Grammatical Lacunae of the Associative-Verbal Network: Data of Children's Speech and the National Russian Corpus

Krasnoperova Ye.S., Konovalova N.I. *

Ural State Pedagogical University, Yekaterinburg, Russian Federation
Corresponding author. Email: sakralist@mail.ru

ABSTRACT

The article analyzes the process of formation of grammatical competence in early ontogenesis. The methodological basis of the research is the theory of associative-network model of language acquisition developed by Yu. N. Karaulov, and the theory of gradual change of tense grammatical systems in the process of ontogenesis. Children's speech data are studied from the point of view of various disciplines: ontological linguistics, psycholinguistics, and corpus linguistics. Preliminary observations on the grammar patterns used by the child allowed us to put forward a hypothesis about the possible correlation of genetically early grammatical forms in the child's speech and their frequency at subsequent stages of the development of their grammatical system. In order to verify this hypothesis, we identified the most frequent grammatical forms in the speech of children from two to five years old and the role of these forms in the process of constructing the child's own language system. The results obtained are confirmed by statistical analysis using digital data from the National Russian corpus. The conclusions are illustrated by the analysis of the functioning of verb inflectional paradigm units in speech. The authors pay special attention to the "marginal' parts of the paradigm, in particular, the forms of the 2nd person of the present tense of the verb, marked by low frequency in the associative-verbal network of the native speaker. Such forms are considered in the linguopragmatic aspect. The active grammatical minimum revealed in the course of the research can become the basis for developing new methods of teaching Russian as a native and non-native language when determining the "boundaries" of the active grammar range at different stages of the formation and development of grammatical competence.

Keywords: ontogenesis, associative grammar, the National Russian corpus, digitalization, interdisciplinarity

1. INTRODUCTION

The problem of studying the universal mechanisms of language acquisition and individual strategies of its native speaker in the course of assigning and using (both native and non-native) language is one of the most relevant ones in modern ontological linguistics and psycholinguistics. Attempts of researchers to explain the processes of speech perception and speech generation hidden from direct observation, clustering and storage of language knowledge are presented in various models, among which the most operational model is the associative-verbal network (hereinafter ABC), [Karaulov 1993, 2006, Karaulov, Ruzhitsky 2015]. The peculiarity of ABC is that "... grammatical forms and meanings are stored in models (in separate patterns), lexicalized (in the form of specific words and word forms) and dissipated, i.e. individual elements of the set (paradigms, declensions of the conjugation) are distributed between different patterns and between different associative fields" [Karaulov 1993: 6-7].

These features are reflected both in the distribution of components of the inflectional paradigm by ABC, their different frequency, and in the lacunarity of some paradigms - the absence of separate grammatical forms in the concordance of associative fields represented in dictionaries.

2. PROBLEM STATEMENT

The uneven distribution of grammatical forms by ABC can be considered as an argument in favor of the fact that more frequent forms that are most in demand in speech activity should be learned by children (extracted from input) as a matter of priority.

Systematic analysis of the grammatical component of language ability shows that by about three years of age, almost all core means of expressing morphological categories are represented in the child's speech. Such a rapid development of the grammar of the native language

is explained by generativists by the presence of some ready-made forms in the child at birth, which are only modified later, adapting to the language of the people around the child. These forms are called "universal grammar". Representatives of another, the system-centric approach (N. I. Lepskaya, D. Slobin, S. N. Tseitlin, K. I. Chukovsky) believe that the language system is formed on the basis of processing speech input, independently performed by a child in the process of natural communication with adults. It is noted that "... the absolute discovery of the system-centric direction in the study of children's speech was that the language is learned by the child as an operational mechanism, and not just as a set of its implementations" [Gridina 2012: 5-6]. This approach explains the gradual dynamics of the child's language ability by changing the forms and meanings that are formed in their minds during ontogenetic development of "... particular dynamic systems of forms and meanings that allow the child to satisfy their communicative and cognitive needs in conditions of nominative deficit" [Gridina 2012: 6]. Cf. the idea of the possibility of observing the change of language systems at each stage of ontogenetic development: "... language is not given to a child at birth ..., but they are given a unique ability to independently build in their own minds all the components of the language system (phonological, morphological, syntactic, lexical, and a number of others), as well as to rebuild and improve this system throughout life when interacting with other people" [Tseitlin 2009: 14]. The child deduces its own rules for constructing and connecting forms based on the input. The grammatical system that the child creates is individual, but as it becomes more complex, it approaches the average language system of an adult native speaker. If children first learn most general rules, which is reflected, in particular, in the mechanisms of supergeneralization ("matching" the used forms of different words to the one that has already been mastered), then there is a tendency to gradually differentiate them, to nuance the forms of expression of grammatical meanings, to acquire exceptions. Cf. in this regard, observations on children's speech, where "... the principle of symmetry prevails, which in the language of "adults" is often violated by a bizarre and illogical norm" [Gridina 2016: 6].

To describe the language system, arrayed in the mind of the child, refer to the model of language proficiency proposed by Yu. N. Karaulov, cf.: "Rules of inflection, word compounding and word-formation, i.e. grammar, which is used by instantaneous native speakers, is completely lexicalized, tied to individual lexical items, as if distributed between them and wholly "spread out", spilt through the associative-verbal network" [Karaulov 1993: 6-7].

This method of presenting grammar in the mind of a native speaker in an internalized form that does not coincide with the results of its linguistic description is called associative grammar: "Associative grammar is a grammar enclosed in the associative-verbal network of a native speaker, a grammar of speech activity, intentions, tendencies and readiness of the speaker, a grammar "in its live, ready-to-use form" [Karaulov 1993: 7].

3. RESEARCH QUESTIONS

The main question that can be solved in this study is to identify the sequence of acquiring grammatical forms for expressing grammatical meanings (based on ontolinguists' published diary entries and longitudinal observations on children's speech) and further statistical comparison with the materials of the National Russian corpus (RNC), which will allow us to speak about the specific vs universal constituents in the process of forming grammatical competence.

According to observations of A. N. Gvozdev, the first in the grammatical paradigm of the noun in the speech of the child from 1year 10 months to two years of age are forms of Nominative Case singular and plural, Accusative, then Genitive and Locative start to be formed. As for Instrumental and Dative, they are absorbed last.

From the paradigmatic of the adjective in the child's speech, the forms of Nominative masculine and feminine are acquired first, but without agreement with nouns; the neuter form is learned last.

Verbal paradigmatic is initially presented in children's speech using the forms of the imperative (2nd person singular) and 3rd person singular of the indicative mood, then we note the development of the forms of the 1st person singular and, finally, forms of the 1st and 3rd person plural. The absence of the forms of the 2nd person plural in the child's speech at this age is logical: the operation of abstraction on behalf of a specific participant of the communicative situation is not yet formed.

4. PURPOSE OF THE STUDY

Defining forms of grammatical categories of noun, adjective and verb as genetically earlier and more relevant in early ontogeny, and in little or no demand on the child's behalf in a language system designed by them, requires further investigation with the involvement of digital data.

The purpose of the research is to assume that the genetically primary grammatical forms that a child learns remain the most frequent at the further stages of speech development, form the core of the language system that a child constructs, and in the future bear the main functional load in their speech activity.

5. METHODS OF RESEARCH

Taking into account the frequency of occurrence of specific grammatical forms of nouns, adjectives and verbs in children's speech, the child's grammar-range invariant was modeled during early ontogenesis. Objectification of fragments of ABC of a native speaker was carried out by the method of word usage analysis. As the material for the

initial study, we used diary entries (see the list of references) of the speech of children from two to five years old, as well as data from the RNC (the subcorpus with the disambiguated grammatical homonymy). This subcorpus of the RNC includes more than 6 million words out of about 280 million of the entire corpus. To collect data, we used a lexical-grammatical search with the specified grammatical and lexical-semantic features.

6. THE RESULTS OF THE STUDY

Analysis of children's speech data generally correlates with the hierarchy of grammatical forms of nouns, verbs, and adjectives by their frequency in ABC. To confirm the obtained results, we present the data of the RNC with the grammatical homonymy disambiguated

(http://www.ruscorpora.ru). Corpus linguistics provides researchers with new opportunities to analyze the material [Ventsov, Grudeva 2009, Boriskina 2009, Viimaranta 2014, Lyashevskaya 2016]. Digitalization of linguistic research methods allows for extensive, including interdisciplinary, research and verification of research results on a wide range of materials.

6.1. In the paradigm of the noun, the most frequent and most popular forms of speech are nominative, accusative and genitive cases. Forms of the dative case are not just the last to be mastered, but also remain the least popular.

These same correspondences are preserved in spoken speech. To confirm this, we present data on the use of these forms in the subcorpus of oral speech (with the grammatical homonymy disambiguated), which is a transcript of recordings of spoken speech (table 1).

Table 1 Frequency of use of case forms of nouns in oral speech

Grammatical form	Number of inputs	% of the total number of uses*	Grammatical form	Number of inputs	% of the total number of uses*
Nominative case, singular (S, nom, sg)	14234	6,90	Nominative case, plural (S, nom, pl)	3354	1,63
Genitive, singular (S, gen, sg)	6541	3,17	Genitive, plural (S, gen, pl)	3227	1,57
Accusative, singular (S, acc, sg)	6235	3,02	Accusative, plural (S, acc, pl)	2067	1,00
Locative case, singular (S, loc, sg)	4057	1,97	Locative case, plural (S, loc, pl)	720	0,35
Instrumental case, singular (S, ins, sg)	1979	0,96	Instrumental, plural (S, ins, pl)	640	0,31
Dative case, singular (S, dat, sg)	1647	0,80	Dative, plural (S, dat, pl)	540	0,26

* the composition of the corpus with disambiguated grammatical homonymy - 205 994 words.

The predominance of the nominative case is absolutely predictable. In the corpus, it is presented primarily in its central meanings: subject, determinative, and nominative (nominative themes) [Russian grammar 1980].

Let's consider in more detail the cases of using the dative case, since it turns out to be the least frequent both in the singular and plural. The main meanings of the dative case: the addressee and the subject of perception. The low frequency of use of dative forms can be explained by the fact that pointing to the addressee of the action (the main function of the case form) is in demand in a limited number of speech situations and grammatical constructions. And in peripheral functions, this case form competes with more frequent forms of accusative and locative cases, for which the expression of a spatial or temporal meaning is a core function.

6.2. Adjectives in the child's speech also appear very early and the forms of gender, number, and case of adjectives are mastered without much difficulty. Possessive adjectives are an exception. Naming a characteristic through the relation of belonging to a certain person, these adjectives form a small and practically non-replenished group of words with their own special declension system. Forms of possessive adjectives are rarely used in colloquial speech and experience constant competition from synonymous constructions with indirect cases.

Let's turn to the data of the RNC. Table 2 shows the distribution of the general use of adjectives by category.

Table 2 Categories of adjectives

Lexico-grammatical class	Number of inputs	% of the total number of uses*
Qualitative adjectives	207691	3,46
Relative adjectives	271390	4,52
Possessive adjectives	4041	0,07

* the composition of the corpus with disambiguated grammatical homonymy - 6 003 397 words.

Relative adjectives predominate over quantitative ones, and not many forms of possessive adjectives in some contexts are used in a meaning close to that of relative or qualitative adjectives.

Considering the dependence of the use of adjectives of different categories on the sphere of functioning of the text, we can note several features.

In the literary text, the use of qualitative and relative adjectives is observed in approximately equal proportions, the share of possessive adjectives is still insignificant.

In the sphere of oral speech and electronic communication, close to oral speech, the proportion of possessive adjectives is reduced by half, while maintaining the frequency of use of qualitative and relative adjectives.

Possessive adjectives in texts of this sphere are more often replaced by case constructions, and the forms of possessive adjectives themselves are often used figuratively to enhance expressiveness.

A decrease in the proportion of possessive adjectives is observed in official business, educational, and scientific texts. Pointing to a material, relation to another place, time, or feature is in demand as a way to establish various types of relationships, which explains the increase in the proportion of relative adjectives.

The sparseness of the category of possessive adjectives and its closeness lead to the fact that the grammatical semantics of belonging characteristic of these forms is blurred. Forms of possessive adjectives begin to be used in the wrong meaning, which makes it even more difficult for children to learn them.

6.3. On the example of verb forms, it can be noted that according to associative dictionaries, the 2nd person of the indicative mood is rarely represented in the network, which is quite correlated with the data of children's speech. In the process of learning the native language for children from two to five years of age, the forms of the 2nd person indicative mood of verbs are irrelevant. We present the frequency of use of forms of the 2nd person indicative mood of the present tense of the verb in relation to the total number of word uses in table 3.

Table 3 Frequency of use of finite verb forms of the present tense

Grammatical form	Number of inputs	% of the total number of uses*
1st person (V, indic,praes,1p)	32 009	0,53
2nd person (V, indic,praes,2p)	15 054	0,25
3rd person (V, indic,praes,3p	145 711	2,43

* the composition of the corpus with disambiguated grammatical homonymy - 6 003 397 words.

The predominance of the 3rd person forms of the verb over the other finite forms in ABC is confirmed by the data of the RNC on the example of finite verb forms of the present tense.

The use of present-tense 2nd person forms in texts related to different functional areas is shown in table 4.

Table 4 Frequency of use of forms of the 2nd person of the present time

Subcorpus (functional sphere)	Capasity of the subcorpus (number of words)	Number of inputs	% of the total number of uses
Electronic communication	120294	429	0,36
Literary text	3557693	12189	0,34
Church and theological	63242	134	0,21
Everyday	67225	116	0,17
Publicism	1546367	1969	0,13
Advertising	24296	28	0,12
Official and business	74216	27	0,04
Educational and scientific	546513	162	0,03

The lower demand for forms of the 2nd person of the present tense of the verb, noted in ABC, is confirmed by the data provided in the RNC. The function of maintaining communication (or its illusion) determines for the speaker

the choice of these forms when generating a text of a certain functional sphere.

Thus, the hypothesis about the correlation of genetically early grammatical forms and their high frequency in different speech practices of an adult native carrier of language consciousness was confirmed by the analysis of materials from associative dictionaries and the National Russian corpus.

Psycholinguistic approach to the study of the process of language acquisition and use can contribute to the creation of new methods of teaching the Russian language. In particular, the established set of genetically basic grammatical forms that the child learns, and the forms that they do not need at the initial stage allows them to:

- identify the most frequent grammatical forms that make up the core of the language system, which bear the main speech load at the further stages of speech development;

- create the basis of active grammar minimum for learning the Russian language on the basis of the identified forms.

REFERENCES

[1] Viimaranta, J. (2014), "The meaning and use of podportit' on the basis of the Russian National Corpus" ["Znachenie i upotreblenie glagola podportit' na osnove Natsionalnogo korpusa russkogo yazyka"], Russkoye yazykoznaniye, No. 38, pp. 63-87.

[2] Boriskina, O. O. (2009), "Corpus vs. psycholinguistics: confrontation or cooperation", Series: Linguistics and intercultural communication ["Korpusnaya vs. psikholingvistika: konfrontatsiya ili kooperatsiya"], Vestnik VGU. Ser.: Lingvistika i mezhkulturnaya kommunikatsiya, No. 2, pp. 34-38.

[3] Ventsov, A.V., Grudeva, E.V. (2009), "On the Corpus of Standard Written Russian" ["O korpuse russkogo literaturnogo yazyka"], Russkoe yazykoznanie, No. 33, pp. 195-209.

[4] Gridina, T.A. (2016), "Your own game": a child in the world of language ["Svoya igra": rebenok v mire yazyka], Ural State Pedagogical University, Yekaterinburg, 177 p.

[5] Gridina, T. A. (2012), "Logic of the language paradox in children's speech" ["Logika yazykovogo paradoksa v detskoy rechi"], Lingvistika kreativa-2, Ural State Pedagogical University, Yekaterinburg, pp. 5-33.

[6] Gvozdev, A. N. (2007), Questions of studying children's speech [Voprosy izucheniya detskoy rechi], Detstvo-Press, SPb, 470 p.

[7] Gvozdev, A.N. (2005), From the first words to the first class. Journal of scientific observations [Ot pervykh slov do pervogo klassa. Dnevnik nauchnykh nablyudeniy], KoMKniga, Moscow, 323 p.

[8] Karaulov, Yu. N. (2006), Dictionary of Pushkin and the evolution of Russian language ability [Slovar' Pushkina i evolyutsiya russkyi yazykovoy sposobnosti], KoMKniga, Moscow, 167 p.

[9] Karaulov, Yu. N. (2010), Associative grammar of the Russian language [Assotsiativnaya grammatika russkogo yazyka], Izd-vo LKI, Moscow, 328 p.

[10] Karaulov, Yu.N., Ruzhitsky, I. V. (2015), "From the dictionary of the writer's language to the knowledge of his world: some basic concepts in the work of F. M. Dostoevsky" ["Ot slovarya yazyka pisatelya k poznaniyu yego mira: o nekotorykh bazovykh kontseptah v tvorchestve F. M. Dostoevskogo"], Voprosy Kognitivnoy lingvistiki, No. 4 (45), pp. 16-22.

[11] Lyashevskaya, O. N. (2016), Corpus tools in grammatical research of the Russian language [Korpusniye instrumenty v grammaticheskih issledovaniyah russkogo yazyka], YASK, Moscow, 518 p.

[12] Shvedova, N. Yu. (ed.) (1980), Russian grammar [Russkaya grammatika], Nauka, Moscow, Vol. 1, 717 p.

[13] Kharchenko, V.K. (1994), Dictionary of children's speech [Slovar' detskoy rechi], BGPI, Belgorod, 254 p.

[14] Kharchenko, V.K. (2012), The Corpus of the children's saying [Korpus detskih vyskazyvaniy], Izd-vo Literaturnogo instituta im. A. M. Gor'kogo, Moscow, 520 p.

[15] Tseitlin, S.N. (ed.) (2001), Dictionary of children's word-forming innovations [Slovar' detskih slovoobrazovatelnyh innovatsiy], Verlag Otto Sagner, Munchen, 202 p.

[16] Tseitlin, S.N. (2009), Essays on word formation and form formation in children's speech [Ocherki po slovoobrazovaniyu i formoobrazovaniyu v detskoy rechi], Znak, Moscow, 592 p.