NATURAL SPEAKING AND HOW TO ASSESS IT

Hille Pajupuu[1], Krista Kerge[2], Lya Meister[3], Eva Liina Asu[4], and Pilvi Alp[5]

[1]*Institute of the Estonian Language, Tallinn,* [2]*Tallinn University,* [3]*Institute of Cybernetics at Tallinn University of Technology,* [4]*University of Tartu, and* [5]*National Examinations and Qualifications Centre, Tallinn*

Abstract. One of the problems in testing the proficiency of Estonian as a first or second language is that high-stake exams are assessed against the standards of the written language. Given this, we set out to describe the features of the actual use of educated language in different types of text. The goal was to develop L1 and L2 teaching and testing through models of educated language use which a language learner can approach step by step. To achieve this goal we compared the following features of educated use of Estonian as L1 and L2 in different situations: (1) lexical richness and vocabulary range; (2) contextuality and formality of the text; (3) syntactic complicacy; (4) temporal characteristics of the dialogue; (5) strength and disruptiveness of the foreign accent; (6) sentence intonation. The results show that educated language use is mainly genre-dependent. This moves the focus of language learning onto texts of specific genres and confirms the suitability of an action-based approach centred on genres in L1 and L2 teaching and testing, and the need for regular assessor training.

Keywords: L1, L2, language teaching and learning, language testing, natural language use, genres, vocabulary, formality, accent, intonation, syntactic complicacy

1. Introduction

In Europe, the standard set of levels of language proficiency (CEFR) has been widely accepted as a common standard to help linguistic communities establish within and between themselves (a) language learning objectives for learners needing to manage in typical work, social, personal, or educational situations (the action-oriented approach), and (b) proficiency exams to measure how learners cope with these situations.

As test takers' chances of getting a job in future or continuing their studies depend on the results of any official exam (so-called high-stakes exams), language

testing should ideally be balanced and fair. At the same time, because there are no benchmarks to measure how native speakers cope with the situations as compared to the L2-speakers of the language[1], it is easy for the assessors to give too much credit to standard (i.e. normative) language use or to features of speaking or writing that are considered to be correct in school-grammar, but never measured to be characteristic of modern natural speech. Our goal is to measure some parameters of real-life use of the educated Estonian language in order to put these aspects in the right proportion. As some features of L2 use may be irrelevant in managing particular situations, and should not be taken into account in testing under the action-oriented approach, we also collected L2 data and compared it to L1 data.

Until the autumn of 2008, Estonian language proficiency was tested nationally on three levels. After this date the proficiency exams of the Council of Europe were adopted (Table 1). To harmonise national and CE language exams, new guidelines for assessing writing and speaking skills were needed.

This article looks at the speaking proficiency of the proficient or effective user, the level where all essential language skills and competences can be tested (see CEFR, or in Estonian, Kerge 2008).

Table 1. The approximate correspondence of proficiency exams of Estonian as the official language to the proficiency levels of the Council of Europe*

Proficiency levels of the Council of Europe		Proficiency exams of Estonian as the official language	
		until autumn 2008	**since autumn 2008**
Proficient User	C2	(neither nationally tested nor required in Estonia)	(neither nationally tested nor required in Estonia)
	C1	High Level Estonian proficiency exam	C1-level Estonian exam (Proficient User)
Independent User	B2+		
	B2	Medium Level Estonian proficiency exam	B2-level Estonian exam (Advanced User)
	B1+	Lower Level Estonian proficiency exam	B1-level exam (Independent User)
	B1		
Basic User	A2+		A2-level Estonian exam (Beginner)
	A2	(neither nationally tested nor required in Estonia)	
	A1	(neither nationally tested nor required in Estonia)	(neither nationally tested nor required in Estonia)

* The striped area in the Table shows the overlapping area of B2-level exam and C1-level exam at B2+ level.

---

[1] Some Estonian text examples (L1 and L2 oral dialogues and presentations, L1 and L2 written essays, L2 written answers at exam of history, two translations into Estonian) are analysed in Kerge 2008: 203–235.

Speaking involves two essential skills: oral interaction (dialogue) and oral pre-sentation (monologue). Both skills were tested in high-level language proficiency exams until 2008, and are now tested in C1-level exams. In this study we look at both monologues and dialogues. We believe that the test taker's performance should not be compared to the standard or officially correct usage, sometimes unattainable even for a native L1 linguist, but to natural usage, which focuses on the message, undisturbed by the sound of speaking or the choice, form, order and binding of lexical units (see also Ratcliff *et al.* 2002).

We take natural language to mean the language that is usually spoken and written by a university-educated non-linguist. The language use of such a person can be considered the reference model of natural language. Our research task is to establish which aspects are important in defining natural speech, and which aspects should be taken into account when assessing speaking skills at a C1-level exam. (Our research questions are presented below under the parameters studied).

Depending on the text type (monologue or dialogue)[2], we will look at the follow-ing aspects: (1) lexical richness (the Uber index of the balance of words and tokens) and vocabulary range (proportion of basic vs. rare words); (2) contextuality and formality (F-index relating context-free vs. context-bound vocabulary, inversely proportional to text ambiguity); (3) complicacy of syntax (sentence length and complexity, plus degree of nominalisation); (4) temporal characteristics of the dialogue (culture-specific length of turn, pausing, simultaneous talking); (5) strength and disruptiveness of the L2 foreign accent (relationship between the perceived strength and disruptiveness); (6) aspects of L2 intonation (patterns of rising intonation).

In order to establish a foundation for the L2 proficiency assessment, we will compare L1 and L2 dialogues and monologues produced at language exams, and present comparative data on spontaneous Estonian speech in dialogues. We will describe some features of language use by looking at both oral text types and written text data. However, as can be seen below, some features lack earlier data that would allow comparisons to be made.

2. Materials and method

The text material for this study was collected in a standardised situation in 2006: for non-natives at the Estonian High Proficiency exams, and for natives in an exam-like situation (same examiner, same time-limit, same task).

There were three tasks: writing an essay (subjective discussion expressing opinions, 250 words, 60 minutes); speaking with another test taker (oral dialogue in the style of a negotiation, 5–7 minutes, see example in Figure 1); and giving a 1–2-minute presentation (oral monologue, see example in Figure 2). The dis-cussion topics were linked by two keywords: environment and society.

---

[2]   In this article, a distinction is made between the terms *text type* (oral or written monologue or dialogue) and *genre* (linguistic expression of register as a contextual semantic configuration).

REAL ESTATE SHARKS VERSUS RESIDENTS

In Estonian urban planning the business interests of real estate developers are often considered above those of the citizens.

Have a conversation and make a joint decision about:

- *What the main faults of the Estonian urban environment are, and*
- *What the city-dwellers could do to improve their own environment*

Figure 1. Conversation slip for a dialogue between two exam candidates.

1. Does the future of Estonia's living environment lie in blocks of flats or detached houses?

2. What are the main shortcomings of Estonia's modern urban environment?

3. How and how much should the state support less successful citizens?

Figure 2. Presentation topics. Candidates were asked to choose one topic and give a short presentation after one minute of preparation.

Although we recorded 24 exam candidates at the high proficiency exam, we chose only eight recordings for in-depth analysis. These eight candidates got the maximum or a near-maximum score for their speaking skills and their total scores were sufficient for them to receive a high proficiency certificate.

The subjects formed two groups that were comparable in terms of language proficiency requirements[3]: 8 native Estonian speakers and 8 native Russian speakers (four women and four men in each group), all fluent in spoken and written Estonian.

The recordings were carried out using a digital recorder (sampling frequency 44.1 kHz, 16 bit, mono) and a high-quality microphone at a distance of about one metre from the candidates.

The oral data was analysed with the speech analysis software PRAAT (Boersma and Weenink 2006) and Sony Sound Forge 9.0, and transcribed syntactically and morphologically. WordSmith Tools 3.0 was used to differentiate between words (i.e. different words, or types) and tokens (word forms) (Scott 1996). Morphological Analyser 3.3 (freeware available from http://www.eki.ee/tarkvara/analyys) was used for parts of speech.

---

[3] Due to the official language requirements in public service, public health, and legal affairs, Estonian language skills were tested at three levels until 2008 and are now tested at four levels. Command of the official language at the highest level (C1) is mostly obligatory for jobs requiring higher education (heads of public institutions, civil servants, lawyers, doctors, teachers of Estonian or students whose language of tuition is Estonian, military officers, etc.).

The spontaneous dialogue data come from Pajupuu (1995).

The methods used will be described in detail under Analysis and Results.

3. Analysis and results

3.1. Lexical richness and vocabulary range

In proficiency testing, vocabulary is mainly viewed in the context of productive skills such as speaking and writing (see Dewaele and Pavlenko 2003). CEFR (2001: 112) presumes that a C1-level language user "*Has a good command of a broad lexical repertoire allowing gaps to be readily overcome with circumlocutions; little obvious searching for expressions or avoidance strategies. Good command of idiomatic expressions and colloquialisms.*"

Both vocabulary use and range are markers of linguistic competence and fluent speech (Little 2005, Read and Chapelle 2001). By subjective assessment it is possible to assess fluency. However, assessing vocabulary range and lexical richness by assessing how well each topic is covered in terms of vocabulary is an extremely complicated task, especially for spoken language.

The question also arises whether it is necessary to deal with these aspects of L2 vocabulary separately when assessing fluency, communicativeness and adequacy of language use in C1-level exams.

We aim to describe lexical richness and vocabulary range as markers of natural speech and consider the importance of these criteria in the subjective assessment of L2 skills. Our research questions are:

- How rich is the vocabulary of educated L1 and L2 users?
- How can the vocabulary range of educated L1 and L2 users be described given the average frequency of words in Estonian?
- Do lexical richness and vocabulary range vary in different forms of language use and text types (spoken dialogue and monologue, written essay as a monologue)?

Lexical richness can be assessed by looking at the number of different words in text. For this we used the Uber index: $U = (\log N)^2/(\log N - \log V)$, where N is the total number of word forms (tokens) and V is the number of different words (types)[4]. This formula is an algebraic transformation of the TTR (type/token ratio), which reduces somewhat the influence of text length on lexical richness assessment, and is suitable for short texts (see Jarvis 2002, Vermeer 2000: 76–79).

We measured lexical richness separately in all the text types under study: oral presentation (monologue), conversation (dialogue) and, as a comparison, written essay (see Table 2 and Figure 3).

---

[4] Only correctly used words were counted as words (i.e. words used in inappropriate contexts or in a way that would impair understanding were excluded). For example we excluded the word *pulpulistlik* as a non-word or strongly deformed version of *populistlik* [*populistic*], and some completely incomprehensible sound combinations.

Table 2. L1 and L2 material by text type and lexical richness index *U*.

Text type	Types (*V*)		Tokens (*N*)		Uber index (*U*) $U = (_{\log}N)^2/(_{\log}N - _{\log}V)$	
	L1	*L2*	*L1*	*L2*	*L1*	*L2*
Oral DIALOGUE	557	542	1745	2337	21.4	17.8
Oral MONOLOGUE	498	379	1315	1343	23.1	17.8
Oral production (dialogue + monologue)	884	745	3060	3680	**22.7**	**18.2**
Written essay (MONOLOGUE)	737	661	1684	1825	**29.0**	**24.2**
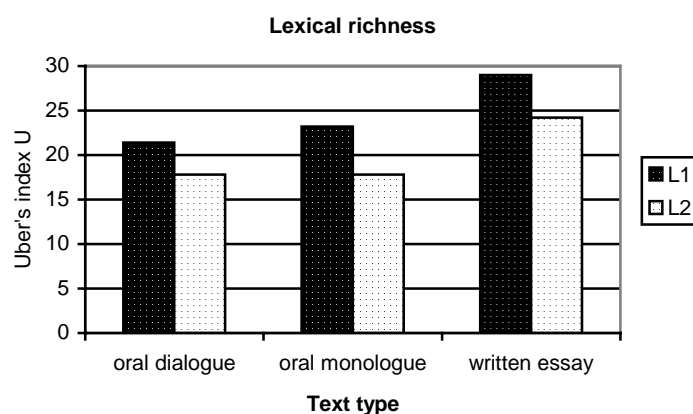
Figure 3. L1 and L2 lexical richness in different text types. The higher the value of *U*, the richer the vocabulary.

The results show that text types hold different levels of lexical richness. There is a natural increase in lexical richness from dialogue to monologue and from oral production to written production.

In oral texts L1 vocabulary is richer than L2 vocabulary, but as this has not stopped the assessors from giving high scores for speaking skills to our candidates, we conclude that listeners are not disturbed by the poorer L2 vocabulary.

The assessment of vocabulary range usually relies on the belief that more frequent words are better known than rarer words (see Laufer 2005).

We compared words from L1 and L2 dialogues and monologues with the 10,000 most frequently used words in public texts in Estonian (Kaalep and Muischnek 2002), as shown in Figure 4.

The share of elementary vocabulary (the 3,000 most frequently used words) was remarkably high in all text types in both L1 and L2. Rarer words (not included in the frequency dictionary) accounted for less than a quarter of the vocabulary use and could be used by researchers to study candidates' language strategies.
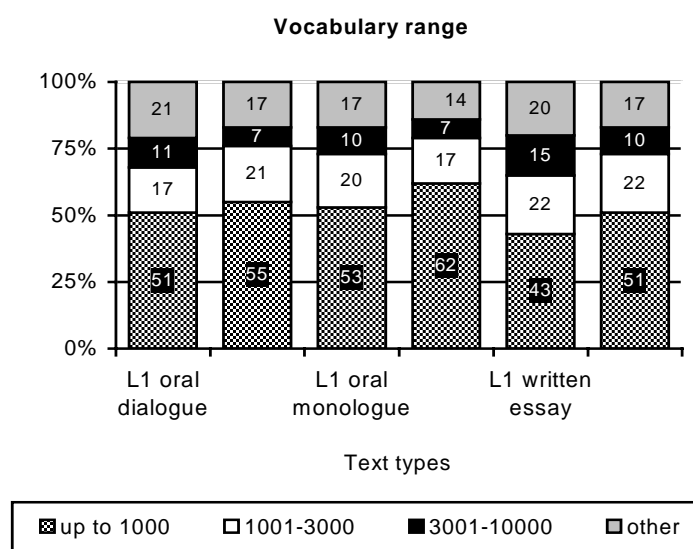
Vocabulary range

Figure 4. L1 and L2 vocabulary range in different text types by range of frequency in the frequency dictionary.

Proper nouns (proper names) account for about 8% of rare words in L1 oral speech and 3.5% in L1 written texts. L2 contains 3–4 times more proper nouns in all text types except oral monologue: 20% in oral dialogue and 13% in written text. The great number of proper nouns in L2 conversation points to a cultural and social awareness and to a wish to demonstrate it. At the same time, the rare vocabulary of all texts contains relatively few numbers and numerals: less than 1% in all oral texts, 2% in L1 written texts and 3.5% in L2 written texts.

Rare vocabulary is mostly thematic vocabulary, principally nouns. Here, nouns make up about 57% of L1 speech and about 63% of written texts. L2 oral monologue is similar to L1. However, nouns make up only around 42% of rare words of L2 dialogue and around 59% of the L2 essay. When we compare the rare thematic vocabulary of L1 and L2, the latter is considerably smaller in dialogue (a difference of 15.5%), the only spontaneous genre in our study. This shows the inability of L2 users to activate rare words as quickly as needed, although spontaneous text is also less demanding.

Strikingly, foreign stems are much more numerous in L2 monologues, with 17% and 28% in the oral presentation and essay respectively in L2 compared to 3.5% and 8% in L1. In dialogue this tendency is reversed, with foreign words accounting for 8.5% of L1 and 2% of L2.

L2 users are also more modest in word formation, something which comes very naturally to native speakers. While compounds and regular formations make up 50–60% of L1 speech and as much as 75% of written text rare words, in L2 they make up 35–40% of speech and less than 50% of written text rare vocabulary.

This material allows us to draw some preliminary conclusions. First, familiarity with foreign stems has been seen as an important support for language learners in the early stages of language acquisition. However, in more demanding contexts the use of foreign stems also increases the freedom and precision of expression at high levels of L2 proficiency. Second, word formation has an important role in natural Estonian syntactic operations. However, given the Indo-European language background of the L2 users, they may never practice word formation to the same extent as L1 users do (see also *Syntactic Complicacy* below).

Although both spoken and written texts mostly contain elementary vocabulary, the choice of words differs in monologue and dialogue (see Figures 5–10). It is worth noticing that despite their more limited vocabulary, L2 users choose words very similar to those chosen by L1 users: the biggest vocabulary overlap occurs in oral language use, both dialogue and monologue, and the overlap is smallest in the written essay and the dialogue.

The overlapping words remain within the 1,000 most frequently used words. We can conclude that natural language is characterised by lexical richness dependent on text type and higher-level language users can choose appropriate words from their vocabulary for each text type.
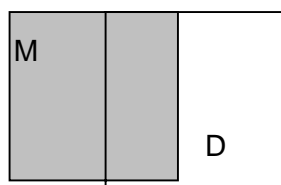
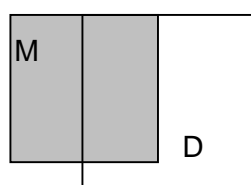**Figure 5.** Overlap in L1 oral monologue and dialogue

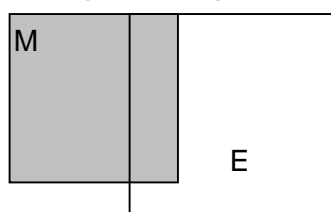**Figure 6.** Overlap in L2 oral monologue and dialogue

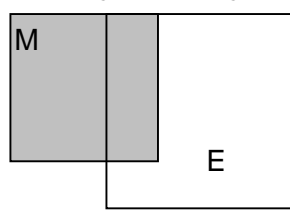**Figure 7.** Overlap in L1 oral monologue and essay

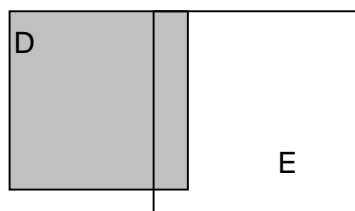**Figure 8.** Overlap in L2 oral monologue essay
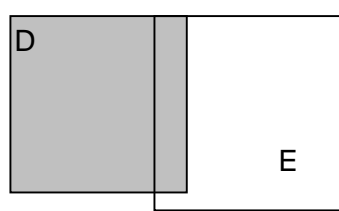
Figure 9. Overlap in L1 dialogue and essay

Figure 10. Overlap in L2 dialogue and essay

Given these results, it would be wrong to assess C1-level vocabulary range in the context of speaking skills. The vocabulary range and richness of C1-level language users should be assessed instead by testing their receptive skills.

3.2. Contextuality and formality of the text

A difference in the balance of parts of speech in oral and written production has been noticed for quite a long time and by many researchers (for an overview since the early 1930s see Chafe and Tannen 1987), and Heylighen and Dewaele based the theory of the contextuality–formality continuum on this balance. Their hypothesis states that nouns and other parts of speech (articles, prepositions, adjectives) bound to them make texts more exact and thus more formal, while more contextual words have the opposite effect. Based on this balance, a formula called the F-index was worked out to index the level of formality of a text (Heylighen and Dewaele 2002).

We measured the contextuality-formality of our texts, after adapting Heylighen and Dewaele's (2002) formula[5] to the Estonian language, which has no articles, very few prepositions and numerous postpositions:

$$F = 0.5*[(noun\ frequency + adjective\ freq. + adposition\ freq.) - (pronoun\ freq. + verb\ freq. + adverb\ freq. + interjection\ freq.) + 100]$$

The formula is based on the assumption that the frequency of such parts of speech as pronouns, verbs, adverbs and interjections makes the text more contextual and thus more ambiguous, while the frequency of others such as nouns, adjectives, adpositions and articles decreases contextuality, making the text less ambiguous and more formal (see Figure 11).

We asked whether F differs significantly in different text types and genres and, if so, whether this feature helps in distinguishing between L1 and L2 texts.

We determined parts of speech in L1 and L2 dialogues, monologues and essays using certain rules of our own, for example we excluded conjunctions and numerals from the formula, although we did count conjunctions to determine the

Contextual style

ambiguity ➔ misinterpretation

Formal style

ambiguity avoidance

more
pronouns
verbs
adverbs
interjections

more
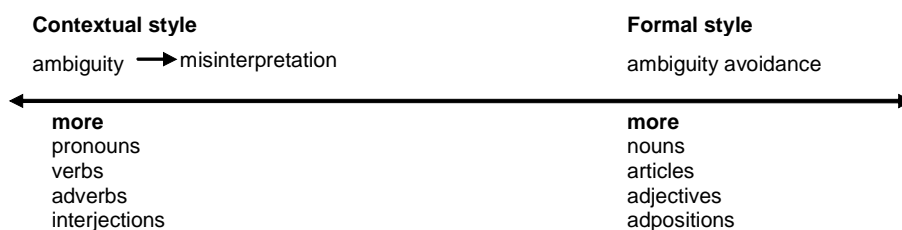nouns
articles
adjectives
adpositions

Figure 11. The Contextuality-Formality continuum

---

[5]   According to the authors of the formula their research shows that it works equally reliably for a text consisting of a few hundred words and for longer texts that they have studied or interpreted (Heylighen and Dewaele 2002: 321).

total number of words. We mostly divided numerals into adjectives (ordinals) and nouns (cardinals), by looking at contextual use, so words like *first* and *second* were marked either as pronouns or adjectives, depending on the contextual meaning. To measure contextuality precisely, we counted all cases of pro-form use (pronouns, pronumerals, proadverbs) and other contextually defined references such as *afore, certain* or *given*. As the candidates used only well-known proper nouns like *Russia*, these were also counted as nouns.

The results are shown in Table 3 and Figure 12.

We can see that contextuality-formality in L1 and L2 use is similar across all text types, as any difference of less than three is not statistically significant (Heylighen and Dewaele 2002: 328). With both L1 and L2 the formality of

Table 3. Part-of-speech percentages and formality scores for L1 and L2 text-types.

| Text-types	L1 or L2	Tokens (N)	Formal categories (%)			Contextual categories (%)				Conjunctions	F-index
			Nouns	Adjectives	Adpositions	Pronouns	Verbs	Adverbs	Interjections		
Oral DIALOGUE	L1	1693	19.0	3.4	2.2	23.7	21.8	16.7	1.1	11.1	30.2
	L2	2265	18.0	4.1	1.5	24.9	21.1	18.6	1.1	10.5	29.0
Oral MONOLOGUE	L1	1291	20.8	2.2	6.2	22.2	19.8	16.7	0.3	11.8	35.2
	L2	1324	22.1	4.5	1.2	22.9	21.9	13.1	1.2	13.1	34.4
Oral production (dialogue+monologue)	L1	2984	19.8	4.6	2.2	23.6	20.9	16.7	0.8	11.4	32.3
	L2	3589	19.5	4.3	1.4	24.2	21.4	16.6	1.2	11.5	30.9
Written essay (MONOLOGUE)	L1	1658	33.5	9.3	2.5	12.0	21.7	12.6	0.0	8.4	49.5
	L2	1798	33.9	8.6	3.2	14.1	21.7	10.4	0.2	7.9	49.7
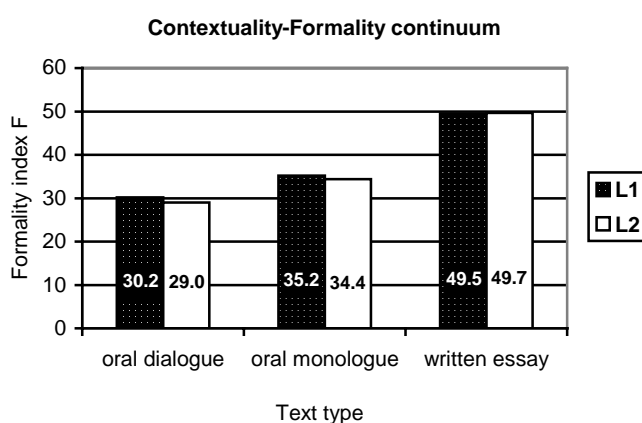
Figure 12. Degrees of formality in L1 and L2 texts. The greater the value of F, the more formal the text.

communication increases from oral speech to written text and from dialogue to monologue. The distribution of parts of speech by text type is also similar in L1 and L2.

Such similarity between L1 and L2 use in the degree of formality of different text types has convinced us that in testing C1-level proficiency, the most important skill to test is the ability of candidates to use language appropriate to a specific genre and register. This includes their ability to choose words, expressions and collocations from their elementary and relatively poor vocabulary that are appropriate to a specific genre and register, as seen in the results in part 3.1, which also referred to differences in the choice of words depending on the text-type. If the degree of formality is suitable for the text type, differences between other aspects of L1 and L2 use go unnoticed. The high scores given to our candidates for their speaking skills also confirm this.

3.3. Syntactic complicacy

Syntactic features such as typological features of word order, phrase ordering, etc vary between languages. According to Stamatatos et al. (2000), the syntax of functional styles also tends to differ in its parameters, as has been noted in many comparable corpus studies.

For Estonian, the only descriptions have been of the syntactic complicacy of written texts in various fields of language use, and three parameters have been used, each of which covers many features. These parameters are sentence length, which looks at the number of tokens; sentence complexity, which measures the total number of punctuation marks and conjunctions without punctuation; and the level of abstractness of a text, derived from the percentage of regular deverbal nouns complicating the text, while nominalised phrases are more abstract than verbal ones (see Kerge 2003).

We asked whether oral texts differ in their characteristics here.

As there is no clear research evidence of the same kind of sentence complexity in oral speech, syntactic complicacy can be measured only by sentence-length and level of abstractness (see Table 4).

The oral data given in Table 4 can be compared to earlier written data (Kerge, op. cit.) only in mean values because the latter does not operate with medians. Earlier measures from analysis of e-mail correspondence between colleagues and similar sources show sentences to be about 9.2 tokens long (the mean for modern Estonian text-types is 14) and the level of text abstractness to be about 0.9 (the mean is 3.4). This means that in any oral language use, sentences tend to be longer, and – being less nominalised – also much less abstract than in writing.

The median values in Table 4 show that the longest sentences – equally long in L1 and L2 – appear in oral monologue. Sentences in L2 dialogue are longer than those of native speakers; the only difference in oral monologue is the maximum value of sentence length in L2. In the essay genre, the values are reversed, and on average L1 sentences are longer. When L2 is used in writing, there is a clear tendency to use shorter sentences, probably in order to avoid mistakes. The same

Table 4. **Distribution data for sentence length (in tokens) and abstractness (degree of nominalisation as % of tokens of text) in different types of text.**

Text type	L1 and L2	Min	Q1	Median	Tokens per sentence (Mean)	Sentences[6]	Tokens	Nominalisation (%)	Q3	Max
DIALOGUE	L1	1.0	6.0	11.0	16.5 (sd 13.3)	107	1745	0.6	23.5	55.0
	L2	1.0	7.0	14.5	18.4 (sd 15.8)	128	2337	0.3	24.5	80.0
MONO-LOGUE	L1	3.0	11.0	21.0	23.9 (sd 16.9)	55	1315	0.5	33.5	80.0
	L2	2.0	11.0	21.0	27.5 (sd 25.2)	49	1343	0.4	36.0	129.0
ESSAY	L1	2.0	9.0	13.0	14.4 (sd 7.2)	118	1684	2.8	18.0	37.0
	L2	2.0	6.0	10.0	11.0 (sd 6.3)	166	1825	1.5	14.0	42.0

seems to apply to nominalisation, the percentage of which in L2 is always lower than in L1 where the highest value of nominalisation appears in writing (L1 2.8%; L2 only half of the L1 value). This shows again the tendency identified in the context of rare words, that word formation in L2 is not easy to learn and the skill also needs to be polished at higher levels of language learning. On the other hand, it may not be very relevant in testing oral, or even writing skills, as there are other syntactic possibilities for conveying the same content, such as subordinate clause or other types of non-finite verb phrases not studied here.

3.4. Temporal characteristics of dialogue

According to CEFR 2001 (p 86), an important skill in discourse is turn-taking. A high-level language user, from B2 level upwards, "*can initiate, maintain and end discourse appropriately with effective turn-taking. Can initiate discourse, take his/her turn when appropriate and end conversation when he/she needs to, though he/she may not always do this elegantly.*"

The length of turn, pauses in speech, simultaneous speaking, etc. are greatly culture-dependent (see ten Bosch *et al.* 2005). In spontaneous Estonian dialogue, turns and pauses between them are short, and simultaneous speaking is often used to give feedback and take turns (Pajupuu 1995).

In the dialogues under study we were interested in the following questions:

---

[6] In oral spontaneous speech, sentence length was determined by looking at the movement of the fundamental frequency, pauses and content of the text.

- How can educated L1 dialogue be described in terms of its temporal characteristics?
- Are the temporal characteristics of L1 and L2 similar?

In order to get answers, we measured the durations of the components of L1 and L2 turns:

ab   – turn duration
sp   – intra-turn pause (after which the turn continues)
lp   – turn-final pause (giving up the turn)
sab – simultaneous intra-turn speaking (failed attempt to take turn or give feedback)
lab – turn-taking by simultaneous speaking

The results are shown in Figure 13.

We can see that the main difference between spontaneous Estonian dialogue and L1 exam dialogue lies in the turn duration, as the turns are longer at the exam. Turns are not taken by speaking simultaneously at the exam, though simultaneous speaking is used to give feedback. Exam dialogues also contain more intra-turn pauses than spontaneous speech, while in spontaneous speech turns are taken quickly. This is not the case at the exam, where one dialogue party tends to wait for the other to continue. Because of these differences, exam dialogue and spontaneous dialogue can be considered two distinct sub-genres.

L2 dialogue is distinguishable by its long turns without pauses. There are also some cases of ultra-long monologue-like turns (see Table 5).

Assessors of C1-level proficiency should be familiar with the nature of exam dialogues, so that the standard should be L1 exam dialogue and not spontaneous speech. Even if assessors accept longer turns at exams, they should bear in mind that monologue-like turns may indicate that the speaker has not mastered culture-specific turn-taking rules.
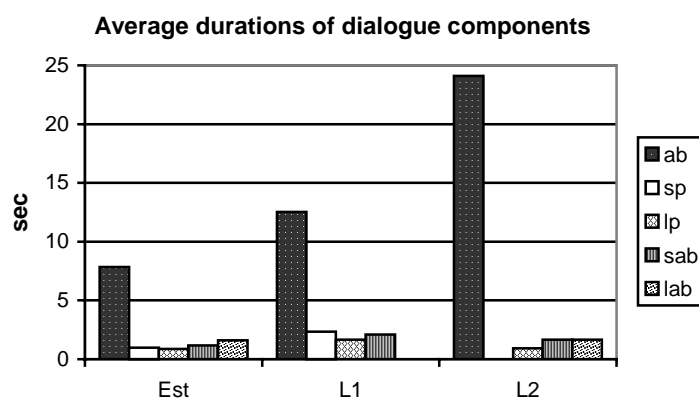

Figure 13. Durations of spontaneous Estonian (Est), L1 and L2 dialogue components (ab – turn duration, sp – intra-turn pause, lp – turn-final pause, sab – simultaneous intra-turn speaking, lab – turn-taking by simultaneous speaking).

Table 5. L1 and L2 turn distribution data, seconds

	Min.	Q1	Median	Mean	Q3	Max
L1	0.21	2.83	8.46	12.54	19.60	39.17
L2	0.49	4.20	10.46	24.09	38.51	104.90

3.5. Strength and disruptiveness of the foreign accent

CEFR says of foreign accent at B1-level (p 117) that *"Pronunciation is clearly intelligible even if a foreign accent is sometimes evident and occasional mispronunciations occur."* At higher levels, CEFR dictates that there should be no noticeable foreign accent.

Earlier, training accent-free pronunciation was an important part of language teaching. It was believed that "… the student should learn to speak the language as naturally as possible, free of any indication that the speaker is not a clinically normal native" (Griffen 1980/1991). Munro and Derwing (1999), who conducted empirical studies on accent, established that a strong foreign accent did not necessarily cause L2 speech to be low in comprehensibility or intelligibility, and therefore it would be impracticable to measure the strength of accent in the assessment of pronunciation. Research results also showed that listeners assessed the strength of accent subjectively.

We asked two questions:

- Do assessors register an accent in the pronunciation of high-level language users?
- Does the accent disturb the assessors and to what extent?

We conducted two experiments to find the answers. We asked all 8 assessors of the region to listen to a 20-second recording from each candidate. Research has shown that this might be the optimal length for recordings to assess accent (see Meister 2006, Meister and Meister 2007). In the first experiment, assessors had to score the strength of accent on a 6-point scale (1 = accent-free, 6 = very strong accent). In the second, conducted a month later, the assessors were asked how disruptive the accent was (1 = not disruptive at all, 6 = very disruptive).

We also counted the number of pronunciation mistakes and language mistakes in each recording. By correlating the strength and disruptiveness of the accent with the score given for speaking skills, the number of pronunciation mistakes and the number of language mistakes, we got the result we had expected. The strength of the accent and disruptiveness of the accent are directly related, and the number of pronunciation errors is related to both. Curiously, scores are affected by the number of pronunciation errors but not by grammar mistakes. This shows that accent may have a greater effect on the scores of speaking skills than other mistakes, as illustrated in Table 6.

When we compare the scores given to the strength and disruptiveness of the accents, we can see that not all professional assessors perceive the strength of accent in the same way, as for example, there is a significant difference in how

assessors A1 and A3 registered the strength of the accents (Figure 14). The disruptiveness of accent is even more subjective, as some assessors are not at all disturbed by accents (A6), while some others clearly are (A1 and A8) (see Figure 15).

Table 6. Correlations between the strength and disruptiveness of accent and other indicators

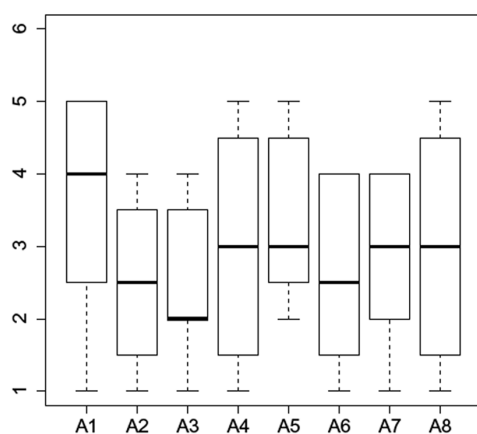|				
Strength of accent	0.937 (p<0.0006)	–0.829 (p<0.0109)	0.975 (p<0.00004)	0.680 (p<0.064)
	Disruptive-ness of accent	–0.885 (p<0.003)	0.903 (p<0.002)	0.683 (p<0.062)
		Score for speaking skills	–0.841 (p<0.009)	–0.396 (p<0.332)
			Number of pronunciation mistakes	0.655 (p<0.078)
				Number of language mistakes

Figure 14. Assessors' (A) ratings of accent strength on a 6-point scale[7]
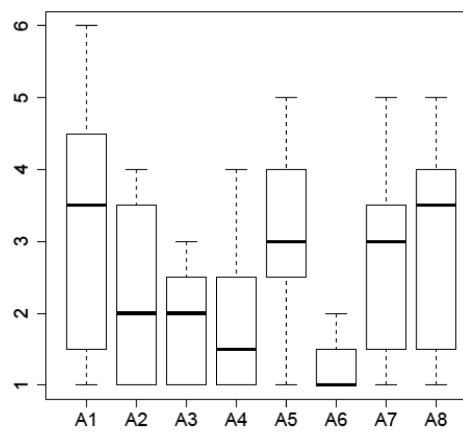

Figure 15. Assessors' (A) ratings of the disruptiveness of accent on a 6-point scale

---

[7] These boxplots display whether the data is symmetric or has suspected outliers. The boxplot has a box with lines at the lower quartile, the median and the upper quartile and whiskers which extend to the min and max (see Verzani 2004).

Although C-1 level language users may have quite strong accents, these should not be considered in the assessment of language use because of the high level of subjectivity, unless the accent impairs understanding. Accent only deserves attention in the assessment of C1 fluency if there are unnatural pauses.

3.6. Intonation

Another aspect that contributes to the perception of foreign accent is intonation (e.g. Magen 1998, Wennerstrom 2001). Being extremely language-specific, intonation is one of the first features acquired in the process of L1 acquisition, and among the last to be mastered in adult L2 acquisition. Within the present study we were above all interested in the issue of language transfer in patterns of rising intonation which occurred phrase- or utterance-finally before pauses. We asked two questions:

- Which L1 patterns of rising intonation are transferred to L2?
- How is the transfer of intonation patterns connected to the degree of perceived foreign accent?

Recent experimental accounts of Estonian intonation (e.g. Keevallik 2003, Asu 2006) show that phrase-final rises are very common in colloquial Estonian, and are mainly used for signalling continuation and incompleteness. Thus, rising intonation forms part of the Estonian intonational phonology and should be included in the model of natural spoken Estonian. A rise in Estonian can start at a low level on an accented syllable and end high at the phrase boundary (see Estonian contour (b) in Figure 16), or a rise starts on or immediately after the accented syllable and is followed by a high plateau (Estonian contour (a) in Figure 16).

Rising intonation is even more common in Russian. The transcription system of Russian intonation (ToRI) developed by Odé (2008) includes three intonation patterns involving a rise. In one of them, the intonation contour forms a high plateau after a rise to the accented syllable (Russian contour (a) in Figure 16). Another rising contour consists of a low accented syllable preceded by a fall and followed by a rise (Russian contour (b) in Figure 16). This pattern is described in the ToRI system as a fall-rise, which among other discourse functions conveys incompleteness. The third pattern is a rise-fall characterised by an intonation contour which ends at a mid-level rather than at the bottom of the speaker's range (Russian contour (c) in Figure 16). Thus, both languages contain rises but there are important differences in the realisation of Estonian and Russian contours.

For the purposes of this study, the eight L2 speakers were divided into three groups on the basis of the results of an error analysis carried out by Meister and Meister (2007), who showed that a higher degree of accent was associated with a higher score of pronunciation errors. The three proficiency levels, according to the degree of perceived Russian accent were as follows: (1) near-native (error rate 0), (2) moderate accent (error rate 0.10–0.32), and (3) strong accent (error rate 0.67–0.70). We would expect L1 intonation patterns to transfer to L2 particularly in case of those speakers who have a strong perceived foreign accent.

The analysis shows that the frequency of rising intonation and the intonation patterns of the L2 speakers in the near-native group are similar to those of the L1 speakers of Estonian. No rising intonation patterns seem to be transferred from L1 to L2, which suggests that these speakers have mastered this aspect of L2 intonation. Interestingly also, the Estonian rise ending in a high boundary (Estonian pattern (b)) occurs only in the speech of these speakers and not in the case of the two other groups.

In L2 Estonian with moderate or strong accent, there appeared a straightforward transfer of at least one rising pitch accent from the intonation system of L1 into L2; the L1 intonation pattern with a high plateau (Russian pattern (a)) was often used instead of the Estonian pattern ending in a plateau (Estonian pattern (a)) in order to signal continuation from the part of the speaker. Both rises are similar in that they end in a high plateau, but differ in that in the Estonian pattern, the rise starts after the low accented syllable, whereas in the Russian pattern, the rise occurs already before the high accented syllable (see Figure 16).

The transfer of this intonation pattern into L2 contributes to the impression of foreign accent and makes a speaker sound more non-native. Such transfers were more noticeable in the speech of these L2 speakers who were youngest (18–23 years old) and whose acquisition period of L2 had been shortest. The contour only present in the data of the L2 speakers from the strong accent group was the rising-falling pattern ending at a mid-level (Russian pattern (c)). Such a pattern is typical of Russian but does not occur in Estonian, where a fall always reaches the bottom of the speaker's range.

The results of the study stress the fact that intonation is one of the most challenging aspects of pronunciation to master in a foreign language, and imply that transfer of intonation patterns contributes to the perception of foreign accent.
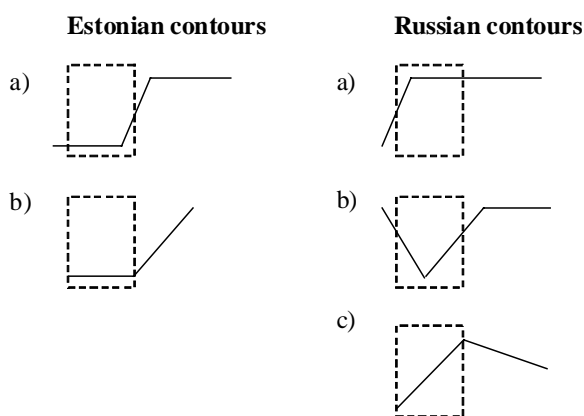

Figure 16. A schematic comparison of Estonian and Russian rising intonation patterns. Accented syllables are marked by the dashed boxes.

4. Conclusion

Our research showed that in natural or L1 use of Estonian, oral monologue is slightly richer in vocabulary than dialogue, and written text is considerably richer in vocabulary than oral production. We established that although the vocabulary range of L1 users mostly remains within the 3,000 most frequently used words in Estonian, the share of rare words (almost 20%) is remarkably equal in the different types of oral production and similarly high in written texts. Most L1 rare words are nouns used as thematic vocabulary, and the share of them is two to three times higher within this vocabulary range of texts than in the average.

When looking at the vocabulary range and degree of nominalisation, we noticed the great importance of word-formation in natural speech: in general, rare words contain many compounds and derivations and deverbal nominalisation plays an especially important role in the written text.

In the natural language use, the oral monologue is more formal than the dialogue and the written text is more formal than the oral production.

C1-level L2 users of Estonian differ from L1 users in most aspects that were researched: L2 vocabulary is poorer and the role of word-formation smaller; turns are longer; and the written text is far less nominalised, all of which is accompanied by accents and foreign intonation. However, there is something L1 and L2 users have in common, and this is the perception of text-type. L1 and L2 texts have a similar degree of formality, vocabulary range and variability, which means that L2 users can choose words that are appropriate to the expected degree of formality of the text-type and their speech contains a natural weighting of words from different ranges of frequency. It is clear that in neither the oral monologue and dialogue, nor in oral and written production does their choice of words overlap more than that of native speakers.

At the same time, L2 exam results show that the candidates coped with the language tasks they had been given. Regardless of the differences we established, the assessors considered the language proficiency of all our subjects sufficient for a certificate, which means that their production must have seemed quite natural. We can assume that the mistakes found in the pronunciation of L2 users did not stop them from completing their tasks, as their choice of words and use of parts-of-speech were natural.

It is important that the guidelines for assessing speaking skills in C1-level language exams take this into consideration (for now we are leaving aside the conclusions on written production which we only researched for the purpose of comparison).

Earlier analyses of assessors indicate that they focus on different aspects of proficiency: some of them expect grammatical correctness or fluency, while others measure vocabulary range, and so forth. This explains why candidates may receive different scores from different assessors. Incidentally, it has been proven that assessor training has a relatively short-term effect (Bonk and Ockey 2003, Eckes 2008, Lumley and McNamara 1995, Orr 2002). To reduce the effect of assessors,

assessment guidelines should clearly indicate the criteria that C1-level language users must meet in the light of our research results, i.e. their production must be appropriate to the text-type. Mistakes that do not inhibit the completion of the task should go unnoticed.

We should also highlight our results on the assessment of speech parameters, where accent deserves special attention. First, our research results challenge the presumption of CEFR 2001 (see 5.2.1.4) that from B1-level upwards, L2 users can already speak accent-free. Second, we established that some assessors tend to assess the whole language proficiency of a candidate based on his or her accent. It appeared that assessors were especially subjective in assessing the accent, so to ensure objectivity, the C1-level assessment guidelines should emphasise that an accent should not be decisive in the assessment of the overall proficiency of a candidate, unless it stops him or her from completing the task. This also applies to intonation which is directly related to the accent. Assessors need to bear in mind that rising intonation is also natural in Estonian, and the use of it should not be considered a mistake in L2, even if the exact patterns of rising intonation of L2 users might be different.

As for other results, we would like to point out that the great share of foreign stems in speech is not only an understanding strategy used by beginner-level L2 speakers, but also describes a coping strategy that C-level users employ. Task-specific L2 rare vocabulary contains considerably more foreign stems than L1 (see CEFR 2001 part 4.4.2.2 for comparison). The great importance of word-formation in L1 use is also significant, and we can conclude that the importance of word-formation should be stressed in C1-level language learning objectives. However, because of their different linguistic background, L2 users may never practice word formation to the same extent that L1 users do, and therefore this skill should not be assessed separately.

Address:
　　Hille Pajupuu
　　Eesti Keele Instituut
　　Roosikrantsi 6
　　10119 Tallinn
Phone: +372 6177 500
E-mail: hille@eki.ee

References

Asu, Eva Liina (2006) "Rising intonation in Estonian: an analysis of map task dialogues and spontaneous conversations". Fonetiikan Päivät 2006 / The Phonetics Symposium 2006. Helsinki University, 1–8.

Boersma, Paul and David Weenink (2006) Praat: doing phonetics by computer (Version 4.5.08). Computer program. http://www.praat.org (3.01.2007).

Bonk, William J. and Gary J. Ockey (2003) "A many-facet Rasch analysis of the second language group oral discussion task". *Language Testing* 20, 1, 89–110.

CEFR = Common European Framework of Reference for Languages: Learning, Teaching, Assessment. (2001). Strasbourg: Council of Europe; Cambridge University Press. [Internet document available at http://www.coe.int/t/dg4/linguistic/Source/Framework_EN.pdf] (12.12.2009)

Chafe, Wallace and Deborah Tannen (1987) "The relation between written and spoken language". *Annual Review of Anthropology* 163, 83–407.

Dewaele, Jean-Marc and Aneta Pavlenko (2003) "Productivity and lexical diversity in native and non-native speech: a study of cross-cultural effects". In *Effects of the second language on the first*, 120–141. (Second Language Aacquisition.) Vivian Cook, ed. Tonawanda, NY: Multilingual Matters.

Eckes, Thomas (2008) "Rater types in writing performance assessments: a classification approach to rater variability" . *Language Testing* 25, 2, 155–185.

Griffen, Toby D. (1991) "A nonsegmental approach to the teaching of pronunciation". In *Teaching English pronunciation: a book of readings*, 178–190. Adam Brown, ed., London: Routledge. (Reprinted from *Revue de Phonétique Appliquée* 54, 81–94, 1980)

Heylighen, Francis and Jean-Marc Dewaele (2002) "Variation in the contextuality of language: an empirical measure". *Foundations of Science* 7, 293–340.

Jarvis, S. (2002) "Short texts, best-fitting curves and new measures of lexical diversity". *Language Testing* 19, 1, 57–84.

Kaalep, Heiki-Jaan and Kadri Muischnek (2002) *Eesti kirjakeele sagedussõnastik.* [Frequency dictionary of standard Estonian.] Tartu: Tartu Ülikooli Kirjastus.

Keevallik, Leelo (2003) "Terminally rising pitch contours of response tokens in Estonian". *Crossroads of Language, Interaction, and Culture* 5, 49–65.

Kerge, Krista (2003) "Keele variatiivsus ja mine-tuletus allkeelte süntaktilise keerukuse tegurina". [Language variation and *mine*-derivation as a factor of sublanguage syntactic complexity.] (Tallinna Pedagoogikaülikooli humanitaarteaduste dissertatsioonid, 10.) Tallinn: Tallinna Pedagoogikaülikooli Kirjastus.

Kerge, Krista (2008) *Vilunud keelekasutaja. C1-taseme eesti keele oskus.* [Proficient user. C1-level proficiency of Estonian.] Tallinn: Eesti Keele Sihtasutus.

Laufer, B. (2005) "Lexical frequency profiles: from Monte Carlo to the Real World: a response to Meare". *Applied Linguistics* 26, 4, 582–588.

Little, Dave (2005) "The common European framework and the European language portfolio: involving learners and their judgements in the assessment process". *Language Testing* 22, 3, 321–336.

Lumley, Tom and Tim F. McNamara (1995) "Rater characteristics and Rater bias: implications for training". *Language Testing* 12, 1, 54–71.

Magen, Harriet S. (1998) "The perception of foreign accented speech". *Journal of Phonetics* 26, 381–400.

Meister, Lya (2006) "Assessment of the degree of foreign accent: a pilot study". In *Fonetiikan päivät 2006 = The phonetics symposium 2006. Reijo Aulanko, Leena Wahlbergand Martti Vainio, eds., University of Helsinki 30.–31.8.2006*, 113–119. (Publications of the Department of Speech Sciences, University of Helsinki, 53.) Helsinki.

Meister, Lya and Einar Meister (2007) "Perceptual assessment of Russian-accented Estonian". In *ICPhS XVI: proceedings of the 16th international congress of phonetic sciences, 6–10 August 2007, Saarbrücken Germany,* 1717–1720. Saarbrücken: Universität des Saarlandes.

Munro, Murray J. and M. Derwing Tracey (1999) "Foreign accent, comprehensibility, and intelligibility in the speech of second language learners". *Language Learning* 49, Supplement 1, 285–310.

Odé, Cecilia (2008) "Communicative functions and prosodic labelling of three Russian rising pitch accents". In *Evidence and counter-evidence: essays in honour of Frederik Kortlandt*, Vol.1: Balto-Slavic and Indo-European linguistics, 377–401. (Studies in Slavic and General Linguistics, 32.) Alexander Lubotsky, Jos Schaeken, and Jeroen Wiedenhof, eds. Amsterdam: Rodopi.

Orr, Michael (2002) "The FCE speaking test: using Rater reports to help interpret test scores". *System* 30, 2, 143–154.

Pajupuu, Hille (1995) *Cultural context, dialogue, time*. Tallinn: EAS. Institute of the Estonian Language. http://www.eki.ee/teemad/kultuur/context/context.html (01.11.2009).

Ratcliff, Ann, Sue Coughlin, and Mark Lehman (2002) "Factors influencing ratings of speech naturalness in augmentative and alternative communication". *AAC: Augmentative & Alternative Communication* 18, 11–19.

Read, John and Carol A. Chapelle (2001) "A framework for second language vocabulary assessment". *Language Testing* 18, 1, 1–32.

Scott, Michael (1996) *WordSmith Tools 3.0*. Oxford: OUP.

Stamatatos, Efstathios, Nikos Fakotakis, and George Kokkinakis (2000) "Automatic text categorization in terms of genre and author". *Computational Linguistics* 26, 4, 471–495.

ten Bosch, Louis, Nelleke Oostdijk, and Lou Boves (2005) "On temporal aspects of turn taking in conversational dialogues". *Speech Communication* 47, 80–86.

Vermeer, Anne (2000) "Coming to grips with lexical richness in spontaneous speech data". *Language Testing* 17, 1, 65–83.

Verzani, John (2004). *Using R for introductory statistics.* Boca Raton, FL: Chapman Hall and CRC Press.

Wennerstrom, Ann (2001) *The music of everyday speech prosody and discourse analysis*. New York: Oxford University Press.