasub kaaluda ressursside levitamist üksnes keeletehnoloogide siseriingis korralikult autenditud kasutajatele ja korralikult vormistatud lepingute kaudu.

### **5. Kokkuvõte**

Keeleressursside arendamine seondub komplitseeritud õiguslike küsimustega, mis kuuluvad nii eraõiguse (autoriõiguse) kui ka avaliku õiguse (põhiseaduslike õiguste, isikuandmete kaitse, jälitustegevuse) valdkondadesse. Tegemist on temaatikaga, mille uurimise vajadus tuleneb otseselt praktikast. Praktika ootab probleemide kaardistamist, kehtivale õigusele tuginevaid lahendeid, aga juhul kui viimased puuduvad, siis teoreetilisi ja praktilisi tõlgendusi ning ettepanekuid õigusaktide täiendamiseks.

Autoriõiguslikud probleemid on keeleressursside arendamisel ühed kesksemad. Kuigi keeleressursside loomisel kasutatakse autoriõiguslikult kaitstavaid teoseid või nende osi, on nende kasutus erinev teose tavapärasest kasutusest. Erinevate teoste abil analüüsitakse eesti keelt ning luuakse keelekirjeldusi, mis on sisendiks keeletehnoloogiale. Teos kui selline on keeleressursside seisukohalt eelkõige “toormaterjal”.

Praegune autoriõiguse süsteem ei arvesta keeleressursside spetsiifika. Ei ole välistatud, et tulevikus on vajalik kehtestada eriregulatsioon ehk täiendav vaba kasutuse alus, mis soodustab enam keeleressursside loomist ja levitamist. Samas on ka kehtiva autoriõiguse süsteemi tingimustes võimalik keeleressursside arendada. Eelkõige omavad keeleressursside puhul tähtsust kaks teose vaba kasutuse juhtu: tsiteerimine ja nn teadustöö erand. Nende üldtunnused on kohaldatavad ka keeleressursside puhul. Alati tuleb silmas pidada nn kolmeastmelist testi ja hinnata vaba kasutamise riske.

Suulise ja spontaanses olukorras salvestatud kõne puhul tõusetuvad päevakorda isikuandmete ja privaatsuse küsimused. Selleks, et suulise kõne kogumine ja töötlemine oleks õiguspärased, on vaja küsida nõusolek isikult, kelle kõnet salvestatakse. Kuna loomulik kõne eeldab, et isik ei tea, kui tema kõnet lindistatakse, siis tuleb nõusolek küsida kohe pärast salvestuse lõppu. Nii isikuandmete töötlemise kui ka spontaanses olukorras salvestatud kõne kasutamise nõusoleku andmisel peab isik aru saama, millistel eesmärkidel ja kuidas tema kõnet kasutatakse. Kogu keeleressursside loomise protsessi käigus tuleb austada ja tagada isikute õigused.

## Viidatud kirjandus

- AutÕS = Autoriõiguse seadus. RT I 1992, 49, 615 ... RT I, 28.12.2011, 1.  
Berni kirjandus- ja kunstiteoste kaitse konventsioon. RT II, 1994, 16, 49.  
Consultation on copyright 2012. HM Government; <http://www.ipo.gov.uk/consult-2011-copyright.pdf> (24.09.2012).  
Correa, Carlos María 2007. Trade Related Aspects of Intellectual Property Rights: A Commentary on the TRIPS Agreement. Oxford: Oxford University Press.  
IKS = Isikuandmete kaitse seadus. RT I 2007, 24, 127 ... RT I, 30.12.2010, 2.  
Infoühiskonna direktiiv = Euroopa Parlamendi ja nõukogu direktiiv 2001/29/EU 22.05.2001 autoriõiguse ja sellega kaasnevate õiguste teatavate aspektide ühtlustamise kohta infoühiskonnas. – ELT L 167, 22.06.2001.  
Intellektuaalomandi õiguste kaubandusaspektide leping. RT II, 1999, 22, 123.  
JTS = Jälitustegevuse seadus. RT I 1994, 16, 290 ... RT I, 29.06.2012, 2.  
KarS = Karistusseadustik. RT I 2001, 61, 364 ... RT I, 04.04.2012, 1.  
Kelli, Aleksei; Tavast, Arvi; Pisuke, Heiki 2012. Copyright and Constitutional aspects of digital language resources: The Estonian approach. – Juridica International: Law Review University of Tartu, XIX, 40–48.  
Maailma Intellektuaalse Omandi Organisatsiooni autoriõiguse leping. RT II, 2006, 14, 39.  
Pisuke, Heiki 2006. Autoriõiguse alused. Tallinn: KPMS & Partnerid.  
Ricketson, Sam; Ginsburg, Jane C. 2006. International Copyright And Neighbouring Rights: The Berne Convention And Beyond. Oxford: Oxford University Press.  
Riigikogu = X Riigikogu stenogramm. VII istungjärk. Neljapäev, 18. mai 2006, kell 10:00; <http://www.riigikogu.ee/?op=steno&stcommand=steno&date=1147935600> (23.09.2012).

Riigikohus = Riigikohtu kriminaalkolleegiumi otsus nr 3-1-1-5-09 (26. märts 2009); <https://www.riigiteataja.ee/akt/13167445> (23.09.2012).

Sootak, Jaan; Pikamäe, Priit 2009. Karistusseadustik: kommenteeritud väljaanne. 3., täiendatud ja ümbertöötatud väljaanne. Tallinn: Juura.

Walter, Michel M.; von Lewinski, Silke (Eds.) 2010. European Copyright Law: A Commentary. Oxford, New York: Oxford University Press.

Vider, Kadri; Kelli, Aleksei; Pisuke, Heiki; Mõtsküla, Peeter P.; Tuulik, Triin 2013. Keeleressursside litsentsimise õiguslikud kaalutlused. – Eesti Rakenduslingvistika Ühingu aastaraamat, 9, 33–347. <http://dx.doi.org/10.5128/ERYa9.21>

WTO = WTO. Report of the Panel. WT/DS160/R (15 June 2000); [http://www.wto.org/english/news\\_e/news00\\_e/1234da.pdf](http://www.wto.org/english/news_e/news00_e/1234da.pdf) (28.09.2011).

Ülemaailmse Intellektuaalse Omandi Organisatsiooni asutamise konventsioon. RT II, 1993, 25, 55.

**Arvi Tavast** (Eesti Keele Instituut, Tartu Ülikool, Tallinna Ülikool) juhatab EKI keeletehnoloogia osakonda; uurimisvaldkond on oskussuhtlus ja selle keelefilosoofilised eeldused.  
[arvi.tavast@eki.ee](mailto:arvi.tavast@eki.ee)

**Heiki Pisuke** (Euroopa Liidu Nõukogu, Tartu Ülikool), põhilised uurimissuunad on intellektuaalne omand (autoriõigus ja autoriõigusega kaasnevad õigused), tõlkimise õiguslikud küsimused, ülikoolide intellektuaalse omandi küsimused.  
[heiki.pisuke@ut.ee](mailto:heiki.pisuke@ut.ee)

**Aleksei Kelli** (Tartu Ülikool) põhilised uurimissuunad on intellektuaalne omand, innovatsioon, teadmispõhine majandus, digitaalsete keeleressursside õiguslikud küsimused.  
[aleksei.kelli@ut.ee](mailto:aleksei.kelli@ut.ee)

# LEGAL CHALLENGES AND POSSIBLE SOLUTIONS IN DEVELOPING LANGUAGE RESOURCES

Arvi Tavast<sup>1,2,3</sup>, Heiki Pisuke<sup>4,1</sup>, Aleksei Kelli<sup>2</sup>

Institute of the Estonian Language<sup>1</sup>, University of Tartu<sup>2</sup>, Tallinn University<sup>3</sup>,  
Council of the European Union<sup>4</sup>

The development of language resources entails a number of legal challenges in both private (copyright) and public law (constitutional rights, personal data protection, privacy). The topic is motivated by practical issues. In creating language resources and sharing them through international research frameworks, there is a pressing need for raised awareness of the challenges as well as solutions based on current legislation or, in their absence, theoretical and practical analysis and proposals for updated legislation.

A central issue in developing language resources is copyright. Although copyrighted works or parts thereof are used in language resources, this differs from their normal use. The works are only used to analyse language and compile various language descriptions to serve as input for language technology, while their creative and expressive purposes are irrelevant.

While the current system of copyright law does not specifically regulate language resources, their development is still possible to a certain degree. In addition to explicit contracts, applicable solutions include citing and the research exception. It is still always necessary to observe the three-step test and to assess the risks of free use. In the long-term perspective, it may be necessary to establish an additional basis for free use that would facilitate the creation and sharing of language resources.

Resources of spoken language and especially spontaneous language usage bring about the additional issues of personal data protection and privacy. To gather and process samples of spoken language legally, permission from the subject is needed. For spontaneous speech, a prerequisite is that the subject is unaware of being recorded – in that case, permission must be sought immediately after completing the recording. In both cases, subjects need to understand how and why their speech will be used. Rights of other persons must be respected and protected during the whole process of creating language resources.

**Keywords:** intellectual property, copyright, licenses, personal data protection, free use, research exception, citing

## KEELERESSURSSIDE LITSENTSIMISE ÕIGUSLIKKE KÜSIMUSI

Kadri Vider, Heiki Pisuke, Peeter P. Mötsküla,  
Triin Tuulik, Aleksei Kelli

**Ülevaade.** Artikkel käsitleb teaduspoliitilisi ja õiguslikke küsimusi, mida tuleks arvestada keeleressursside litsentsimisel, ja määratleb sobivaimad litsentsimise tingimused, pöörates tähelepanu eelkõige lepinguõiguse ja intellektuaalomandi litsentsimise küsimustele. Samuti võrreldakse artiklis USA, Euroopa riikide ja Eesti keeleressursside litsentsimise õiguslikku raamistikku ning tutvustatakse levinumaid praktikaid avatud juurdepääsu (ingl *open access*) ja vabade (avatud) litsentside vaatepunktist. Artikli praktiliseks eesmärgiks on selgitada Eestis keeleressursside loomise ja kasutamisega seotud teadlastele ja teadusasutustele intellektuaalomandi (IO) objektide litsentsimise põhi-aluseid ja jagada näpunäiteid litsentsitingimuste valimiseks. Artikkel on sündinud keeleteadlaste ja juristide koostöös ning mõeldud esmajoones keeleteadlastest lugejaskonnale.\*

**Võtmesõnad:** keeleressursid, keeletehnoloogia, intellektuaalomand, autoriõigus, litsentsimine, avatud juurdepääs, META-SHARE, CLARIN

### 1. Sissejuhatus

Keeleuuringute ja keeletehnoloogia edukaks arendamiseks on vaja hulgaliselt keelelist lähtematerjali. Uuringute käigus lähtematerjali töödeldakse ja analüüsitakse ning selle tulemusena valmivad ka mitmesugused digitaalsed keeleressursid. Neid ressursse võidakse kasutada uutes teadusuuringutes, kuid nad võivad leida rakendust muudese avalikes huvides ning olla sisendiks uute toodete ja teenuste väljatöötamisel erasektoris. Sellise lähtematerjali ja loodud digitaalsete keeleressursside kasutamise üheks teaduspoliitiliseks ja õiguslikuks nurgakiviks on autoriõigusest tulenevad erandid ja litsentsilepingud.

\* Artikkel on valminud Euroopa Sotsiaalfondi kaasabil teadus- ja innovatsioonipoliitika seire programmi raames.

Problemaatilisemad keeleressursid on õiguslikus tähenduses andmebaasid. Selliste andmebaaside loomisel ja kasutamisel kerkivad mitmed keerulised õiguslikud küsimused seoses autoriõiguse ja sellega kaasnevate õigustega.

Eestis ja Euroopas on mitmed laiahaardelised algatused keeleressursside loomiseks ja laialdaselt kättesaadavaks tegemiseks, näiteks META-NET ja CLARIN, mille raames on loodud ka keeleressursside tüüplitsentse. Nii META-SHAREi hartas<sup>1</sup> kui ka CLARIN ERICu põhikirjas<sup>2</sup> rõhutatakse avatud juurdepääsu ja avatud lähtekoodi põhimõtete järgimise olulisust. Autorid selgitavad, mida avatud juurdepääs keeleressursside litsentsimisel tähendab ja mille poolest vastavad tüüplitsentsid omavahel erinevad.

## 2. Keeleressursid ja nende litsentsimise vajadus

Lingvistide ja keeletehnoloogide vaatenurgast jagunevad digitaalsed keeleressursid iseloomu poolest kaheks: keeleteaduslikuks andmestikuks ja tarkvaraks.

- 1) Keeleteaduslik andmestik:
  - a) korpused (tekstilised või helilised/pildilised keelematerjali sisaldavad ja elavat keelekasutust esindavad kogumid) – nende probleemistikuga on siiani kõige rohkem tegeldud (Kelli jt 2012: 40–48, Tavast jt 2013 siinses kogumikus), sest sisaldavad juba loodud tekste ja kõneldud keelt, mis on sageli autoriõiguse objektiks;
  - b) leksikonid ehk sõnavara kogumid;
  - c) keelemudelid (näiteks puudepank).
- 2) Tarkvara, mis midagi keeleandmestikuga teeb, seda suuremal või vähemal määral töötleb ja muudab (näiteks annoteerib, loob sisendfaile järgmise masintöötamise etapi jaoks, loob väljundit lõppkasutaja jaoks, olgu lõppkasutajaks IT-süsteem või inimene):
  - d) iseseisvad programmid või lausa infotehnoloogilised terviksisüsteemid;
  - e) tarkvarajupid või teegid, mida ühendatakse umbes nagu legoklotse erinevaid ülesandeid täitvateks struktuurideks.

Keeleressursside loojal võib tekkida küsimus, miks on litsentsimine vajalik? Seni on ju keeleteadlased suhteliselt muretuks kasutanud digitaalselt saadaolevaid keeleressursse oma teadustöös andmestiku või tehnoloogiana. Keeleteadlase jaoks on andmestiku puhul erinevus ainult meediumis, mis on muutunud digitaalseks ja saanud juurde ohtralt analüüsiprotsessi abistavat tarkvara. Seinatäis kartoteegikaste erinevatest allikatest välja kirjutatud sedelitega oli eelmise põlvkonna tekstikorpused, keelejuhtide salvestised erinevatelt murdealadelt aga kõnekorpused. Kordades rohkem aega võttis toona teadlast huvitava keelematerjali ülesleidmine, andmete hulk oli väiksem ja analüüs aeglasem, sest kõikide keeletesandite analüüs tuli teha käsitsi ja paberi peal. Umbes 1995. aastast alates on eesti keeleteadlased aga harjunud üha enam ja enam kasutama digitaalselt saadaolevaid keeleandmestikke ja analüüsitarkvara, õppinud koostama üha keerulisemaid päringuid huvipakkuva ainese korjamiseks kogu digimaailmas leiduva keelematerjali pealt.

Euroopa Komisjoni rahastatud tippteadmiste võrgustiku META-NET läbi viidud keeleressursside ja -tehnoloogiate alast uurimusest (Liin jt 2012) selgus, et

<sup>1</sup> Vt [http://www.meta-net.eu/meta-share/METASHARE\\_Charter.pdf](http://www.meta-net.eu/meta-share/METASHARE_Charter.pdf) (14.12.2012).

<sup>2</sup> CLARIN ERIC (Euroopa teaduse infrastruktuuri konsortsium), vt <http://www-sk.let.uu.nl/u/D85-5a.pdf> (14.12.2012).

enamikku Euroopa keeltest ootab ees “digitaalne hääbumine”. Eksperdid leidsid, et vahe “suurte” ja “väikeste” keelte vahel üha kasvab ning väiksematele ja vähemate ressurssidega keeltele tuleb digitaalse väljasuremise ennetamiseks tagada esmatähtsad baastehnoloogiad.<sup>3</sup> Ka eesti keel paigutub META-NETi uuringus ohustatud keelte hulka.<sup>4</sup> Arvestades eesti keele- ja kultuuriruumi väiksust on see mõistetav, sest eestikeelse kõnelejaskonnaga sihtturg ei paku ettevõtjatele nii avaraid ärilisi võimalusi nagu ulatuslikuma kasutajaskonnaga keelte puhul. Enamikul maailma rahvastikust pole huvi tagada eesti keele digitaalset kestmajäämist ega lootust teenida maailma mõistes arvestatavat kasumit, varustades eesti keele toega eri seadmeid ja tarkvara. Riiklikud investeeringud keeleressursside ja -tehnoloogiate loomiseks, arendamiseks ning kättesaadavaks tegemiseks on üks otsesemaid viise, kuidas seda turutõket ületada<sup>5</sup>.

Eestis on keeletehnooloogilise toe vajadust teadvustatud riiklikus eesti keele arengukavas, mille mõjul on juba alates 2006. aastast keeletehnoologia riiklike programmide kaudu rahastatud valdkondlikku teadus- ja arendustegevust (Liin jt 2012: 25).

Keeleressursse luuaksegi Eestis praegu peamiselt riikliku teadus- ja arendustegevuse raames (Liin jt 2012: 25). Eesti Keeleressursside Keskus (EKRK) on kutsutud ja seatud hoolitsema selle eest, et riiklike keeletehnoologia teadusprogrammide tulemused ei jääks hajutatult ripakile, vaid oleks kättesaadavad ja kasutatavad ühes raamistikus erinevatele huviliste rühmadele ja selgelt sõnastatud tingimustel. Tingimuste hulka kuuluvad iga keeleressursi kasutamise kaasnivad õigused ja kohustused. Nende õiguste ja kohustuste raames keeleressursside kasutada andmist nimetataksegi litsentsimiseks (VÕS § 368). Kuigi programmides on üldisemalt sõnastatud Eesti riigi raha eest tehtava teadus- ja arendustöö tulemuste intellektuaalomandi kuuluvus<sup>6</sup> ja levitamise poliitika<sup>7</sup>, ei piisa sellest iga üksiku keeleressursi kasutusse andmiseks maailmas, kus tehnoloogia kiire arengu taustal on keeleressursside kasutamise vastu huvi mitte ainult teadlastel, vaid ka laiemal avalikkusel ja kasumitaotlusega erasektoril. Selged kasutustingimused litsentsi näol on vajalikud ka ressursside loojatele, kes saavad selle kaudu väljendada oma soove, kuidas nende loomingu tohib kasutada.

Nii nagu muud riigi toel loodud teadustulemused, on ka keeleressursid oma olemuselt avalik hüve. Seega peaks nende loomine, arendamine ja levitamine alluma samadele poliitikakaalutlustele nagu teiste teadustulemuste levitamine. Et vältida olukorda, kus keeleressursid on vaid kitsa teadlaskogukonna uurimisobjekt, tuleks kehtestada poliitika, mis lubab nende kättesaadavaks tegemist ja kasutamist ka laiema avalikkuse poolt. Sellisel juhul tekib paratamatult küsimus, kas riiklikest vahenditest rahastatud keeleressursid peaks jääma kasutamiseks ainult avalikes huvides või tohib neid rakendada ka ärielistel eesmärkidel, näiteks uute toodete ja teenuste arendamiseks ja müügiiks.

<sup>3</sup> META-NET. Vähemalt 21 Euroopa keelt on digihääbumise ohus. META-NET Valge raamatu sari: Pressiteade, 20.09.2012. Vt <http://www.meta-net.eu/whitepapers/press-release-et> (23.09.2012).

<sup>4</sup> *Ibid.*

<sup>5</sup> Käesolevas artiklis ei käsitleta tarbijakaitseõiguse küsimusi, millest tuleneb kohustus teatud juhtudel anda toote ja teenuse kohta informatsiooni eesti keeles.

<sup>6</sup> Vt <http://www.keeletehnoologia.ee/ekkt-1> (14.12.2012).

<sup>7</sup> Vt <http://www.keeletehnoologia.ee/> (14.12.2012).

### 3. Avatud juurdepääs

Eestis ei ole hetkel kehtestatud ühtset avatud juurdepääsu (ingl *open access*) poliitikat teadusasutustes ega riiklikul tasemel. Samas on see teema aktuaalne teadus- ja innovatsioonipoliitika seire programmis, mis on sisendiks haridus- ja teadusministeeriumile vastava poliitika kujundamisel.<sup>8</sup> Eesti juhtivatel ülikoolidel on küll olemas intellektuaalomandi käsitlemise dokumendid<sup>9</sup>, kuid need käsitlevad peamiselt töötajate ja ülikooli vahelist kokkulepet töötaja intellektuaalomandi õiguste üleminekuks ning kasutamiseks ülikooli, mitte aga avalikkuse poolt. Sellele on juba tähelepanu juhitud erialakirjanduses ning tehtud ettepanek kehtestada ülikoolides loodud intellektuaalomandi kasutamise ühtsed riiklikud põhimõtted ning edendada avatud litsentsimise poliitikat (Pisuke, Kelli 2007: 171).

Avatud juurdepääs ei ole õiguslik mõiste (Tuulik 2012). Nagu nimetuski viitab, tähendab see teatud õiguslike ja administratiivsete tõkete kõrvaldamist. Siin peetakse esmajoones silmas kasutuspiiranguid, mis tulenevad autoriõigusest. Näiteks on autoriõiguse üldpõhimõteteks, et teose õiguspäraseks kasutamiseks peab olema autori nõusolek, tuleb maksta autoritasu, teost ei tohi autori nõusolekuta muuta ja autorile tuleb alati viidata. Vaid autor või autoriõiguste omaja (tööandja, teose lepingu alusel kasutaja, pärija vms) saab otsustada, kas lubada avatud juurdepääsu, andes selleks oma teose kasutamiseks vastava loa ehk litsentsi. Kaudselt edendatakse avatud juurdepääsu võimaldamist ka poliitiliste vahenditega, kehtestades vastavaid rahastamis- ja publitseerimis põhimõtteid või korraldades üldharivaid kampaaniaid. Näiteks riiklikus teadus- ja arendustegevuse programmis “Eesti keele tehnoloogia 2011–2017”<sup>10</sup> on loodavate keeleressursside ja tarkvara litsentsimise põhimõtete seas välja toodud, et eelistatakse avatud juurdepääsuga litsentsimist.

Avatud juurdepääsu mõiste on kasutusel nii kitsas kui ka laias tähenduses. Kitsamalt tähendab see näiteks eelretsenseeritavate teadustulemuste tasuta kättesaadavaks tegemist internetis.

Laiemas tähenduses ei ole avatud juurdepääs piiratud ainult teadustulemuste kättesaadavaks tegemisega internetis. Juba 1980. aastate alguses tekkis nn vaba tarkvara liikumine ning võeti kasutusele vaba tarkvara (ingl *free software*) mõiste. Vaba tarkvaraga on tegemist, kui kasutajale on tagatud neli põhivabadust: vabadus programmi mistahes eesmärgil kasutada, seda uurida ja muuta ning programmi ennast ja selle modifikatsioone levitada.<sup>11</sup> 1990ndate lõpus arenes selle kõrval välja avatud lähtekoodiga tarkvara (ingl *open source software*) mõiste.<sup>12</sup> Kuigi “vaba tarkvara” ja “avatud lähtekoodiga tarkvara” ideoloogiline taust on vägagi erinev, kasutatakse neid sageli sünonüümidenä.<sup>13</sup> Hiljem hakati neid lisaks tarkvarale laiendama ka muudele teose liikidele ning tänaseks on välja pakutud mitmeid

<sup>8</sup> Eestis on valminud teadus- ja innovatsioonipoliitika seire programmi raames vastav analüüs koos poliitika-soovitustega. – Aleksei Kelli, Tõnis Mets, Merit Burenkov. Autoriõiguse ning avatud juurdepääsu (*open access*) küsimused teadus- ja arendustegevuses. Vt <http://www.tips.ut.ee/index.php?module=32&op=1&id=3527> (21.12.2012).

<sup>9</sup> Vt nt Tartu Ülikool. Intellektuaalse omandi käsitlemise põhimõtted Tartu Ülikoolis. KINNITATUD Tartu Ülikooli nõukogu 18. novembri 2003. a määrusega nr 17 (jõustunud 28.11.2003). Muudetud Tartu Ülikooli nõukogu 19. detsembri 2008. aasta määrusega nr 25 (jõustunud 29.12.2008) – vt [http://www.ut.ee/sites/default/files/livelink\\_files/intellektuaalse\\_omandi\\_ksitlemise\\_phimtted\\_tartu\\_liikoolis\\_tervikekst.pdf](http://www.ut.ee/sites/default/files/livelink_files/intellektuaalse_omandi_ksitlemise_phimtted_tartu_liikoolis_tervikekst.pdf) (23.09.2012); Tallinna Ülikool. Intellektuaalse omandi õiguskaitsese eeskiri Tallinna Ülikoolis. KEHTESTATUD TPÜ nõukogu 14.06.2004. a määrusega nr 9 – vt <http://www.tlu.ee/files/arts/314/intel2e2435b4d99b60a2481cd7c54d2f93a7.pdf> (23.09.2012).

<sup>10</sup> Kinnitatud haridus- ja teadusministri 25. jaanuari 2011. a käskkirjaga nr 71; allalaetav <http://www.keeletehnoloogia.ee/EKT2011-2017-programm-uuendet.pdf> (14.12.2012).

<sup>11</sup> Free Software Definition. Free Software Foundation. Vt <http://www.gnu.org/philosophy/free-sw.html> (14.12.2012).

<sup>12</sup> <http://opensource.org> (14.12.2012).

<sup>13</sup> Vt nt E. S. Raymond. Goodbye, “free software”; hello, “open source”, <http://www.catb.org/~esr/open-source.html> (21.12.2012) ja R. Stallman. Why Open Source misses the point of Free Software, <http://www.gnu.org/philosophy/open-source-misses-the-point.html> (21.12.2012).



erinevaid avatud juurdepääsu määratlusi sõltumata objektist, mida selle mõistega kirjeldatakse.<sup>14</sup> Seni puudub aga üksmeel selle osas, milline määratlus on domineeriv. Kuna puudub selge kokkulepe, kui vähe piiranguid on võimalik avatud juurdepääsu litsentsiga kehtestada, tutvustavad autorid lühidalt erinevaid lähenemisviise.

Hetkel on kasutusel rohkem kui 100 infotehnoloogia valdkonnas tunnustatud vaba tarkvara litsentsi.<sup>15</sup> Neist kõige tuntum on Richard M. Stallman'i poolt 1989. aastal kirjutatud ning hiljem kahel korral uuendatud GNU<sup>16</sup> üldine avalik litsents (GNU General Public License ehk GPL). Selle erinevaid versioone kasutatakse hinnanguliselt rohkem kui pooltes kõigist vaba tarkvara projektidest (Mõtsküla 2009: 19). Kuigi GPLi alusel võib litsentsida mistahes teoseid, mille "lähtekood" on selgelt määratletav<sup>17</sup>, sobib see, nagu ka muud vaba tarkvara litsentsid, ennekõike arvutiprogrammide litsentsimiseks.

Lihtsustamaks vaba tarkvara põhimõtete rakendamist muudele teostele, asutati professor Lawrence Lessigi juhtimisel 2001. aastal mittetulundusühing Creative Commons<sup>18</sup> (CC), mis avaldas 2002. aastal esimese litsentside komplekti, mis sobib väga laiale teosteringile. Siiski ei soovita CC oma litsentse kasutada arvutiprogrammide litsentsimiseks.<sup>19</sup> Andmebaaside puhul on CC litsentsid kasutatavad niivõrd, kuivõrd andmebaas on kaitstud autoriõigusega. Euroopa Liidu õigusest tulenevaid andmebaasi tegija iseseisvaid ehk *sui generis* õiguseid<sup>20</sup> CC litsentside kehtivais versioonides ei puudutata. Kuid järgmises, praegu väljatöötamisel olevas versioonis (4.0), võib see muutuda ja CC litsentse saaks sellisel juhul kasutada ka *sui generis* õiguste litsentsimiseks.<sup>21</sup> Erandiks on siinkohal CC Zero (CCo)<sup>22</sup>, mis on kasutatav nii lähtekoodi, andmebaaside kui ka mistahes muude teoste andmiseks avalikku kasutusse. Sellist kõigile isikutele vaba kasutamist nimetatakse üldsusele avatud kasutamiseks (ingl *public domain*). Hiljem on CC oma litsentse korduvalt täiustanud ning kohandanud rohkem kui 50 riigi õigussüsteemi jaoks.<sup>23</sup> Täna-seks on CC litsentsisüsteem kujunemas avatud sisu rahvusvaheliselt tunnustatud litsentsistandardiks (Kreutzer 2011: 135).

Nähes, et ei vaba tarkvara ega muud avatud sisu litsentsid ei ole optimaalselt kohaldatavad andmebaasidele, käivitas Jordan Hatcher 2007. aastal projekti Open Data Commons (ODC)<sup>24</sup>, mille eesmärk oli töötada välja vabad (avatud) andmebaasilitsentsid. Alates 2009. aastast tegutseb ODC Open Knowledge Foundation'i egiidi all. ODC projekt pakub uudse lahendusena välja "soovituslike tingimustega vabakasutuse". Nimelt annab ODC avalikku kasutusse andmise avaldus ja litsents (*Public Domain Dedication and License*, PDDL)<sup>25</sup> selle kasutajale võimaluse viidata heale tavale, mis on sätestatud ODC kogukonna normides.<sup>26</sup> Need normid ei oma

<sup>14</sup> Open Knowledge Foundation'i Open Definition (<http://opendefinition.org>, 21.12.2012) ning Vabade Kultuuriliste Teoste Definitsioon (Definition of Free Cultural Works, <http://freedomdefined.org>, 21.12.2012), mida tunnustab ka mittetulundusühing Creative Commons (CC) (<http://creativecommons.org/weblog/entry/8051>, 21.12.2012).

<sup>15</sup> Vt <http://www.gnu.org/licenses/license-list.html> (21.12.2012) ja <http://opensource.org/licenses> (21.12.2012).

<sup>16</sup> GNU (GNU's Not Unix, rekursiivne lühend) on täielikult vabast tarkvarast koosnev UNIXi-laadne operatsioonisüsteem.

<sup>17</sup> Frequently Asked Questions about the GNU Licenses. Free Software Foundation. <http://www.gnu.org/licenses/gpl-faq.html#GPLOtherThanSoftware> (21.12.2012).

<sup>18</sup> Creative Commons <http://creativecommons.org> (21.12.2012).

<sup>19</sup> Vt [http://wiki.creativecommons.org/FAQ#Can\\_I\\_apply\\_a\\_Creative\\_Commons\\_license\\_to\\_software.3F](http://wiki.creativecommons.org/FAQ#Can_I_apply_a_Creative_Commons_license_to_software.3F) (21.12.2012).

<sup>20</sup> *Sui generis* andmebaaside olemust käsitletakse punktis 4.1.

<sup>21</sup> Vt [http://wiki.creativecommons.org/FAQ#Can\\_I\\_apply\\_a\\_Creative\\_Commons\\_license\\_to\\_data\\_or\\_a\\_database.3F](http://wiki.creativecommons.org/FAQ#Can_I_apply_a_Creative_Commons_license_to_data_or_a_database.3F) (21.12.2012).

<sup>22</sup> Vt <http://creativecommons.org/choose/zero/> (21.12.2012).

<sup>23</sup> CC litsentside versiooni 3.0 eestikeelseid ja Eesti õigusele kohandatud variandid avaldati 2010. aasta oktoobris.

Vt <http://creativecommons.ee> (21.12.2012).

<sup>24</sup> Vt <http://opendatacommons.org> (21.12.2012).

<sup>25</sup> Vt <http://opendatacommons.org/licenses/pddl/> (21.12.2012).

<sup>26</sup> Vt <http://opendatacommons.org/norms/odc-by-sa/> (21.12.2012).

iseseisvat õiguslikku sisu ning nad on soovituslikud vaid nende teoste kasutajate suhtes, mis on vabasse kasutusse antud PDDL alusel ning mille vabasse kasutusse andja on kogukonna normidele viidanud.

Eelpool käsitletud litsentside puhul on tegemist nn kasutuskogukonna normidega, mida kohaldavad omavahel selle kogukonna liikmed või isikud, kes on nõus neid soovitusi teatud tüüptingimustena vabatahtlikult järgima. Kogukonnad võivad anda oma liikmetele soovitusi selle osas, milliseid litsentse kasutada (või isegi seada teatava litsentsi kasutamine kogukonna ühisvaramusse panustamise eeltingimuseks, nt Wikipedia). Siiski on GPL, ODC-ODbL (*Open Data Commons Open Database License*)<sup>27</sup> ja CC litsentsid, juhul kui autor on oma teose nende alusel avalikkusele litsentsinud, teoste kasutajate jaoks õiguslikult siduvad (kui kasutaja just ei võta vaevaks autorilt endale eritingimusi välja kaubelda). Seda kinnitab asjaolu, et Saksa kohtutes on GPL litsentside tingimusi korduvalt jõustatud (Mötsküla 2009: 46–47). Kui näiteks CC litsentsi kasutustingimustega nõustumine ja teatud litsentsitüübi kasutamine ei oleks isikule õiguslikult siduv, siis ei saaks me rääkida ühiskonna õiguskindlusest, kui iga hetk võiks oma lubadust tagasi võtta.

Vaba ja avatud tarkvara liikumine sai alguse Ameerika Ühendriikidest. USA õigussüsteem, õigusteooria ja -praktika erinevad väga suurel määral kontinentaal-Euroopa riikide tsiviilõigusel põhinevast tavast.<sup>28</sup> Erinevad on õigusdoktriinid ja isegi lepinguõiguses kasutatav terminoloogia. Nende probleemide ületamiseks töötati Euroopa Liidu (EL) Komisjoni otsesel initsiatiivil välja Euroopa Liidu tarkvara vaba kasutuse litsents (*European Union Public License, EUPL*), mis avaldati 2007. aastal. EUPL põhineb suures ulatuses Euroopas kasutusel olevatel litsentsidel ning kasutab Euroopa õigussüsteemile omast mõisteaparaati ja terminoloogiat. Oluline uudne lähenemine oli ka see, et nimetatud litsents tõlgiti ja toimetati EL Komisjoni initsiatiivil vastavaks kõigi EL ametlike keelte õigussüsteemidega, sealhulgas eesti keeles Eesti õigussüsteemiga.<sup>29</sup> Kuigi see litsents on välja töötatud konkreetselt tarkvara kasutust silmas pidades, väärrib see erilist tähelepanu kui võimalik eeskujuna muude teose liikide kasutamiseks.

Avatud juurdepääsu põhimõtteid on püütud määratleda ka teistes dokumentides ja initsiatiivides.<sup>30</sup>

Arvestades erinevaid välja pakutud avatud juurdepääsu käsitusi ja tüüplitsentside tingimusi, on autorid seisukohal, et avatud juurdepääsuga laias tähenduses on tegemist siis, kui autor on andnud litsentsilepingu alusel kasutajale õiguse teost mistahes eesmärgil (sh nii isiklikul otstarbel, avalikes huvides kui ka ärielistel eesmärkidel) kasutada ja taaskasutada, uurida ja modifitseerida ning reprodutseerida ja levitada, kusjuures nende õiguste teostamine on piiratud ainult kohustusega viidata algele autorile.

<sup>27</sup> Vt <http://opendatacommons.org/licenses/odbl/> (23.02.2013)

<sup>28</sup> Euroopas on USAga samas õigusperekonnas Ühendkuningriik ja Iirimaa, mille õigus põhineb niinimetatud üldise õiguse (*Common Law*) traditsioonidel.

<sup>29</sup> Euroopa Liidu tarkvara vaba kasutuse litsents (1.0) <http://ec.europa.eu/idabc/servlets/Doc3cb1.pdf?id=29923> (22.12.2012); vt ka European Union Public Licence – EUPL v.1.1 <http://ec.europa.eu/idabc/en/document/7774.html> (23.12.2012).

<sup>30</sup> UNESCO. Open Access to scientific information: Policy guidelines released. 06.04.2012 – vt [http://www.unesco.org/new/en/communication-and-information/resources/news-and-in-focus-articles/all-news/news/open\\_access\\_to\\_scientific\\_information\\_policy\\_guidelines\\_for\\_open\\_access\\_released/](http://www.unesco.org/new/en/communication-and-information/resources/news-and-in-focus-articles/all-news/news/open_access_to_scientific_information_policy_guidelines_for_open_access_released/) (23.09.2012); Bethesda Statement on Open Access Publishing. Released 20.06.2003 – vt <http://www.earlham.edu/~peters/fos/bethesda.htm> (23.09.2012); Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities – vt <http://oa.mpg.de/berlin-prozess/berliner-erklarung/> (23.09.2012); OECD Principles and Guidelines for Access to Research Data from Public Funding. OECD, 2007 – vt <http://www.oecd.org/science/scienceandtechnologypolicy/38500813.pdf> (23.09.2012); Policy Statements on Data Management for Global Change Research. U.S. Global Change Research Program. Executive Office of the President, Office of Science and Technology Policy, Washington, D.C. 20506, July 2, 1991 – vt <http://www.gcrio.org/USGCRP/DataPolicy.html> (23.09.2012).

## 4. Autoriõigus kui litsentsimise õiguslik raamistik

### 4.1. Litsentsilepingu objekt ja subjekt keeleressursside puhul

Keeleressursside litsentsimisel on vajalik määratleda litsentsilepingu objekt. Nagu eelnevalt öeldud, võib keeleressursid jagada laias laastus kaheks: keeleteaduslik andmestik ja tarkvara (vt ptk 2). Arvutiprogramm on autoriõiguslikult kaitstav kui kirjandusteos ning selle litsentsimisega seonduvaid küsimusi antud artiklis pikemalt ei käsitleta. Keeleteaduslik andmestik on samuti kaitstav autoriõigusega, kuid selle õigusliku olemuse täpsem avamine on komplitseeritum. Keeleteaduslikule andmestikule alljärgnevalt keskendutaksegi.

Autoriõigus on loodud kaitsma teose autori õigusi. Autoriõiguse seadus määratleb teosena originaalset tulemust kirjanduse, kunsti ja teaduse valdkonnas, mis on väljendatud objektiivses vormis (AutÕS § 4 lg 2). Teose ühe näitena nimetab autoriõiguse seadus ka andmebaase, mis kujutavad endast “süsteemaatiliselt või meetoodiliselt korrastatud iseseisvate teoste, andmete või muu materjali kogu, mis on individuaalselt kättesaadav elektrooniliste või muude vahendite abil” (AutÕS § 4 lg 3 p 22).

Lisaks teosena kaitstavale andmebaasile, mis eeldavad autori loomingulist panust, kaitseb autoriõigus ka andmebaase, mille kaitstavuse seisukohalt ei ole oluline mitte autori loomingulisus, vaid selle valmimiseks tehtud investeering.<sup>31</sup>

Edaspidi nimetatakse selliseid andmebaase artiklis *sui generis* andmebaasideks (vt ka Tavast jt 2013 siinses kogumikus), sest nende suhtes on kehtestatud autoriõigusest erinev kaitserežiim. *Sui generis* andmebaase määratletakse seaduses teoste, andmete või muu materjali süsteemaatiliselt või meetoodiliselt korraldatud koguna, mille kaitstavuse hindamise seisukohalt on oluline laadilt, väärtuselt või suuruselt olulise investeeringu tegemine selle andmebaasi sisuks olevate andmete kogumiseks, omandamiseks, kontrollimiseks, süstematiseerimiseks või kättesaadavaks tegemiseks (AutÕS § 75<sup>2</sup>, 75<sup>3</sup>).

Eelneva pinnalt tekib küsimus, milline on keeleressursside õiguslik seisund? Artikli autorid on seisukohal, et keeleressursid on olemuselt andmebaasid, mis eeldavad reeglina loomingulist panust<sup>32</sup>, ning seetõttu kvalifitseeruvad need autoriõiguslikult kaitstavate teostena. Samas vajab keeleressursside loomine ka ulatuslikke investeeringuid, mistõttu põhimõtteliselt laieneb keeleressurssidele ka *sui generis* andmebaaside kaitse. Seega on keeleressursid kumulatiivselt kaitstavad nii autoriõiguslike teostena kaitstavate andmebaasidena kui ka *sui generis* andmebaasidena.<sup>33</sup> Kuna teosena kaitstud andmebaas omab ulatuslikumat kaitset, siis reeglina ei teki *sui generis* kaitse järgi vajadust. *Sui generis* kaitse omab keeleressursside kontekstis tähtsust peamiselt siis, kui keeleressursside loomine eeldab ainult investeeringut, kuid mitte loovust.

Keeleressursside litsentsimise muudab keeruliseks olukord, kui ressursside loomisel kasutatakse autoriõiguslikult kaitstavaid tekste ning materjali, mille avaldamine võib olla vastuolus isikuandmete kaitse põhimõtetega. Kuna keeleressursside arendamise küsimusi on uuritud teistes artiklites<sup>34</sup>, siis siinkohal sellel pikemalt ei peatuta. Tuleb siiski välja tuua, et kui keeleressursid on loodud autoriõigusliku

<sup>31</sup> Seadus räägib viimasel juhul just andmebaasi tegijast, mitte andmebaasi loojast. Vt AutÕS § 75<sup>3</sup>.

<sup>32</sup> Loomingulisus võib väljenduda materjali valikus, mida andmebaasi sisestatakse.

<sup>33</sup> Teatud juhtudel võib tõusetuda ka keeleressursside kaitse ärisaladusena. Ärisaladuse kaitse kohta vt Kelli jt 2010.

<sup>34</sup> Vt Kelli jt 2012, Tavast jt 2013 (siinses kogumikus).

teadustöö erandi raames (AutÕS § 19 p 2 ja 3)<sup>35</sup> ilma litsentsilepinguta ning ka isikuandmeid on töödeldud isikuandmete kaitse seaduse teadusuuringu regulatsioonile (IKS § 16) tuginedes, siis ei saa taoliste ressursside kasutus väljuda teadusliku uurimistöö raamidest. Litsentsimisel on oluline lähtuda põhimõttest, et teistele ei saa anda rohkem õigusi kui need, mis endale kuuluvad. See ei tähenda, et ressursse ei saaks kasutada teised teadlased, kuid nende laiem levitamine on õiguslikult problemaatiline.

Samaväärselt oluline keeleressurssides sisalduva materjali õigusliku staatusega on ka keeleressursside kuuluvuse küsimus. Teiste sõnadega, kes on subjektiks litsentsilepingus, mille objektiks on keeleressursid? Keeleressursse on tavaliselt arendanud ülikooli või teadusasutuse töötajad. Autoriõiguse seaduse kohaselt töölepingu alusel oma otseste tööülesannete täitmise korras loodud teose autori varalised õigused lähevad üle tööandjale (AutÕS § 32 lg 1), kuid tööandja ei saa automaatselt õigust kasutada töötaja isiklike õigusi. Eelkõige on see oluline teose puutumatus õiguse kontekstis, mis autori kui teose looja isikliku õigusena tähendab autori õigust teha ise või lubada teistel isikutel teha teoses muudatusi (AutÕS § 12 lg 1 p 3).

Teine probleemide ring võib tekkida asjaolust, et mitte kõik keeleressursside loojad ei ole olnud teadusasutuse koosseisulised töötajad. Osa tööd on tehtud töövõtulepingu alusel. Autoriõiguse seadus ei näe ette autori varaliste õiguste üleminekut töövõtulepingu raames. Sageli puuduvad sellistes lepingutes punktid autoriõiguse kuuluvuse kohta. Seetõttu on õiguslik olukord keerulisem, kuna ei ole selge, kas töövõtu alusel loodud ressursid kuuluvad teadusasutusele ja kas asutus saab neid edasi litsentsida. Samas tuleb antud probleemi puhul lähtuda ka lepinguõiguse üldistest põhimõtetest ning tõlgendada töövõtulepingut lähtudes poolte tahtest ja selle eesmärgist. Kui isiku ülesanne töövõtulepingu kohaselt oli keeleressursside arendamine, siis võib eeldada, et isik andis tellijale ka loa loodud keeleressursside kasutamiseks.

## 4.2. Litsentsitavad õigused

Autoriõigusega kaitstavatele keeleressursside andmebaasidele laienevad autori isiklikud ja varalised õigused.

Üks kesksemaid isiklike õigusi on õigus autorsusele (AutÕS § 12 lg 1 p 1), mida tuleb järgida nii keeleressursside loomisel kui ka kasutamisel. Samas võib praktikas tekkida probleeme juhul, kui teos avaldatakse tööandja nime all. Tööandja saab seda teha juhul, kui töölepingus on selgelt kokku lepitud isiklike õiguste teostamise osas. Keeleressursside, näiteks tekstikorpuste puhul ei ole siiani eriti rõhutatud nende seost ühe või teise isiku kui autoriga, kes loob teose, valides tekstimassiividest korpuse lingvistiliseks sisuks sobivad tekstid, lõigud, laused ja annoteerimisviisi. Teadustulemustes ja artiklites viidatakse praeguses praktikas korpustele enamasti tööandja (näiteks Tartu Ülikooli tasakaalus korpus<sup>36</sup>) või allüksusena uurimisrühma (näiteks TÜ suulise kõne uurimisrühma korpus<sup>37</sup>) kaudu. Kuna keeleressursid on oma olemuselt ka kollektiivsed teosed<sup>38</sup>, siis on taoline praktika õigustatud.

<sup>35</sup> Autoriõiguse seadus (§ 19 p 2 ja 3) lubab kasutada autoriõiguslikult kaitstavaid teoseid ilma autori nõusolekuta ja autoritasu maksimiseta teaduslikel eesmärkidel. Vaatamata asjaolule, et teadustöö erand praegu kehtivas autoriõiguse seaduses saab olla aluseks keeleressursside loomisele, tuleks õigusselguse huvides kaaluda konkreetse keeleressursside loomisega seonduva erandi sätestamist autoriõiguse seaduses.

<sup>36</sup> Vt <http://www.cl.ut.ee/korpused/grammatikakorpus/> (22.12.2012).

<sup>37</sup> Vt <http://www.cl.ut.ee/suuline/Ylevaade> (22.12.2012).

<sup>38</sup> Autoriõiguse seaduse kohaselt "[k]ollektiivseks teoseks loetakse teost, mis koosneb erinevate autorite kaastöödest, mis on ühendatud ühtseks tervikuks füüsilise või juriidilise isiku poolt tema initsiatiivil ja juhtimisel ning mis antakse

Keeleressursside litsentsimisel on isiklikest õigustest kõige suurema praktilise tähtsusega niinimetatud muutmisõigused: õigus teose puutumatusel ja õigus teose lisadele. Põhimõtteliselt lubab autoriõiguse seadus teost muuta vaid autori selgesõnalisel nõusolekul. Jälle on tööandja huvides see, et töölepingus oleks muutmisõiguste teostamise osas saavutatud eelnev kokkulepe.<sup>39</sup>

Keeleressursside loomisel ja kasutamisel saab kohaldada järgmisi autori varalisi õigusi: õigust teose reprodutseerimisele, õigust teose levitamisele, õigust teose tõlkimisele, õigust teose töötlemisele, õigust teoste kogumikele, õigust teose avalikule esitamisele, õigust eksponeerimisele, õigust teose edastamisele ja õigust teose üldsusele kättesaadavaks tegemisele (AutÕS § 13). Juba selle loetelu pikkus näitab, kui võrd keeruliseks võib praktikas kujuneda andmebaasi kasutamine. Tööandja omandab selle paketi töötaja töökohustuste korras loodud teostele seaduse alusel (AutÕS § 32). Töövõtulepingus tuleks nende õiguste omandamises või kasutamises kokku leppida kas üksikute õiguste kaupa või nn pakatina tervikuna – viimane on kasutajale soodsam<sup>40</sup>. Samas võib keeleressursside omaja olla ise valmis sõlmima litsentsilepinguid teiste kasutajatega just sellesama varaliste õiguste paketi piirides.

Litsentsimise juures kerkib veel üks oluline küsimus – milline on täienduste õiguslik staatus. Kui konkreetne keeleressurss ise on tehtud üldsusele vabalt kättesaadavaks, siis peaks litsentsilepingus nägema ette tingimuse, et ka selle täiendusi levitatakse samadel tingimustel. Kui keeleressursside omajal on huvi litsentsitavat ressursi ka ise edasi arendada, siis võib leppida kokku, et litsentsisaaja kohustub litsentsiandjale omakorda andma loa tema poolt tehtud täienduste kasutamiseks.

## 5. Võimalikud litsentsilepingute eeskujud keeleressursside litsentsimisel

### 5.1. Vabade (avatud) litsentside toimemehhanismid

Sõltumata sellest, kas tegemist on koodi, muu sisu või andmebaaside jaoks mõeldud avatud litsentsidega, toimivad need üldjoontes ühtmoodi. Litsentsiandjaks on reeglina autor (või iga kaasautor või mõni muu autoriõiguste omaja) ja litsentsisaajaks teose või *sui generis* andmebaasi kasutaja. Kõik avatud litsentsid lubavad litsentsisaajal teost reprodutseerida ja oma tarbeks muuta ning muutmata koopiaid levitada. Avatud litsentsid võivad litsentsisaajale seada ka teatavaid kohustavaid tingimusi, millest levinuimad on

- 1) teose (või sellest tuletatud teoste<sup>41</sup>) levitamisel viitamine algsele autorile ja teosele ning
- 2) teose või sellest tuletatud teoste levitamine algse litsentsiga samadel või sarnastel tingimustel.

Tähelepanelik tuleb olla samadel tingimustel jagamise kohustust ('jagamine samadel tingimustel', ingl *share-alike*) sisaldava litsentsiga tähistatud materjalide

---

välja selle füüsilise või juriidilise isiku nime või nimetuse all (teatmeteosed, teaduslik kogumik, ajaleht, ajakiri ja teised perioodilised või jätkuväljaanded jms)" (AutÕS § 31 lg 1).

<sup>39</sup> Käimasoleva intellektuaalse omandi reformi kontekstis on kaalumisel piirata autori isiklike õiguste ulatust viies muutmisõigused autori varaliste õiguste kataloogi. Täiendav info kättesaadav arvutivõrgus: <https://ajaveeb.just.ee/intellektuaalneomand/> (22.12.2012).

<sup>40</sup> Rõhutame, et esimesel juhul on töösuhte aluseks tööleping. Töövõtulepingu puhul on aga tegemist iseseisva tsiviilõigusliku lepingulise suhtega.

<sup>41</sup> Autoriõiguse seadus (§ 4 lg 3 p 21) määratleb tuletatud teosena teose tõlget, algse teose kohandust (adaptsiooni), töötlust (arrangeeringut) ja teose muud töötlust. Minemata sügavuti tuletatud teose olemusse, võib öelda, et tuletatud teose loomine tähendab algteose vähemal või suuremal määral muutmist.

kasutamisel. Sõltuvalt konkreetse litsentsi tingimuste sõnastusest võib see tingimus nõuda litsentsitud teosest tuletatud teoste levitamisel täpselt sama litsentsi kasutamist. See tähendab, et juhul, kui loodav teos on tuletatud teoseks mitme erinevate “jagamine samadel tingimustel”-nõuet sisaldava litsentsi alusel levitatava komponentteose suhtes, ei pruugi uue teose levitamine olla võimalik ilma ühe või enama komponentteose autoriõiguste valdajailt selleks eraldi loa saamist.

Mõned teose reprodutseerimist ja levitamist lubavad litsentsid võivad piirata ka litsentsitava teose kasutusviise (nt lubades selle kasutamist ainult mitteäriksel eesmärgil või keelates sellest tuletatud teoste levitamist) või kasutajaskonda (nt lubades teose ja sellest tuletatud teoste koopiade levitamist ainult kindlaksmääratud kogukonna piires nagu META-SHARE Commons litsentsid<sup>42</sup>). Tõsi, sellisel juhul ei ole enam tegemist vabade (avatud) litsentsidega Free Software Definition’i<sup>43</sup>, Open Source Definition’i<sup>44</sup> ega Open Definition’i<sup>45</sup> mõttes.

## 5.2. META-SHARE ja CLARIN võrgustikud ja nendes kasutatavad litsentsid

Mitmekeelses Euroopas on kaks suuremat algatust, mis püüavad koguda ja teha kättesaadavaks paljude Euroopa keelte keeleressursse: META-NET<sup>46</sup> ja CLARIN<sup>47</sup>.

META-NET on kujundanud keeleressursside litsentsimiseks põhjaliku poliitika ja avalikustanud oma veebiportaalis üksikasjaliku ülevaate, miks ja millised litsentsid on välja töötatud<sup>48</sup>. Tartu Ülikool osaleb selles algatuses projekti META-NORD<sup>49</sup> kaudu ja on kohustatud seetõttu töös hoidma META-SHARE<sup>50</sup> registrit ja repositooriumi (nn võrgustikusõlme ehk *node*’i kujul), milles hoitakse ka Eesti keeleressursse. Registreeritud hoitakse ressursside kirjeldusi ehk metaandmeid ja repositooriumis ressursside endid. Metaandmete hulka kuulub ka teave ressursi omaniku kohta, omaniku nõusolek (ingl *Depositor’s Agreement*) ressursi võrgustiku sõlme kaudu jagada ning erinevatele kasutajatele mõeldud kasutusõigused vastavate litsentside kujul.

META-SHARE üldine poliitika nõuab, et repositooriumis kättesaadavaks võib teha vaid neid ressursside, mille kasutusõigused on selged ning puuduvad autoriõiguslikud või isikuandmete kaitsest tulenevad piirangud. Sellepärast on vajalik saavutada juba kasutusel olevate ja mitmel pool veebi kaudu kättesaadavate keeleressursside litsentsimine.

META-SHARE lähenemine sarnaneb CC omale: litsentsiandja ei ole litsentsitingimuste valikul täiesti vaba, vaid peab valima olemasolevate tüüplitsentside hulgast talle sobivaima. Valida on kolme tüüpi litsentside vahel:

- 1) CC litsentsid: võimalusel CCO, vajadusel koos täiendavate piirangutega (autorile viitamise või “jagamine samadel tingimustel” kohustus; äriksel kasutuse või tuletatud teoste levitamise keeld);
- 2) META-SHARE Commons litsentsid (lubavad levitamist META-SHARE võrgustiku siseselt; piirangute valik on sarnane CC litsentsidega);

<sup>42</sup> Vt <http://www.meta-net.eu/meta-share/licenses> (21.12.2012).

<sup>43</sup> Free Software Definition <http://www.gnu.org/philosophy/free-sw.html> (21.12.2012).

<sup>44</sup> Open Source Definition <http://opensource.org/docs/osd> (21.12.2012).

<sup>45</sup> Open Definition <http://opendefinition.org> (21.12.2012).

<sup>46</sup> META-NET <http://www.meta-net.eu/> (21.12.2012).

<sup>47</sup> CLARIN <http://www.clarin.eu/> (21.12.2012).

<sup>48</sup> Vt <http://www.meta-net.eu/meta-share/licenses> (21.12.2012).

<sup>49</sup> META-NORD <http://www.meta-nord.eu/> (21.12.2012).

<sup>50</sup> META-SHARE [http://www.meta-net.eu/meta-share/index.html-et?set\\_language=et](http://www.meta-net.eu/meta-share/index.html-et?set_language=et) (21.12.2012).

- 3) META-SHARE NoRedistribution litsentsid (keelavad litsentsitavate teoste modifitseerimata koopiade levitamise, muud piirangud on sarnased CC litsentsidega).

META-NETi omanikuleppe kohta (ingl *Depositor's Agreement*, DA) on Eesti keeleteadlastel juba üks positiivne näide olemas. META-NORDi projekti raames pidid kõik osalevad keeled looma puudepanga (vt keeleressursside liigitust ptk 2 alapunkt c) ja et kõikide osalevate keelte puudepangad oleks võrreldavad, kasutati puudepangas analüüsiks üht ja sama teksti, Norra kirjaniku Jostein Gaarderi teose "Sophie maailm" esimese peatüki lauseid. Seega oli vaja saada DA mitte ainult originaalkeeles teosele, vaid ka selle kõikidele tõlgetele.

CLARINit ehk üldist keeleressursside ja -tehnoloogia infrastruktuuri ettevalmistavas projektis tegeldi samuti keeleressursside õiguslike teemadega ja selle kohta valmis raport.<sup>51</sup> Raporti koostajad, Helsingi ülikooli keeleteadlased, on hiljem koostöös juristidega avaldanud samal teemal ka artikleid (Oksanen jt 2010, Oksanen, Lindén 2011) keeletehnoloogia ja -ressursside konverentsidel, kus keeleressursside õiguslikud aspektid tekitavad viimastel aastatel väga elavat huvi.

Eestil CLARIN ERICu asutajaliikmena on kohustus järgida ERICu põhikirjas<sup>52</sup> märgitud raamtingimusi ligipääsu, intellektuaalomandi ja litsentsimise osas. Eesti Keeleressursside Keskusel kui CLARIN ERICu riiklikul konsortsiumil on kohustus hallata keeleressursse ka kui intellektuaalset omandit ja järgida seejuures CLARIN ERICu raamtingimusi.

CLARINis kasutatakse uute ressursside litsentsimisel standardiseeritud tüüplepingut, mis sisaldab rida erinevaid valikulisi tingimusi, mille hulgast litsentsiandja valib endale sobivad. Olemasolevate ressursside litsentside klassifitseerimiseks on CLARINi jaoks välja töötatud nn "pesumärkide" süsteem (Oksanen, Lindén 2012), mille peamine eesmärk on hõlbustada ressursside kasutajail ülevaate saamist olulistest litsentsitingimustest. Pesumärkidega tähistatakse esmalt iga ressursi maksimaalne leviala (PUB – avalik, ACA – akadeemiline, RES – piiratud) ning täiendavate tähistega null kuni kolm lisapiirangut (NC – ärilise kasutuse keeld, INF – kohustus informeerida ressursi autoriõiguste omajat selle kasutamisest publitseeritud artiklites, ReD – kohustus ressursi modifitseeritud versioonid CLARINi repositooriumisse laadida).

Sarnaselt META-SHAREiga soovitatakse ka CLARINis litsentsida sinna lisatavaid uusi keeleressursse nii väheste piirangutega kui võimalik, tagamaks nende vaba kasutatavus vähemalt vastava võrgustiku piires. Mõlemas võrgustikus on esimeseks eelistuseks CCo.

## 6. Järeldused ja ettepanekud

Mida vabamalt saab kasutada avalike vahendite baasil loodud keeleressursse ja -tehnoloogiaid, seda parem eesti keelele, kultuurile ja rahvale. Avatud juurdepääsu põhimõtteid tuleks laiendada kõigile avalikest vahenditest finantseeritud teadustulemuste vormidele ja kasutusviisidele keeleressursside kontekstis.

Võimalusel on soovitatav valida keeleressursside- ja tehnoloogiade levitamiseks litsentsid, mis toetavad avatud juurdepääsu põhimõtteid. Erinevate litsentside

<sup>51</sup> A report including Model Licensing Templates and Authorization and Authentication Scheme. Vt <http://www-sk.let.uu.nl/u/D75-2.1.pdf> (21.12.2012).

<sup>52</sup> ET 3.3.2012 Euroopa Liidu Teataja L 64, vt art 2.2 (g), art 19, 22, 25.

kasutamisel tuleb arvestada võimalike ühildumisprobleemidega. Lihtsaim võimalus ületada ühildumisprobleeme on valida litsents, mis kõrvaldab kõige enam autori ainuõigustest tulenevaid piiranguid, lubades muuhulgas ärilisel otstarbel kasutamist avalikes huvides oleval eesmärgil ning kohustades seejuures üksnes autorile viitama. Selleks sobiksid näiteks Creative Commons autorile viitamise kohustusega (CC BY) või CCo, millega keeleressurss antaks üldiseks kasutamiseks.

Keeleressursside litsentsimisel on oluline arvestada nii andmebaasi, selle sisu kui ka selle kasutamiseks vajalike arvutiprogrammide õigusliku režiimiga. Samuti on tähtis, et kõigi isikutega, kes keeleressursside loomisesse panustavad, oleksid sõlmitud korrektsed lepingud.

Asjaolu, mis võib ulatuslikult kitsendada keeleressursside (eelkõige korpuste) litsentsimise võimalusi avatud juurdepääsu põhimõtete kohaselt, on see, kui nende loomiseks vajalikke teoseid kasutati tuginedes autoriõiguse teose vaba kasutamise eranditele. Sellisel juhul peab ka kasutus piirduma teadustöö eesmärkidega.

Litsentsiraamistiku kujundamisel Eesti keeleressursside jaoks tuleb arvestada rahvusvahelise dimensiooniga. Vajalik oleks võtta kasutusele litsentsiraamistik, mis tagab Eesti keeleressursside kasutustingimuste ühildumise rahvusvahelises kontekstis. Seetõttu ei ole otstarbekas töötada välja Eesti-spetsiifilisi tüüptingimusi, vaid soovitav on valida juba olemasolevate litsentside vahel. Praeguses seisus tunduvad olevat kõige sobilikumad Creative Commons'i litsentsid. Täiendavalt tuleks koostada keeleressursside loojate jaoks praktiline juhend, milliseid litsentsitüüpe konkreetsel juhul kasutada.

## Viidatud kirjandus

AutÕS = Autoriõiguse seadus. RT I 1992, 49, 615 ... RT I, 28.12.2011, 1.

IKS = Isikuandmete kaitse seadus. RT I 2007, 24, 127 ... RT I, 30.12.2010, 2.

Kelli, Aleksei; Mets, Tõnis; Pisuke, Heiki; Vasamäe, Elise; Värvi, Age 2010. Trade secrets in entrepreneurs' intellectual property strategies: The Estonian experience. – Review of Central and East European Law, 35 (4), 315–339. <http://dx.doi.org/10.1163/157303510X12650378240476>

Kelli, Aleksei; Tavast, Arvi; Pisuke, Heiki 2012. Copyright and Constitutional aspects of digital language resources: The Estonian approach. – Juridica International, 19, 40–48.

Kreutzer, T. 2011. User-related assets and drawbacks of open content licensing. – L. Guibault, C. Angelopoulos (Eds.). Open Content Licensing: From Theory to Practice. Amsterdam University Press.

Liin, Krista; Muischnek, Kadri; Müürisep, Kaili; Vider, Kadri 2012. Eesti keel digiajastul. META-NET Valge raamatu sari. Toim: Hans Uszkoreit, Georg Rehm. Springer. <http://www.meta-net.eu/whitepapers/volumes/estonian> (23.09.2012).

Mõtsküla, Peeter P. 2009. Vaba tarkvara litsentside süstemaatiline analüüs. Magistritöö. Tartu Ülikooli õigusteaduskond.

Oksanen, Ville; Lindén, Krister; Westerlund, Hanna 2010. Laundry symbols and license management: Practical considerations for the distribution of LRs based on experiences from CLARIN. – Proceedings of LREC 2010: Workshop on Language Resources: From Storyboard to Sustainability and LR Lifecycle Management. <https://helda.helsinki.fi/bitstream/handle/10138/29359/LREC2010.pdf?sequence=2> (21.12.2012).

Oksanen, Ville; Lindén, Krister 2011. Open Content Licenses: How to choose the right one. – Workshop on Visibility and Availability of LT Resources, NODALIDA 2011.



[https://helda.helsinki.fi/bitstream/handle/10138/29355/Open\\_Content\\_Licenses.pdf?sequence=2](https://helda.helsinki.fi/bitstream/handle/10138/29355/Open_Content_Licenses.pdf?sequence=2) (21.12.2012).

- Oksanen, Ville; Lindén, Krister 2012. Building shared language research environments inside European Union – how to optimize the system based on experiences from real life. Presented at The 1st Global Thematic IASC Conference on the Knowledge Commons. [http://biogov.uclouvain.be/iasc/doc/full\\_papers/Ville-Linden.pdf](http://biogov.uclouvain.be/iasc/doc/full_papers/Ville-Linden.pdf) (21.12.2012).
- Pisuke, Heiki; Kelli, Aleksei 2007. Some issues regarding entrepreneurial universities and intellectual property. – Juridica International 12, 161–172.
- Tavast, Arvi; Kelli, Aleksei; Pisuke, Heiki 2013. Õiguslikud väljakutsed ja võimalikud lahendused keeleressursside arendamisel. – Eesti Rakenduslingvistika Ühingu aastaraamat, 9, 317–332. <http://dx.doi.org/10.5128/ERYa9.20>
- Tuulik, Triin 2012. Avatud juurdepääsu kontseptsiooni mõned õiguslikud aspektid. – Juridica, VII, 522–528.
- VõS = Völaõigusseadus. RT I 2001, 81, 487 ... RT I, 08.07.2011, 6.
- CLARIN ERIC põhikiri. ET 3.3.2012 Euroopa Liidu Teataja L 64. <http://www-sk.let.uu.nl/u/D8S-5a.pdf> (14.12.2012).

## Võrgumaterjalid

- Avatud juurdepääsu määratlusi. Open Knowledge Foundation'i Open Definition. <http://opendefinition.org> (21.12.2012); Vabade Kultuuriliste Teoste Definiitsioon. Definition of Free Cultural Works. <http://freedomdefined.org> (21.12.2012).
- CLARIN <http://www.clarin.eu/> (21.12.2012).
- CLARIN ERICu põhikiri. ET 3.3.2012 Euroopa Liidu Teataja L 64. <http://www-sk.let.uu.nl/u/D8S-5a.pdf> (14.12.2012).
- CLARINi õigusasjade raport. (A report including Model Licensing Templates and Authorization and Authentication Scheme.) <http://www-sk.let.uu.nl/u/D7S-2.1.pdf> (21.12.2012).
- Creative Commons <http://creativecommons.org> (21.12.2012).
- CC Eesti <http://creativecommons.ee> (21.12.2012).
- CCo. <http://creativecommons.org/choose/zero/> (21.12.2012).
- EKT riikliku programmi tekst <http://www.keeletehnoloogia.ee/EKT2011-2017-programm-uuendet.pdf> (14.12.2012).
- EUPL <http://ec.europa.eu/idabc/servlets/Doc3cb1.pdf?id=29923> (22.12.2012); <http://ec.europa.eu/idabc/en/document/7774.html> (23.12.2012).
- Free Software Definition <http://www.gnu.org/philosophy/free-sw.html> (14.12.2012).
- META-NET <http://www.meta-net.eu/> (21.12.2012).
- META-NET pressiteade <http://www.meta-net.eu/whitepapers/press-release-et> (23.09.2012).
- META-NORD <http://www.meta-nord.eu/> (21.12.2012).
- META-SHARE [http://www.meta-net.eu/meta-share/index\\_html-et?set\\_language=et](http://www.meta-net.eu/meta-share/index_html-et?set_language=et) (21.12.2012).
- META-SHARE Commons litsentsid <http://www.meta-net.eu/meta-share/licenses> (21.12.2012).
- META-SHARE harta [http://www.meta-net.eu/meta-share/METASHARE\\_Charter.pdf](http://www.meta-net.eu/meta-share/METASHARE_Charter.pdf) (14.12.2012).
- ODC PDDL <http://opendatacommons.org/licenses/pddl/> (21.12.2012).
- Open Data Commons <http://opendatacommons.org> (21.12.2012).
- Open Source Definition <http://opensource.org/docs/osd> (21.12.2012).
- Open Source Software <http://opensource.org> (14.12.2012).
- Riiklik programm “Eesti keele keeletehnoloogiline tugi 2006–2010” (EKKT) <http://www.keeletehnoloogia.ee/ekkt-1> (14.12.2012).
- Riiklik programm “Eesti keeletehnoloogia 2011–2017” (EKT) <http://www.keeletehnoloogia.ee> (14.12.2012).

- Tartu Ülikooli tasakaalus korpus <http://www.cl.ut.ee/korpused/grammatikakorpus/> (22.12.2012).
- Tartu Ülikooli suulise kõne uurimisrühma korpus <http://www.cl.ut.ee/suuline/Ylevaade> (22.12.2012).
- Tartu Ülikooli intellektuaalomandi kord [http://www.ut.ee/sites/default/files/livelink\\_files/intellektuaalse\\_omandi\\_ksitlemise\\_phimtted\\_tartu\\_likoolis\\_terviktekst.pdf](http://www.ut.ee/sites/default/files/livelink_files/intellektuaalse_omandi_ksitlemise_phimtted_tartu_likoolis_terviktekst.pdf) (23.09.2012).
- Tallinna Ülikooli intellektuaalomandi kord <http://www.tlu.ee/files/arts/314/intel2e2435b-4d99b60a2481cd7c54d2f93a7.pdf> (23.09.2012).
- Vaba tarkvara litsentsid <http://www.gnu.org/licenses/license-list.html> (21.12.2012); <http://opensource.org/licenses> (21.12.2012).

**Kadri Vider** (Eesti Keeleressursside Keskus, Tartu Ülikool), üks uurimissuundi on keeletehnoloogiliste ressursside ja tarkvara haldamine, kättesaadavaks tegemine ning litsentsimine või sellealane konsulteerimine.  
[kadri.vider@ut.ee](mailto:kadri.vider@ut.ee)

**Heiki Pisuke** (Euroopa Liidu Nõukogu, Tartu Ülikool), põhilised uurimissuunad on intellektuaalne omand (autoriõigus ja autoriõigusega kaasnevad õigused), tõlkimise õiguslikud küsimused, ülikoolide intellektuaalse omandi küsimused.  
[heiki.pisuke@ut.ee](mailto:heiki.pisuke@ut.ee)

**Peeter P. Mõtsküla** (Tartu Ülikool), põhilised uurimissuunad on intellektuaalne omand, autoriõigus (eriti tarkvara õiguskaitse) ning litsentsimine.  
[peeterpaul@motskula.net](mailto:peeterpaul@motskula.net)

**Triin Tuulik** (Tartu Ülikool), põhilised uurimissuunad on intellektuaalne omand (eriti looja-keskne intellektuaalomand ettevõtjate perspektiivist).  
[tuultriinik@gmail.com](mailto:tuultriinik@gmail.com)

**Aleksei Kelli** (Tartu Ülikool) põhilised uurimissuunad on intellektuaalne omand, innovatsioon, teadmispõhine majandus, digitaalsete keeleressursside õiguslikud küsimused.  
[aleksei.kelli@ut.ee](mailto:aleksei.kelli@ut.ee)

# LEGAL CONSIDERATIONS IN LICENSING LANGUAGE RESOURCES

**Kadri Vider<sup>1,2</sup>, Heiki Pisuke<sup>3,2</sup>, Peeter P. Mõtsküla<sup>2</sup>,  
Triin Tuulik<sup>2</sup>, Aleksei Kelli<sup>2</sup>**

Center of Estonian Language Resources<sup>1</sup>, University of Tartu<sup>2</sup>, Council of the European Union<sup>3</sup>

The purpose of this article is to present the underlying legal considerations for licensing language resources. The authors concentrate on analysing language resources in the form of data bases, addressing the complex legal issues concerning copyright and related rights. Different frameworks for licensing language resources, such as META-NET, CLARIN and Creative Commons standard licences, are explained and compared in the context of Open Access.

The article has been written in cooperation between linguists and lawyers and is practically oriented. Its purpose is to address the scholars and scholarly institutions associated with Estonian language resources and explain the fundamental principles of licensing objects of intellectual property. The authors highlight risks and problems, which can become serious obstacles, when combining and processing several language resources, each subjected to a different licence regime.

The article provides an overview of the current situation regarding licensing of language resources in Estonia, establishes the main legal characteristics of the data base as an object of intellectual property rights, and offers guidance regarding the selection of a suitable set of standard rules for licensing such data bases.

**Keywords:** human language resources, language technology, copyright, licensing, open access, META-SHARE, CLARIN