



Vocabulary Growth in Lexical Categories Between Ages 13 and 24 Months as a Function of the Child's Sex, Child, and Family Factors

Annette Nylund^{1*}, Piia af Ursin², Pirjo Korpilahti³ and Pirkko Rautakoski¹

¹Department of Speech and Language Pathology, Abo Akademi University, Turku, Finland, ²Department of Education, University of Turku, Turku, Finland, ³Department of Psychology and Speech-Language Pathology, University of Turku, Turku, Finland

OPEN ACCESS

Edited by:

Juhani Järvikivi,
University of Alberta, Canada

Reviewed by:

Virve-Anneli Vihman,
University of Tartu, Estonia
Jelena Kuvač Kraljević,
University of Zagreb, Croatia

*Correspondence:

Annette Nylund
annette.nylund@abo.fi

Specialty section:

This article was submitted to
Language Sciences,
a section of the journal
Frontiers in Communication

Received: 13 May 2021

Accepted: 21 July 2021

Published: 02 August 2021

Citation:

Nylund A, Ursin P af, Korpilahti P and Rautakoski P (2021) Vocabulary Growth in Lexical Categories Between Ages 13 and 24 Months as a Function of the Child's Sex, Child, and Family Factors.
Front. Commun. 6:709045.
doi: 10.3389/fcomm.2021.709045

We examined the vocabulary growth of lexical categories in 719 children (age 13–24 months) as part of a longitudinal cohort study (the STEPS Study) and found a discrepancy in how these categories were affected depending on the child's sex. In girls, attending day care at 24 months of age predicted a positive vocabulary growth in the lexical categories sound effects, nouns, people, and games and routines, compared to girls staying at home. Firstborn girls had a greater vocabulary growth in descriptive and function words, in contrast to those born later. A boy attending day care at age 24 months was likely to have greater growth in sound effects and animal sounds, compared to boys not in day care. A family history of late onset of speech predicted less vocabulary growth in all lexical categories in boys, except for sound effects and animal sounds. Early vocabulary is of importance for later language and literacy development. Vocabulary is not an impenetrable entirety but consists of various types of words (lexical categories) developing at different tempos as they contribute to the developing language. Factors influencing early vocabulary development in boys and girls have been painstakingly studied, but fewer have examined these factors across lexical categories, let alone whether they have an equal effect in both sexes. More knowledge of what affects the variation in early vocabulary in boys and girls is needed for clinical practice and preventive purposes. Vocabulary was measured with the Finnish version of the MacArthur Communicative Development Inventory. The effect of child and family factors on vocabulary growth in various lexical categories was analyzed separately for boys and girls using structural equational modelling. The results of the present study indicate that vocabulary development in the lexical categories is affected differently by child and parental factors in girls and boys as early as the second year of life, which gives new insights into the factors that need consideration in clinical practice and preventive work.

Keywords: child's sex, day care attendance, late onset of speech, lexical categories, MacArthur-Bates communicative development inventory, vocabulary development, lexical category growth

INTRODUCTION

Broad variations are typical of early language development, which is commonly measured in vocabulary size. Such variation is especially true of expressive vocabulary size already at 1 year of age (Bates et al., 1994; Fenson et al., 2007), and the range appears to increase between ages 1 and 2 years as the vocabulary grows (Stolt et al., 2008). Investigating the factors that influence variation in early language is important, as there is evidence that a small vocabulary size at 2 years of age is often related to later language impairments or persistent language delay and can predict language and literacy ability up into school age (Armstrong et al., 2017; Lee, 2011; Rescorla and Dale, 2013; Torppa et al., 2010). Moreover, there are indications that vocabulary size at age 18 months is associated not only with vocabulary size 2 years later, but with development in all language domains (Vehkavuori and Stolt, 2019).

Variation in early language development has been studied comprehensively in relation to various factors related to the child and the child's parents. The most common factors are mentioned here. The parents' own background of late onset of speech has been related to late language emergence or limited language in the child in many studies (e.g., Bavin and Bretherton, 2013; Reilly et al., 2007; Zubrick et al., 2007), and a family history of late onset of speech has been found to triple the risk of persistent delayed language in the child (Zambrana et al., 2014). However, there are also studies where this association has not been found in early language development up to 36 months of age (e.g., Korpilahti et al., 2016; Lankinen et al., 2018). Strong correlations have additionally been detected between parental socioeconomic level (SES) and the child's early language development, where a high level of education and occupation have been associated with larger vocabulary size in countries like the United States, Australia, and Estonia (e.g., Fernald et al., 2013; Hart and Risley, 1995; Reilly et al., 2007; Tulviste and Schults, 2020; Urm and Tulviste, 2016). On the other hand, no associations have been found between high parental education and early expressive vocabulary size in children from the Netherlands (Henrichs et al., 2011) or in a cross-linguistic study from Croatia, Estonia, and Finland (Kuvač-Kraljević et al., 2021). Especially the mother's high educational and/or occupational level has been associated with greater early language development (Gilkerson et al., 2018; Hart and Risley 1995; Hoff, 2003; Letts et al., 2013). However, more recent studies have also found associations between paternal educational and occupational level and early language development in children (e.g., Armstrong et al., 2017; Barbu et al., 2015; Lankinen et al., 2018). There are also studies where positive associations between parental education and or occupation have not been found or have been identified only in particular language domains or lexical categories. Some associations have been found between parental high education and number of nouns, predicates, and closed class words in the child's early language (Bates et al., 1994; Schults et al., 2012). Feldman et al. (2000) again found no relation between maternal educational level and expressive vocabulary scores in 2-year-old children, but rather in other language domains. In addition, at the age of 1 year, they found that children of low educated mothers

had higher vocabulary scores compared to children of highly educated mothers (Feldman et al., 2000). In a few studies, the SES factors have been analyzed separately in boys and girls in relation to language development, showing different effects of the parents' education and occupation depending on the child's sex (Barbu et al., 2015; Lankinen et al., 2018). The methods used to measure early language development vary in different studies, making it difficult to analyze in depth what causes the differences in the results. Most of the studies mentioned above have used the MacArthur Developmental Inventories (CDI), a parental report which has been adapted to many languages (Bates et al., 1994; Bavin and Bretherton, 2013; Feldman et al., 2000; Fernald et al., 2013; Henrichs et al., 2011; Korpilahti et al., 2016; Kuvač-Kraljević et al., 2021; Lankinen et al., 2018; Reilly et al., 2007; Schults et al., 2012; Tulviste and Schults, 2020; Urm and Tulviste, 2016). Even in studies which have used this instrument, results indicating the effect of background variables on the development of children's vocabulary differ from each other.

Child factors often associated with early language development are birth order, day care attendance, and the child's sex; the most evident is the latter, most studies suggesting a discrepancy in favor of girls between the ages of 1–3 years (Andersson et al., 2011; Feldman et al., 2000; Henrichs et al., 2011; Schults and Tulviste, 2016; Simonsen et al., 2014; Tulviste and Schults, 2020). In a study with children from 10 different language backgrounds, Eriksson et al. (2012) found this discrepancy in favor of girls to increase up to the age of 30 months. Differences between boys and girls in early vocabulary development have not been found in all studies. Some studies point out that at an earlier age, under 12 months, differences between expressive vocabulary size in boys and girls are small or non-existent (Simonsen et al., 2014; Stolt et al., 2008), but also that no sex differences have been found at later ages either (Andersson et al., 2011; Bornstein et al., 2004; Hadley et al., 2016). Therefore, even if most studies propose a sex difference in early expressive vocabulary in favor of girls, there are results that indicate the opposite. A greater expansion in early vocabulary development and skills has also, according to some studies, been linked to birth order, with firstborn children having an advantage over those born later (Berglund et al., 2005; Hoff-Ginsberg, 1998; Urm and Tulviste 2016). It has further been suggested that the "word spurt" is more common in firstborns than in children born later, the vocabulary growing faster in the former but more steadily and slowly in the latter (Goldfield and Reznick, 1990). There are studies contradicting these results. Kuvač-Kraljević et al. (2021) found no advantage to being firstborn in terms of vocabulary size in 2-year-old Croatian, Estonian, and Finnish children, while Tulviste and Schults (2020) found a larger vocabulary size at the age of 36 months in children with older siblings. Another child factor that is often studied, with inconsistent results, is the association between day care attendance outside the home and early language development. Some studies suggest positive language effects of attending day care in the Netherlands (Keegstra et al., 2007), Norway (Lekhal et al., 2011), and the United States (Scheffner Hammer et al., 2017). However, there is also evidence of positive short-range effects, but smaller long-

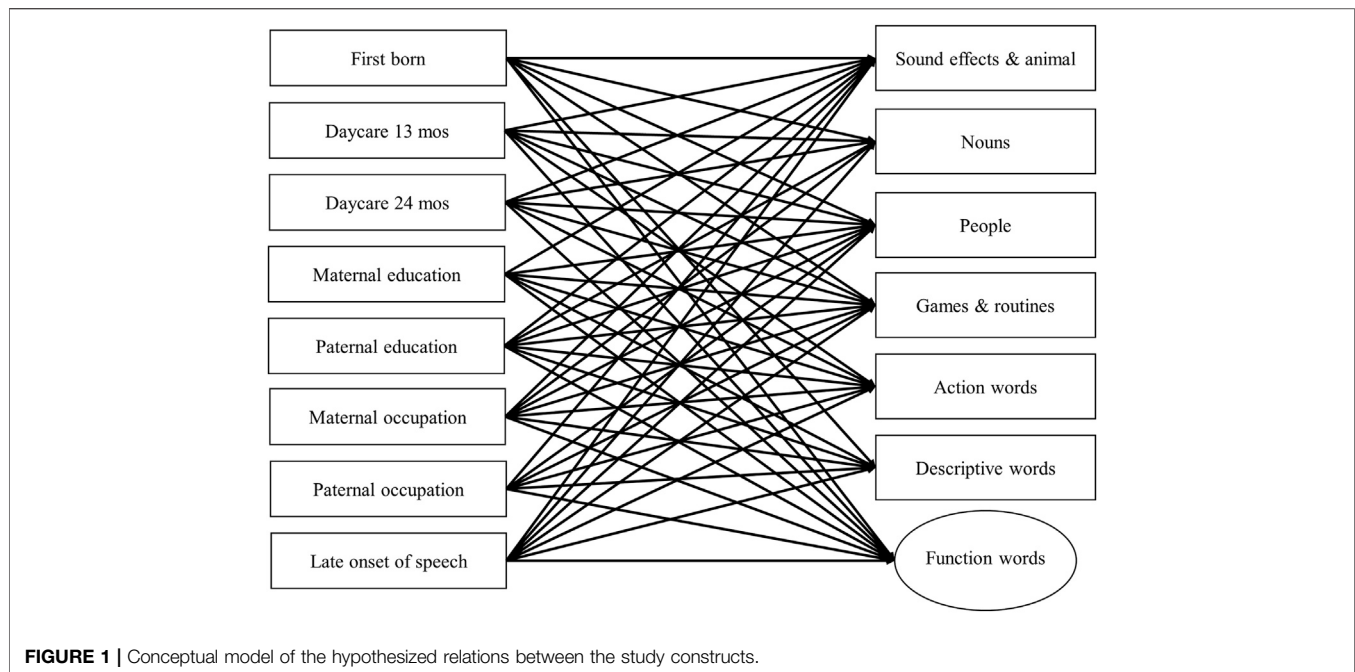
range effects, of attending day care on early language development in Europe, America, and Asia (review by Burger 2010). In studies conducted in the Netherlands (Luijk et al., 2015), Germany (Stolarova et al., 2016), and Estonia (Urm and Tulviste, 2016), negative correlations have been found between early language development and the number of hours spent in day care or the age the child entered day care. These in-depth studies have largely focused on the total vocabulary size at various child ages, but only a few of them look at separate lexical categories.

Studies on the composition of vocabulary (here referred to as lexical categories) have divided vocabulary into broader vs narrower lexical categories. In studies that focus on broader lexical categories, these have usually included social terms (sound effects, games and routines, and people), common nouns, predicates (verbs and descriptive words), and grammatical function words (i.e., Cadime et al., 2018; Schults et al., 2012; Stolt et al., 2007). Studies with a narrow classification of lexical categories have investigated sound effects, nouns, people, games and routines, verbs, descriptive words, and function words (including time words, pronouns, question words, prepositions, and quantifiers) (i.e., Fenson, 2007; Schults & Tulviste, 2016; Wehberg et al., 2007; 2008).

Trajectories of growth in lexical categories can be observed in early vocabulary development. The first word-like utterances are usually sound effects and routine words. These kinds of words, often referred to as onomatopoeic words, can be linked to both the sounds and meanings of the concepts they refer to (Imai et al., 2008; Laing, 2014). They are common in the early vocabulary but later give way to other words (Wehberg et al., 2007). As the vocabulary grows, nouns establishing reference are added more quickly, followed by action, descriptive and function words (Caselli, et al., 1999; Stolt et al., 2008; Wehberg et al., 2007). In a longitudinal study, Wehberg et al. (2007) followed Danish children with the CDI inventory every month between the ages of 8 and 30 months. They found that initial words used by the children were predominantly from the categories of sound effects and games and routines. These words decreased in use as the vocabulary grew and nouns became more predominant (Wehberg et al., 2007). The lexical groups of action, descriptive and function words grew more slowly in these initial stages of development with descriptive words as the last group to develop. The number of people words was quite constant in the first 100 words. Similar trajectories were found in Finnish and Estonian children. Stolt et al. (2008) studied early vocabulary growth in 35 Finnish children aged 0;9–2;0, using the CDI questionnaires. They examined social terms (sound effects, people names, and games and routine words), nouns, verbs, adjectives, and function words. As in the study of Wehberg et al. (2007), social terms were proportionally the largest group in the initial vocabulary. The developmental trajectory of the other lexical categories was similar to that in the study of Wehberg et al. (2007), with a larger growth of nouns followed by action words. Function words and adjectives were the lexical categories that increased more slowly. Schults et al. (2012) investigated vocabulary development, with special focus on broader lexical categories. Social terms (including the categories of sound effects, people words, and games and

routines) was the largest lexical group at the beginning of expressive vocabulary development but decreased proportionally as the vocabulary grew (Schults et al., 2012). As in the study of Wehberg et al. (2007) and Stolt et al. (2008), the category of nouns grew fast in Estonian children and was the dominant lexical category, followed by predicates (verbs and adjectives) and function words (Schults et al., 2012). The category of nouns in early vocabulary increased fast until it reached about 50% of the words in the early vocabulary, after which the growth slowed (Stolt et al., 2008; Wehberg et al., 2007). These referential words function as building blocks for verbs (e.g., Bates et al., 1994; Caselli et al., 1995). Longobardi et al. (2017) found connections between the development of nouns and verbs, the percentage of nouns in the child's vocabulary at 1;4 predicting the number of verbs at 1;8 years. There are indications of mutual relationships between the lexical categories, which are more dependent upon the size of the vocabulary than on the age of the child (McGregor et al., 2005). This means that for the function words to emerge and be taken into communicative use, a larger total vocabulary is needed (Bates et al., 1994; Caselli, et al., 1999). There is evidence that both predicates and function words are central to the development of grammar and syntax (McGregor et al., 2005), and that the number of action words at 24 months of age better predicts grammatical complexity than nouns (Hadley et al., 2016). This implies that precursors to limitations in sentence building and grammar can be found already in the analysis of early vocabulary content. Given that the sizes of various lexical categories are of importance for later language development, there is a need to dig deeper into factors influencing their growth. However, studies focusing on early growth of lexical categories in relation to parental and child factors are scarce.

Although the development of lexical categories has been studied extensively, studies on the effects of background factors (like the child's sex, birth order, maternal educational level, and day care attendance) on the development of early lexical categories are scarce and relate only to some of these factors. None of these studies have, furthermore, examined the effect of background factors on lexical categories separately in boys and girls. There is evidence from earlier studies that girls outpace boys in the number of words in lexical categories, but the distribution and developmental phase of the categories seems to be equivalent between boys and girls (Schults and Tulviste, 2016; Wehberg et al., 2008). Furthermore, there seem to be more content differences within the categories than between them in boys vs girls (Schults and Tulviste, 2016; Wehberg et al., 2008). Schults et al. (2012) found effects of the child's sex only in the category of social terms, girls producing more social words than boys. Which lexical categories are considered to develop more in firstborn children varies between studies. Wehberg et al. (2008) studied lexical development in the first 100 words between ages 0;8 and 1;3 and found that firstborns used more people names than children born later. Schults et al. (2012), on the other hand, found that firstborn children of the same ages had more nouns in their vocabulary. In a study by Wehberg et al. (2008), no differences according to firstborn status were found in the categories of nominals and non-nominals. Richer use of nouns in children between the ages of 8 and 23 months has been associated with



high maternal education (Cadime et al., 2018; Schults et al., 2012). In older children up to 30 months of age, a high maternal educational level has been associated with a greater number of words in all broader lexical categories including social terms, common nouns, predicates, and function words (Cadime et al., 2018). Cadime et al. (2018) found no association between attending day care outside the home and increased vocabulary in any of the lexical categories in children aged 16–23 and 24–30 months.

Only a few studies have examined the association between child and parental factors and early development of lexical categories (Cadime et al., 2018; Schults et al., 2012; Schults and Tulviste, 2016). To our knowledge, there is a lack of studies which focus on the impact of parental SES levels and family burden of late onset of speech on the development of these categories in boys and girls separately. There is reason to believe that the trajectory of vocabulary growth of lexical categories differs in boys and girls in relation to child and parental factors, as has been suggested for early language development and total vocabulary size (Barbu et al., 2015; Lankinen et al., 2018). In the search for explanations for the variation in early vocabulary, some associations could possibly be hidden when composite scores are used for both sexes.

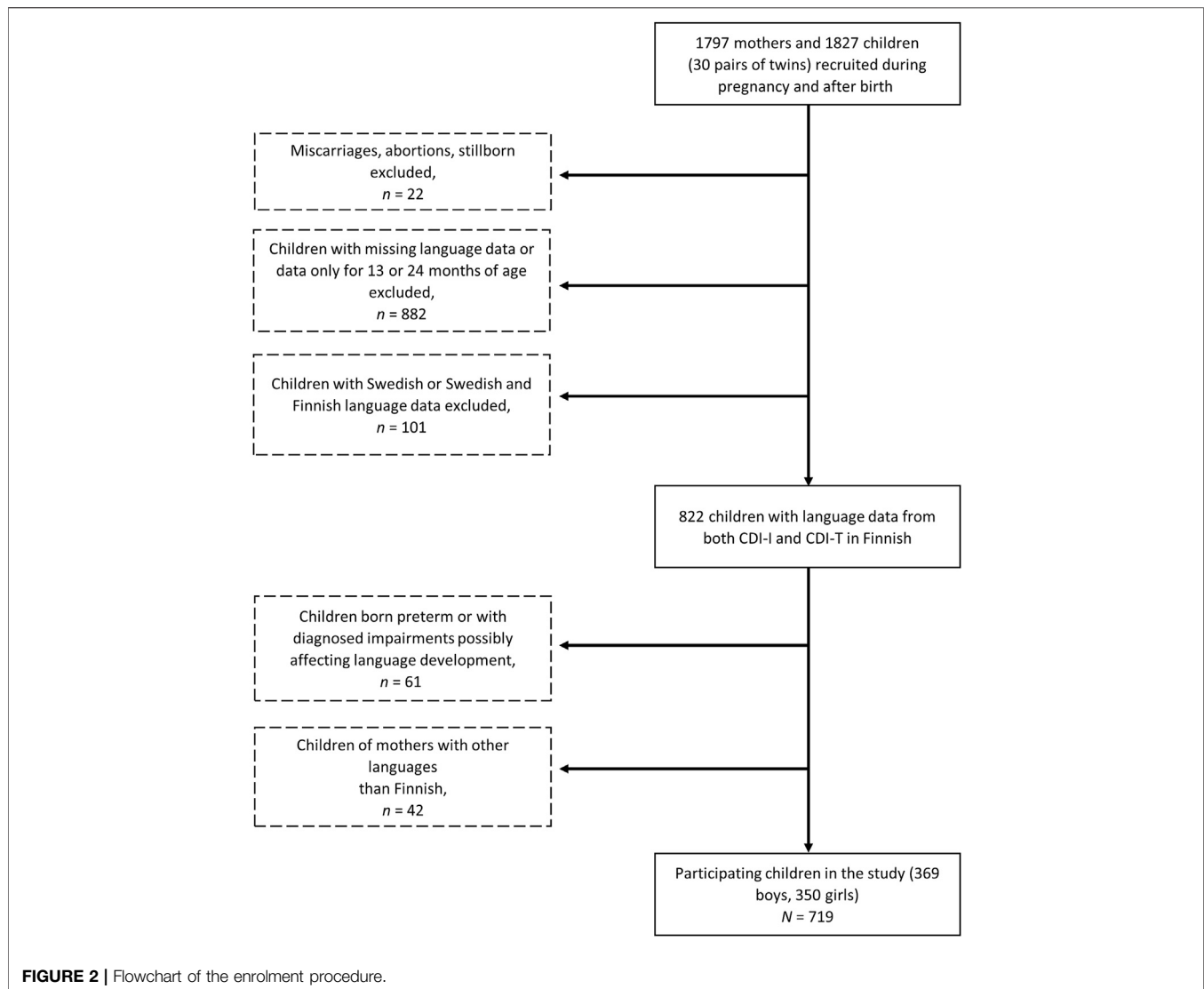
As the studies above show, there is a discrepancy in the results concerning factors predicting early language development, and too few studies have investigated these factors separately in boys and girls. The aim of the present study was to broaden the perspective on early vocabulary growth of lexical categories in boys and girls during the second year of life, and in doing so reduce the inconsistency regarding factors influencing early lexical development. We endeavored also to share more light especially on early vocabulary development of lexical categories in boys vs girls in relation to child and family factors. We aimed to

examine 1) whether child factors (being firstborn, day care attendance at ages 13 and 24 months) and family factors (level of parental education and occupation, a family burden of late onset of speech) predict early vocabulary growth differently in lexical categories; and 2) whether these child and family factors predict vocabulary growth differently among boys and girls. To address these aims, we analyzed vocabulary growth between the ages of 13 and 24 months in 719 children from a cohort from the study “Steps to the Healthy Development and Well-being of Children” (the STEPS study), with structural equational modelling (SEM). Based on previous, we hypothesized that high parental education and occupational level, family history without late onset of speech, firstborn status, and not attending day care at 13 and 24 months of age would predict larger vocabulary growth in the lexical categories. We also hypothesized that there would be different effects of child and family factors in the development of lexical categories in boys and girls, as suggested for total vocabulary or language development by Barbu et al. (2015) and Lankinen et al. (2018). **Figure 1** illustrates the conceptual model depicting the hypothesized predictions between the key study variables.

MATERIALS AND METHODS

Participants

The participants were part of a longitudinal birth cohort study, Steps to the Healthy Development and Well-being of Children (the STEPS study), in the Hospital District of Southwest Finland (Lagström et al., 2013). The Finnish Ministry of Social Affairs and Health and the Ethics Committee of the Hospital District of Southwest Finland approved the STEPS Study in 2007. Participants were enrolled from an eligible cohort of 9,811



Finnish- and Swedish-speaking mothers at maternity clinics during pregnancy or on the delivery ward at birth. From the eligible cohort, 1,797 mothers provided written consent to participate in the study and were informed of the possibility to withdraw at any time. The recruited children were born between January 2008 and April 2010.

Inclusion criteria for the present study were language data in Finnish for the child at both ages 13 and 24 months. Children with language data in Swedish or both Finnish and Swedish were not included, either because Finnish was not the child's first language or its status in the family was unknown. Exclusion criteria were children born preterm (<259 days) or with missing gestational data, and children with diagnosed impairments possibly affecting language development (i.e., epilepsy, cleft palate). Families with a mother with another home language than Finnish but who had completed the language questionnaires in Finnish were also excluded. The final sample was 369 boys and 350 girls ($N = 719$, two twins). **Figure 2** summarizes the enrolment procedure.

Data Collection

Expressive vocabulary growth between 13 and 24 months of age was assessed using the parental report of the Finnish MacArthur Communicative Development Inventory (CDI): Words and Gestures (CDI-I) at the child's age of 13 months, and Words and Sentences (CDI-T) at age 24 months (Lyytinen, 1999; Test information cf.; Fenson et al., 2007). The parents could complete the questionnaires on the study's website or on paper returned with a stamped envelope. A new questionnaire was sent if there was no answer within 2 weeks. The evaluation of vocabulary production by parental reports has been found to correlate highly with standardized testing (Fenson et al., 2007) and has been considered a valid tool to examine early vocabulary (Feldman et al., 2005; Korpilahti et al., 2016; Thal et al., 1999).

The outcome measure was vocabulary growth between the ages of 13 and 24 months in the various lexical categories comprised in the measured expressive vocabulary. The Finnish CDI-I and CDI-T inventories consist of 19 and 20 semantic or lexical categories, respectively, which were analyzed in the

following lexical categories: sound effects and animal sounds (13, 13 items), common nouns (207, 293 items), people (16, 24 items), games and routines (18, 22 items), action words (60, 106 items), descriptive words (26, 54 items), time words (8, 12 items), pronouns (8, 24 items), questions (7, 8 items), prepositions (11, 20 items), amount (6, 9 items), and particles (10 items only in the CDI-T). Growth between the measure points 13 and 24 months was analyzed for all lexical categories except particles, which was only available in the CDI-T. We analyzed the vocabulary in number of words at ages 13 and 24 months. To be comparable, the growth between 13 and 24 months was analyzed in percentage words of the total vocabulary in each lexical category, as numbers of words in the lexical categories differ between the CDI-I and CDI-T. A function factor was constructed of five items measuring different kinds of grammatical categories. These were time words, pronouns, questions, prepositions, and amount. All these categories express different kinds of grammatical function words that start to develop in the latter part of the second year.

Child and Family Factors

The families completed parental and child questionnaires about demographic factors such as educational and occupational background, family structure, health history, and day care attendance at ages 13 and 24 months. The questionnaires were answered at gestational weeks 10–15 by the mothers, at gestational weeks 20 and 30 separately by both parents, and when the child was aged 13 and 24 months by one of the parents. Mothers recruited on the delivery ward completed the first questionnaire at that time.

Independent child factors analyzed in the study were being the firstborn and reported day care attendance at 13 and 24 months of age. Family factors were parental educational level, occupational level, and a family history of late onset of speech. All child and family variables were analyzed separately for lexical growth in boys and girls. The firstborn variable was analyzed as dichotomous (yes or no). The child's attendance in day care outside the home was asked for at 13 and 24 months and analyzed as dichotomous variables (yes or no). At age 13 months, only 21.1% of the children were in day care outside the home compared to 53.15% at age 24 months. Educational level was analyzed in four categories for the mothers and fathers: no occupational education ($n = 101$ [14.5%] and $n = 196$ [28.7%], respectively), college degree ($n = 146$ [20.9%] and $n = 173$ [25.3%], respectively), lower university degree ($n = 233$ [33.4%] and $n = 164$ [24.0%], respectively), and higher university degree ($n = 217$ [31.1%] and $n = 150$ [22.0%], respectively). In cases where the mother or father had chosen the alternative “other education” in the questionnaire, this was regarded as a missing value ($n = 22$ and $n = 36$, respectively), as there was no way to know the level of the education the parent had participated in. The occupational status of the mothers and fathers was analyzed in three categories: low (including farmers, construction, process or transport workers, $n = 55$ [8.7%] and $n = 183$ [31.5%], respectively), medium (including office and service workers, $n = 187$ [29.5%] and $n = 70$ [12.0%], respectively), and high (including managers, specialists and

professionals, $n = 391$ [61.8%] and $n = 328$ [56.5%], respectively). A family burden of late onset of speech was analyzed as a sum variable of parental self-reports of late onset in their own speech, in a sibling's speech or in the speech of a near relative and was categorized as a dichotomous variable (yes or no). There were no significant differences in background variables between boys and girls. It is therefore feasible to assume that the two groups were comparable. See **Table 1** for a descriptive overview of the participants.

Analysis

The analyses for the present study were carried out using the IBM Statistics SPSS 25–26 and Mplus 8.0 software with Maximum Likelihood estimator (Muthén and Muthén, 1998–2011). Missing data on the dependent variables (0.1–0.3% per item) were handled with the Expectation Maximization procedure. In the descriptive analysis of the present study, mean and standard deviation (*SD*) for continuous variables and number and percent for categorial variables were used to describe the study participants. The differences in the mean size of lexical categories and in vocabulary growth of lexical categories between boys and girls were analyzed using an independent two tailed *t*-test. Pearson's correlation analysis was used in comparing the relationship between growth in the lexical categories in boys and girls. *p*-values of less than 0.05 were considered statistically significant.

To test the hypothesized model of differences in early vocabulary growth of lexical categories, SEM analysis was performed (See **Figure 1**). SEM provides the possibility to use multiple indicators to represent a latent variable leading to a reduction in measurement error, analyses of multiple dependent variables simultaneously, flexible handling of missing values on the dependent variables, and testing coefficients across multiple between-subjects groups (e.g., Bollen, 1989; Byrne, 2012). Analyses were conducted using, first, confirmatory factor analysis (CFA) to examine the factor structure of the latent outcome variable, and second, SEM to assess the regressions, including the CFA models and observed variables. In all cases, factors were allowed to correlate, and errors were assumed to be uncorrelated. The fit of the models was evaluated by the chi-square test statistic and fit indices including root mean square error of approximation (RMSEA), standardized root mean square residual (SRMR), Tucker-Lewis index (TLI), and comparative fit index (CFI). The following cutoff values were applied: RMSEA values under 0.08 and TLI and CFI values preferable over 0.95 (e.g., Hu and Bentler (1999)) but acceptable if over 0.90 (e.g., Kline, 2011; Metsämuuronen, 2009). While the ratios of the chi-square statistic and degrees of freedom were carefully considered, statistical significance of the chi-square value alone was not interpreted to indicate an inadequate fit (Hu and Bentler, 1995; Byrne, 2012).

To examine the extent to which possible differential item functioning of the outcome factor “function words” may affect the group differences in focus in the present study, a comparison of multigroup CFA (MGCFA) models across the child's sex was performed, starting with the less restricted model to the more

TABLE 1 | Descriptive overview of participants in the study. Comparison between background variables as a function of the child's sex. *t*-test (conducted for continuous variables) denoted by *mean* (*SD*), and Chi-square test (conducted for categorical variables) is presented by *n* (%).

Variable	Boys (<i>n</i> = 369)	Girls (<i>n</i> = 350)	<i>p</i>
Child			
Apgar, <i>M</i> (<i>SD</i>)	9.0 (0.8)	9.1 (0.8)	0.852
Firstborn, <i>M</i> (<i>SD</i>)	201 (54.5)	187 (53.4)	0.779
Day care attendance			
at 13 months	78 (21.4)	73 (20.9)	0.882
at 24 months	182 (52.6)	178 (53.6)	0.792
Family			
Age at birth (mother), <i>M</i> (<i>SD</i>)	31.1 (4.2)	31.2 (4.3)	0.887
Age at birth (father), <i>M</i> (<i>SD</i>)	33.1 (5.2)	33.1 (5.4)	0.900
Educational level (mother), <i>n</i> (%)	—	—	0.406
Higher university degree	118 (32.8)	99 (29.4)	
Lower university degree	120 (33.3)	113 (33.5)	
College degree	67 (18.6)	79 (23.4)	
No occupational education	55 (15.3)	46 (13.6)	
Educational level (father), <i>n</i> (%)	—	—	0.469
Higher university degree	85 (23.8)	65 (19.9)	
Lower university degree	79 (22.1)	85 (26.1)	
College degree	88 (24.6)	85 (26.1)	
No occupational education	105 (29.4)	91 (27.9)	
Occupational level (mother), <i>n</i> (%)	—	—	0.143
High	204 (62.6)	187 (60.9)	
Medium	88 (27.0)	99 (32.2)	
Low	34 (10.4)	21 (6.8)	
Occupational level (father), <i>n</i> (%)	—	—	0.571
High	166 (55.7)	167 (57.2)	
Medium	33 (11.1)	37 (13.1)	
Low	99 (33.2)	84 (29.7)	
Family history of late onset of speech, <i>n</i> (%)	—	—	0.541
Yes	32 (8.7)	26 (7.4)	

constrained model as suggested by Brown (2015). We applied a hierarchy employing two levels of measurement invariance: configural and metric. The metric invariance is considered sufficient for the purposes of the present study, as it meets the prerequisite to examine structural relationships between variables (e.g., Kline, 2011). In determining the invariance of nested models, we examined the differences in multiple fit indices including the RMSEA, CFI, and TLI. In comparing the metric model with a less restricted configural one, we used the critical values of 0.01 and 0.015 for the differences in CFI and RMSEA, and 0.022 for TLI (e.g., Chen, 2007). To examine the between-group differences in the hypothesized regressions, a multigroup SEM (MGSEM) was applied. While the invariance of the regressions would indicate similar effects of background variables on language development in boys and girls, the non-invariance confirms the differential associations.

RESULTS

Vocabulary Growth Between the Ages of 13 and 24 months in Lexical Categories

There was a large variation in total vocabulary size at 13 and 24 months of age in and between boys and girls. Mean vocabulary size at 13 and 24 months of age in boys was 7.8 (*SD* = 10.9, *Range* = 0–116) and 262.2 (*SD* = 168.1, *Range* = 4–593), respectively,

and in girls 11.8 (*SD* = 22.8, *Range* = 0–297) and 340.3 (*SD* = 152.5, *Range* = 4–595), respectively. Total vocabulary growth between the ages of 13 and 24 months in boys was *mean%* 42.0 (*SD* = 27.5) and in girls *mean%* 54.1 (*SD* = 24.6).

Mean sizes of the various lexical categories varied widely between the children. Mean number of words in the category of common nouns at age 13 months ranged from 0 to 71 in boys and 0 to 179 in girls. Forty-two percent of the boys and 35% of the girls had no words in the common noun category. At 24 months of age, the variation in vocabulary size of nouns was still large, varying from 0 to 291 words in boys and 0 to 293 words in girls. At 2 years of age, 1.7% (*n* = 7) of the boys and 0.6% (*n* = 2) of the girls had no noun in their vocabulary. The girls had a significantly larger vocabulary size than boys at 13 months of age in four of the lexical categories: sound effects and animal sounds, common nouns, people, and games and routines. At 24 months of age, the girls outperformed the boys in all lexical categories (*p* < 0.001).

The percentual growth of vocabulary between the ages of 13 and 24 months was *mean%* 0.42 (*SD* = 0.27) in boys and *mean%* 0.54 (*SD* = 0.25) in girls. Vocabulary growth in lexical categories varied substantially in and between the categories. Some children had fewer words in some categories at 24 months than at 13 months of age, which explains the negative numbers in the ranges (Table 2). Girls outperformed boys in vocabulary growth in all lexical categories except sound effects and animal sounds (*p* < 0.001). However, the effect sizes were small to moderate (*d* = –0.25 to –0.47).

TABLE 2 | Descriptive statistics of mean vocabulary growth (in %) between ages 13 and 24 months for boys ($n = 369$) and girls ($n = 350$). Independent two-tailed t -test conducted when comparing mean differences between boys and girls.

Lexical growth 13–24 months %	Boys Mean (SD)	Girls	P	Cohen's d (95% CI)
Sound effects and animal sounds (13, 13)	0.62 (0.22)	0.61 (0.23)	0.772	0.02 (–0.12, 0.168)
Common nouns (207, 293)	0.47 (0.29)	0.60 (0.26)	<0.001	–0.47 (–0.62, –0.32)
People (16, 24)	0.38 (0.24)	0.44 (0.21)	0.001	–0.25 (–0.40, –0.10)
Games and routines (18, 22)	0.54 (0.29)	0.67 (0.25)	<0.001	–0.44 (–0.59, –0.30)
Action words (60, 106)	0.43 (0.35)	0.59 (0.32)	<0.001	–0.47 (–0.62, –0.32)
Descriptive words (26, 54)	0.30 (0.27)	0.40 (0.27)	<0.001	–0.38 (–0.53, –0.23)
Time words (8, 12)	0.27 (0.30)	0.37 (0.32)	<0.001	–0.30 (–0.45, –0.16)
Pronouns (8, 24)	0.18 (0.21)	0.24 (0.21)	<0.001	–0.29 (–0.44, –0.15)
Questions (7, 8)	0.29 (0.30)	0.40 (0.30)	<0.001	–0.36 (–0.51, –0.22)
Prepositions (11, 20)	0.43 (0.32)	0.54 (0.29)	<0.001	–0.38 (–0.53, –0.23)
Amount (6, 9)	0.33 (0.28)	0.44 (0.28)	<0.001	–0.40 (–0.55, –0.25)

In parentheses, total number of items in each category at age 13 months, 24 months.

TABLE 3 | Correlations between growth in lexical categories in boys and girls aged 13–24 months (Pearson correlation).

Boys	1	2	3	4	5	6	7	8	9	10	11
	Girls										
1. Sound effects	—	0.365***	0.365***	0.358***	0.278***	0.223***	0.196***	0.182**	0.164**	0.221***	0.197***
2. Nouns	0.439***	—	0.766***	0.867***	0.921***	0.803***	0.683***	0.592***	0.686***	0.805***	0.713***
3. People	0.426***	0.854***	—	0.739***	0.737***	0.682***	0.624***	0.601***	0.594***	0.641***	0.609***
4. Games and routines	0.447***	0.902***	0.841***	—	0.865***	0.730***	0.647***	0.564***	0.675***	0.743***	0.640***
5. Action words	0.361***	0.932***	0.824***	0.881***	—	0.868***	0.759***	0.656***	0.755***	0.856***	0.778***
6. Descriptive words	0.341***	0.875***	0.812***	0.822***	0.906***	—	0.846***	0.743***	0.751***	0.843***	0.814***
7. Time words	0.261***	0.773***	0.706***	0.741***	0.817***	0.842***	—	0.665***	0.689***	0.705***	0.730***
8. Pronouns	0.237***	0.678***	0.668***	0.665***	0.740***	0.777***	0.754***	—	0.729***	0.701***	0.715***
9. Questions	0.257***	0.752***	0.708***	0.710***	0.806***	0.802***	0.741***	0.769***	—	0.750***	0.729***
10. Prepositions	0.339***	0.878***	0.777***	0.838***	0.897***	0.858***	0.792***	0.727***	0.787***	—	0.788***
11. Amount	0.274***	0.775***	0.707***	0.751***	0.806***	0.840***	0.755***	0.730***	0.753***	0.823***	—

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The result of the Pearson correlation indicated a significant positive association ($p < 0.001$) between lexical growth in the lexical categories in boys ($r = 0.237$ – 0.932) and in girls ($r = 0.196$ – 0.921), except for associations between sound effects vs pronouns and questions ($r = 0.182$, $p = 0.001$ and $r = 0.164$, $p = 0.002$, respectively) (Table 3). The correlations between sound effects and the other lexical categories were weak in both boys and girls ($r = 237$ – 447 and $r = 182$ – 365 , respectively).

Psychometric Analyses and Measurement Invariance for the Function Words Factor

Before predicting study outcomes, we examined the structure and invariance of the function factor across the sexes. The internal consistencies of the scale varied between $0.923 \leq \alpha \leq 0.937$, showing excellent measurement reliability (e.g., Nunnally and Bernstein, 1994). With regard to normality, the study items' univariate distributions were all within a reasonable range (skewness ± 2 , kurtosis ± 7 ; see Curran et al., 1996). In addition, across both sexes the freely estimated factor loadings were significant ($p < 0.001$) and the coefficients were salient, ranging from 0.808 to 0.930. Table 4 shows the tests for

measurement invariance across the child's sex. The first step involved testing the same factor structure with boys' and girls' empirical covariance matrix separately. This produced very well-fitting models in each group.

A configural invariance model also showed an excellent fit to the data (Table 4, model M1). Further, constraining corresponding factor loadings to be equal across the sexes produced the metric invariance model which exhibited a good overall fit (M2 in Table 4). The minor changes in GFIs indicated full metric invariance of the function words scale across the study groups.

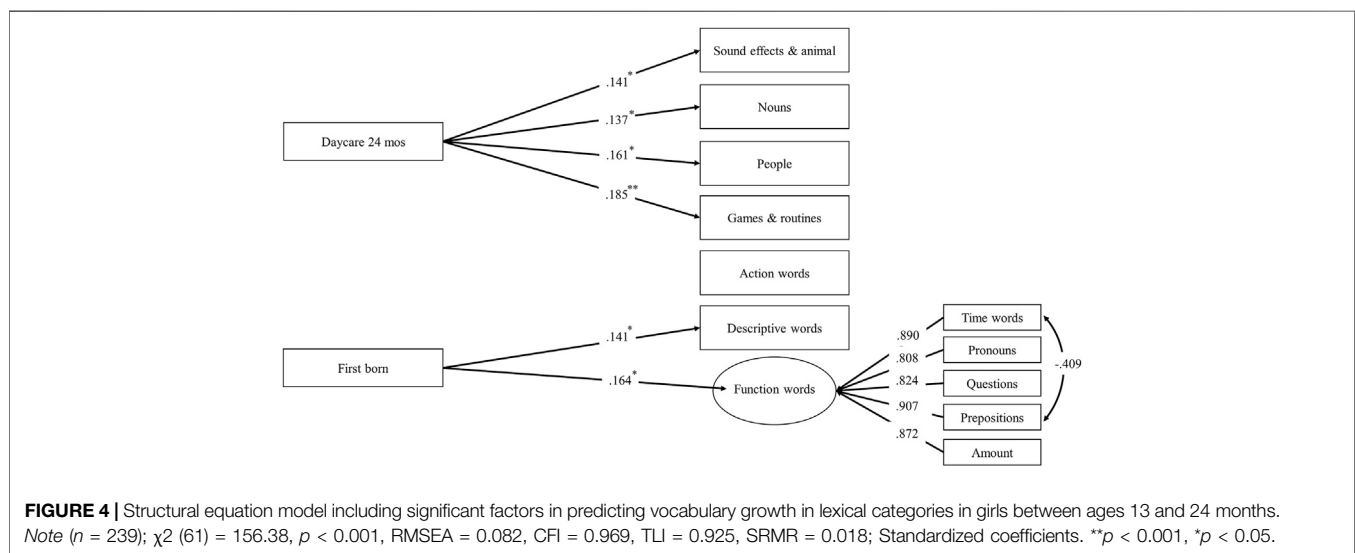
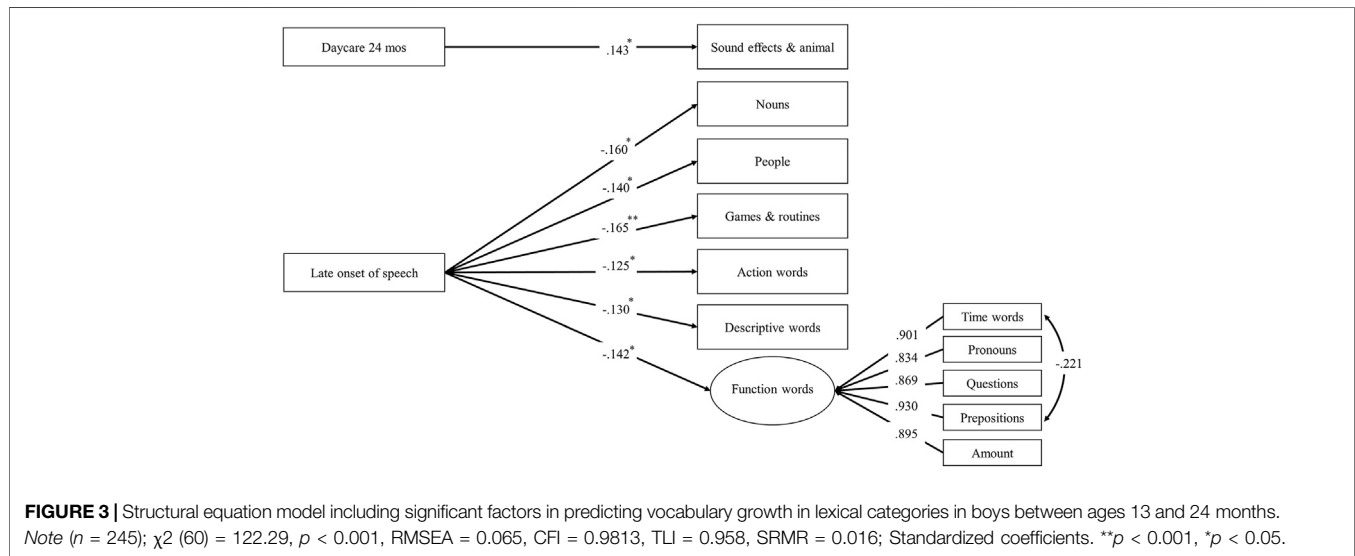
Predicting Differences in Early Vocabulary Growth Among Boys and Girls (MGSEM)

The initial structural model including the hypothesized associations between independent predictors and vocabulary growth outcomes (see Figure 1) demonstrated a well-fitting multigroup model to the data: $\chi^2 (128) = 285.13$, RMSEA = 0.071, CFI = 0.976, TLI = 0.947, SRMR = 0.018. However, constraining the regression coefficients to be equal across the sexes resulted in a misfit of the data. This suggests that child and

TABLE 4 | Tests for measurement invariance for function words factor across boys and girls.

Model	χ^2	df	P	RMSEA	CFI	TLI	SRMR	Δ RMSEA	Δ CFI	Δ TLI
Single group solutions										
Boys1	5.91	4	0.206	0.036	0.998	0.996	0.009	—	—	—
Girls	11.12	5	0.049	0.059	0.995	0.991	0.011	—	—	—
Measurement invariance ^a										
M1 configural	16.65	8	0.034	0.055	0.997	0.993	0.009	—	—	—
M2 metric	24.41	12	0.018	0.054	0.996	0.993	0.024	0.001	0.001	0.000

¹Boys (n = 368), girls (n = 350); The following residual error was allowed to correlate: (Pronouns with Prepositions; $r = -0.39$);^aModels: M1, configural invariance (equal form); M2, metric invariance (factor loading constrained to be equal across groups).



family factors affecting early vocabulary growth vary between boys and girls. **Figures 3, 4** report the findings of the fully estimated models separately for both groups. Only statistically significant associations are included in the models.

In boys, a family burden of late onset of speech predicted less vocabulary growth in all lexical categories except for the category of sound effects and animal sounds. However, this category was the only lexical category with larger growth in those boys attending day care at 24 months of age. In girls, being firstborn predicted positive growth in descriptive and function words. In girls, attending day care predicted larger growth of words in the categories sound effects, nouns, people, and games and routines. (**Figures 3, 4**). However, the amount of explained variance was small.

DISCUSSION

The aim of the present study was to examine if there are differences in how child factors and family factors predict vocabulary growth in various lexical categories, and whether these factors predict vocabulary growth differently in boys vs girls. We found differences in how these factors predicted vocabulary growth in the lexical categories. The main predictors of vocabulary growth between 13 and 24 months of age were a family burden of late onset of speech, firstborn status, and day care experience at 24 months of age. Moreover, there was a difference in how these factors predicted growth in the lexical categories as a function of the child's sex. The comparison of the results with previous studies is not straightforward, as earlier studies have analyzed the effect of factors on development of lexical categories in boys and girls together, and not separately as in the present study. Analyzing vocabulary growth as combined scores for boys and girls can possibly hide significant scores for either sexes.

Vocabulary Growth in Lexical Categories

The largest lexical group in both boys and girls at 13 months of age was common nouns, followed by sound effects, people words, and games and routines. The predominance of nouns in early vocabulary acquisition as the vocabulary size increases is congruent with earlier studies (e.g., Bates et al., 1994; Caselli et al., 1995; Stolt et al., 2008). At 24 months of age, nouns still formed the largest group in the vocabulary of both boys and girls, followed by action words, descriptive words, and games and routines. When the vocabulary reaches over 50 words, action and descriptive words start to develop (Wehberg et al., 2007), as was similarly observed at group level in both boys and girls at 24 months of age in our study. We found that the individual variability of vocabulary size within the lexical categories increased with age, as has been suggested in earlier studies with English, Italian, Slovenian, and Estonian children measured with CDI (Caselli et al., 1999; Marjanovič-Umek et al., 2016; Schults and Tulviste, 2016). The girls outperformed the boys at age 13 months in the lexical categories of sound effects, common nouns, people, and games and routines words, and at 24 months in all categories. This is comparable with the results of Schults and Tulviste (2016), who found significant differences between boys and girls in vocabulary size at ages 1;2 and 1;4 ($N = 903$) in all categories

except descriptive and function words. None of the children in the present study reached the maximum ceiling of available words in a category at 13 months of age, whereas at 24 months the ceiling of all lexical categories was reached by some of the girls. Among the boys, some reached the ceiling in all categories except nouns, where the ceiling was not reached.

Considering the vocabulary growth, calculated in percentage, in the lexical categories between 13 and 24 months of age in boys and girls, the same five categories were the fastest growing but in a slightly different order. In boys, sound effects showed the largest growth, followed by games and routines, nouns, action words, and prepositions, whereas in girls, games and routines was the largest growing category followed by sound effects, nouns and action words and prepositions. As for the growth size in the various lexical categories, the girls outpaced the boys in all lexical categories except for sound effects. The discrepancy between total vocabulary size in boys and girls has been shown to increase up to 30 months of age in a cross-linguistic study with ten different European languages (Eriksson et al., 2012). The same discrepancy should apparently be found in the size of and growth within the lexical categories, which was confirmed in the present study. The correlations between growth in the lexical categories were strong, except between sound effects and the other categories, where the correlation was small or medium-sized. The category sound effects and animal sounds is perhaps the one that differs most from the others, as it is large and transitory in early expressive vocabulary development and paves the way to more conventional words (Caselli et al., 1995).

Family Factors and Vocabulary Growth in Lexical Categories

A family burden of late onset of speech predicted in boys a smaller growth in all lexical categories except for sound effects and animal sounds. This is in line with our hypothesis and with previous studies suggesting an effect of late onset of speech in the family. The result also supports the second hypothesis of different effects in boys and girls. The effect of a family history of late onset of speech and language problems on early language development has been reported in many previous studies (Bishop et al., 2012; Keegstra, 2007; Reilly et al., 2010; Zambrana et al., 2014; Zubrick, 2007). Reilly et al. (2010) and Bishop et al. (2012) reported that a family history of speech and language problems in children with late onset of speech at 18–24 months of age predicted more persisting language difficulties compared to children without this background. In the studies mentioned above, the focus has been more on the relation between a family history of late talking, reading, and writing difficulties and language impairment. However, in the present study, the parents only reported a late onset of speech of their own, in siblings, and/or in the near family. Language impairments or problems with reading and writing were not accounted for. A parental report of late onset of speech in the family, without information on diagnosed language problems, was enough to predict slower growth in lexical categories between 13 and 24 months of age in boys. This confirms that already at this early age we can detect potential later language problems, which is of significance as it has been suggested that almost 50% of late talking children aged

2 years still have poorer language skills at 10 years of age (Armstrong et al., 2017). The only lexical category not affected by a family burden of late onset of speech in the present study was that of sound effects and animal sounds. This category is often large in the beginning of vocabulary growth (the first five words) and contains sounds resembling word forms which later give way to more conventional words (Caselli et al., 1995; Schults and Tulviste, 2016; Wehberg et al., 2007). A possible explanation as to why this category was not affected by a family burden of late onset of speech could be its role as a more transitional category in the lexical development. Laing (2014) suggests that these first sound-resembling words, called onomatopoeic words, serve as a bridge to more conventional words. In the present study, paternal late onset of speech affected boys but not girls. One reason for this could be that heritability for expressive vocabulary has been found to be more than two times higher in boys than in girls (Van Hulle, et al., 2004).

In the present study, neither educational nor occupational status of the parents predicted lexical growth in boys or girls, which contradicted both of the hypotheses. The result was contrary to earlier studies by Fernald et al. (2013), Gilkerson et al. (2018), and Hart and Risley (1995), but confirmed the studies by Henrichs et al. (2011) and Kuvač-Kraljević et al. (2021). The lack of effect of high education or occupational status on early vocabulary growth can reflect cultural differences in educational system and how this has been measured in different studies. The studies of Fernald et al. (2013), Gilkerson et al. (2018), and Hart and Risley (1995) were performed in English-speaking communities, while for example those of Henrichs et al. (2011) and Kuvač-Kraljević et al. (2021) were conducted in European countries. However, Lankinen et al. (2018) found high education of Finnish fathers to be associated with expressive vocabulary size at 24 months of age. In our study, over 60% of the participants' (boys and girls) mothers had at least a lower university education, and at least 45% of the fathers. The occupational level was also high in mothers and fathers (over 60 and 55%, respectively). This should be taken into consideration when interpreting the results, as the large percentage of highly educated parents with advanced occupational status could have skewed the results. It is also possible that the effects of educational and occupational status would have come across through process variables, such as the habit of reading to the child, if they had been analyzed in this study. Vocabulary growth in lexical categories in both boys and girls were equally unaffected by the parents' educational or occupational level. This contradicts the studies of Barbu et al. (2015) and Lankinen et al. (2018), where early language in boys was more affected by SES factors than in girls. The reason for this could be the high educational and occupational level of both parents and that the differences would be more obvious in low-SES families, as suggested by Barbu et al. (2015).

Child Factors and Vocabulary Growth in Lexical Categories

Being the firstborn predicted larger vocabulary growth in the lexical categories of descriptive and function words in girls, which was partly in line with our first hypothesis of the advantage of being firstborn, and in line with our hypothesis of different effects

in boys and girls. There are indications that mothers generally talk more to their daughters than to their sons, even if both parents seem to talk somewhat more to their firstborn sons (Gilkerson and Richards, 2009). The vocabulary growth in boys was not affected by being firstborn or not. One explanation could be that boys lag behind girls in grammar development at this age and cannot take advantage of heard speech. Another explanation could be that even if firstborn boys hear more talk than those born later, the language addressed to them is more restricted than what girls hear (Gilkerson and Richards, 2009). Previous studies have suggested that firstborn children have an advantage over those born later (Hoff-Ginsberg, 1998; Urm and Tulviste, 2016). However, those studies focused on total vocabulary size, whereas in our study, only the growth in descriptive and function words was predicted by birth order. Descriptive and function words emerge more slowly and do not increase until the latter half of the second year, as the total vocabulary needed for developing grammar expands (Bates et al., 1994; Caselli et al., 1995; Marchman and Bates, 1994; Stolt et al., 2008; Wehberg et al., 2007). This was also found in our study, where descriptive and function words (time words, pronouns, questions, prepositions, and amount) were few at 13 months of age in the girls' vocabulary but had increased at 24 months. The growth rate in the categories of descriptive and function words showed a substantial increase between 13 and 24 months of age. As it has been suggested by Hoff-Ginsberg (1998) that both vocabulary and grammar development are more advanced in firstborn children, it could be that of the whole vocabulary, descriptive and function words are the grammar-building blocks that are most influenced by being born first. The language the mothers use with their firstborn children is, furthermore, more advanced, with longer sentences and fewer questions, than language used with those born later Hoff-Ginsberg (1998), which could also explain the predictive function of being firstborn on the growth of descriptive and function words.

Day care attendance at age 24 months predicted vocabulary growth positively in the lexical category of sound effects in both boys and girls. However, only in girls did day care attendance predict vocabulary growth also in nouns, people words, and games and routine words. This was in line with our hypothesis of different effects in boys and girls. The results were not in line with our hypothesis that day care attendance does not predict vocabulary growth, and they contradict the results of Cadime et al. (2018), where no effects of time spent at day care were found in vocabulary size in any of the lexical categories at ages 16–23 or 24–30 months in Portuguese children. Cadime et al. (2018) used a broader distribution of lexical categories, with social terms (sound effects and animal sounds, people words, games and routines), nouns, predicates (verbs, descriptive words), and function words. In the present study, the growth of sound effects was greater in both boys and girls attending day care at 24 months of age compared to children not attending. This is the lexical group of word forms resembling animal sounds, car sounds, etc., and which will for the most part develop into conventional words towards the end of the second year (Caselli et al., 1995). One could argue that this is a category

mostly used in play and group activities in day care, and that at home other categories are used more when talking to the child. This is, however, not applicable, because vocabulary growth in girls in day care at 24 months was also found in the lexical categories of nouns, people words, and games and routines. The outcome furthermore somewhat contradicts the results of Stolarova et al. (2016), who found a larger vocabulary in German girls under 2 years of age not attending day care outside the home. There is evidence that spending a great many hours in day care at a young age can negatively affect vocabulary size (Urm and Tulviste, 2016). As the vocabulary size of girls is more advanced than that of boys (Andersson et al., 2011; Eriksson et al., 2012), one explanation could be that girls attending day care at 24 months of age have a greater capacity than boys of the same age to take advantage of the language spoken in group situations. We could also assume that the speech children hear at home is more child-directed than at day care, where instructions are given to a group. There is in fact evidence that child-directed speech is associated with larger vocabulary size (Hart and Risley 1995; Rowe, 2008; Weisleder and Fernald, 2013), which could explain why attending day care only predicted the growth of sound effects in boys. Sound effects are usually onomatopoeic words which are possibly much used in play between young children and not so much in child-directed speech at home. Girls with a stronger vocabulary can presumably develop vocabulary also in day care settings without speech directed specifically at them. Among the girls there were also lexical categories that predicted no growth in day care attendance at 24 months of age. The categories affected by day care were sound effects and animal sounds, nouns, people words, and games and routines, which were already more developed at 13 months of age in the girls. All these categories, except for people words, belonged to the four largest developing categories between 13 and 24 months of age. This could suggest that the focus and language used in day care at this early age is related more to the first developing categories of sounds, people, and play words and nouns than to action, descriptive, and function words.

In the present study, the predictive value of day care attendance “overrides” the educational and occupational status of the parents in relation to early vocabulary growth. This is a surprising result. Day care in Finland falls under early childhood education and care legislation to secure lifelong learning, starting from day care. For each child attending day care, educational plans are created, including individual goals for the development of the child (Ministry of Education and Culture). Of the children in the present study, only 53% were in day care at 24 months of age. It has been suggested in international studies that children of parents with a higher educational level attend fewer hours at day care at a younger age (Urm and Tulviste, 2016), but also that children attending day care in the first year of life have parents who are more highly educated (de Hoog et al., 2014). As discussed above, it is possible that the effects of highly educated parents come across other variables such as day care attendance. Formal day care arrangements and the use of these services varies greatly in different countries. Only in Europe is there a large discrepancy in provision and use of formal day care services at the ages of

0–2 years, where 73% of Danish children attend day care compared to only 2% of Czech and Polish children (European Commission, 2009).

It can also be argued that cultural differences in environmental factors (such as day care arrangements, educational systems) influence and cause variations in the results of different studies. However, Eriksson et al. (2012), Bleses et al. (2008), Caselli et al. (1995), Braginsky et al. (2019), and Kuvač-Kraljević et al. (2021) demonstrate in their studies including language societies inside and outside Europe that cultural differences do not supersede the basic process in early vocabulary development.

Different Effects on Lexical Growth Between Boys and Girls

The results of the present study suggest that lexical growth is affected differently by child and family factors as a function of the child’s sex. Various factors predicted the growth in lexical categories in boys and girls in diverging ways. As there were no significant differences between the background variables in boys and girls, the differences we found in lexical growth were analyzed as a function of the child’s sex. These findings are in line with the hypothesis that there would be different effects of child and family factors on the development of lexical categories in boys and girls, and with previous studies by Barbu et al. (2015) and Lankinen et al. (2018) where sex differences in language development have been found to relate to family factors.

One explanation for the differences in how vocabulary development was predicted by background factors in boys and girls could be different susceptibilities in boys vs girls to the factors influencing early language development. It has been suggested in twin studies that early language development is influenced differently by biological and environmental factors (Galsworthy et al., 2000; Van Hulle, et al., 2004). There is a possibility that when analyzing boys and girls together in relation to language development and the factors affecting it, factors that would enlighten the gender variations in early language development are overlooked. This variation as a function of the child’s sex is important and needs to be studied more, as it brings new perspectives to early language development and is thus of significance in supporting early vocabulary development and children at risk.

Strengths and Limits

A strength of the present study is the prospective longitudinal birth cohort design. Moreover, it presents new detailed information on the early vocabulary growth of lexical categories and how vocabulary growth is predicted by background factors as a function of the child’s sex. However, there are some limitations that need to be considered when interpreting the results. The effect sizes were small, meaning that the studied factors explained only part of the vocabulary growth in lexical categories between the ages of 13 and 24 months. This was true of both boys and girls. In the present study, vocabulary size at 13 months of age was not considered in the factor analysis; it could have added the aspect of where in the vocabulary development the child was

at the beginning of the growth period. For some children with a larger vocabulary size, the growth could have continued on the same trajectory. On the other hand, for a child with a larger vocabulary already at 13 months, the growth may not have seemed as big as for a child with less vocabulary in the beginning. Another restriction was that the date of enrolment in day care was not asked for. This limited the interpretation of the effects concerning day care attendance, as we lacked detailed information on how long the child had attended day care.

CONCLUSION

The present study provides detailed information of vocabulary growth in lexical categories between the ages of 13 and 24 months. It shows that in all the lexical categories, the growth in girls outpaced that in boys. Furthermore, our results suggest differences between vocabulary growth in boys and girls in relation to child and family factors. Boys with a family risk of late onset of speech had slower growth in all lexical categories except for sound effects and animal sounds. Being a girl and the firstborn enhanced vocabulary growth in the lexical classes of descriptive and function words. Day care attendance at 24 months of age affected vocabulary growth positively in girls more than in boys. In boys, only the lexical category of sound effects was enhanced, whereas the categories of sound effects, nouns, people words, and games and routines grew larger in girls attending day care at age 24 months compared to children staying at home. The present study emphasizes the need to be aware that not all factors influence early vocabulary development in the same way in boys and girls and suggests that studies of effects on early language development should consider examining vocabulary growth separately in boys and girls.

DATA AVAILABILITY STATEMENT

The datasets presented in this article are not readily available because: The data has not been collected in the whole country, but in a part of Finland and includes sensitive information that could potentially identify participants. The Clinical Research Centre of

REFERENCES

- Andersson, I., Gauding, J., Graca, A., Holm, K., Öhlin, L., Marklund, U., et al. (2011). Productive vocabulary Size Development in Children Aged 18-24 Months—Gender Differences. *Speech, Music Hearing Q. Prog. Status Rep.* 51 (1), 109–112.
- Armstrong, R., Scott, J. G., Whitehouse, A. J. O., Copland, D. A., McMahon, K. L., and Arnott, W. (2017). Late Talkers and Later Language Outcomes: Predicting the Different Language Trajectories. *Int. J. Speech-Language Pathol.* 19, 237–250. doi:10.1080/17549507.2017.1296191
- Barbu, S., Nardy, A., Chevrot, J.-P., Guellai, B., Glas, L., Juhel, J., et al. (2015). Sex Differences in Language across Early Childhood: Family Socioeconomic Status Does Not Impact Boys and Girls Equally. *Front. Psychol.* 6:1874. doi:10.3389/fpsyg.2015.01874
- Bates, E., Marchman, V., Thal, D., Fenson, L., Dale, P., Reznick, J. S., et al. (1994). Developmental and Stylistic Variation in the Composition of Early Vocabulary. *J. Child Lang.* 21, 85–123. doi:10.1017/s0305000900008680

the Hospital District of Southwest Finland has specified that legal and ethical restrictions prevent public sharing of deidentified individual participant data. Requests for data are handled by the directory board of the STEPS Study and can be sent to Hanna LagstroÅm (hanlag@utu.fi). Requests to access the datasets should be directed to; Hanna LagstroÅm (hanna.lagstrom@utu.fi).

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by The Finnish Ministry of Social Affairs and Health and the Ethics Committee of the Hospital District of Southwest Finland approved the STEPS Study in 2007. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

AN was responsible for writing the manuscript, reviewing the literature, and performing the initial analyses. PU was responsible for performing the SEM analysis and describing it in the manuscript. PK and PR were responsible in planning the study design and the questionnaires. PK was responsible for collecting the language data. PR was the main supervisor of the project, and PU, PK and PR revised the manuscript.

FUNDING

This research was supported in part by Kommunalrådet C G Sundells stiftelse.

ACKNOWLEDGMENTS

The authors are grateful to all the families who took part in this study, the midwives for their help in recruiting them, and the whole STEPS study team.

- Bavin, E. L., and Bretherton, L. (2013). “The Early Language in Victoria Study” in Late Talkers,” in *Language Development, Interventions, and Outcomes*. Editors L. A. Rescorla and P. S. Dale (Paul H. Brookes Publishing Co), Baltimore, MD, USA, 3–21.
- Berglund, E., Eriksson, M., and Westerlund, M. (2005). Communicative Skills in Relation to Gender, Birth Order, Childcare and Socioeconomic Status in 18-Month-Old Children. *Scand. J. Psychol.* 46, 485–491. doi:10.1111/j.1467-9450.2005.00480.x
- Bishop, D. V. M., Holt, G., Line, E., McDonald, D., McDonald, S., and Watt, H. (2012). Parental Phonological Memory Contributes to Prediction of Outcome of Late Talkers from 20 Months to 4 years. A Longitudinal Study of Precursors of Specific Language Impairment. *J. Neurodevelopmental Disord.* 4, 3. doi:10.1186/1866-1955-4-3
- Bleses, D., Vach, W., Slott, M., Wehberg, S., Thomsen, P., Madsen, T. O., et al. (2008). Early Vocabulary Development in Danish and Other Languages: A CDI-Based Comparison. *J. Child Lang.* 35, 619–650. doi:10.1017/S0305000908008714

- Bollen, K. A. (1989). *Structural Equations with Latent Variables*. Hoboken NJ, USA: Wiley Press. doi:10.1002/9781118619179
- Bornstein, M. H., Hahn, C.-S., and Haynes, O. M. (2004). Specific and General Language Performance across Early Childhood: Stability and Gender Considerations. *First Lang.* 24 (3), 267–304. doi:10.1177/0142723704045681
- Braginsky, M., Yurovsky, D., Marchman, V. A., and Frank, M. C. (2019). Consistency and Variability in Children's Word Learning Across Languages. *Open Mind: Dis. Cognit. Sci.* 3, 52–67. doi:10.1162/opmi_a_00026
- Brown, T. A. (2015). *Confirmatory Factor Analysis for Applied Research*. New York, NY, USA: Guilford. doi:10.14325/mississippi/9781628462265.001.0001
- Burger, K. (2010). How Does Early Childhood Care and Education Affect Cognitive Development? an International Review of the Effects of Early Interventions for Children from Different Social Backgrounds. *Early Child. Res. Q.*, 25, 140–165. doi:10.1016/j.jecresq.2009.11.001
- Byrne, B. M. (2012). *Structural Equation Modeling with Mplus Basic Concepts, Applications, and Programming*. New York, NY, USA: Routledge.
- Cadime, I., Silva, C., Ribeiro, I., and Viana, F. L. (2018). Early Lexical Development: Do Day Care Attendance and Maternal Education Matter. *First Lang.* 38 (5), 503–519. doi:10.1177/0142723718778916
- Caselli, M. C., Bates, E., Casadio, P., Fenson, J., Fenson, L., Sanderl, L., et al. (1995). A Cross-Linguistic Study of Early Lexical Development. *Cogn. Develop.* 10, 159–199. doi:10.1016/0885-2014(95)90008-x
- Caselli, C., Casadio, P., and Bates, E. (1999). A Comparison of the Transition from First Words to Grammar in English and Italian. *J. Child Lang.* 26, 69–111. doi:10.1017/s0305000998003687
- Chen, F. F. (2007). Sensitivity of Goodness of Fit Indexes to Lack of Measurement Invariance. *Struct. Equation Model.* 14, 464–504. doi:10.1080/10705510701301834
- Curran, P. J., West, S. G., and Finch, J. F. (1996). The Robustness of Test Statistics to Nonnormality and Specification Error in Confirmatory Factor Analysis. *Psycholo. Methods* 1, 16–29.
- de Hoog, M. L. A., Venekamp, R. P., van der Ent, C. K., Schilder, A., Sanders, E. A. M., Damoiseaux, R. A. M. J., et al. (2014). Impact of Early Daycare on Healthcare Resource Use Related to Upper Respiratory Tract Infections During Childhood: Prospective WHISTLER Cohort Study. *BMC Med.* 12 (2). doi:10.1186/1741-7015-12-107
- Eriksson, M., Marschik, P. B., Tulviste, T., Almgren, M., Pereira, M. P., Wehberg, S., et al. (2012). Differences between Girls and Boys in Emerging Language Skills: Evidence from 10 Language Communities. *Br. J. Develop. Psychol.* 30, 326–343. doi:10.1111/j.2044-835X.2011.02042.x
- European Commission (2009). The Provision of Childcare Services. A Comparative Review of 30 European Countries. Available from: <https://ec.europa.eu/social/main.jsp?catId=89&newsId=545&furtherNews=yes&langId=en&> (Accessed June 22, 2021).
- Feldman, H. M., Dale, P. S., Campbell, T. F., Colborn, D. K., Kurss-Lasky, M., Rockette, H. E., et al. (2005). Concurrent and Predictive Validity of Parent Reports of Child Language at Ages 2 and 3 Years. *Child Devel.* 76 (14), 856–868.
- Feldman, H. M., Dollaghan, C. A., Campbell, T. F., Kurs-Larsky, M., Janosky, J. E., and Paradise, J. L. (2000). Measurement Properties of the MacArthur Communicative Development Inventories at Ages One and Two Years. *Child. Develop.*, 71(2), 310–322. doi:10.1111/1467-8624.00146
- Fenson, L., Marchman, V., Thal, D., Dale, P., Reznick, J., and Bates, E. (2007). *MacArthur-Bates Communicative Development Inventories*. 2nd ed., 66. Baltimore, MA, USA: Paul H Brookes Publishing Co., 72.
- Fernald, A., Marchman, A., and Weisleder, A. (2013). SES Differences in Language Processing Skill and Vocabulary Are Evident at 18 Months. *Develop. Sci.* 16 (2), 234–248. doi:10.1111/desc.12019
- Galsworthy, M. J., Dionne, G., Dale, P. S., and Plomin, R. (2000). Sex Differences in Early Verbal and Non-verbal Cognitive Development. *Develop. Sci.* 3 (2), 206–215. doi:10.1111/1467-7687.00114
- Gilkerson, J., and Richards, J. A. (2009). *The Power of Talk. Impact of Adult Talk, Conversational Turns and TV during the Critical 0–4 Years of Child Development*. Boulder, CO, USA: LENA Foundation.
- Gilkerson, J., Richards, J. A., Warren, S. F., Oller, K., Russo, R., and Vohr, B. (2018). Language Experience in the Second Year of Life and Language Outcomes in Late Childhood. *Pediatrics* 142, 4. doi:10.1542/peds.2017-4276
- Goldfield, B. A., and Reznick, J. S. (1990). Early Lexical Acquisition: Rate, Content, and the Vocabulary Spurt. *J. Child Lang.* 17, 171–183. doi:10.1017/s0305000900013167
- Hadley, P. A., Rispoli, M., and Hsu, N. (2016). Toddlers' Verb Lexicon Diversity and Grammatical Outcomes. *Lang. Speech, Hearing Serv. Schools* 47, 44–58. doi:10.1044/2015_lshss-15-0018
- Hart, B., and Risley, T. R. (1995). *Meaningful Differences in the Everyday Experience of Young American Children*. Baltimore, MD, USA: Brookes.
- Henrichs, J., Rescorla, L., Schenk, J. J., Schmidt, H. G., Jaddoe, V. W. V., Hofman, A., et al. (2011). Examining Continuity of Early Expressive Vocabulary Development: The Generation R Study. *J. Speech, Lang. Hearing Res.* 54 (5), 854–869. doi:10.1044/1092-4388(2010/09-0255)
- Hoff, E. (2003). The Specificity of Environmental Influence: Socioeconomic Status Affects Early Vocabulary Development via Maternal Speech. *Child. Develop.* 74 (5), 1368–1378. doi:10.1111/1467-8624.00612
- Hoff-Ginsberg, R. (1998). The Relation of Birth Order and Socioeconomic Status to Children's Language Experience and Language Development. *Appl. Psycholinguistics* 19, 603–629. doi:10.1017/s0142716400010389
- Hu, L., and Bentler, P. M., and (1999). Cutoff Criteria for Fit Indexes in Covariance Structure analysis Conventional Criteria versus New Alternatives. *Struct. Equation Model. A Multidisciplinary J.* 6, 1–55. doi:10.1080/10705519909540118
- Hu, L., and Bentler, P. M. (1995). "Evaluating Model Fit," in *Structural Equation Modeling. Concepts, Issues, and Applications*. Editor R. H. Hoyle (London, UK: Sage), 76–99.
- Imai, M., Kita, S., Nagumo, M., and Okada, H. (2008). Sound Symbolism Facilitates Early Verb Learning. *Cognition* 109, 54–65. doi:10.1016/j.cognition.2008.07.015
- Keegstra, A., Knijff, W., Post, W., and Goorhuis-Brouwer, S. (2007). Children with Language in a Speech and Hearing Clinic: Background Variables and Extent of Language Problems. *Int. J. Pediatr. Otorhinolaryngol.* 71, 815–821. doi:10.1016/j.ijporl.2007.02.001
- Kline, R. B. (2011). *Principles and Practice of Structural Equation Modeling*. New York, NY, USA: The Guilford Press.
- Korpilahti, P., Kaljonen, A., and Jansson-Verkasalo, E. (2016). Identification of Biological and Environmental Risk Factors for Language Delay: the Let's Talk STEPS Study. *Infant Behav. Develop.* 42, 27–35. doi:10.1016/j.infbeh.2015.08.008.
- Kuvač-Kraljević, J., Blazi, A., Schults, A., Tulviste, T., and Stolt, S. (2021). Influence of Internal and External Factors on Early Language Skills: A Cross-Linguistic Study. *Infant Behav. Develop.* 63. doi:10.1016/j.infbeh.2021.101552
- Lagström, H., Rautava, P., Kaljonen, A., Riihämä, H., Pihlaja, P., Korpilahti, P., et al. (2013). Cohort Profile: Steps to the Healthy Development and Well-Being of Children (The STEPS Study). *Int. J. Epidemiol.* 42 (5), 1273–1284. doi:10.1093/ije/dys150
- Laing, C. E. (2014). A Phonological Analysis of Onomatopoeia in Early Word Production. *First Lang.* 34 (5), 387–405. doi:10.1177/0142723714550110
- Lankinen, V., Lähteenmäki, M., Kaljonen, A., and Korpilahti, P. (2018). Father-child Activities and Paternal Attitudes in Early Child Language Development: the STEPS Study. *Early Child. Development Care*. doi:10.1080/03004430.2018.1557160
- Lee, J. (2011). Size Matters: Early Vocabulary as a Predictor of Language and Literacy Competence. *Appl. Psycholinguistics* 32, 69–92. doi:10.1017/S0142716410000299
- Lekhal, R., Zachrisson, H. D., Wang, M. V., Schjølberg, S., and von Soest, T. (2011). Does Universally Accessible Child Care Protect Children from Late Talking? Results from a Norwegian Population-Based Prospective Study. *Early Child. Develop. Care* (8), 1007–1019. doi:10.1080/03004430.2010.508558
- Letts, C., Edwards, S., Sinka, I., Schaefer, B., and Gibbons, W. (2013). Socio-economic Status and Language Acquisition: Children's Performance on the New Reynell Developmental Language Scales. *Int. J. Lang. Commun. Disord.*, 48:2, 131–143. doi:10.1111/1460-6984.12004
- Longobardi, E., Spataro, P., Putnick, D. L., and Bornstein, M. H. (2017). Do o Early Nouns and Verb Production Predict Later Verb and Noun Production? Theoretical Implications. *J. Child Lang.* 44 (2), 480–495. doi:10.1017/S0305000916000064
- Luijk, M. P. C. M., Linting, M., Henrichs, J., Herba, C. M., Verhage, M. L., Schenk, J. J., et al. (2015). Hours in Non-parental Child Care Are Related to Language Development in a Longitudinal Cohort Study. *Child. Care, Health, Development* 41 (6), 1188–1198. doi:10.1111/cch.12238

- Lyytinen, P. (1999). Varhaisen Kommunikaation Ja Kielen Kehityksen Arviointimenetelmä (The Finnish Version of the MacArthur Communicative Development Inventory, CDI). *Jyväskylä: Niilo Mäki Instituutti*, 7.
- Marchman, V. A., and Bates, E. (1994). Continuity in Lexical and Morphological Development: a Test of the Critical Mass Hypothesis*. *J. Child Languages* 21, 339–366. doi:10.1017/s0305000900009302
- Marjanovič-Umek, L., Božin, N., Čermak Hauko, N., Štiglic, N., Bajc, S., and Fekonja Peklaj, U. (2016). Early Language Development: Vocabulary Comparison of Slovenian Boys and Girls. *Sodobna Pedagogika* 67, 12–6. doi:10.1007/s11199-016-0734-7
- McGregor, K. K., Sheng, L., and Smith, B. (2005). The Precocious Two-Year-Old: Status of the Lexicon and Links to the Grammar. *J. Child Lang.* 32, 563–585. doi:10.1017/s0305000905006926
- Metsämuuronen, J. (2009). Tutkimuksen Tekemisen Perusteet Ihmistieteissä 4 [The Basis for Conducting Research in Human Sciences 4]. *Jyväskylä: Gummerus Kirjapaino Oy*, 629–708.
- Ministry of Education, and Culture Early Childhood Education and Care. Available from: <https://minedu.fi/en/early-childhood-education-and-care>.
- Muthén, L. K., and Muthén, B. O. (1998–2011). *Mplus User's Guide*. Los Angeles, CA: Muthén and Muthén.
- Nunnally, J. C., and Bernstein, I. H. (1994). *Psychometric Theory*. 3rd ed.. New York, NY, USA: McGraw-Hill.
- Reilly, S., Wake, M., Bavin, E. L., Prior, M., Williams, J., Bretherton, L., et al. (2007). Predicting Language at 2 Years of Age: A Prospective Community Study. *Pediatrics* 120. doi:10.1542/peds.2007-0045
- Reilly, S., Wake, M., Ukoumunne, O. C., Bavin, E., Prior, M., Cini, E., et al. (2010). Predicting Language Outcomes at 4 Years of Age: Findings From Early Language in Victoria Study. *Pediatrics* 126. doi:10.1542/peds.2010-0254
- Rescorla, L. A., and Dale, P. S. (2013). *Late Talkers: Language Development, Interventions, and Outcomes*. Baltimore, MD, USA: Brookes.
- Rowe, M. (2008). Child-directed Speech: Relation to Socioeconomic Status, Knowledge of Child Development and Child Vocabulary Skill. *J. Child Lang.* 35, 185–205. doi:10.1017/S0305000907008343
- Scheffner Hammer, C., Morgan, P., Farkas, G., Hillemeuer, M., Bitetti, D., and Maczuga, S. (2017). Late Talkers: a Population-Based Study of Risk Factors and School Readiness Consequences. *J. Speech, Lang. Hearing Res.* 60, 607–626. doi:10.1044/2016_JSLHR-L-15-0417
- Schults, A., and Tulviste, T. (2016). Composition of Estonian Infants' Expressive Lexicon According to the Adaptation of CDI/Words and Gestures. *First Lang.* 36 (5), 485–504. doi:10.1177/0142723716648864
- Schults, A., Tulviste, T., and Konstabel, K. (2012). Early Vocabulary and Gestures in Estonian Children. *J. Child Lang.* 39, 664–686. doi:10.1017/S0305000911000225
- Simonsen, H. G., Kristoffersen, K. E., Bleses, D., Wehberg, S., and Jørgensen, R. N. (2014). The Norwegian Communicative Development Inventories: Reliability, Main Developmental Trends and Gender Differences. *First Lang.* 34 (1), 3–23. doi:10.1177/0142723713510997
- Stolarova, M., Briellmann, A. A., Wolf, C., Burke, T., and Baayen, H. (2016). Early Vocabulary in Relation to Gender, Bilingualism, Type, and Duration of Childcare. *Adv. Cogn. Psychol.* 12 (3), 130–144. doi:10.5709/acp-0192-6
- Stolt, S., Haataja, L., Lapinleimu, H., and Lehtonen, L. (2008). Early Lexical Development of Finnish Children: A Longitudinal Study. *First Lang.*, 28:3, 259–279. doi:10.1177/0142723708091051
- Stolt, S., Klippi, A., Launonen, K., Munck, P., Lehtonen, L., Lapinleimu, H., et al. (2007). Size and Composition of the Lexicon in Prematurely Born Very-Low-Birth-Weight and Full-Term Finnish Children at Two Years of Age. *J. Child Lang.* 34 (2), 283–310. doi:10.1017/S0305000906007902
- Thal, D. J., O'Hanlon, L., Clemmons, M., and Fralin, L. (1999). Validity of a Parent Report Measure of Vocabulary and Syntax for Preschool Children with Language Impairment. *J. Speech Lang. Hearing Res.* 42, 482–496. doi:10.1044/jslhr.4202.482
- Torppa, M., Lyytinen, P., Erskine, J., Eklund, K., and Lyytinen, H. (2010). Language Development, Literacy Skills, and Predictive Connections to reading in Finnish Children with and without Familial Risk for Dyslexia. *J. Learn. Disabilities* 43 (4), 308–321. doi:10.1177/0022219410369096
- Tulviste, T., and Schults, A. (2020). Parental Reports of Communicative Development at the Age of 36 Months. *Estonian Cdi-iii. First Lang.* 40 (1), 64–83. doi:10.1177/0142723719887313
- Urm, A., and Tulviste, T. (2016). Sources of Individual Variation in Estonian Toddlers' Expressive Vocabulary. *First Lang.* 36 (6), 580–600. doi:10.1177/0142723716673951
- Van Hulle, C. A., Goldsmith, H. H., and Lemery, K. S. (2004). Genetic, Environmental, and Gender Effects on Individual Differences in Toddler Expressive Language. *J. Speech, Lang. Hearing Res.* 47, 904–912. doi:10.1044/1092-4388(2004)067
- Vehkavuori, S.-M., and Stolt, S. (2019). Early Lexicon and Language Skills at 42 Months. *Clin. Linguistics Phonetics.* 33, 854–868. doi:10.1080/02699206.2019.1584721
- Wehberg, S., Vach, W., Bleses, D., Thomsen, P., Madsen, T. O., and Basbøll, H. (2007). Danish Children's First Words: Analysing Longitudinal Data Based on Monthly CDI Parental Reports. *First Lang.* 27 (4), 361–383. doi:10.1177/0142723707081723
- Wehberg, S., Vach, W., Bleses, D., Thomsen, P., Madsen, T. O., and Basbøll, H. (2008). Girls Talk about Dolls and Boys about Cars? Analyses of Group and Individual Variation in Danish Children's First Words. *First Lang.* 28 (1), 71–85. doi:10.1177/0142723707081729
- Weisleder, A., and Fernald, A. (2013). Talking to Children Matters: Early Language Experience Strengthens Processing and Builds Vocabulary. *Psychol. Sci.* 24 (11), 2143–2152. doi:10.1177/0956797613488145
- Zambrana, I. M., Pons, F., Eadie, P., and Ystrom, E. (2014). Trajectories of Language Delay from Age 3 to 5: Persistence, Recovery and Late Onset. *Int. J. Lang. Commun. Disord.* 49 (3), 304–316. doi:10.1111/1460-6984.12073
- Zubrick, R. S., Taylor, C. L., Rice, M. L., and Slegers, D. W. (2007). Late Language Emergence at 24 Months: An Epidemiological Study of Prevalence, Predictors, and Covariates. *J. Speech, Lang. Hearing Res.* 50, 1563–1592. doi:10.1044/1092-4388(2007)106

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Nylund, Ursin, Korpilahti and Rautakoski. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.