Input-output Relations in Early Epistemic Modality in Russian: a Corpus-based Approach*

Victoria V. Kazakovskaya

Russian Academy of Sciences, Institute for Linguistic Studies, Tuchkov Per., 9, St. Petersburg, Russia

Abstract - The report discusses the acquisition of epistemic semantics and the basic means of its expression in the early stages of speech ontogenesis in L1. The mechanism of epistemic marking is also considered. The target-system and communicative factors, which contribute to this process are identified and corpus data of modern Russian speech is included. The results are compared with results of previous studies of epistemic modality and evidentiality.

Index Terms - First language acquisition, grammar, epistemic modality, Russian, longitudinal corpus data.

1. Introduction

No research has previously been carried out on the acquisition of epistemic modality in Russian based on longitudinal corpus data and presented in Child Language Exchange Data System (CHILDES, http://childes.psy.cmu.edu) [1]. Our study is the first attempt to discuss the early phases of acquisition of this phenomenon within a morphologically rich Slavic language (cf. [2-6]). Moreover, data of adult-directed speech from the Russian National Corpora (RNC, http://www.ruscorpora) were included, which has also not been done before.

The theoretical background of this problem includes at least two aspects: previous studies of epistemic modality in Russian child speech and the so-called target-system aspect of this grammatical category.

A. Previous Investigations of Epistemic Modality in Russian Child Language

There have been few previous papers devoted to epistemic modality in Russian ontogenesis. Firstly, there is some evidence from the prosodic experiments, including the problem of epistemic marker comprehension, which were conducted with pre-schoolers [7]. Secondly, there are some observations based on various diary data [8-10]. Finally, there are our own investigations realised within the framework of several projects [11, 12], where epistemic modality was considered within the larger problem of development of semantic structure of child utterances, namely, *dictum* and *modus* structure, according to [13].

We analysed the initial ways of children expressing their point of view and their 'subjective / propositional attitude' to the statement in adult — child dialogue [14]. These results were based on all Russian data available, including tape recordings and numerous parental diaries, and took into consideration the relevant results of a modality study of Russian colloquial speech (adult-directed speech, *ADS*)

focusing on the problems of authority and point of view, in the terms of [15] and [16].

The results of our previous studies have shown that although expression of epistemic modality — propositional attitude to any objective situation via special linguistic tools — happens later and less frequently (than, for example, deontic modality [3]), this process is very important since it is significant not only for the proper language development of children, but also for their cognitive development, because it allows them to overcome so-called 'egocentric thinking', according to [17] and [18]. It also contributes to the formation of "theory of mind" [19 -22], which is characterized by numerous concepts of reality and varying viewpoints, so epistemic marking is the first step in this direction: cf. *I think / believe / consider* etc. *that* (…) — *In my opinion / personally* etc. (…) — *According to N* etc. (…) — *Of course / maybe / probably* etc. (…) [14].

B. Target-System Aspect of Epistemic Modality in Russian

Before moving onto the discussion of our latest results, let us look briefly at the target-system aspect of Russian epistemic modality, including its grammar status, semantic structure, syntax features and functions [16, 23 -27].

1) Grammar Status of Epistemic Modality. This is one of the groups within such parts of speech as modal words and which is constantly supplemented by words from different grammatical classes: e.g., from nouns: *pravda* 'truth', *fakt* 'fact'; from adjectives and adverbs: *vozmožno* 'perhaps', *verojatno* '(most) likely', *nesomnenno* 'undoubtedly'; from verb forms: *kažetsja* 'seems', *vidimo* 'probably', *razumeetsja* 'needless to say', and finally, from pronouns: *samo soboj* 'it goes without saying'. Acting as modal words (or "playing the role of") they are rethinking semantically and grammatically.

2) A Semantic Structure of Epistemic Modality. The semantics of epistemic modality is abstract: it indicates the degree of the speaker's certainty of reality, reliability, authenticity and accuracy of what he / she says.

The semantics of epistemic modality is represented by two polar micro fields of *certainty* (confidence, truth) and *uncertainty* (probability). Epistemic markers are placed on a scale in order of decreasing proportion of uncertainty and, consequently, increasing degree of certainty in the speaker's information, in other words, from the greatest doubt (*vidimo* 'probably', *možet byt'* 'maybe', *vozmožno* 'perhaps', *po-vidimomu* 'apparently, appear to', *očevidno* 'obviously,

clearly', *navernoe* 'certainly', *kažetsja* 'seems', *verojatno* '(most) likely, probably', *po vsej verojatnosti* 'in all probability, in all likelihood', *po vsej vidimosti* etc.) — to the greatest confidence (*konečno* 'of course', *nesomnenno* 'undoubtedly', *dejstvitel'no* 'indeed', *razumeetsja* 'needless to say', *v samom dele* 'actually', *bessporno* 'undubitably', *bezuslovno* 'certainly', *bez somnenija* 'no doubt', *verno* 'faithfully', *samo soboj razumeetsja* 'it goes without saying' etc.).

At the same time, each semantic field has some structure in ADS. Particularly, the micro field of uncertainty is presented by three degrees. A high degree of uncertainty is expressed by *navernoe* 'probably', *vidimo* 'probably', *po-vidimomu* 'apparently, appear to', *očevidno* 'obviously, clearly' etc. in the conclusions and predictions of speakers. The middle degree is expressed by *možet byt'* 'maybe', *možet* 'may' and *vozmožno* 'perhaps' in suggestions, assumptions and suppositions of the speakers. A low level is expressed by *kažetsja* 'it seems' and *požaluj* 'very likely' in weakly reasoned and unmotivated statements, which are already very close to a low degree of certainty.

The micro field of certainty is organised in another way. The point is that so-called 'simple certainty' is characterised by the absence of modifiers in the statement. It is knowledge: *I know that P*, which would usually be presented implicitly. For the appearance of these markers in an utterance, special conditions are needed, at least the intention of the speaker to underline her / his certainty. Categorical certainty, or underlined reliability (so-called "epistemic intensity") is expressed by *konečno* 'of course, certainly', *dejstvitel'no* 'really, indeed', *na samom dele* 'in fact', *pravda* 'truly' etc. It borders on knowledge, but can't be considered equal to it. The next level is 'the problematic certainty' 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. *it seems* and some particles).

It is important to emphasize that the semantics of certainty can be expressed both explicitly and implicitly, whereas uncertainty is expressed only explicitly.

Indicators of authority (or authorization) connect epistemic modality with the source of information (i. e. the category of evidentiality [28, 5]): *according to N / to rumours, in my opinion* etc. Such a way evidentiality finds its expression in Russian language, where this category is not grammaticalised.

3) Some Syntax Characteristics of Epistemic Markers. The operators of epistemic modality are indeclinable words and collocations along with some modal particles, which can occupy any position within an utterance, although they are not its members. To emphasise this special state, a term 'parenthetical words' is used [16, 23, 25]. Significantly, in a colloquial (or spontaneous) dialogue, epistemic markers often take the reactive position within an adjacent pair: *Would you like to have a young husband? — Of course. [An old husband is cold.]* (RNC).

4) The Functions of Epistemic Markers. The primary (basic) functions of the epistemic markers are related to the implementation of the epistemic value (i. e. with the qualification information in terms of reliability). The secondary (non-basic) functions are connected with the actualization of the figure of the speaker with the full or partial neutralization of the basic values, complication of its semantics by additional semantic shades and pragmatic nuances [12].

2. Data and Method

The data of four typically developing monolingual Russian children (Vanja, Liza, Vitja and Philip) from Saint Petersburg and their main caregivers (165 hours of recorded speech) were transcribed (CHILDES) [1] and analyzed according to a) the first emergence of epistemic modality operators, b) further development of both semantic types, namely *certainty* and *uncertainty*, c) their frequency in child speech (*CS, output*), d) the impact of quality and quantity of "modality input" (child-directed speech, *CDS*) on output, and e) general correlations between CS and ADS. Finally, special linguistic tools which facilitate epistemic marking development were studied.

3. Results and Their Discussion

A. General Results

The total number of epistemic operators in the data under observation is more than nine hundred items. The largest corpora here belong to Vanja (353 markers) and Liza (318 markers). Characteristics of lexical diversity in the sphere of epistemic marking are summarised as follows. The number of epistemic operators in lemmas varies from 5 (in Vanja's and Philip's data) to 7 (in Liza's speech) in CS, and from 6 (in Philip's data) to 10 (in Vanja's corpus) in CDS. Thus, the more there are of them in the adult's utterances, the more of them there are in the child's, which, however, is not supported in respect of frequency (in tokens). The size of epistemic modality repertoire of the three children approaches the size of the epistemic markers in the speech of their caregivers.

B. The Main Tendencies of Epistemic Marking Acquisition

Most children start to acquire epistemic semantics from the micro fields of uncertainty at no definite time within the third year of their life. Philip begins before anyone (quite early, at the age of 2;11), Vanja does so later (at the age of 2;8). In turn, children start to perceive epistemic markers from the middle of the second year of their life.

The "period of exposure" (i.e. the time between the first usage of epistemic operators by a caregiver and the first occurrence in the speech of the child) may range from 3 (in Philip's corpus) to 15 months (in Vanja's data) for the semantics of uncertainty, and from 6 months (in Liza's corpus) to 17 (in Vanja's data) — for the micro field of certainty.

Significantly, the uncertainty operators are dominant: their number is twice greater than the operators of certainty (in

---

[1] It means two years and one month.

relation to all data observed) (66%); and, in addition, the uncertainty repertoire is much more diverse. However, the ratio between uncertainty and certainty within the three corpora differs slightly: from 74% in Liza's speech to 96% in Vitja's production.

C. Similarities in the Acquisition of Epistemic Modality

Let's describe some tendencies which were common to either to all corpora or, at least, to three of them. The first important correlation is that the first operator in each semantic type tends, in all data under observation, to become the most frequent in epistemic modality repertoire. Within the modality of uncertainty, it is mainly navernoe 'probably'; while konečno 'of course, certainly' has the highest frequency within the micro field of certainty.

The next common feature is that the sequence of emergence of different operators within both semantic types coincides with their frequency in our corpora, and can be represented in the following way. Acquiring the modality of uncertainty, Russian children start with navernoe 'probably', and after that možet 'may' which is from 'may (be)', po-moemu 'in my opinion', možet byt' 'maybe', kažetsja 'it seems' — emerge subsequently. Acquisition of certainty begins with konečno 'of course, certainly' followed by dejstvitel'no 'really, indeed', na samom dele 'in fact' and pravda 'truly'. In addition, it needs to be emphasised that children use these items correctly.

With a relatively small repertoire of epistemic operators in CS, their tendency of usage can be characterized as a gradual expansion of the children's reflection in respect of subjects, objects and realities of "surrounding reality": in particular, their presence or absence, nomination or identification of attributes (color, size etc.), and finally, their activities and different circumstances (location, time, causality, intentions etc.). In general, they are used in the judgments of a child about a situation, its components and third persons, in so-called "HE / SHE / IT-utterances":

2;8   Child: There may [be] a little mouse will run.
2;9   Child: This is probably the grandfather said.

From the age of 2;10 epistemic operators are used by children in utterances about themselves (in "I-utterances"): in relation to their own actions, proposals and joint actions. This is some evidence of the beginning of introspection, and a pre-requisite for mental decentration (i. e. overcoming of "egocentric thinking"):

2;11   Child: Ah, I probably threw it.
3;4   Child: Or maybe I will buy you frogs?

From the age of 3;10 children use the operators of epistemic modality in the utterances about their partner in a dialogue and mental actions (in "YOU-utterances"). This is the first step to another consciousness:

3;10   Child: You probably dreamed it.
4;0   Child: You probably know.

So, the order of emergence represents an increase of situations where children express their subjective (or propositional) attitude.

D. Differences in the Acquisition of Epistemic Modality

However, some differences in the process of epistemic marking development were revealed in the following aspects such as a) simultaneity of appearance of epistemic markers, b) a dominating marker, and c) type / token ratio (TTR index).

1) Aspect of simultaneity in the appearance of uncertainty/certainty operators. Despite the strong sequence in the process of occurrence of epistemic markers in the data of most of the children (explicit uncertainty occurs before explicit certainty), in the girl's corpus it was a simultaneous process which occurred in both CDS and CS. The second feature of Liza's data is that the modality of certainty was more frequent in input (in tokens).

2) Aspect of a dominating epistemic marker. The first uncertainty operator in Vitja's corpus is not navernoe 'probably' (as in other corpora) but po-moemy 'in my opinion', which, however, coincides totally with his input. Development of epistemic marking in Vitja's speech also differs in the aspect of prevailing semantics. Uncertainty and certainty are equal, their ratios are 48% and 52% respectively; however, this ratio is not reflected in CDS (79% vs 21%).

3) Aspect of TTR in the sphere of epistemic marking. The highest TTR amongst all the data of CS belongs to Vanja (his index is 0.667), the lowest one belongs to Liza (0.226). However, in their CDS these results are similar (0.036 and 0.031), as well as the number of lemmas / types (10 and 9).

E. Epistemic Modality in Terms of Correlation ADS — CS

According to the analysis of the Russian National Corpus, the most frequent epistemic operators are navernoe 'probably', možet byt' 'maybe', možet 'may', and kažetsja 'it seems' within the semantics of uncertainty, and konečno 'of course' within the certainty micro field.

The graduated scale of epistemic modality is reflected in CS in the following way. Each degree (high, middle and low) within both micro fields (certainty and uncertainty) has its own representatives emerging sequentially from high degree to low. So a high degree of uncertainty is expressed by only one marker which is navernoe 'probably' (cf. 2.2). A middle degree is expressed by možet byt' 'maybe' and možet 'may'. Finally, a low level is expressed by kažetsja 'it seems' and požaluj 'very likely'.

F. CDS: Caregiver's Contributions towards Epistemic Marking Development

To elicit some epistemic markers the caregivers may use special linguistic tools such as modus questions. By that we mean such types of caregiver's questions as What do you think / believe / consider etc. …? How do you think / believe / consider ...? Why do you think / believe / consider ... etc.? [29]:

1;9   Adult: Is it raining here? What do you think?
       Child: It was raining.
       Adult: In my opinion, it has stopped. How do you think?
       Child: Walk!

Our observations initially based on numerous dairy data have shown that semantic types of questions appealing to children's mental advancement might be considered as a trigger for the development of their 'subjective sphere' since they play an important role in development of the "model of the psychic" (the theory of mind) [22, 30].

However, modus questions as well as epistemic markers in a dialogue with a child are a certain feature of caregiver's communicative strategy. This is typical for so-called declarative mothers rather than imperative ones. But, in any case, these questions are significant for a child's cognitive and linguistic development since they launch the process of decentration in a child's mentality and speech [17]: the child is no longer focused exclusively on his own viewpoint, but is capable of acknowledging and appreciating another person's point of view. Thus, modus questions facilitate development of a new field in a child's mentality, enhance objectivity of his/her perception of reality and help reflect multiple viewpoints in his / her consciousness.

Significantly, rapid development of symbolic functions of the game falls in the third year of life [18], and after the age of three and a half children begin to understand that people may be mistaken in their suppositions and expectations.

4. Conclusions

Speaking generally, acquisition of epistemic modality in Russian starts with a marked (labelled) member of modal paradigm, which is semantics of uncertainty (cf. [6]). Interestingly, a child needs to express his/ her uncertainty or doubt more than underlining or exaggerating its confidence.

Comparing epistemic marking in CS and ADS reveals that the prototypicality of an operator in a target-system influences its occurrence and further development.

Modal input, being a "mediator" between ADS and CS is important for this process, not only in terms of quality and quantity, but also from the point of view of its special organisation in "adult — child" dialogue.

References

[1] B. MacWhinney, The CHILDES Project: Tools for Analyzing Talk, Mahwah; NJ. 2000.

[2] Aksu-Koc, The Acquisition of Aspect and Modality: the Case of Past Perfect in Turkish, Cambridge: CUP, 1988.

[3] U. Stephany, "Modality in First Language Acquisition: The State of the Art," in Modality in Language Acqusition, N. Dittmar and A. Reich, Eds. Berlin/NY: Walter de Gruyter, 1993, pp. 133–144.

[4] S. Fitneva, "Epistemic Marking and Reliability Judgments: Evidence from Bulgarian," Journal of Pragmatics, vol. 33, pp. 401–420, 2001.

[5] Evidentiality: A Window into Language and Cognitive development, S. Fitneva and T. Matsui, Eds. Wiley Periodicals, 2009.

[6] T. Matsui, "Children's Understanding of Linguistic Expressions of Certainty and Evidentiality," in Pragmatic Development in First Language Acquisition, D. Matthews, Ed. Amsterdam: John Benjamins Publishing Company, 2014, pp. 295–316.

[7] M. Krauze, "Modal'nye markery v reči detej: Stanovlenie funkcij i sistemy," in Detskaya reč kak predmet lingvističeskogo issledovanija, S. N. Cejtlin, Ed. Sankt-Peterburg: Nauka, 2004, pp. 136–139 [in Russian].

[8] A. N. Gvozdev, Voprosy izučenija detskoj reči. Moskva: APN, 1961 [in Russian].

[9] I. V. Stolyarova, "Vyraženie sub'ektivno-modal'nykh otnošenij v detskoj reči," in Detskaja reč: Lingvističeskij aspect, S. N. Cejtlin, Ed. Sankt-Peterburg: Obrazovanije, 1992, pp. 86–95 [in Russian].

[10] V. M. Švec, Sub'ektivnaja (epistemičeskaja) modal'nost' i jejo vyraženie v detskoj reči, in Semantičeskie kategorii v detskoj reči, S. N. Cejtlin, Ed. Sankt-Peterburg: Nestor, 2007, pp. 161–180 [in Russian].

[11] V. V. Kazakovskaya, Questions and Answers in "Adult — Child" Dialogue, Moscow: URSS, 2011 [Vopros i otvet v dialoge "vzroslyj — rebenok"].š Š

[12] V. V. Kazakovskaya, Sposoby vyraženija avtorstva v structure predloženija. Moskva: MGPU, 1996 [in Russian].

[13] Ch. Balli, Obščaja lingvistika i voprosy francuzskogo jazyka, Moskva: URSS, 2001 [in Russian].

[14] V. V. Kazakovskaya, "K voprosu ob ontogeneze ramočnykh sredstv," in Voprosy russkogo jazykoznanija. Grammatika i tekst (K jubileju Galiny Aleksandrovny Zolotovoj), N. K. Onipenko, Ed. Moskva: MGU, 2011, pp. 258–282 [in Russian].

[15] M. M. Bakhtin, Problemy poetiki Dostoevskogo, Moskva: Sov. Rossija, 1979 [in Russian].

[16] V. V. Vinogradov, Russkij jazyk (grammatičeskoe učenie o slove, Moskva: Vysshaja shkola, 1986 [in Russian].

[17] Piaget, Reči myšlenie rebenka, Moskva: Pedagogika-Press, 1994.

[18] S. Vygotskij, Izbrannye psikhologičeskie issledovanija, Moskva: Labirint, 1996 [in Russian].

[19] H. M. Wellman and K. H. Lagatuta, "Developing Understandings of Mind," in Understanding Other Minds: Perspectives from Developmental Cognitive Neuroscience, 2nd ed., Oxford: OUP, 2000, pp. 21–49.

[20] J. H. Flavell, "Development of Children Knowledge about the Mental World", International journal of behavioral development, vol. 24, no. 1, pp. 15–23, 2000.

[21] J. M. Mandler, "Thought before Language," TRENDS in cognitive sciences, vol. 8, no. 11, pp. 508–513, 2004.

[22] V. V. Kazakovskaya, "Functions of Questions in Adult — Child Dialogues (based on Russian)," X IASCL, p. 58–59, July 2005.

[23] Russian Grammar, N. Yu. Švedova, Ed. Moscow: Nauka, 2005 [in Russian].

[24] Teorija funkcional'noj grammatiki: Temporal'nost'. Modal'nost', A. V. Bondarko, Ed. Leningrad: Nauka, 1990 [in Russian].

[25] G. A. Zolotova, "Vvodno-modal'nye slova v predloženii i v tekste," Československa rusistika, vol. 8, no 5, pp. 207–213, 1983 [in Russian].

[26] P. A. Lekant, Očerki po grammatike russkogo jazyka, Moskva: MGOU, 2002 [in Russian].

[27] E. S. Yakovleva, Fragmenty russkoj jazykovoj kartiny mira (modeli prostranstva, vremeni i vosprijatija), Moskva: Gnozis, 1994 [in Russian].

[28] A. Aikhenvald, Evidentiality. Oxford: OUP, 2005.

[29] V. V. Kazakovskaya, "Modusnye voprosy (k probleme izbytočnosti)," Acta Linguistica Petropolitana, vol. 6, pp. 225–242, 2010 [in Russian].

[30] V. V. Kazakovskaya, "What helps children to acquire "the language of ToM"?," Budapest Central European University Conference on Cognitive Development, pp. 123–124, January 2013.