

# Early Stages of the Acquisition of Verbal Grammar by Russian-speaking 1-to-3-year-old Children (Based on the CHILDES Corpus)

Valeriya Lelik, Anastasiya Lopukhina, Irina Korkina

## Introduction

The purpose of the present study is to investigate the timeline of children's acquisition of verbal grammatical categories during the first three years of life. We analyze longitudinal data of two children: a boy Y., whose speech was recorded from 1 year 5 months to 3 years and a girl T., whose speech was recorded from 10 months to 3 years. The novelty of our study consists, firstly, in the description of two new cases of longitudinal recordings of children's speech that have not been described before. Secondly, we pay attention to the input and the relationship between usage of different grammatical categories in the input and the output.

## Methods

The data was video-recorded and transcribed according to the principles of the CHILDES corpus. Then the data was transliterated into Cyrillic using Python code and morphologically parsed using the pymystem3 module. The homonymy (for Y. corpus) was removed manually.

For each verb after the morphological parsing we received its infinitive, aspect (perfect or imperfect), tense (past, present or future), mood (indicative or imperative), person (1, 2 or 3) and number (if exist).

### Total volume of word forms for Y.:

Output: 24270 (2957 verbs)  
Input: 32914 (6228 verbs)

### Periods for Y.:

- 1) 1 year 5 months – 1 year 10 months
- 2) 1 year 11 months – 2 years 4 months
- 3) 2 years 5 months – 3 years

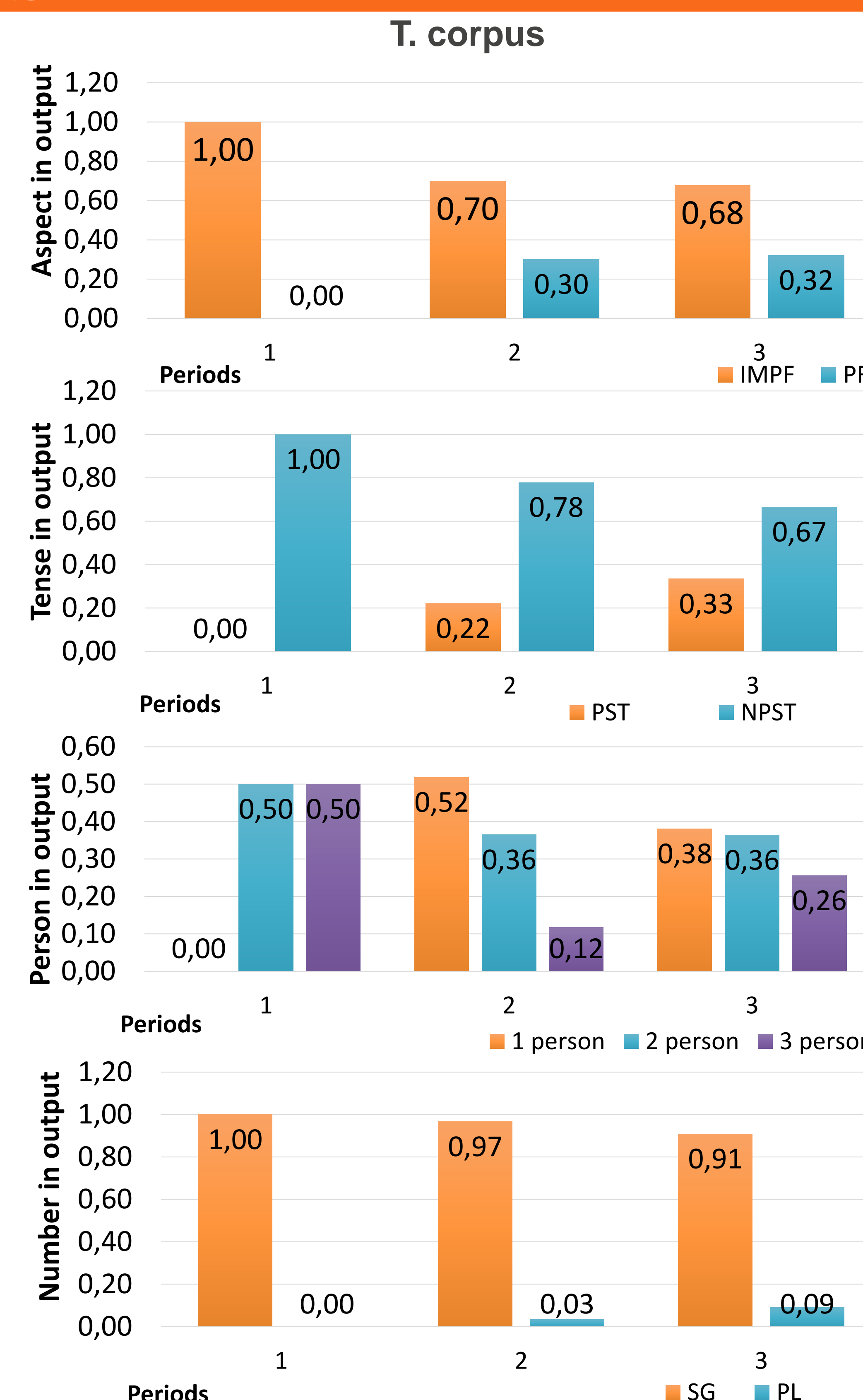
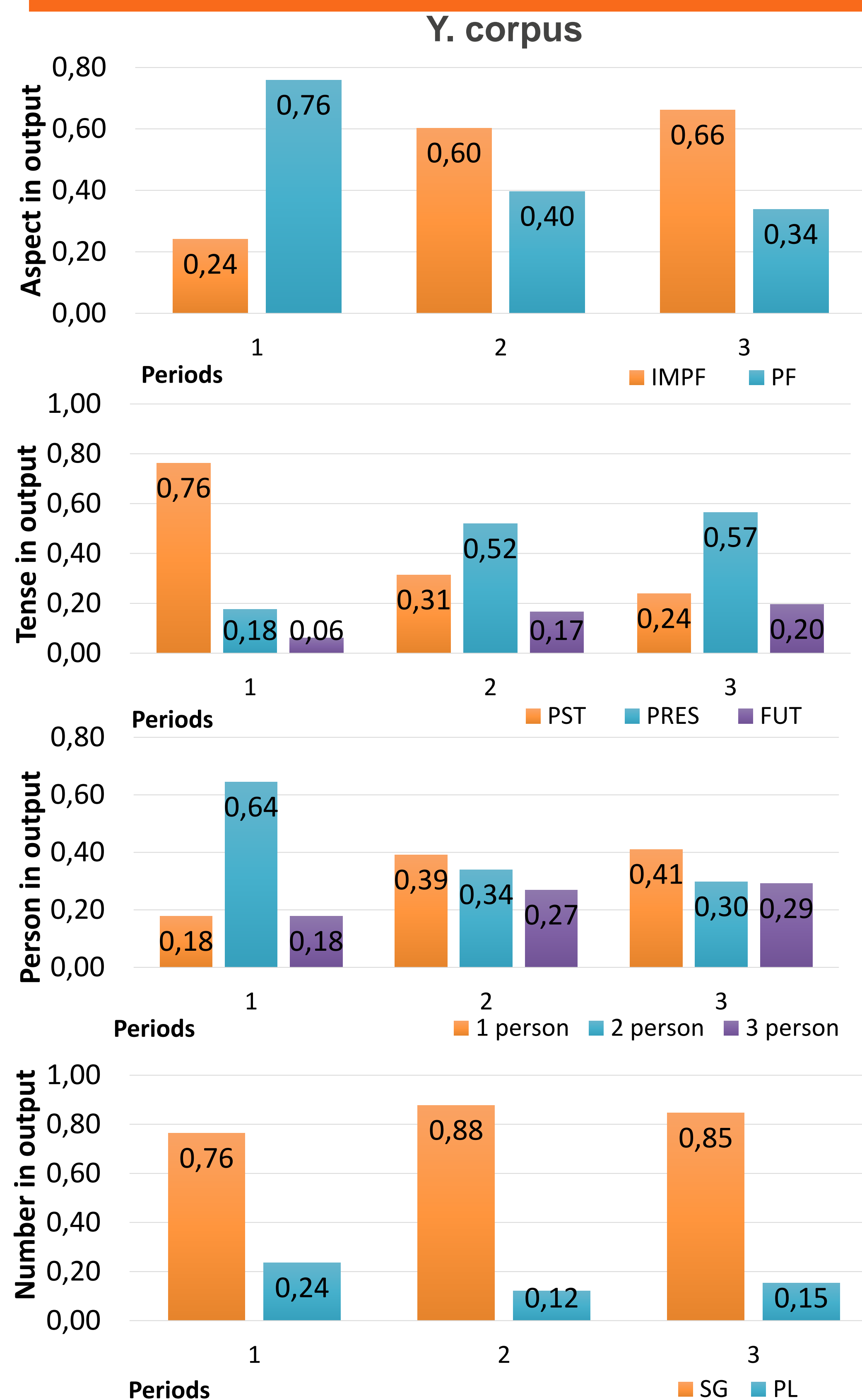
### Total volume of word forms for T.:

Output: 9377 (1384 verbs)  
Input: 27875 (5389 verbs)

### Periods for T.:

- 1) 10 months – 1 year 1 month
- 2) 2 years 4 months – 2 years 6 months
- 3) 2 years 7 months – 2 years 9 months

## Results



## Infinitives and imperatives

### Y. corpus

	1 period		2 period		3 period	
	Infinitives (%)	Imperatives (%)	Infinitives (%)	Imperatives (%)	Infinitives (%)	Imperatives (%)
output	8	28	15,8	26,1	18,3	18,4
input	18,1	30,5	19,3	28,4	15,3	28,1

### T. corpus

	1 period		2 period		3 period	
	Infinitives (%)	Imperatives (%)	Infinitives (%)	Imperatives (%)	Infinitives (%)	Imperatives (%)
output	0	33	12,4	20,2	17,6	17,1
input	12,4	45,7	17,1	24,4	35,4	19,6

## Conclusions

- (1) Aspect: the perfect verbs were more difficult for both children to produce, during most of the periods in each of the corpora, imperfect verbs prevailed over perfect verbs, but during the first period in Y. corpus a predominance of perfect verbs over imperfect verbs was observed. It is in line with previous literature.
- (2) Tense: during the first period in Y. corpus the verbs in past tense are most common, on the other two stages – the verbs in the present tense are most frequent. The future is the least common in all the periods. For T. corpus it is impossible to make reliable conclusions about the production of tense because of the lack of manual parsing. The point from the previous literature that the future tense will be the most difficult to produce has been confirmed, we suppose that it can also mean the latest acquisition.
- (3) Personal forms: the order of production of persons for Y. is: 1, 3, 2, which is not compatible with the following order of acquisition of persons: 3, 1, 2 from literature. For T. the order is also 1, 3, 2. However, the results confirm the point that 2 person is the most complex to produce for children.
- (4) Number: in all the periods in both corpora, there are more verbs in singular form than verbs in plural, which may indicate that the plural form is learned later, which is in line with the previous literature and can possibly mean the later acquisition of plural form than the singular one.
- (5) The point from the literature about a big number of infinitives and imperatives in children's speech was confirmed. In the Y. corpus, the number of imperatives in the output is gradually reduced by the third period, and the same is observed in the T. corpus. The number of infinitives in both corpora has increased. The point about a gradual reduction of number of imperatives and infinitives in children's speech was partly confirmed.
- (6) In both corpora input remains comparatively unchanged. According to the most frequent word forms, we can talk about its influence on output: a lot of lexemes from the top-15 verbal word forms are similar for input and output. In addition, a big number of infinitives and imperatives in input is reflected in both children's corpora.