ACQUISITION OF EPISTEMIC MARKING IN ESTONIAN AND RUSSIAN

Victoria V. Kazakovskaya, Reili Argus

Abstract. The article compares the acquisition of epistemic modality in typologically different languages, Estonian and Russian. The longitudinal data of 4 children was used to analyse the first emergence and the further development of lexical markers expressing certainty and uncertainty. Developmental analysis of micro fields of epistemic modality has shown that in both languages the semantics of uncertainty starts to be acquired slightly earlier than certainty. The main tendencies in the usage of epistemic markers in utterances are similar in our corpora. Since the acquisition of epistemic modality is similar in the two languages, it would be reasonable to assume that some categories, like epistemic modality, rely more on general cognitive development than others.*

Keywords: first language acquisition, epistemic modality, certainty, uncertainty, Estonian, Russian

1. Introduction

No contrastive research has previously been carried out on the acquisition of epistemic modality (EM) in typologically different languages. Our study is the first attempt to make a comparative investigation of Finno-Ugric (Estonian) and East Slavic (Russian) languages, in which the repertoire of epistemic markers (or epistemic modals) is quite similar (see Õispuu 1999: 153–154), but the category of evidentiality (which is connected with EM) has a varying degree of grammaticalisation.

Evidentiality is subsumed under a wider category of EM (Palmer 1986, Boye 2006, Egan, Weatherson 2011, Auwera, Plungian 1998). The speaker has greater doubts about the information if she has received it indirectly (e.g. in the case of hearsay, being one kind of evidentiality). Based on this, we initially believed that evidentiality in Estonian (i.e. the quotative) can cause infrequent expression of EM, and uncertainty could be less represented by different lexical means of EM in Estonian child speech (CS, output). At the same time, due to the lack of a grammaticalised

* The research was carried out with the financial support of the Russian National Foundation (grant 14-18-03668). We are grateful to the two anonymous reviewers for their comments.
category of evidentiality (the possibility of expressing indirect information by special grammatical tools) in Russian, EM can be well represented in this language. However, recent studies have shown that such a dependence could not exist, and it has been found by experiment that Estonian evidentiality is acquired late, after the age of six (Tamm et al. 2015). Additionally, the lexical markers of EM emerge in Estonian CS quite early, at approximately the same time as they occur in Russian CS (e.g., Kazakovskaya 2011).

These results allowed us to assume that the availability of evidentials, and the degree of their grammaticalisation, are less important for EM acquisition than the cognitive (semantic) complexity of the modality domain and the influence of child-directed speech (CDS, input), along with the frequency of prototypical means (expressions) in the modern adult-directed speech (ADS) in a particular language.

Research on the acquisition of EM is important not only for analysis of the early communicative competence of children (e.g., Fitneva 2001, Fitneva, Matsui 2009, Matsui 2014, Öztürk, Papafragou 2015), but also for investigating their cognitive development. In particular, children’s ability to comprehend and produce concepts of possibility and necessity is one of the main questions of developmental psychology (e.g., Piaget, Inhelder 1975, Shtulman, Carey 2007). Moreover, self-awareness and marking of their point of view allows children to overcome egocentric thinking (Vygotskij 1986, Piaget 1994). Epistemic modals representing the speaker’s point of view with respect to an assessment of the certainty / uncertainty of the proposition is the initial step in this direction (e.g., Kazakovskaya 1996, 2011: cf. of course > maybe > probably > personally etc. (...) – I think / believe / consider etc. that (...). Furthermore, the relationship between theory of mind (characterised by numerous concepts of reality and the existence of varying viewpoints apart from one’s own (e.g., Flavell 2004, Wellman 1990) and language acquisition has been discussed intensively (e.g., Miller 2006, Milligan et al. 2007). Specifically, children’s abilities to use different epistemic markers are connected with the development of so-called modal language (including mental state vocabulary), belonging to an explicit theory of mind (Brooks, Meltzoff 2015, Moore et al. 1990, Moore, Furrow 1991, Öztürk, Papafragou 2015, Papafragou 1998, Ruffman et al. 2006, Taumoepeau, Ruffman 2016). However, only the children’s understanding of the expression of the speaker’s certainty and uncertainty has been examined (usually experimentally). Additionally, previous research has shown that children between the ages of 4 and 6 should be able to understand mediated information as far as their cognitive development is concerned (de Villiers et al. 2009). However, Estonian children start inferring the relationship between another source and the lower reliability of information that is conveyed by evidentials only at the age of 6 (Argus et al. 2014, Tamm et al. 2015), but the data of younger children, as studied here, has not previously been available. Finally, our results could provide further evidence in the debate about the relationships between categories of EM and evidentiality in different languages.

On the basis of the naturalistic data mentioned above (see also p. 3), the following research goals can be formulated after the contrastive presentation of the category of EM in Estonian and Russian:

• to compare the acquisition of initial epistemic modals, including their positional aspect and the degree of expression of semantics;
to analyse the further development of certainty and uncertainty in children’s utterances, focusing on their semantic structure and communicative features;

to reveal the similarities and differences in early epistemic development;

to investigate the influence of input frequency of different markers on their acquisition;

based on the acquisition of ‘modal language’, to supply some possible data on when children overcome ‘egocentric thinking’.

2. Epistemic modality in Estonian and Russian

The semantics of EM is abstract: it indicates the degree of the speaker’s certainty of reality, reliability, authenticity and accuracy of what she says. Epistemic modal meanings can be represented by polar semantic spheres, or ‘micro fields’ of certainty (confidence) and uncertainty (probability) (Bondarko 1990: 167). At the same time, epistemic markers can be placed on a ‘virtual’ modal scale in order of decreasing proportion of uncertainty and, consequently, increasing degree of certainty in the speaker’s information, so this category has a scaled structure.

Each epistemic sphere has some organisation. The micro field of uncertainty can be presented by three smoothly transitioning degrees. A high degree of uncertainty (HDU) is, for example, expressed by vaevalt ‘hardly’ in Estonian, and by naverno ‘probably’, vidimo ‘evidently’, po-vidimomu ‘apparently’, očevidno ‘obviously, clearly’ in Russian. The middle degree of uncertainty (MDU) is expressed by äkki and võib-olla ‘maybe’, vist and tõenäoliselt ‘probably’ etc. in Estonian, and možet byt’ ‘maybe’, vozmožno ‘perhaps’ etc. in Russian. A low level (LDU) is expressed by paistab, et ‘it seems’ etc. in Estonian, and požaluj ‘very likely’, kažetsja ‘it seems’ in Russian.

The semantic field of certainty is organised in another way, since simple certainty is characterised by the absence of special markers in the statement. It is knowledge: I know that P, which would usually be presented implicitly, i.e. expressed without modal lexical or grammatical means (see also “implicit modality of certainty” (Bondarko 1990), or even “the bare, unmodalized proposition” (Öztürk, Papafourgou 2015)). For the appearance of these markers, special conditions are needed: at least, the speaker’s intention to underline her certainty. Categorical certainty (CC) is expressed by Est kindlasti, muidugi / Rus konečno ‘of course, certainly’, Est tõepoolest / Rus dejstvitel’no ‘really, indeed’, Est tegelikult / Rus na samom dele ‘in fact’, Est tõesti / Rus pravda ‘truly’ etc.² It borders on knowledge, but can’t be considered equal to it. The next level is the problematic certainty (PC) that arises in conditions of incomplete knowledge, which may be due to an imperfection of memory, a lack of perception or incomplete information, e.g. Est tundub, et / Rus kažetsja ‘it seems’ and some modal particles.

In Estonian, EM is mainly expressed lexically – with a predicate (Ma arvan, et ta on haige ‘I think that she is sick’), a predicative (On kindel, et ta on haige ‘(It) is sure that she is sick’) or with a sentence adverbial (Kindlasti on ta haige ‘Surely she is sick’). Among lexico-grammatical means, also modal verbs can be used for expressing epistemic meanings (e.g., võima ‘can’). Finally, some evidentials (the

² Division of EM markers relies to a large extent on the monograph of Mati Erelt (2013: 115).
quotative and a construction with the verb pidama ‘have’) carry this semantics (Erelt 2013: 115, Tamm 2009: 395, Argus et al. 2014).

In Russian, epistemic semantics is also mainly expressed by lexical and grammatical means. Firstly, it is a group of ‘modal words’ (or ‘modal adjuncts’), supplemented by words from nouns (pravda ‘truth’), adjectives, adverbs (vozmožno ‘perhaps’), verbs (kažetsja ‘it seems’), and pronouns (samo soboj ‘it goes without’). Acting as modal words, they are rethinking semantically and grammatically (Vinogradov 1986, Russkaja Grammatika 2005). Russian modal words are indeclinable words and combinations thereof, which can occupy any position within an utterance, not being its members (i.e. ‘parenthetical words’), as well as being reactions within a dialogue unit:

Ty hotela by, ţoby u tebya byl molodoj muţ?
‘Would you like to have a young husband?’

– Konečno.
‘Of course.’ (RNC)

Moreover, in both languages epistemic semantics is expressed by intonational and syntactic means, namely by so-called modus frames (Bally 1965, Arutjunova 1988), as e.g. I am sure / confident etc. that (...) – I am not convinced / persuaded etc. that (...), which are not considered in the present study.3

It is important to emphasise that in Estonian and Russian, the semantics of certainty can be expressed both explicitly and implicitly (i.e. ‘implicit / simple certainty’), whereas uncertainty is expressed only explicitly. Table 1 shows the 17 main epistemic lexemes used in these languages in general which are, according to our analysis, similar.

<table>
<thead>
<tr>
<th>Epistemic modality</th>
<th>Estonian</th>
<th>Russian</th>
</tr>
</thead>
</table>
| Certainty          | selge ‘clearly’
                   | muidugi ‘of course’
                   | kindlasti ‘certainly, surely’
                   | loomulikul ‘naturally’
                   | tegelikult ‘actually’
                   | tõöpoolet ‘indeed’
                   | tõesti ‘really, truly’
                   | kahtlemata ‘for sure, no doubt’
                   | vaike, mõjutatult ‘indisputably’ etc. |
|                    | vaevalt ‘hardly’
                   | äkki ‘maybe’
                   | võib-olla ‘maybe’
                   | vist ‘probably’
                   | võimalik ‘(it is) possible’
                   | ilmselt ‘apparently’
                   | paistab, et ‘seems that’
                   | tõenäoliselt ‘probably’ etc. |
| Uncertainty        | konečno ‘of course’
                   | nesomnenno ‘undoubtedly’
                   | dejstvitel’ ‘no indeed’
                   | razumeetsja ‘needless to say’
                   | v samom dele ‘actually’
                   | bessporno ‘indubitably’
                   | bezuslovno ‘certainly’
                   | bez somnenia ‘no doubt’
                   | verno ‘faithfully’ etc. |
|                    | vidimo ‘probably’
                   | možet byt ‘maybe’
                   | vozmožno ‘perhaps’
                   | po-vidimomu ‘apparently’
                   | očevidno ‘obviously, clearly’
                   | navernoe ‘certainly’
                   | kažetsja ‘seems’
                   | verojatno ‘(most) likely’ etc. |

Table 1. EM markers in Estonian and Russian

3 Predicates expressing epistemic meanings are extremely rare in the Estonian child language corpora: no usage of predicates with the verbs arvama ‘think’, tunduma ‘feel’, näima ‘seem’, paistma ‘seem’ etc. were found.
3. Data and method

The recordings (151.8 hours, see Table 2) of spontaneous dialogues of four typically developing monolingual children from Saint Petersburg (the Baltic region of northwest Russia) and Tallinn and Tartu (Estonia), were studied. The conversations of two boys (Andreas and Filipp) and two girls (Martina and Liza) and their main caregivers were transcribed (CHILDES) and then analysed by hand according to the following parameters: a) the emergence of epistemic modals and their frequency, b) further development of certainty and uncertainty, c) the impact of quality (types/lemmas) and quantity (tokens) of ‘modality input’ on output. In interpretation of utterances consisting of EM markers we relied on the dialogue context.

Table 2. Data analysed

<table>
<thead>
<tr>
<th>Age</th>
<th>Length of recordings (min.)</th>
<th>MLU in CS</th>
<th>MLU in CDS</th>
<th>Length of recordings (min.)</th>
<th>MLU in CS</th>
<th>MLU in CDS</th>
<th>Length of recordings (min.)</th>
<th>MLU in CS</th>
<th>MLU in CDS</th>
<th>Length of recordings (min.)</th>
<th>MLU in CS</th>
<th>MLU in CDS</th>
<th>Length of recordings (min.)</th>
<th>MLU in CS</th>
<th>MLU in CDS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1:3</td>
<td>-</td>
<td>60</td>
<td>2.3</td>
<td>6.3</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1:5</td>
<td>-</td>
<td>60</td>
<td>2.5</td>
<td>7.0</td>
<td>-</td>
<td>-</td>
<td>120</td>
<td>2.0</td>
<td>4.2</td>
<td>120</td>
<td>2.0</td>
<td>4.2</td>
<td>120</td>
<td>2.0</td>
<td>4.2</td>
</tr>
<tr>
<td>1:6</td>
<td>-</td>
<td>60</td>
<td>2.7</td>
<td>7.5</td>
<td>21</td>
<td>1.4</td>
<td>47</td>
<td>120</td>
<td>1.8</td>
<td>45</td>
<td>120</td>
<td>1.8</td>
<td>45</td>
<td>120</td>
<td>1.8</td>
</tr>
<tr>
<td>1:7</td>
<td>45</td>
<td>1.1</td>
<td>3.2</td>
<td>-</td>
<td>93</td>
<td>1.2</td>
<td>5.7</td>
<td>120</td>
<td>1.9</td>
<td>4.9</td>
<td>120</td>
<td>1.9</td>
<td>4.9</td>
<td>120</td>
<td>1.9</td>
</tr>
<tr>
<td>1:8</td>
<td>90</td>
<td>1.2</td>
<td>3.4</td>
<td>-</td>
<td>34</td>
<td>1.2</td>
<td>5.2</td>
<td>120</td>
<td>2.6</td>
<td>4.8</td>
<td>120</td>
<td>2.6</td>
<td>4.8</td>
<td>120</td>
<td>2.6</td>
</tr>
<tr>
<td>1:9</td>
<td>45</td>
<td>1.5</td>
<td>3.9</td>
<td>60</td>
<td>4.4</td>
<td>8.1</td>
<td>193</td>
<td>1.1</td>
<td>5.4</td>
<td>120</td>
<td>2.5</td>
<td>4.9</td>
<td>120</td>
<td>2.5</td>
<td>4.9</td>
</tr>
<tr>
<td>1:10</td>
<td>90</td>
<td>1.6</td>
<td>3.8</td>
<td>60</td>
<td>4.7</td>
<td>7.5</td>
<td>79</td>
<td>2.1</td>
<td>6.1</td>
<td>120</td>
<td>2.5</td>
<td>4.9</td>
<td>120</td>
<td>2.5</td>
<td>4.9</td>
</tr>
<tr>
<td>1:11</td>
<td>60</td>
<td>1.8</td>
<td>3.9</td>
<td>60</td>
<td>4.0</td>
<td>8.5</td>
<td>58</td>
<td>1.5</td>
<td>4.6</td>
<td>120</td>
<td>2.7</td>
<td>5.9</td>
<td>120</td>
<td>2.7</td>
<td>5.9</td>
</tr>
<tr>
<td>2:0</td>
<td>1200</td>
<td>3.6</td>
<td>4.4</td>
<td>-</td>
<td>83</td>
<td>2.0</td>
<td>4.8</td>
<td>120</td>
<td>3.2</td>
<td>5.8</td>
<td>120</td>
<td>3.2</td>
<td>5.8</td>
<td>120</td>
<td>3.2</td>
</tr>
<tr>
<td>2:1</td>
<td>600</td>
<td>2.7</td>
<td>4.3</td>
<td>60</td>
<td>4.9</td>
<td>9.1</td>
<td>75</td>
<td>3.2</td>
<td>4.6</td>
<td>120</td>
<td>3.3</td>
<td>6.3</td>
<td>120</td>
<td>3.3</td>
<td>6.3</td>
</tr>
<tr>
<td>2:2</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>85</td>
<td>3.2</td>
<td>5.0</td>
<td>120</td>
<td>3.7</td>
<td>6.1</td>
<td>120</td>
<td>3.7</td>
</tr>
<tr>
<td>2:3</td>
<td>45</td>
<td>2.4</td>
<td>4.3</td>
<td>60</td>
<td>8.0</td>
<td>8.9</td>
<td>54</td>
<td>3.3</td>
<td>4.8</td>
<td>120</td>
<td>3.6</td>
<td>6.0</td>
<td>120</td>
<td>3.6</td>
<td>6.0</td>
</tr>
<tr>
<td>2:4</td>
<td>45</td>
<td>4.1</td>
<td>4.2</td>
<td>60</td>
<td>4.7</td>
<td>7.8</td>
<td>72</td>
<td>3.3</td>
<td>5.3</td>
<td>120</td>
<td>3.6</td>
<td>5.6</td>
<td>120</td>
<td>3.6</td>
<td>5.6</td>
</tr>
<tr>
<td>2:5</td>
<td>45</td>
<td>4.1</td>
<td>4.3</td>
<td>-</td>
<td>57</td>
<td>3.2</td>
<td>5.1</td>
<td>120</td>
<td>3.9</td>
<td>5.6</td>
<td>120</td>
<td>3.9</td>
<td>5.6</td>
<td>120</td>
<td>3.9</td>
</tr>
<tr>
<td>2:6</td>
<td>60</td>
<td>3.8</td>
<td>4.6</td>
<td>-</td>
<td>64</td>
<td>3.2</td>
<td>4.8</td>
<td>120</td>
<td>4.2</td>
<td>6.0</td>
<td>120</td>
<td>4.2</td>
<td>6.0</td>
<td>120</td>
<td>4.2</td>
</tr>
<tr>
<td>2:7</td>
<td>60</td>
<td>3.9</td>
<td>4.9</td>
<td>60</td>
<td>8.6</td>
<td>8.9</td>
<td>14</td>
<td>3.0</td>
<td>4.5</td>
<td>120</td>
<td>4.2</td>
<td>6.7</td>
<td>120</td>
<td>4.2</td>
<td>6.7</td>
</tr>
<tr>
<td>2:8</td>
<td>60</td>
<td>4.6</td>
<td>4.6</td>
<td>-</td>
<td>31</td>
<td>4.0</td>
<td>5.2</td>
<td>120</td>
<td>4.7</td>
<td>6.9</td>
<td>120</td>
<td>4.7</td>
<td>6.9</td>
<td>120</td>
<td>4.7</td>
</tr>
<tr>
<td>2:9</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>129</td>
<td>4.2</td>
<td>4.8</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2:10</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>74</td>
<td>4.4</td>
<td>5.1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2:11</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>73</td>
<td>6.2</td>
<td>7.5</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3:0</td>
<td>1200</td>
<td>3.9</td>
<td>4.3</td>
<td>-</td>
<td>85</td>
<td>5.1</td>
<td>8.4</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3:1</td>
<td>600</td>
<td>5.0</td>
<td>5.3</td>
<td>60</td>
<td>5.7</td>
<td>7.8</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3:5</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>17</td>
<td>6.9</td>
<td>7.8</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3:6</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>57</td>
<td>7.0</td>
<td>7.2</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3:7</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>33</td>
<td>7.2</td>
<td>8.2</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3:10</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>23</td>
<td>7.7</td>
<td>7.6</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>4:1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>30</td>
<td>7.2</td>
<td>7.2</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Total</td>
<td>4245</td>
<td>660</td>
<td>2280</td>
<td>1920</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

4 The Estonian corpora were collected by M. Vija and A. Kapanen; the Russian corpora were collected by M. B. Eliseeva and T. V. Pranova. All data has been used in several previous studies as well.
5 There were no instances where the child used a marker incorrectly.
6 Mean length of utterances in words (using the automatic calculation available in CHILDES).
Although the speech samples are of different sizes as a result of being from different children of different ages, we still decided to use all the data available. Hence the usage of epistemic markers was fairly limited (in several cases, only 1 or 2 occurrences), but by excluding some dialogues we could have lost a lot of important data.

4. Results and their discussion

4.1. The first emergence of epistemic markers in child language

The Estonian girl Martina first used the marker of certainty tõesti ‘really’ at age 1;6 (MLU 2.7), but it was used only once in the whole corpus (not even imitations were found). Five months later, at age 1;11 (when her MLU was already 4.0), she used the first marker of uncertainty vist ‘probably’:

(1) Mother: on (valmis) või? ‘Is ready?’
   Martina: amis [valmis] küll, peab vist puhuma ‘Ready yes, must probably
   blow (to get cooled)’

The usual position of vist is in the middle of CS utterances. The girl uses it to refer to some kind of activity of herself or somebody else, positioning the marker before or after the verb.

The marker of uncertainty võib-olla ‘maybe’ emerged in the same recording, at age 1;11 (4 tokens):

(2) Mother: jaa sajab lund ‘Yes it’s snowing’
   Martina: võibolla [võibolla] ei sajagi ‘Maybe it does not snow (running
to the window to check)’

It appears only in an utterance-initial position and refers mostly to the possibility of the situation.

Her third uncertainty marker äkki ‘maybe’ (2 tokens) emerged 2 months later (2;1):

(3) Mother: ma usun, et ju ta täna ikka tuleb see linnukene siia ‘I believe that
   the bird will come here today’
   Martina: äkki ta tuleb sööma ka ‘Maybe it will come and eat also’

Initially Martina used this marker in the same position and for referring to some kind of activity (usually in the third person). However, from 2;3 she started to also use it in the middle of utterances and for referring to noun phrases.

The second Estonian child, Andreas, started with the marker of uncertainty vist ‘probably’ (2 tokens, more in later recordings) at 2;0, when his MLU was 3.6:

(4) Andreas: mis sa noonistad [joonistad]? ‘What are you drawing?’
   Father: jõehobu ‘A hippo’
   Andreas: liblika vist ‘Butterfly probably’

At the beginning Andreas used this marker only in an utterance-final position and in contexts where the uncertainty is focused on the existence or non-existence of an object referred to by a noun (4). After 3;0 he also began to use the marker in
the middle of the phrase, and with verbs. At 2;3 the second marker of uncertainty, äkki ‘maybe’ (1 token) appeared. Markers of certainty started to emerge a bit later: tegelikult ‘really’ emerged at 2;3 (1 token, as an imitation), and tõesti ‘really/truly’ at 2;5 (1 token):

(5) Mother: aga tegelikult? ‘But really?’
Andreas: tegelikult ka ei olnud (midagi kõrvas) ‘(It) was really nothing (in my ear)’

(6) Andreas: meil silmad liiguvad, aga meil muumil ei liigu ‘Our eyes are moving but the eyes of our Muumi do not move’
Andreas: tõesti ei liigu ‘Really/truly do not move’

Both children used epistemic markers only in non-interrogative utterances during the whole observation period.

The Russian boy Filipp begins to use EM markers at 2;1 (at the stage when his MLU is 3.3) with navernoe ‘probably’ (3 tokens) for the expression of uncertainty in trying to determine the name of a fictional animal:

(7) Filipp: èto lošadka naverno ‘This is a horse-DIM, probably (about Konek-Gorbunok)’

Four months later, at 2;5 (his MLU index at this time is 3.9), his first marker of certainty konečno ‘of course’ (1 token) occurs. It refers to an adjective, namely to the identification of an object’s attribute (its size):

(8) Mother: a rozovaja pugovica bol’šaja ili malen’kaja? ‘Is a pink button big or small?’
Filipp: bol’šaja, konečno ‘Big of course’

The boy’s utterances containing initial EM markers are non-interrogatives. They are located at the very end of place within utterances.

Liza begins with two markers of uncertainty, navernoe ‘probably’ (3 tokens) and po-moemu ‘in my opinion, personally’ (1 token) at 2;2, also within non-interrogative utterances. Her MLU in this time is slightly less than Filipp’s (3.1):

(9) Liza: èto, ina [navernoe], guljajut (s) Lizočkoj ‘This is (children) walking probably with Liza-DIM’

(10) Mother: davaj, rasskaži etu sčitaločku ‘Come on, tell me this rhyme-DIM’
Liza: ne ta, pamamy [po-moemu] ‘It is not that one in my opinion’

The initial uncertainty markers in Liza’s utterances are mainly referring to the naming of participants in some situation. At the same time, when the girl is included in this situation, she talks about herself as a third person:

(11) Liza: Liza, inna [navernoe], kupat’ja (v) vannočke ‘Liza (will) probably bathe in the bath-DIM’

Liza’s first certainty marker konečno ‘of course’ (1 token) emerges simultaneously, at the age of 2;2, the marker occurred in her CDS as well:

(12) Liza: èto tramvaj, kanesja [konečno] ‘This is a tram, of course’
Liza’s initial attempt to express her confidence was noted in the utterance of nominalisation of an inanimate object, a means of transport. EM markers are used in non-interrogatives, where they are not placed at the beginning of utterances. However, although starting to use both uncertainty and certainty markers at the same time, the girl does not return to the labelling of confidence until she is nearly four years old (3;10) when her MLU was 7.7.

Based on the above, we can conclude that the semantics of uncertainty and its markers is expressed first in CS in both languages. Moreover, Estonian and Russian children use the uncertainty markers expressing high (‘probably’) and middle (‘maybe’) degrees, and markers of categorical certainty (‘really’, ‘of course’), which can be correlated with the degree of HDU.

Characterising the level of communicative competence within initial epistemic marking events (on the whole, from the end of the 2nd to the beginning of the 3rd year), one could state that in the Russian CS data the process of uncertainty marking occurs at the beginning of the stage of three-word utterances, whereas Estonian children need more words in an utterance when they start to express EM.

4.2. Development of certainty and uncertainty and their expression

The frequency and the variety of EM markers is not great in the speech of both Estonian children. The semantics of uncertainty is represented by 4 markers. The repertoire of Andreas consisted of vist ‘probably’, äkki ‘maybe, suddenly’, võib-olla ‘maybe’ and minu meelest ‘in my opinion’. Although Martina acquired uncertainty markers in a different order (vist > võib-olla > äkki) she had exactly the same first three markers as Andreas.

Both children used 3 markers for expressing certainty. Andreas started to express this semantics with tegelikult ‘really, indeed’ at 2;3, later he used tõesti ‘really’ at 2;5 and muidugi ‘of course’ at 3;0. There was only one marker of certainty, tõesti ‘really’, in Martina’s speech, which could be caused by the limited nature of her data.

Acquiring the modality of uncertainty, Russian children start, as mentioned, with navernoje ‘probably’, and po-moemu ‘in my opinion’ (Liza’s data). After that možet ‘may’ which is from ‘may (be)’ (at 2;4 in Filipp’s speech, at 2;5 in Liza’s data) and možet byt’ ‘maybe’ (at 2;7, and at 2;9 respectively), kažetsja ‘(it) seems’ (at 2;8 in Filipp’s production) emerged subsequently.

Acquisition of certainty begins with konečno ‘of course, certainly’ followed by dejstvitel’no ‘really, indeed’ and na samom dele ‘in fact’ (at 3;10 in Liza’s utterances).

Thus, the earlier semantics of uncertainty is represented by a bigger choice of markers in comparison with certainty in CS of both languages. However, it needs to be mentioned that Martina’s data is smaller at a later age and that is probably why she had only one certainty marker.

7 The word äkki has two meanings in Estonian: ‘maybe’ and ‘suddenly’. All usages of the äkki in non-epistemic meaning were excluded from the data, the second meaning in aphostrophes indicates that the marker is not absolute synonym with another marker, võib-olla.
4.3. Frequency of epistemic markers of in child speech

Table 3 illustrates the lexical diversity along with the quantity aspect of EM markers in our corpora. The number of markers varies from 4 to 7 in Estonian and from 5 to 7 in Russian CS, and from 9 to 11 in Estonian CDS and from 6 to 9 in Russian CDS. That is, the variety of epistemic markers is somewhat bigger in CDS than in CS. However, the correlation between CDS and CS is not significant \( r = 0.512 \).

Moreover, there is a strong correlation between EM tokens in CDS and CS \( r = 0.837, p < 0.01 \). Thus, a high significance level means that the more there are in the caregiver’s utterances, the more of them there are in the child’s phrases.

Finally, the type/token ratio (TTR) reflecting the breadth of vocabulary, according to Phillips (1973: 183), shows that Filipp has the most diverse epistemic vocabulary, whereas Andreas has the least diverse one. A high level of correlation in TTR between CDS and CS is present in both languages \( r = 0.887, p < 0.001 \).

Table 3. Lexical diversity of EM markers and their frequency in Estonian and Russian

<table>
<thead>
<tr>
<th></th>
<th>CS</th>
<th></th>
<th>CDS</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Types/lemmas</td>
<td>Tokens</td>
<td>TTR</td>
<td>Types/lemmas</td>
</tr>
<tr>
<td>Martina</td>
<td>4</td>
<td>14</td>
<td>0.28</td>
<td>9</td>
</tr>
<tr>
<td>Andreas</td>
<td>7</td>
<td>59</td>
<td>0.12</td>
<td>11</td>
</tr>
<tr>
<td>Liza</td>
<td>7</td>
<td>31</td>
<td>0.24</td>
<td>9</td>
</tr>
<tr>
<td>Filipp</td>
<td>5</td>
<td>13</td>
<td>0.38</td>
<td>6</td>
</tr>
</tbody>
</table>

Thus the Estonian and Russian children and adults are more similar in EM repertoire than in the frequency of use of markers; nevertheless, at the same time there is a strong correlation between the frequency of EM markers in CDS and CS.

In Estonian the most frequent epistemic marker (in tokens) in CS is the marker of uncertainty *vist* ‘probably’ (49% of all occurrences); while the most frequent certainty marker is *tegelikult* ‘really, actually, in fact’ (19%) which occurs in the speech of only one child, Andreas (Table 4).

Table 4. The repertoire of epistemic markers in Estonian CS (in tokens)

<table>
<thead>
<tr>
<th>Epistemic markers</th>
<th>Martina</th>
<th>Andreas</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 vist ‘probably’</td>
<td>3</td>
<td>33</td>
<td>36</td>
</tr>
<tr>
<td>2 äkki ‘maybe/suddenly’</td>
<td>6</td>
<td>6</td>
<td>11</td>
</tr>
<tr>
<td>3 võib-olla ‘maybe’</td>
<td>4</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>4 tegelikult ‘actually, in fact’</td>
<td>–</td>
<td>14</td>
<td>14</td>
</tr>
<tr>
<td>5 muidugi ‘of course’</td>
<td>–</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>6 tõesti ‘really’</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>7 minu meelest ‘in my opinion’</td>
<td>–</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Total</td>
<td>14</td>
<td>59</td>
<td>72</td>
</tr>
</tbody>
</table>

Table 5 represents the most frequent markers in Russian data. Within uncertainty there are *navernoe* ‘probably’ (35% in all epistemic tokens) and *możet byt* ‘maybe’
(16%), and within certainty there is *konečno* ‘of course, certainly’ the most frequent marker (16%).

**Table 5.** The repertoire of epistemic markers in Russian CS (in tokens)

<table>
<thead>
<tr>
<th>Epistemic markers</th>
<th>Filipp</th>
<th>Liza</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 navernoe ‘probably’</td>
<td>8</td>
<td>18</td>
<td>26</td>
</tr>
<tr>
<td>2 možet byt’ and možet (from možet byt’) ‘maybe’</td>
<td>3</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>3 konečno ‘of course’</td>
<td>1</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>4 dejstvitel’no ‘really’</td>
<td>–</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>5 po-moemu ‘in my opinion, personally’</td>
<td>–</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>6 kažetsja ‘(it) seems’</td>
<td>1</td>
<td>–</td>
<td>1</td>
</tr>
<tr>
<td>7 na samom dele ‘in fact, actually’</td>
<td>–</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>13</td>
<td>31</td>
<td>44</td>
</tr>
</tbody>
</table>

In comparative analyses it was revealed that initial epistemic repertoires coincide in the two languages. Moreover, in both cases the diversity of markers is the same (7 markers); however, some of them are used more frequently, whereas others are used only once or twice. The most frequent uncertainty markers for both languages are ‘probably’ (Est *vist* / Rus *navernoe*) and ‘maybe’ (Est *äkki* / Rus *možet byt*, *možет*); among certainty markers, the most common are ‘of course’ (*konečno*) in Russian and ‘actually’ (*tegelikult*) in Estonian.

Rare markers (e.g., Rus *dejstvitel’no* / Est *tõesti* ‘really’, Rus *po-moemu* / Est *minu meelest* ‘in my opinion’) coincide in Russian (13) and Estonian (14) CS:

(13) Liza (3;10): *polnaja skovorodka, dejstvitel’no.* ‘The pan is full really’

(14) Andreas (3;0): *siin vahel on minu meelest muna.* ‘Here between (cake layers) is in my opinion the egg’

On the whole, uncertainty is dominant in CS in the early stages of language development (Fig. 1).

![Figure 1. Ratio of certainty and uncertainty in Estonian and Russian CS (in tokens)](image)
4.4. The impact of quality and quantity of input on output

The marker of uncertainty *vist* ‘probably’ is the most frequent in CDS in Estonian corpora (Table 6). The effect of frequency can be noticed in the order of acquisition on epistemic modals: this marker was acquired first by both children. Other frequent markers in CDS were all among the children’s early epistemic markers.

**Table 6. The repertoire of epistemic markers in Estonian CDS (in tokens)**

<table>
<thead>
<tr>
<th>Epistemic markers</th>
<th>Martina</th>
<th>Andreas</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 <em>vist</em> ‘probably’</td>
<td>37</td>
<td>132</td>
<td>169</td>
</tr>
<tr>
<td>2 <em>muidugi</em> ‘of course’</td>
<td>18</td>
<td>60</td>
<td>78</td>
</tr>
<tr>
<td>3 <em>võib-olla</em> ‘maybe’</td>
<td>35</td>
<td>40</td>
<td>75</td>
</tr>
<tr>
<td>4 <em>ääki</em> ‘maybe’</td>
<td>33</td>
<td>13</td>
<td>46</td>
</tr>
<tr>
<td>5 <em>tõesti</em> ‘really’</td>
<td>11</td>
<td>19</td>
<td>30</td>
</tr>
<tr>
<td>6 <em>tegelikult</em> ‘actually’</td>
<td>–</td>
<td>27</td>
<td>27</td>
</tr>
<tr>
<td>7 <em>kindlasti</em> ‘for sure’</td>
<td>8</td>
<td>5</td>
<td>13</td>
</tr>
<tr>
<td>8 <em>ilmselt</em> ‘apparently’</td>
<td>2</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>9 <em>selgelt</em> ‘clearly’</td>
<td>6</td>
<td>–</td>
<td>6</td>
</tr>
<tr>
<td>10 <em>minu arvates, minu meelest</em>  ‘in my opinion’</td>
<td>5</td>
<td>4</td>
<td>9</td>
</tr>
<tr>
<td>11 <em>tõendoliselt</em> ‘probably’</td>
<td>–</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>155</td>
<td>306</td>
<td>461</td>
</tr>
</tbody>
</table>

Additionally, uncertainty markers are predominant in Estonian subcorpora (Fig. 2).

![Certainty and uncertainty in Estonian data](image)

**Figure 2.** Certainty and uncertainty in Estonian data

Fig. 3 and 4 illustrate the dynamics of uncertainty as being more frequent in all data. The number of uncertainty markers is much higher in Martina’s CDS, but the number of them in CS is still quite low (Fig. 3). The amount of uncertainty markers in CDS rises during the whole period and especially at child age 1;11. But we still can’t say that the relative frequency of EM markers in CDS increases: at child age 2;1 (when the first markers of uncertainty emerged), 1.2% of utterances in CDS contained epistemic markers, compared to only 0.7% at age 3;1.
There are two age points in Andreas’ data (2;0 and 3;0) where there are more uncertainty markers in CDS as presented in Fig. 4; at both age points, the child also had more markers in his speech. Although there is much more data at these age points, an increase in the relative frequency of EM markers in Andreas’ data is noticeable: he has 0.67% of utterances containing EM markers at age 2;0 and 7.23% at age 3;0. At age point 2;3 (only one hour of recorded speech) 2% of utterances contained EM markers. Hence, despite the limited data from age 2;3, there is a clear correlation between the age of the child and the frequency of usage of EM markers.

The same could be argued for the usage of certainty markers. The impact of Russian ‘modal input’ on output is supported by the data reflected in Table 7, cf. Table 5. We can observe the same preferences with regard to the frequency of EM markers in CDS: the most frequent of them, novernoe ‘probably’, možet byt ‘maybe’, and konečno ‘of course’, were acquired by children first (see 4.1 above).

---

8 The larger amount of data shows that adults actually use EM markers, and not sparingly, in their speech. Taking into consideration that at these age points the amount of data was approximately 10 times larger, this quite proportionate increase is not unexpected.

9 As stated before, the number of certainty markers is not as high as those of uncertainty in both Andreas’ CS and CDS.
Table 7. The repertoire of epistemic markers in Russian CDS (in tokens)

<table>
<thead>
<tr>
<th>Epistemic markers</th>
<th>Filipp</th>
<th>Liza</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 naverno 'probably'</td>
<td>33</td>
<td>50</td>
<td>83</td>
</tr>
<tr>
<td>2 možet byt', možet 'maybe'</td>
<td>15</td>
<td>28</td>
<td>43</td>
</tr>
<tr>
<td>3 konečno 'of course'</td>
<td>5</td>
<td>85</td>
<td>90</td>
</tr>
<tr>
<td>4 dejstvitel'no 'really'</td>
<td>–</td>
<td>47</td>
<td>47</td>
</tr>
<tr>
<td>5 po-moemu 'in my opinion'</td>
<td>2</td>
<td>52</td>
<td>54</td>
</tr>
<tr>
<td>6 kažetsja 'it) seems'</td>
<td>1</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>7 na samom dele 'in fact'</td>
<td>1</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>8 pravda 'truly'</td>
<td>4</td>
<td>14</td>
<td>18</td>
</tr>
<tr>
<td>Total</td>
<td>61</td>
<td>287</td>
<td>348</td>
</tr>
</tbody>
</table>

As was expected, the EM repertoire in CDS is more diverse than in CS, mainly due to the indicators of low degree and/or problematic certainty kažetsja ‘(it) seems’ and pravda ‘truly, in truth’:

(15) 2;1 Filipp
Mother: (oni) krasivye, pravda ‘(They are) nice, truly’

(16) 3;10 Liza
Mother: po nazvaniju kažetsja, čto možno est’ syrymi, da? ‘By title it seems that it’s possible to eat them raw? (about mushrooms)’

There is also a big difference in the total amount of markers in the speech of the two caregivers. One reason could be the different amount of speech material. Still, the difference in the amount of data is not proportionally as big as the difference in the amount of epistemic markers and could be more connected with the individual communication style of the caregivers.

Generally, markers of uncertainty are predominant in all Russian subcorpora but in varying degrees (Fig. 5). In Liza’s CDS there are more certainty markers and this can influence the child’s strategy: she has more certainty markers than does Filipp.

Figure 5. Certainty and uncertainty in Russian data
The results of distributional analysis reveal some relationships between input-output in the sphere of marking of the prevailing semantics of uncertainty (Fig. 6 and 7).

**Figure 6.** Dynamics of uncertainty in Liza’s corpus (number in tokens)

In the development of uncertainty in Liza’s data, 3 simultaneous peaks in the usage of markers by both the child and her caregiver were noted at 2;2, 2;9, and 3;10. These peaks do not depend on the length of recordings (see Table 2), but are connected with the development of personal pronouns in CS. So the first spurt coincides with the beginning of usage of personal pronouns, whilst the second and third correlate with the peaks in their usage in CS (95 and 86 tokens, respectively). Thus epistemic markers occur earlier than personal pronouns in Liza’s speech, while their spurt is supported (or even ‘provoked’) by the usage of personal pronouns.

**Figure 7.** Dynamics of uncertainty in Filipp’s corpus (number of tokens)

The general dynamics of EM acquisition in the Filipp – adult dyad can be described as a progressive movement, in which the subsequent peak (at 2;8) is higher than the previous one (2;1). The first peak in CDS ‘provokes’ the initial appearance of EM in CS and the second spurt in CDS seems to be simultaneous in CS. Interestingly, EM acquisition is also connected with personal pronouns: an abundance of pronouns in CS (2;0, 2;2, 2;7) facilitates subsequent increasing of EM (2;1, 2;4, 2;8).

Thus, in the communication of 2 dyads (between Liza and Andreas and their caregivers) the so-called epistemic peaks coincide. This demonstrates the direct
influence of the quantity of modal input on the occurrence of epistemic markers in output. However, in Martina’s and Filipp’s data this picture is somewhat different with regard to the simultaneity of these peaks: the appearance of markers in CS is delayed. In turn, one could state that the quality of modal input directly affects the modal repertoire of children.

4.5. The relationship between adult-directed speech, child-directed speech and child speech

In order to a) reveal whether there is correlation between the EM markers in colloquial adult speech and in caregiver speech and b) to define the distribution of EM markers, we have compared the relevant data. Such analysis contributes to the judgement of the prototypicality question of EM as well.

For the analysis of the distribution of epistemic markers in Estonian ADS and CDS, 2 hours of recordings of spontaneous speech were used (Estonian child language database, CHILDES, dialogues between two adults from the Korgesaar subcorpus). The big split between CDS and ADS can already be observed in the general number of EM markers: there were 164 markers (115 markers in one and 49 markers in the other recording) in the data consisting of only 120 minutes of dialogue. Thus, EM markers in Estonian ADS are very frequent. The average number of utterances in one hour of recorded material was 322, and 31% of utterances contained EM markers.

The ‘virtual scale’ of EM is reflected in Estonian CS in the following way: each degree within the semantics of uncertainty, except HDU, has its own representatives emerging sequentially from a high degree to a low (Table 8).

### Table 8. Epistemic markers in Estonian CS, CDS and ADS (% of all markers in each subcorpus)

<table>
<thead>
<tr>
<th>Epistemic markers</th>
<th>Degree/sphere of EM</th>
<th>Martina CS</th>
<th>Martina CDS</th>
<th>Andreas CS</th>
<th>Andreas CDS</th>
<th>Total CS</th>
<th>Total CDS</th>
<th>ADS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 vaevalt ‘hardly’</td>
<td>HDU</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>2 võib-olla ‘maybe’</td>
<td>MDU</td>
<td>29</td>
<td>24</td>
<td>2</td>
<td>13</td>
<td>7</td>
<td>16</td>
<td>14</td>
</tr>
<tr>
<td>3 äkki ‘maybe, suddenly’</td>
<td>MDU</td>
<td>43</td>
<td>21</td>
<td>10</td>
<td>4</td>
<td>15</td>
<td>10</td>
<td>3</td>
</tr>
<tr>
<td>4 vist ‘probably’</td>
<td>MDU</td>
<td>21</td>
<td>24</td>
<td>57</td>
<td>43</td>
<td>48</td>
<td>37</td>
<td>19</td>
</tr>
<tr>
<td>5 minu meelest/arvates ‘in my opinion’</td>
<td>LDU</td>
<td>0</td>
<td>5</td>
<td>2</td>
<td>4</td>
<td>2</td>
<td>9</td>
<td>4</td>
</tr>
<tr>
<td>6 ilmelt ‘(it) seems’</td>
<td>LDU/PC</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>15</td>
</tr>
<tr>
<td>7 tõenäoliselt ‘probably’</td>
<td>PC</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.3</td>
<td>0</td>
<td>0.2</td>
<td>0</td>
</tr>
<tr>
<td>8 selgelt ‘clearly’</td>
<td>CC</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>9 muidugi ‘of course’</td>
<td>CC</td>
<td>0</td>
<td>12</td>
<td>3</td>
<td>20</td>
<td>5</td>
<td>17</td>
<td>12</td>
</tr>
<tr>
<td>10 kindlasti ‘surely’</td>
<td>CC</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>11 loomullikult ‘naturally’</td>
<td>CC</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>12 tegelikult ‘in fact, actually’</td>
<td>CC</td>
<td>0</td>
<td>0</td>
<td>23</td>
<td>9</td>
<td>19</td>
<td>6</td>
<td>21</td>
</tr>
<tr>
<td>14 tõesti ‘truly’</td>
<td>CC</td>
<td>7</td>
<td>7</td>
<td>2</td>
<td>6</td>
<td>0</td>
<td>6</td>
<td>6</td>
</tr>
</tbody>
</table>

Hence there are a significant number of conversations which differ from the situations existing in the recordings in CS corpora (like phone conversations between a client and a salesman etc.). The Corpus of Spoken Estonian was not used here.
The most frequent marker used in ADS was, as in CDS, the uncertainty marker *vist* ‘probably’ (36% of all uncertainty markers) and *ilmelt* ‘apparently’ (29%). Within the second micro field, the most frequent marker of CC, *muidugi* ‘of course’, is ‘reflected’ in CS, while the second most frequent marker, *tegelikult* ‘really’, is not and is characteristic of ADS. The distribution of markers of certainty and uncertainty is almost equal in ADS: there were only slightly more markers of uncertainty (52%) than of certainty (48%). MDU is expressed by *võib-olla* and *äkki* ‘maybe’. LDU is expressed by *ilmelt* ‘(it) seems’ and *minu meelest* ‘in my opinion’.

The correlation between CS and CDS is more significant than the correlation between CDS and ADS. That is, not all the markers used in ADS are present in CDS and CS, but when the marker is frequent in CDS, it is also frequent in CS. Still, the most frequent ADS marker is also frequent in CDS and CS (Fig. 8).

According to the analysis of ADS presented in the Russian National Corpus (the total is 330,086), the distribution of EM markers is equal (53% of epistemic markers are uncertainty markers, 47% certainty markers). The most frequent uncertainty markers are *možet byt* ‘maybe’ and its colloquial version *možet* ‘may’ (19%), *kažetsja* ‘it seems’ (9%), *naverno* and its colloquial variant *naverno* ‘probably’ (5%), *vozmožno* ‘perhaps’ (4%), whereas *konečno* ‘of course’ (16%), *pravda* ‘true’ (11%), *dejstvitel’no* ‘indeed’ (7%) are dominant within the semantics of certainty.

The ‘virtual scale’ of EM mentioned above is reflected in Russian CS in the following way. Each degree within the semantics of uncertainty has its own representatives emerging sequentially from a high degree to low (Table 9). HDU is expressed by only one marker which is *naverno* ‘probably’. MDU is expressed by *možet byt* and *možet* ‘maybe’. Finally, LDU is expressed by *kažetsja* ‘(it) seems’ and *po-moemu* ‘in my opinion’. Within the second micro field, the most frequent marker of CC *konečno* ‘of course’ is ‘reflected’ in CS, along with the less frequent *dejstvitel’no* ‘indeed’ and *pravda* ‘true’.

![Figure 8. CS–CDS–ADS relations in Estonian EM](image-url)
Table 9. EM markers in Russian CS, CDS and ADS (% in all markers in each subcorpus)

<table>
<thead>
<tr>
<th>Epistemic markers</th>
<th>Degree/ sphere of EM</th>
<th>Filipp CS</th>
<th>Filipp CDS</th>
<th>Liza CS</th>
<th>Liza CDS</th>
<th>Total CS</th>
<th>Total CDS</th>
<th>ADS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 navernoе ‘probably’</td>
<td>HDU</td>
<td>61</td>
<td>52</td>
<td>58</td>
<td>17</td>
<td>59</td>
<td>24</td>
<td>5</td>
</tr>
<tr>
<td>2 моет быть, може ‘maybe’</td>
<td>MDU</td>
<td>23</td>
<td>25</td>
<td>13</td>
<td>10</td>
<td>16</td>
<td>12</td>
<td>19</td>
</tr>
<tr>
<td>3 по-моему ‘in my opinion’</td>
<td>LDU</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td>18</td>
<td>2</td>
<td>16</td>
<td>2</td>
</tr>
<tr>
<td>4 кацетса ‘(it) seems’</td>
<td>LDU/PC</td>
<td>8</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>5 koneчно ‘of course’</td>
<td>CC</td>
<td>8</td>
<td>8</td>
<td>19</td>
<td>30</td>
<td>16</td>
<td>26</td>
<td>16</td>
</tr>
<tr>
<td>6 dejствително ‘really’</td>
<td>CC</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>16</td>
<td>2</td>
<td>13</td>
<td>7</td>
</tr>
<tr>
<td>7 на самом деле ‘in fact, actually’</td>
<td>CC</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>0,2</td>
</tr>
<tr>
<td>8 правда ‘truly’</td>
<td>CC</td>
<td>0</td>
<td>7</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>5</td>
<td>5</td>
</tr>
</tbody>
</table>

Results of quantitative analysis have revealed only a correlation between the level of general frequency of EM markers in ADS and their possibility ‘of getting into’ adult–child communication (i.e. mainly the most frequent markers from ADS get into CDS, and then in CS). However, stronger correlations were not noted in Russian data (Fig. 9).

Figure 9. CS–CDS–ADS relations in Russian EM

Thus our comparative study shows that, firstly, the ratio between uncertainty and certainty is the same in Estonian and Russian ADS, which does not depend on differences in the corpora under observation. Secondly, the coincidence of the marker frequency in ADS, CDS and CS was identified partially. So, the most frequent for the conversation and the most prototypical for the target-system epistemic markers are more likely ‘to get into’ CS.

4.6. Development of epistemic marking within children’s utterances

In both languages, early EM markers are used in the children’s judgments about a situation, its components and third persons, i.e. in so-called HE/SHE/IT-utterances. Moreover, in Russian it happens even when a child talks about herself as if it were about a third person using her name, e.g. (11).
Liza, inna [navernoe], kupat’sja (v) vannočke ‘Liza (will) probably bathe in the bath—dim’

Epistemic marking shows a gradual expansion of the children’s reflection in respect of surrounding realities: subjects (17), (18) and objects (e.g. their presence/absence (19), naming or identification (7), (13), such as their attributes of size (8) and colour (20), actions (10) and their circumstances (location (21), time, causality etc.).

(17) 2;1 Filipp
Mother: a kto tebe ego podaril? ‘Who gave it to you? (about a toy tractor)’
Filipp: naverno, papa ‘Probably, daddy’

(18) Liza (2;9): aj, kakoe bezobrazie formennoe. ‘(It’s) a downright outrage’
Liza: èto, naenja [navernoe], deduška skazal ‘This probably grandfather said’

(19) Andreas (2;3): tegelt [tegelikult] ka e i olnud (midagi kõrvas) ‘Actually, there was (nothing in my ear)’

(7) Filipp (2;1): èto lošadka naverno ‘This is a horse—dim, probably (about Konek-Gorbunok)’

(13) Liza (3;10): polnaja skovorodka, dejstvitel’no. ‘The pan is full really’

(8) 2;5 Filipp
Mother: a rozovaja pugovica bol’šaja ili malen’kaja? ‘Is the pink button big or small?’
Filipp: bol’šaja, konečno ‘Big of course’

(20) 2;4 Filipp
Mother: a kakogo on cveta? ‘What color is it? (about a toy bear which is a yellow-brown color)’
Filipp: naverno, želtyj ‘Probably, yellow’

(10) 2;2 Liza
Mother: davaj, rasskaži etu ščitaločku ‘Come on, tell me this rhyme—dim’
Liza: ne ta, pamamy [po-moemu] ‘It is not that one in my opinion’

(21) Andreas (2;0): mängutoas vist (on klaver) ‘(The piano is) probably in the playing room’

From approximately the end of the third year of life (2;8 in Russian and 3;0 in Estonian), epistemic marking is used in children’s utterances about themselves: their own actions including mental ones and inner (psychological) states, and also proposals for joint actions:

(22) 2;8 Filipp
Mother: a počemu on plačet? ‘Why is he crying? (about Noddi)’
Filipp: upal ‘(He) fell down’
Filipp: kažetsja, i ja plaču ‘So it seems, and I am crying’

---

11 Since the colour was not just yellow but between yellow and brown, the child could not decide if this colour was yellow or not.
(23) 2:9 Liza
   Mother: pojdem domoj, a to xot’ i malen’kij, no doždik uže ‘Let’s go home, the rain is coming already, although it is small’
   Liza: <…> ja poka polju@err, naena [navernoe] ‘I am still weeding, probably’

(24) Andreas (3;1): ma praegu vist ei kopsi ‘I probably will not knock now’

(25) Andreas (3;0): rebane, kas sa tahad joonistada minuga ‘fox, would you like to draw with me? (asks the fox)’
   Andreas: muidugi tahan ‘Of course I want (answers instead of the fox)’

Such kinds of i-utterances are some evidence of the beginning of child introspection, which can be interpreted as a prerequisite for future mental decen-
tration or overcoming ‘egocentric thinking’ at its verbal/explicit level. We can’t say that this is strongly correlated with the length of utterances (MLU increases steadily during the whole observation period and any sudden increase can be observed at the time when i-utterances with EM emerge) and also with the development of first person marking in general. First person marking starts to develop in both languages much earlier: e.g., in Andreas’ speech, the frequency of the pronoun ma ‘I’ is only 1–2 occurrences per hour of recording at age 2;3; it further increases suddenly to 21 tokens during the same time period, and only at age 3;0 do EM markers in i-utterances emerge (see also 4.3).

From the end of the fourth year (in Russian) EM markers are used in the children’s utterances about the dialogue partner and her mental actions (YOU-utterances):

(26) Liza (3;10): i tebe, naaverno, prisnilos’ èto ‘You probably dreamed it’
Such types of you-utterances we tend to regard as the first step to another conscious-
ness (namely as the child’s theory of mind). Interestingly, the first you-sentences reflecting not mental actions but physical ones, such as moving, are earlier (at 2;1 in Russian, at 2;7 in Estonian) and are represented more widely in Estonian CS:

(27) Andreas (2;7): istu võib-olla siia ‘Sit-IMP maybe here’
Although situations of epistemic evaluation of mental activity of a dialogue partner were not revealed in Estonian CS, some examples were found where the child uses the EM marker in you-sentences with the verb olem ‘to be’:

(28) Andreas (3;0): sa oled vist väga haige laps ‘You are probably a very sick child’
Importantly, from 3;1 such utterances with the modal verb tahtma ‘want’ occur:

(29) Andreas (3;1): tegelikult sina, kas sa tahad nii palju mulle anda? ‘Actually you, do you want to give me so much?’
Using this modal verb for describing the partner’s desires or mental states could demonstrate the next step in the child’s development, the overcoming of so-called egocentric thinking.

12 The child is not sure which seat would be best.
A comparison of our data allows us to reach a preliminary conclusion that the development of epistemic marking in Estonian and Russian CS is quite similar (see also Table 10). Some discrepancies are observed in the last stage: Russian children begin to use EM in *you*-utterances with mental/propositional verbs, whereas Estonian children prefer to use them with modal verbs, i.e. with deontic modality markers.

Table 10. Epistemic marking in children's utterances

<table>
<thead>
<tr>
<th></th>
<th>HE/SHE/IT-utterances</th>
<th>I-utterances</th>
<th>YOU-utterances</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estonian CS</td>
<td>73.5% (58 utterances)</td>
<td>20.5% (14 utterances)</td>
<td>6% (8 utterances)</td>
</tr>
<tr>
<td>Martina</td>
<td>75%</td>
<td>25%</td>
<td>0%</td>
</tr>
<tr>
<td>Andreas</td>
<td>72%</td>
<td>16%</td>
<td>12%</td>
</tr>
<tr>
<td>Russian CS</td>
<td>73% (32 utterances)</td>
<td>20% (9 utterances)</td>
<td>7% (3 utterances)</td>
</tr>
<tr>
<td>Liza</td>
<td>71%</td>
<td>23%</td>
<td>6%</td>
</tr>
<tr>
<td>Filipp</td>
<td>77%</td>
<td>15%</td>
<td>8%</td>
</tr>
</tbody>
</table>

5. Conclusions

Our comparative analysis of EM in the languages under observation has shown that despite typological differences, this category is organised quite similarly in the two languages: the semantics of certainty can be expressed both explicitly and implicitly, whereas uncertainty is expressed only explicitly; both lexical and grammatical means can be used for expressing EM; the number of different EM markers is approximately the same, although the markers cannot be mapped one-to-one according to their meaning in all cases.

Our study of EM acquisition has revealed some similarities in the emergence of its markers. Particularly, we have found that uncertainty is an early and dominating semantic sphere of EM in both languages (75% in Estonian, 66% in Russian). This micro field is represented in CS by 4 markers, while the certainty micro field begins to develop later and is represented by 3 markers. The first marker in each semantic field tends to become the most frequent in the EM repertoire. These are *vist* / *naveerno* ‘probably’ in uncertainty, and *muidugi* / *konečno* ‘of course’ in certainty. The sequence of emergence of EM markers in CS coincides with their frequency in CDS (and ADS, to some extent). In both languages, the first method of expressing EM was lexical. The children under observation begin to use EM markers approximately after the age of two. At this time the MLU of Estonian utterances in CS were 3.6 and 4.0, while this index in Russian CS was equal to 3.1 and 3.3. Children used EM markers appropriately. In the micro field of certainty, both Estonian and Russian children start with markers expressing CC, while in the micro field of uncertainty Estonian children start with markers with MDU, but Russian with HDU.

Taking into consideration the semantic structure of the children’s utterances with EM markers, one could state the following order of usage. Initially, epistemic expressions start to be used within HE/SHE/IT-utterances, later in I-sentences, and finally, in YOU-utterances, including ones containing mental and modal verbs. In other words, children begin epistemic marking with the evaluation of some objective
situation and its components, and move onto mental reflection concerning their partner.

The investigation of the relationship between input and output has revealed the influence of the input frequency of different markers on their acquisition. The first EM markers acquired by children are also frequent in CDS in both languages. Analysis of the correlation between ADS, CDS and CS has shown that EM markers, which are frequent in ADS, are, in most cases, also frequent in CDS but with some exceptions. In particular, some of them seem to be more characteristic in ADS than in CDS, e.g. *tegelikult* ‘actually’, *ilmelt* ‘apparently’ in Estonian; in Russian, conversely, *navernoe* and its colloquial variant *naverno* ‘probably’ are more frequent in CDS in comparison with ADS. Moreover, the EM markers that are both frequent and prototypical in the target system of each language are likely to be acquired first.

No striking differences were found in acquisition when comparing the two languages except for some individual features: *po-moemu* ‘in my opinion’ is quite frequent in Russian CDS but its equivalents *minu arvates* and *minu meelest* are not so frequent in Estonian CDS, and are infrequent in CS.

On the basis of the results obtained, we can conclude that several factors are intertwined in the process of EM acquisition in L1: they are, in particular, both the semantic and structural complexity of the category itself (based on a more general cognitive domain of modality), a lexical means of expressing both micro fields and their target system prototypicality and frequency, the influence of input being some kind of mediator between ADS and CS, the level of communicative competence of children (including, to some extent, MLU) and their general cognitive development. On the whole, EM is more complicated in a cognitive sense than deontic modality and moreover requires a special position within utterances. This is because epistemic markers are beyond the syntax structure, which also may explain their late emergence, not only in ontogenesis but also in historical development.

**Abbreviations and symbols**

[ ] correct spelling of the word
ADS adult-directed speech
CC categorical certainty
CDS child-directed speech
CS child speech
DIM diminutive
EM epistemic modality
Est Estonian
HDU high degree of uncertainty
IMP imperative
LDU low degree of uncertainty
MDU middle degree of uncertainty
MLU mean length of utterance
PC problematic certainty
Rus Russian
TTR type/token ratio
References


de Villiers, Jill; Garfield, Jay; Gernet-Girard, Harper; Roeper, Tom; Speas, Margaret 2009. Evidentials in Tibetan: Acquisition, Semantics, and Cognitive Development. – New Directions for Child and Adolescent Development, 125, 29–47.


Moore, Chris; Pure, Kiran; Furrow, David 1990. Children’s understanding of the modal expression of speaker certainty and uncertainty and its development to


Victoria V. Kazakovskaya is a leading researcher at the Department of theoretical grammar at the Institute for linguistic studies (Russian Academy of Sciences). Her research interests include philology, linguistics, first language acquisition.

Reili Argus is Professor of Estonian language in the School of Humanities at Tallinn University at Tallinn University. Her research interests include include philology, linguistics, first language acquisition.

Web references

Victoria V. Kazakovskaya

Reili Argus


Kahe vaadeldud keele episteemilise modaalsuse markerite omandamisprotsessis on kindlasti põimunud mitu olulist tegurit: 1) omandamiskäik võib olla keeltes sarnane seetõttu, et kategooria ise, st leksikaalsete vahendite paigutumine tõenäosus–ebatõenäosus–ebatõenäosus skalaal on üsna sarnane; 2) omandamise vanus on seotud nii üldise kognitiivse arenguga (nt võime tabada ja väljendada vestluspartneri kavatsusi areneb välja suhteliselt hilja); 3) omandamine on seotud üldisema grammatika arenguga (esimesel episteemilise modaalsuse markerile ilmuvad laste keelde sis, kui nende ühes lausungis on keskmiselt juba üle kolme sõna); 4) episteemilise modaalsuse väljendusvahendite omandamist mõjutab ka lapsele suunatud kõne, st et markerite omandamise järjekord sõltub nende sagedusest sisendkeelas.

Võtmesõnad: esimene keele omandamine, episteemiline modaalsus, eesti keel, vene keel