



Parental Report Evidence for Toddlers' Grammar and Vocabulary in Bulgarian

Journal:	<i>First Language</i>
Manuscript ID:	Draft
Manuscript Type:	Original Manuscript
Keywords:	Bulgarian, language, development, MacArthur Bates, CDI
Abstract:	<p>This is the first study of a large sample of Bulgarian-speaking toddlers and their communicative development. A Bulgarian adaptation of the MacArthur Bates Communicative Development Inventory Words and Sentences was used to obtain parent report data on one hundred and fifty-three children between the ages of 20 and 30 months. Lexical and grammatical development were examined as a function of age, gender, and maternal education. While we found evidence for both lexical and grammatical developmental changes with age even in this age-restricted sample, gender had little effect on developmental levels, especially on the grammar components of the instrument. Maternal education was a significant predictor of children's developmental achievements and there was considerable individual variability. Lexical and grammatical development were found to be correlated significantly.</p>

SCHOLARONE™
Manuscripts

Abstract

This is the first study of a large sample of Bulgarian-speaking toddlers and their communicative development. A Bulgarian adaptation of the MacArthur Bates Communicative Development Inventory Words and Sentences was used to obtain parent report data on one hundred and fifty-three children between the ages of 20 and 30 months. Lexical and grammatical development were examined as a function of age, gender, and maternal education. While we found evidence for both lexical and grammatical developmental changes with age even in this age-restricted sample, gender had little effect on developmental levels, especially on the grammar components of the instrument. Maternal education was a significant predictor of children's developmental achievements and there was considerable individual variability. Lexical and grammatical development were found to be correlated significantly.

Keywords

Language acquisition, Bulgarian, MacArthur Bates Communicative Development Inventory

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For Peer Review

Introduction

Language acquisition is an important facet of the cognitive and social development of individuals; it contributes to social integration, professional interaction and growth, emotional and behavioral stability. Research in this field is essential in the pursuit of several goals: broadening opportunities for individuals and societies to contribute and benefit from economic and social transformation, dealing with emerging developmental risks, as well as assessing the results of educational policy.

The timely acquisition of the language of one's community is of particular importance for the further social, academic, and professional development of individuals. That is why much research has been invested in the area in academic communities studying languages such as English, German, Spanish, etc. Less is known, however, about the processes, stages, and mechanisms of communicative development in the languages of Central and Eastern Europe. This disparity needs to be overcome for reasons beyond pure academic interest. The pace with which children acquire language can be indicative of a number of social, pragmatic, cognitive, and academic problems that emerge at a later stage of development (Clegg, Hollis, Mawhood, & Rutter, 2005; McCormack, McLeod, McAllister, & Harrison, 2009). Therefore, early detection and prevention of communicative problems can make a significant change in the lives of individual children and their communities as a whole.

Whereas evidence-based research in the field of communicative development has been in the mainstream of academic endeavor in many developed countries, a clear research gap exists in this respect in the case of Bulgarian. Previous work on early language acquisition for Bulgarian

1
2
3 has relied on a handful of time-limited longitudinal corpora of children, such as, for example,
4 tracing the development of 5 children in child-caregiver interactions in the work of Stoyanova
5 (2011), due to lack of a reliable instrument to examine and assess language and general
6 communicative development that is adapted to the specific characteristics of the Bulgarian
7 language and cultural context.
8
9

10
11
12
13
14
15
16 Many countries faced with a similar problem have found it useful to work on an
17 adaptation of a well-known instrument of research and assessment of communicative
18 development in early childhood first launched in the U.S. in the 1990s called the MacArthur
19 Bates Communicative Development Inventories (CDIs). MBCDI-1 is designed for work with
20 infants in the 8 to 16 month age bracket, and MBCDI-2 is designed for work with toddlers in the
21 16 to 30 month age bracket (Fenson et al., 1994). By now adaptations to the local language and
22 communicative features have been carried out for a large number of languages. These
23 instruments have proven to have high validity and reliability and have become a prerequisite for
24 conducting basic and applied research, including detection of delayed development and its
25 consequences, the design of clinical assessment and interventions, etc.
26
27
28
29
30
31
32
33
34
35
36
37
38
39

40 The Communicative Development Inventories (CDIs) are parent report measures of
41 vocabulary, grammar, and other aspects of communicative development in very young children
42 (Fenson et al., 1994). They comprise the CDI Infant form (Words and Gestures: CDI-1) and the
43 CDI Toddler form (Words and Sentences: CDI-2). The CDI-2 is designed to measure language
44 production in 16-30 month olds and it consists of a 680-word vocabulary production checklist, as
45 well as measures of morphological and syntactic development, including utterance length and
46 grammar complexity.
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

There have been over 40 adaptations of the CDIs for use with different languages worldwide (see http://www.sci.sdsu.edu/cdi/adaptations_ol.htm). In the last several years, work on the adaptation and piloting of a Bulgarian version of the MBCDI resulted in a revised version currently in use for data collection and norming with a view to producing a reliable research and assessment tool.

Here we present a pioneering study of early language acquisition in Bulgarian based on data for a large sample. In order to assess the language development of toddlers, we used an adaptation for Bulgarian of the MacArthur Bates Communicative Inventories, or CDI (Fenson et al., 1994), and in particular the CDI: Words and Sentences scale which is suited to the age under examination here. The Bulgarian CDIs have been constructed so as to adhere to the same principles and main categories of analysis as those originally included in the US version of the MBCDI (Fenson et al., 1994). Culturally and linguistically different adaptations have striven to maintain the general structure and content of the original US version in order to allow for more comparability. The same applied in the case of Bulgarian. Thus, the Bulgarian CDI: Words and Sentences includes two parts. Part 1 consists of a checklist of 637 vocabulary items distributed into 22 categories a section on Past and Future activities and Absent Things and People. Part 2 examines grammatical development in several ways. It includes measures for: (a) the use of noun and verb inflections (grammatical affixation), (b) grammatical complexity of phrases as a binary choice, (c) assessment of mean length of utterance on the basis of the three longest child produced sentences as provided by parents, (d) grammatical complexity on a graded assessment scale, etc.

Bulgarian as a language belongs to the South branch of Slavic and shares many features with other Slavic languages. However, it has also been influenced by neighboring languages

1
2
3 throughout centuries of contact in the so-called Balkan Sprachbund both in terms of grammatical
4 system and lexicon. Unlike other Slavic languages, for example, Bulgarian has lost its nominal
5 case system almost entirely, while at the same time allowing for relatively free word order. In the
6 context of the adaptation of the CDIs for Bulgarian, several critical decisions had to be made
7 regarding the most common citation form to be used for the items listed in the questionnaires.
8 Grammatical gender variation for adjectives had to be eliminated and the Neuter gender forms
9 were used throughout. There are no verb infinitives in Bulgarian, and much consideration went
10 into choosing the best citation form for the CDIs such that would be frequent in mother-child
11 interactions, thus resulting in the use of 1st and 3rd person singular forms for verbs, e.g.,
12 *искам/иска*. After careful consideration and piloting a longer form earlier, verb aspect was also
13 eliminated as a variable of verb forms. The over-generalization of syntax rules section was not
14 extensive because little reliable information exists on typical over-generalizations by Bulgarian
15 children with the exception of an unpublished longitudinal corpus of child language
16 development. In fact, the most challenging part of the adaptation efforts was the construction of
17 the grammar assessment components in Part 2 of the Toddlers' CDI. Otherwise, the adaptation
18 for Bulgarian preserved the structure of the assessment instrument. For example, in Part 1, we
19 had the same number of vocabulary categories, although the exact number in each category and
20 the semantic content of individual items different to a small extent. Culture-specificity and
21 similar considerations led to the replacement of items by culturally appropriate materials.
22 General information on the child, parents and child's environment was also collected.
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54

55 METHOD

56
57
58
59
60

Participants

The sample included parent reports on the language development of 153 toddlers aged from 20 to 30 months. Their mean age was 27.59 months ($SD = 2.39$). Information was provided by the mothers of these children. All were learners of Bulgarian as a mother tongue. The children were closely balanced by gender – there were 78 girls and 75 boys. Maternal education fell into one of two categories: high school graduates ($n = 50$) and mothers who either had or were in the process of obtaining a university degree ($n = 101$). Maternal education level for two of the children was not available. This distribution is not nationally representative; it is closer to the profile of the young urban-dwelling mother.

Procedure

The toddlers' mothers were recruited via preschools, pediatricians, and social contacts. They filled out the parent report questionnaires on a voluntary basis and on printed handouts. Data were subsequently entered in digital form.

Results

Before proceeding to the results of the statistical analyses, four children's data were excluded from the analyses because the estimates in the parent report for their vocabulary size placed them in the category of outliers from the overall distribution. They were at the ages of 24 ($n = 2$) and 30 months ($n = 2$) and were reported to have expressive vocabulary between 0 and 4 words.

Lexical Development

In this section, we look at the development of expressive vocabulary in toddlers aged 20 to 30 months by examining the size of their lexicon as attested by parent report with respect to their age, gender, and maternal education level. On average, parents reported that their children produced a good amount of words at this stage of their development, $M = 315.88$, $SD = 159.74$.

The average number is comparable with previous findings on languages such as American English, Italian, Slovenian and others. Around their second birthday, children produced on average 208 words ($SD = 39$), and at two and a half years of age children produced 392 words ($SD = 25$). Significant individual variation was observed at this stage of toddlers' language development in Bulgarian as well as it has been previously in many other languages and cultures.

There was a ten-fold difference between the child with the lowest vocabulary count ($n = 61$) and the child with the highest vocabulary count ($n = 618$) at the age of 30 months, for example.

Overall vocabulary was found to increase with age even in this rather age-restricted sample (see Figure 1). There was a significant positive partial correlation between age and total vocabulary score after controlling for gender and maternal education, $r = .50$, $p < .001$.

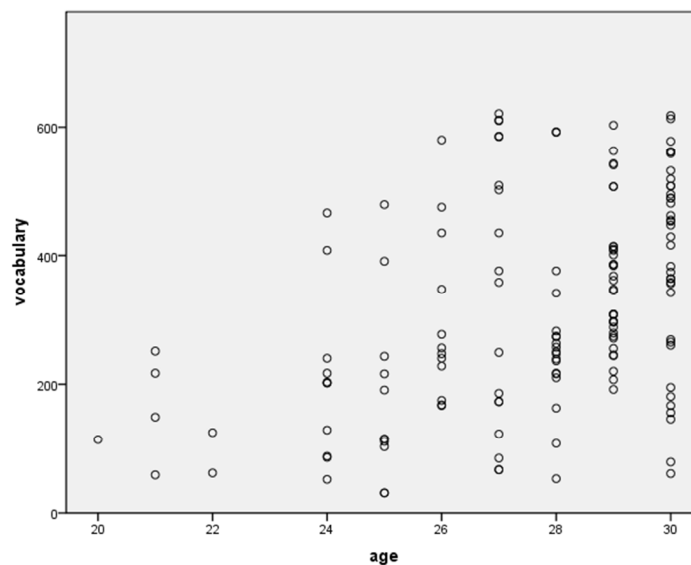


Fig.1. Growth of expressive vocabulary (N words) with age (in months) for toddlers.

The analyses of partial correlations between age and individual vocabulary categories after controlling for gender and maternal education revealed that vocabulary growth was observed in all of the 22 categories with one exception (*animal sounds*). The correlations were positive, moderate, and ranged from $r = .34$ (*connecting words*) to $r = .54$ (*places*); all correlations were significant at $p < .001$.

A t-test for independent samples revealed no gender difference for total vocabulary score. Boys ($n = 71$) produced 295 words on average, $SD = 158$, and girls ($n = 78$) produced 335 words on average, $SD = 159$. At the level of individual vocabulary categories, no significant gender differences were found, either (significance was set at the more conservative level of $p < .01$). Gender differences on vocabulary were examined in a series of t-tests for the individual month-age categories of the children in the sample for which sufficient numbers of observations were available per age/gender combination. At the age of 27 gender differences in favor of girls

1
2
3 emerged on multiple vocabulary categories and on the overall number of words produced. With
4
5 that exception, there were no gender differences at any of the other ages (24, 25, 26, 28, 29, 30
6
7 months).
8
9

10
11 Next, a t-test for independent samples looked at vocabulary scores as a function of
12
13 maternal education. Maternal education in this sample was either high school ($n = 48$) or
14
15 university level ($n = 99$). There was a significant effect of maternal education on expressive
16
17 vocabulary, $t(145) = 3.89, p < .001$. The children of mothers with higher education were
18
19 reported to be producing more words ($M = 352, SD = 162$) than the children of mothers with
20
21 high school diploma ($M = 248, SD = 129$).
22
23
24
25

26 Including all three demographic variables in a combined regression analysis gives a
27
28 clearer understanding of the relative contribution of each to the variability found in toddlers'
29
30 vocabulary size. A standard multiple regression analysis found that they accounted for a
31
32 considerable amount of this variability, $R^2 = .34$. All three variables made a significant unique
33
34 contribution to the prediction of the dependent variable vocabulary size. The largest unique
35
36 contribution was made by age ($\beta = .48$). Maternal education level also contributed to the
37
38 variance in vocabulary size ($\beta = .36$), and gender had a modest unique contribution ($\beta =$
39
40 $.21$). Thus, 23% of the variability was attributed to *age* differences, a result that emphasizes the
41
42 rapid growth during the 20 to 30 month period of toddlers' age. Maternal education accounted
43
44 for 13% of the unique variance in this measure.
45
46
47
48
49
50
51
52
53

54 *Grammatical Development*
55
56
57
58
59
60

1
2
3 The acquisition of Bulgarian grammar by toddlers was assessed by means of several measures.
4
5 The most important are: (a) two complexity measures –one was a binary choice as in the
6
7 MacArthur Bates CDI, and the other was a set of four items with a graded complexity response,
8
9
10 (b) the use of common grammatical affixes, and (c) mean length of utterance (MLU). These
11
12 variables were significantly positively, and moderately to highly correlated, with correlation
13
14 coefficients ranging from .42 to .66, the latter for the two measures of complexity used. This
15
16 pattern of results reveals an underlying common set of related grammatical knowledge in
17
18 toddlers at this age.
19
20
21

22
23 The analyses of partial correlations between age and grammatical development measures
24
25 after controlling for gender and maternal education revealed that grammar development was
26
27 underway for all measures used here. The correlations with child age were significant, positive,
28
29 though relatively weak, ranging from $r = .23$ (*complexity binary choice*) to $r = .39$ (*use of*
30
31 *grammatical affixes*).
32
33
34
35

36 The next round of analyses included all three demographic variables in a series of
37
38 multiple regression analyses for each of the grammar measures.
39
40
41 *Grammatical affixes*. A standard multiple regression analysis found that age, maternal education
42
43 and gender accounted for a modest amount of variability in the use of grammatical affixation, R^2
44
45 = .18. All three variables made a significant, if somewhat modest, unique contribution to the
46
47 prediction of the dependent variable. The largest unique contribution was made by age (beta =
48
49 .39). Maternal education level also contributed (beta = .15), and so did gender (beta = .19). Here
50
51 15% of the variability was accounted for by *age* differences, once again revealing significant
52
53 growth in the use of morphology during the 20 to 30 month period. Maternal education
54
55
56
57
58
59
60

1
2
3 accounted for 2% of the unique variance in this measure, and gender contributed 4% to the
4
5 variance.
6
7

8
9 *Mean length of utterance (MLU)*. Out of the three demographic factors that predicted MLU as
10 well, $R^2 = .10$, it was only age that was a significant contributor by 7% to MLU variance (beta =
11
12 .27). MLU for these children ranged from 2 to 8 words, $M = 3.84$, $SD = 1.53$.
13
14
15

16
17
18 *Grammatical complexity (binary choice)*. Here again, it was only age that made a significant 5%
19 contribution (beta = .23) to the variance in grammatical complexity, in a standard multiple
20 regression, $R^2 = .07$.
21
22
23

24
25
26
27 *Grammatical complexity (graded choice)*. A standard multiple regression analysis found that age
28 (beta = .35) and maternal education (beta = .28) contributed 12% and 8% respectively to the
29 variance in this measure, $R^2 = .19$.
30
31
32
33
34
35
36
37
38

39 *The Relationship between Lexical and Grammatical Development*

40
41
42
43
44

45 Previous studies have found that toddlers' grammatical and lexical development are not
46 independent of each other but rather go hand in hand. This relationship was examined in the
47 Bulgarian sample in terms of vocabulary size and the four different grammatical development
48 measures listed above. Moderate partial correlations were found after controlling for gender, age,
49 and maternal education between toddlers' vocabulary size and their mean length of utterance ($r =$
50
51 .64), grammatical complexity as a binary variable affixes ($r = .65$), graded grammatical
52
53
54
55
56
57
58
59
60

1
2
3 complexity ($r = .53$), and use of grammatical affixes ($r = .61$). All correlations were significant at
4
5 $p < .001$.
6
7

8 9 10 11 **DISCUSSION**

12
13
14 The Bulgarian toddlers' data and analysis presented here support the findings of previous authors in
15 several respects. First, we have shown that an adaptation of the MacArthur Bates CDI for a yet unstudied
16 language/culture has been able to produce internationally comparable data. A direct comparison between
17 the vocabulary scores for the Bulgarian-speaking toddlers and the available data on US English-speaking
18 toddlers in the original CDI reveals similar, if somewhat lower, levels of lexical development. It is hard to
19 draw conclusions on the relatively lower scores, however, as the Bulgarian CDI version relies to a large
20 extent on the structure and content of the 22 vocabulary categories established in the original CDI for
21 comparative ease and assessment purposes. As in other instruments, lower scores for non-US populations
22 may indicate less about the abilities of children being assessed than about the inevitable limitations of the
23 tool derived from cross-cultural and cross-linguistic differences.
24
25
26
27
28
29
30
31
32
33
34

35
36
37 The significant correlations with age indicate that between 20 and 30 months of age there is
38 considerable growth in both vocabulary and grammar acquisition. The correlations are lower for the
39 grammar components of the CDI than for the vocabulary categories. This may be explained in two ways.
40 One possibility is that vocabulary development follows a more uniform developmental trajectory across
41 languages and cultures than the acquisition of specific and quite diverse grammar systems. The second
42 possibility is that instruments measuring grammatical development across languages differ in the degree
43 of suitability and reliability. The findings we see here with the Bulgarian toddlers' sample may be a
44 product of both. Future work will show whether it is possible to find better ways to measure grammatical
45 development, especially with the accumulation of large-scale data which is currently lacking for
46 Bulgarian with the exception of the sample described here.
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 At the same time as we established the general trends for toddlers' communicative development
4 between the ages of 20 and 30 months in Bulgarian-speaking children, we also found considerable
5 individual variation as attested in previous studies (e.g., Bates, Dale, & Thal, 1995). In fact, we had to
6 exclude four children from the analysis on the basis of their poor vocabulary scores. While some of the
7 variability may be due to differences in caregivers' approach to the task of providing information on their
8 children's level of development, this is not sufficient to explain the very real differences in level of
9 attainment across children at this age.
10
11

12
13
14
15
16
17
18
19 As with previous studies on the topic, we found significant correlations between two strands of
20 communicative development – the acquisition of grammar and vocabulary growth (Fenson et al., 1994;
21 Bates & Goodman, 1997, Caselli, Casadio, & Bates, 1999, Devescovi, Caselli, Marchione, Pasqualetti,
22 Reilly, & Bates, 2005, Bleses et al. 2008). Interestingly enough, we found a significant correlation
23 between vocabulary size and a number of different grammatical development measures, specifically
24 designed to address the characteristics of the Bulgarian language and the challenges it poses to children
25 learning to communicate in it, including two measures of complexity, mean length of utterance, and the
26 use of inflectional morphology on nouns and verbs. One other measure was not found to be associated
27 with developmental change in our sample – whether children were reported as having started to combine
28 words. This is possibly due to the limited age brackets of the sample and ceiling effects as only 16
29 children in the sample were reported by parents not to have started combining words.
30
31
32
33
34
35
36
37
38
39
40
41
42
43

44 Apart from age, two main demographic factors were examined in terms of their influence on
45 children's communicative development, i.e., gender and maternal education. Both have been shown to
46 affect language developmental trajectories in early childhood. Despite previous findings from different
47 studies, gender was found to have little influence in this sample. This was particularly valid in the
48 analysis of grammatical development where on 3 out of 4 measures (two complexity measures and MLU),
49 gender made no significant contribution. The one exception was parent estimates of the regularity of use
50 of inflectional morphology by their children. However, at this age and stage of language development of
51
52
53
54
55
56
57
58
59
60

1
2
3 Bulgarian-speaking toddlers it is difficult to say whether use of inflections is based on an underlying
4 understanding of grammatical categories and regularities vs. rote learning. We need more refined tool in
5 order to make finer distinctions. Generally, however, we can conclude that our boys and girls did not
6 differ significantly in their acquisition of grammar. At the same time, gender did make a significant
7 contribution to the assessment of vocabulary development in line with previous studies.
8
9

10
11
12
13
14
15 Finally, maternal education was found to be an important predictor of both lexical and
16 grammatical development of toddlers. This finding deserves future investigation given the somewhat
17 biased sample distribution with respect to education levels of caregivers.
18
19

20
21
22 In sum, we have found the Bulgarian adaptation of the MacArthur Bates CDI-2 Words and
23 Sentences to be a useful tool in the assessment and research on language development of Bulgarian-
24 speaking toddlers. Further efforts in data collection, instrument improvement and development of short
25 forms will help us throw more light on communicative development in a language and cultural
26 environment that have been under-studied so far. Cross-linguistic comparisons will allow for conclusions
27 to be made on the generalizability of findings and to seek explanations of differences that relate to the
28 major developmental task of becoming communicative in one's community.
29
30
31
32
33
34
35
36
37
38
39
40

41 REFERENCES

42
43
44 Bates, E., & Goodman, J. C. (1997). On the inseparability of grammar and the lexicon: evidence from
45 acquisition, aphasia, and real-time processing. *Language and cognitive processes*, 12, 507–584.
46
47

48
49 Bates, E., Dale, P., & Thal, D. (1995). Individual differences and their implications for theories of
50 language development. In P.Fletcher & B.MacWhinney (Eds.), *Handbook of child language* (pp. 96–
51 151). Oxford, UK: Basil Blackwell.
52
53
54
55
56
57
58
59
60

1
2
3 Caselli, C., Bates, E., Casadio, P., Fenson, J., Fenson, L., & Sanderl, L. (1995). A cross-linguistic study of
4 early lexical development. *Cognitive Development*, 10, 159–199.

5
6
7
8 Caselli, C., Casadio, P., & Bates, E. (1999). A comparison of the transition from first words to grammar
9 in English and Italian. *Journal of Child Language*, 26(01), 69-111.

10
11
12
13 Clegg, J., Hollis, C., Mawhood, L., & Rutter, M. (2005). Developmental language disorders—a follow-up
14 in later adult life. Cognitive, language and psychosocial outcomes. *Journal of Child Psychology and*
15
16
17
18 *Psychiatry*, 46(2), 128-149.

19
20
21 Devescovi, A., Caselli, M.C., Marchione, D., Pasqualetti, P., Reilly, J., & Bates, E. (2005). A
22 crosslinguistic study of the relationship between grammar and lexical development. *Journal of Child*
23
24
25
26
27 *Language*, 32, 759–786.

28
29 Fenson, L., Dale, P.S., Reznick, J.S., Bates, E., Thal, D.J., & Pethick, S.J. (1994). Variability in early
30 communicative development. *Monographs of the Society for Research in Child Development*, 59, 1–185.

31
32
33 Fenson, L., Marchman, V., Thal, D.J., Dale, P.S., Reznick, J.S., & Bates, E. (2007). *Communicative*
34
35
36
37 *Development Inventories*. Baltimore: Brookes.

38
39 McCormack, J., McLeod, S., McAllister, L., & Harrison, L. J. (2009). A systematic review of the
40 association between childhood speech impairment and participation across the lifespan. *International*
41
42
43
44
45 *Journal of Speech-Language Pathology*, 11(2), 155-170.

46 Stoyanova, J. (2011). *Ranno ezikovo razvitie*. Sofia: Logopedichen Tsentur Romel.