

Manuscript accepted for publication (March 2021) in Infant Behavior and Development

Influence of internal and external factors on early language skills:

a cross-linguistic study

Jelena Kuvač-Kraljević

Department of Speech and Language Pathology, Faculty of Education and Rehabilitation
Sciences, University of Zagreb, Croatia

Antonija Blaži

Doctoral Program in Hearing, Language and Speech Disorders, Department of Speech and
Language Pathology, University of Zagreb, Croatia

Astra Schults, Tiia Tulviste

University of Tartu, Estonia

Suvi Stolt

Department of Psychology and Logopedics, Faculty of Medicine, University of Helsinki, Finland

Abstract

Cross-linguistic studies can provide information about general and language specific features of language development, but relatively few such studies are available in the literature. The main aim of the present study was to investigate, from a cross-linguistic perspective, the roles of the internal factor of gender and the external factors of birth order and parental education level on development of language in 2-year-old children. We examined 351 children growing up in the three European language contexts of Croatian (N=104), Estonian (N=141) and Finnish (N=106). Information on lexical skills and word combination ability was collected using the short form of the MacArthur-Bates Communicative Development Inventories, and the influence of background factors on these aspects of language development was investigated. No significant differences were found in lexical skills or word combination ability among the three language groups. These aspects of language development varied significantly with gender, but not with the external factors. Our findings suggest that internal factors may influence early language development more than external factors.

Key words: early lexical development, syntactic development, gender, birth order, parental education, cross-linguistic study

1. Introduction

Although contrastive analysis of languages of different typologies has been one of the areas of greatest progress in psycholinguistic research during the last 40 years, large studies have focused on English (Berman, 2014). Cross-linguistic studies can provide more comprehensive information than mono-linguistic ones about general as well as language-specific features in language development. Languages differ in their lexicon and grammar, and these differences may influence language acquisition. Many mono-linguistic studies have shown substantial variation in early language development (e.g. Fenson et al., 1994; Urm & Tulviste, 2016), but given the limited number of cross-linguistic studies (e.g. Caselli et al., 1995; Devescovi et al., 2005), it remains unclear how much that variation is due to specific features of different languages or due to various background factors such as internal (biological) or external (environmental) factors. The main aim of the present study was to investigate, across three European language contexts, the effects of gender, birth order and maternal or paternal education level on language development of 2-year-old children.

1.1. Language skills of 2-year-old children

By the age of 2 years, children have typically acquired basic knowledge of their first lexicon and begun to acquire the grammar of their native language. This stands for all children regardless of language. Namely, according to studies in different language contexts, mean lexicon size at that age is roughly 300 words, although this can vary substantially from one child to another (e.g. Fenson et al., 2007; Kovačević, Kuvač Kraljević & Ceganec, 2006; Kovačević et al., 2007; Stolt et al., 2007). Expressive lexicon size in 2-year old children can vary with language

context: for example, the average 2-year-old Croatian child produces approximately 275 words (Kovačević et al., 2007), compared to 193 for an average Estonian child of the same age (Urm & Tulviste, 2016). The lexicon of 2-year-old children typically includes nouns, verbs and adjectives (Caselli, Casadio & Bates, 1999; Stolt et al., 2007), while children with larger lexicons have usually begun to acquire closed-class words such as articles, pronouns, prepositions, question words, and quantifiers, which are used to express grammatical meaning in sentences. Thus, the acquisition of closed-class words can be considered an early marker of grammatical development (Stolt, 2018).

Approximately 85-90% of 2-year-old children combine words into clauses at least sometimes (Fenson et al., 2007; Stolt et al., 2009), and many children may express themselves using even longer sentences. The typical utterance length of the 2-year-old Croatian child is four words, while the mean value of the three longest utterances is six morphemes (Kovačević et al., 2007). Similarly, 2-year-old children in languages with rich inflectional morphology, such as Estonian and Finnish, have also begun to acquire morphological inflections. For example, Finnish children at that age have acquired roughly six inflectional types (Stolt et al., 2009). The use of morphological inflections enables the increasingly specific expression of different grammatical features.

1.2. Influence of background factors on language development

Many studies have investigated the influence of series of non-linguistic factors on early language development although usually from a mono-linguistic perspective (e.g. Eriksson et al., 2006; Fenson et al., 1994; 2007). These background factors can be biologically (internal) and

environmental (external) determined. Among many of them, the effects of gender, birth order and maternal/paternal education level have been particularly well studied. The consensus seems to be that gender influences early language development: girls tend to reach language milestones faster, producing their first words, reaching a vocabulary spurt and using word combinations earlier than boys (Eriksson, 2006; Eriksson et al., 2011; Frota et al., 2016; Kovačević, Kuvač Kraljević & Capanec, 2006; Silva et al., 2017; Stolt et al., 2008). Around 3 years of age, girls outperform boys on a wide range of syntactic measures, including mean utterance length as well as sentence structure and complexity (Le Normad, Parisse & Cohen, 2008). Three potential explanations for the gender difference in early language development have been proposed (Tse et al., 2002). One explanation is biological: neuroanatomic differences between males and females may enable faster language processing among females. A second explanation is psychological: females are generally more emotionally expressive than males. A third explanation is socio-contextual: girls are encouraged to express themselves more than boys. Whatever the cause of the gender difference in language development, the difference is not observed across all language skills or throughout the entire age span. For example, no gender difference was found in a study of receptive language development (Eriksson et al., 2011; Luijk et al., 2015; Stolt et al., 2008), and another study (Eriksson, 2006) found that gender was a significant predictor of lexical and syntactic development between the ages of 16 to 30 months, but not earlier. Thus, the available literature suggests that at certain ages, girls demonstrate more advanced language skills with respect to expressive vocabulary and morpho-syntax.

Birth order may also influence early language development because first-borns may begin to acquire their language chronologically earlier than later-borns (Berglund, Eriksson &

Westerlund, 2005; Fenson et al., 1994). First-borns may have more opportunities than later-born children to communicate with adults in diverse situations (Hoff, 2003; Brooks & Kempe, 2012). Berglund, Eriksson & Westerlund (2005) reported that first-born children 18 months old performed better on various receptive and expressive language measures than their later-born peers. First-borns have also been reported to produce more complex and more diverse syntactic structures as well as longer utterances than later-borns (Szagun, Stumper & Schramm, 2009). However, differences between first- and later-born children may be limited to certain language skills and certain ages. For example, Schults, Tulviste & Konstabel (2012) found birth-order differences in noun production only among children aged 8 to 16 months. Zambrana, Ystrom & Pons (2012) reported better performance by first-borns on measures of receptive language only between the ages of 18 and 36 months. Other work also suggests that birth order significantly influences language skills only during the first three years of life (Fenson et al., 1994). To complicate things further, several studies have failed to detect differences in language development between first- and later-borns (Tulviste, 2006; Tulviste & Schults 2019; Westerlund & Lagerberg, 2008) and not all studies that detect influence of birth order have reported a first-born advantage (Oshima-Takane, Goodz & Deverensky, 1996). Bornstein, Leach & Haynes (2004) using three different methods - maternal report, sampling of child spontaneous speech and formal language testing - have found that firstborn`s vocabulary competence exceed secondborn`s only in maternal reports but not in two other methods. It seems that because firstborns enjoy more personal interaction with their mother than laterborn children, the mothers think that firstborns have better language skill. Moreover, Oshima-Takana et al. (1996) found that second-born children have opportunity to overhear conversations between caregivers

and their older first-born siblings and learn from it. Bornstein et al (2004) emphasize that laterborn children more often than firstborns participate in multichild interactive contexts. Exposure to all these more sophisticated language situations in multiparty conversation have positive effect on secondborn language development. In conclusion, the literature has not unequivocally determined whether birth order affects language development and, if it does, in what direction, at what ages and with respect to which language skills.

Maternal education level can positively affect a child's language development (Basit et al., 2015; Pace et al., 2016; Stolt et al., 2007; Westerlund & Lagerberg, 2008; Zauche, 2016). The impact of maternal education may even increase with the child's age and continue throughout the child's schooling (Vasilyeva, Waterfall & Huttenlocher, 2008). However, other studies have come to different conclusions. Luijk et al. (2015) failed to identify any significant influence of maternal education level on several language variables among children aged 1 to 6 years. Maternal education level may influence only certain language variables, such as receptive and expressive vocabulary, and it may operate only at younger ages (Berglund, Eriksson & Westerlund 2005). It is possible that such influence may be evident only at the extremes of minimal or maximal education levels (Letts et al., 2013). Even less is known about the potential influence of paternal education level on children's language development. Clarifying the effects of parental education on such development is important, not least because the effects may be cumulative (van Houdt et al, 2019).

1.3. Cross-linguistic perspective on the influence of background factors on language development

In the present study, the roles of the internal factor of gender and the external factors of birth order and parental education level in early language development were investigated from a cross-linguistic perspective. Few cross-linguistic studies have examined the effect of background factors on language development. One study of 13,000 children from ten non-English language communities (Austrian-German, Basque, Croatian, Danish, Estonian, French, Galician, Slovenian, Spanish and Swedish) found that girls outperformed boys in early language and communication development, and the performance difference increased with age (Eriksson et al., 2011). A study of 512 children aged 2 years old in Italy and Finland found that maternal education level was positively associated with the lexical composition of children who had been born at a gestational age of <32 weeks (Stolt et al., 2017). In both mentioned studies show consistency between languages i.e. the analyzed background factors (gender and maternal education) were confirmed as robust predictor of language development.

The present study explored the role of background factors on language development in 2-year-old children in Croatia, Estonia, and Finland. Croatian belongs to the family of South-Slavic languages. Case, number and gender marking are used for nouns and adjectives, while person, number and tense markings are used for verbs. For noun inflection, there are three genders (masculine, feminine and neutral), two numbers (singular and plural) and seven cases (nominative, genitive, dative, accusative, vocative, locative and instrumental). Noun inflections are expressed using suffixes or, occasionally, infixes (Kovačević, Palmović & Hržica, 2009). Verb forms are expressed with the help of six categories: person (first, second and third), number

(singular and plural), tense (simple-present, aorist, imperfect, compound-perfect, plusquamperfect, future I and future II), aspect (perfective or imperfective), mode (infinitive, imperative and two conditionals) and voice (active or passive). The verbal system can be further divided into seven verbal classes based on infinitive and present forms (Barić et al., 1995). The canonical word order is subject-verb-object (SVO), but word order is relatively free.

Both Estonian and Finnish are Finno-Ugric languages sharing many features. Both languages have a rich morphological inflectional system for nominals (i.e. nouns, adjectives, numerals, pronouns) and verbs (Stolt et al., 2009; Toivainen, 1997; Tulviste & Schults, 2019). Cases, of which Estonian has 14 and Finnish 15, are used for nominals, as are singular and plural markings, most of which are suffixes. No articles or gender markings are used for nominals. Both languages use subject-verb agreement for verbs. Tenses include present, preterit, perfect, and plusquamperfect. In addition, the morphology for finite verbs can express voice (active or passive) and mood (indicative, imperative, conditional, potential). Common word order for both languages is SVO, but word order may change according to expressional needs. Despite their many similarities, Estonian and Finnish differ in lexicon and phonology.

Children growing up speaking Croatian, Estonian or Finnish are exposed to different language contexts. For example, speech addressed to Estonian children generally contains more imperatives than speech addressed to American, Finnish or Swedish children (Junefelt & Tulviste, 1997; Tulviste, Mizera & De Geer, 2004). Children growing up in these language contexts are likely to have different numbers of siblings: the fertility rate, which indicates the average number of children born to one woman, is 1.3 in Finland, where the rate has been declining (www.statista.com/statistics/530225/fertility-rate-in-finland/), compared to 1.7 in Estonia

(www.statista.com/statistics/377028/fertility-rate-in-estonia/) or 1.6 in Croatia, where half of all families have at least two children (www.statista.com/statistics/348296/fertility-rate-in-croatia/). Children in these households may have parents who differ substantially in education level: 38% of the general population in Finland has attained some level of tertiary (higher) education, compared to 36% in Estonia and only 17% in Croatia (www.dzs.hr/Hrv_Eng/publication/2016/SI-1583.pdf; www.estonica.org/en, www.stat.fi/til/perh/2015/02/perh_2015_02_2016-11-25_kat_001_en.html).

The main motivation for this study arises from the contradictory evidence of previous studies regarding the role of internal and external factors in early language development. So far only a few papers have aimed to combine multiple factors in a single language (Tulviste, 2006) or explore the role of a single factor from a cross-linguistic perspective (Eriksson et al, 2011). The purpose of this paper is to explore multiple factors from cross-linguistic perspective addressing the following specific research questions:

1. Do lexical skills and word combination ability of 2-year-old children differ across Croatian, Estonian and Finnish language contexts?
2. Does the influence of gender, birth order or parental education level on lexical skills or word combinations of 2-year-old children differ across Croatian, Estonian or Finnish language contexts?

Method

2.1. Participants

Study participants were 351 children with no diagnosed language, hearing, communication or speech pathologies. All participants were native speakers of Croatian ($N = 104$), Estonian ($N = 141$) or Finnish ($N = 106$). Data on children's mean age, gender, and birth order as well as the education levels of mothers and fathers are shown in Table 1. Parental education level was categorized as elementary (8-9 years), secondary (12 years) or high (15+ years). Children whose parents had only elementary education were excluded from the study because too few were recruited to allow comparison across countries.

Table 1. Data on participating children and their parents.

Language	N	Age M (SD)	Gender		Birth order		Maternal education		Paternal education	
			F	M	First	Later	Secondary	High	Secondary	High
Croatian	104	24 (0.8)	49	55	48	56	89	15	67	37
Estonian	141	24 (0.7)	73	68	40	101	53	77	87	43
Finnish	106	24 (0.2)	50	56	60	46	27	79	44	57
Total	351	24 (0.6)	172	179	148	203	169	171	198	137

Values are N or n. M - mean, SD - standard deviation. Education information was missing for 11 mothers and 16 fathers.

The mean age of the children was 24 months. There were no significant differences in gender distribution across the groups ($\chi^2 = 7.24, p > .05$), but there were significant differences in birth order ($\chi^2 = 20.75, p < .001$): the Finnish group contained the highest proportion of first-borns, and the Estonian group the smallest proportion. There were also significant differences in education levels of mothers ($\chi^2 = 82.59, p < .001$) and fathers ($\chi^2 = 14.60, p < .001$): the Croatian sample included the highest proportion of mothers at a secondary education level, and the

Finnish sample included the smallest proportion. Similar proportions of Croatian and Estonian fathers were at a secondary education level, and this proportion was smallest in the Finnish sample.

2.2. Instrument

Data were collected using the short version of the MacArthur-Bates Communicative Development Inventories (Toddler version) (Fenson et al, 2000), hereafter referred to as the short form of the CDI. This measure consists of a vocabulary checklist and a section that asks the parent whether his or her child has begun to combine words. In this second section, parents are given three response options: “not yet”, “sometimes” and “often”. This scale has been adapted and normed to Croatian (Kuvač Kraljević, Capanec & Kovačević, in press), Estonian (Urm and Tulviste, in press) and Finnish (Stolt & Vehkavuori, 2018). All three versions follow the structure of the original measure. However, the vocabulary checklist contains 103 words in the Croatian version but 100 in the Estonian and Finnish versions. The number of words in different semantic categories is comparable across all three language versions (Table 2).

Table 2. Structures of the original English version of the short form of the MacArthur Communicative Development Inventories (Toddler version) and the versions adapted to Croatian, Estonian and Finnish.

Semantic category	English (original)	Language version		
		Croatian	Estonian	Finnish
Sound effects	5	5	2	5
Animals	6	6	10	6
Vehicles	3	3	4	2
Toys	3	3	1	3
Food and drink	8	8	11	8
Clothing	4	4	5	4
Body parts	4	4	3	3

Small household items	7	7	9	9
Furniture	5	5	3	4
Outside things and places to go	7	9	5	7
People	3	3	4	4
Games and routines	5	5	3	5
Action words	15	15	14	14
Helping verbs	3	3	0 *	0 *
Descriptive words	8	9	9	8
Words about time	3	3	1	5
Pronouns	4	4	4	4
Question words	1	1	3	1
Prepositions	3	3	5	3
Quantifiers and articles	2	2	3	2
Connecting words	1	1	1	3
Total	100	103	100	100

Values are n.

* There are no helping verbs in Estonian and Finnish.

4.3. Procedure

This study is based on parental reporting of child language development. Therefore, parents or other caregivers who agreed to participate in the study and provided informed consent, completed the short form of the CDI when their children were a mean of 24 months old. In the Croatian sample, 85% of scales were completed by mothers, 13% by fathers and 2% by a grandmother or other caregiver. In the Estonian sample, 99% of scales were completed by mothers and 1% by fathers. In the Finnish sample, 85% of scaled were completed by mothers, 2% by fathers and 13% by the mother and father together.

4.4. Data analysis

Inspection of the distribution of total number of words produced in each language showed that all three distributions were platycurtic, violated one assumption of a normal distribution. Since other normality assumptions were met and the distributions were symmetrical, parametric analysis was conducted. Differences in the numbers of words between language groups and for different values for internal or external factors were assessed for significance using ANOVA. Children's ability to combine words was scored as 0 ("not yet") or 1 ("sometimes/often"), and score differences across the three languages were assessed using a chi-squared test.

Logistic regression was used to examine whether vocabulary predicts ability to combine and whether gender, birth order, parental education, or language predicts that ability. The predictor variables were tested *a priori* to verify the validity of the assumption of proportionality and absence of multicollinearity.

5. Results

5.1. Lexical development and word combination skills of 2-year-old children in three European languages

Descriptive statistics for the number of words produced across all three languages are presented in Table 3. In all three languages, the 2-year-old-children produced an average of 56 words, and this mean value did not differ significantly among the groups, based on one-way ANOVA [$F(2,$

348) = 0.46; $p = 0.63$]. Therefore, there is no statistically significant difference in the number of words among the languages.

Table 3. Descriptive data on the numbers of words produced by children in each language group

Language	N	Min	Max	M	SD	Skewness		Kurtosis	
						Value	SE	Value	SE
Croatian	104	0	103	57.85	32.92	-0.098	0.237	-1.380	0.469
Estonian	141	0	100	54.70	32.88	-0.200	0.204	-1.430	0.406
Finnish	106	4	100	53.93	28.02	-0.319	0.235	-1.123	0.465

The frequencies and percentages of children who combined words are presented in Table 4. Across the entire sample, 56 children (16%) had not yet begun to combine words, and this percentage did not differ significantly across the languages, based on the chi-squared test ($\chi^2 = 3.15, p=0.21$). This indicates that parents in all three language groups assessed their children to have similar word combination skills.

Table 4. Parental assessment of children's ability to combine words

Language	<i>Not yet combining</i>	<i>Sometimes or often combining</i>	<i>Total</i>
Croatian	12 (11.5%)	92 (88.5%)	104 (100.0%)
Estonian	28 (19.9%)	113 (80.1%)	141 (100.0%)
Finnish*	16 (15.2%)	89 (84.8%)	105 (100.0%)
Total	56 (16.0%)	294 (84.0%)	350 (100.0%)

* Data missing for one child.

The average number of words in children’s vocabulary and their word combination ability are presented in Table 5. Generally, the more words children had acquired, the more often they used word combinations.

Logistic regression for predicting word combining from vocabulary showed that vocabulary is a significant predictor of combining ($\beta = 0.55$; $p < 0.001$) and that model correctly classifies 84% of cases.

Table 5. Numbers of words used by children stratified by their ability to combine words*

Number of words \ Word combining**	N	M	SD	Min	Max
Not yet	56	15.68	14.71	0	81
Sometimes or often	294	62.93	27.97	2	103

* All children in the sample.

** Data missing for one child.

5.2 Effects of internal and external factors

To examine the effects of gender and language on vocabulary, two-way ANOVA 2x3 was used with factors of gender (male, female) and language (Croatian, Estonian, Finnish) on the children’s lexical ability. Results show that the factor gender was statistically significant [$F(1, 345) = 22.93$; $p < 0.001$], but the second factor language was not [$F(2, 345) = 0.43$; $p = 0.65$], and the interaction was not statistically significant [$F(2, 345) = 0.12$; $p = 0.89$]. Then ANOVA of the data for each language separately was conducted. Results showed a significant effect of gender in each language: [Croatian, $F(1, 102) = 5.65$; $p = 0.02$]; [Estonian, $F(1, 139) = 11.26$; $p < 0.001$]; and

Finnish, [$F(1, 104) = 7.46; p=0.01$]). In every language separately, girls used more words than boys.

In none of the languages did birth order [Croatian, $F(60, 103) = 0.67; p=0.92$]; Estonian $F(70, 140) = 0.86; p=0.74$; Finnish, $F(61, 105) = 1.03; p=0.47$] or parental education level significantly affect children's lexical ability [maternal education: Croatian $F(60, 103) = 0.88; p=0.68$]; Estonian, $F(66, 129) = 0.87; p=0.71$]; Finnish, $F(61, 105) = 0.94; p=0.60$]; paternal education: [Croatian, $F(60, 103) = 0.79; p=0.80$]; Estonian, $F(66, 129) = 1.52; p=0.05$]; and Finnish, $F(59, 100) = 1.24; p=0.24$].

Figure 1 presents vocabulary scores by gender for each language. The mean values for boys were 49.90 for Croatian, 46.04 for Estonian, and 46.30 for Finnish, lower than the corresponding values for girls of 64.93, 63.99, and 60.75.

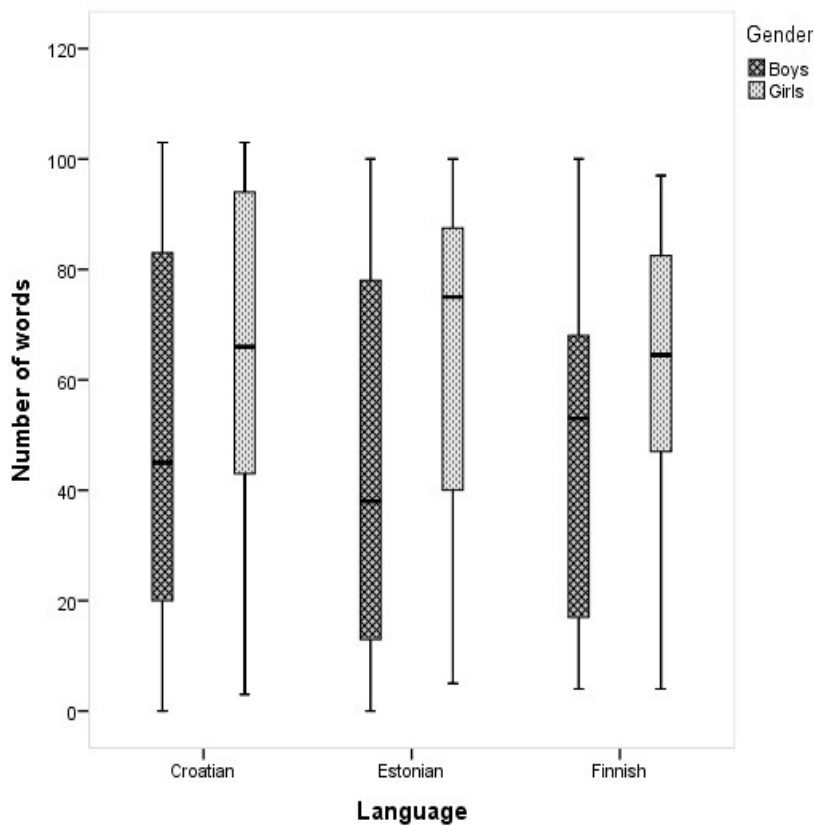


Figure 1. Boxplots showing vocabulary scores in each language by gender.

Logistic regression was conducted to investigate whether gender, birth order, parental education level and language predicted word combination ability. Only gender proved to be a significant predictor: boys were significantly less likely than girls to combine words (odds ratio 0.455, 95%CI 0.289 to 0.781; Wald $\chi^2(1) = 13.167$, $p < 0.001$; Figure 2). All other factors are not significant predictors (birth order [Wald $\chi^2(1) = 0.035$, $p = 0.85$], maternal education [Wald $\chi^2(1) = 0.081$, $p = 0.78$], paternal education [Wald $\chi^2(1) = 0.043$, $p = 0.84$], language [Wald $\chi^2(1) = 0.384$, $p = 0.54$]).

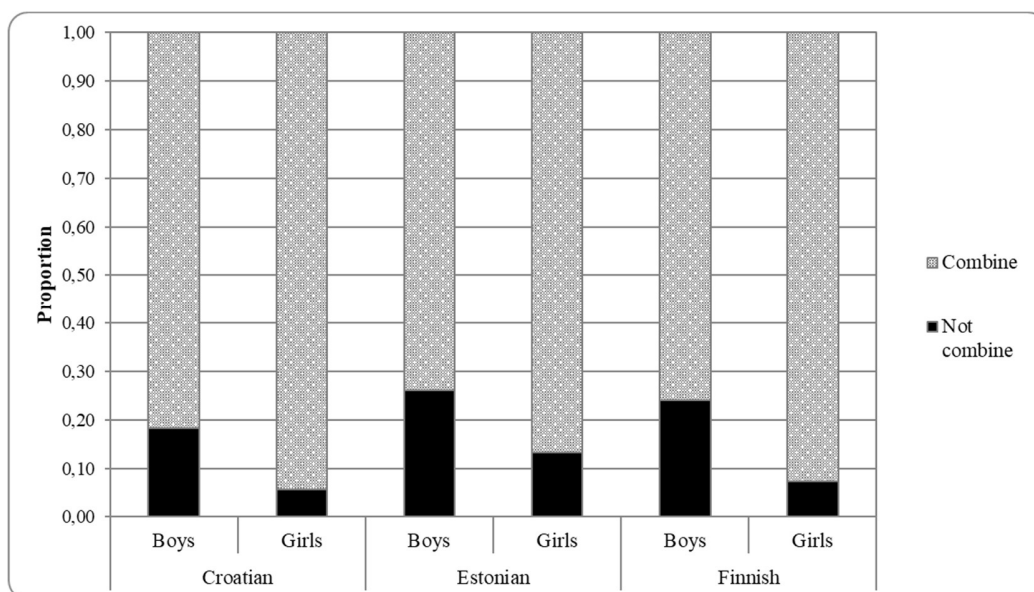


Figure 2. Proportions of the children who are able and not able to combine words by gender and language.

The predictive role of factors on word combination ability with respect to each language is as follows: Croatian: gender [Wald $\chi^2(1) = 3.629, p=0.06$], birth order [Wald $\chi^2(1) = 0.857, p=0.36$], maternal education [Wald $\chi^2(1) = 0.157, p=0.69$], paternal education [Wald $\chi^2(1) = 3.117, p=0.08$]; Estonian: gender [Wald $\chi^2(1) = 3.518, p=0.06$], birth order [Wald $\chi^2(1) = 0.244, p=0.62$], maternal education [Wald $\chi^2(1) = 0.329, p=0.57$], paternal education [Wald $\chi^2(1) = 0.054, p=0.82$] and Finnish: gender [Wald $\chi^2(1) = 3.674, p=0.06$], birth order [Wald $\chi^2(1) = 0.100, p=0.75$], maternal education [Wald $\chi^2(1) = 0.099, p=0.75$] and paternal education [Wald $\chi^2(1) = 1.079, p=0.30$]. None of these factors are significant.

6. Discussion

Cross-linguistic studies can provide information on which features are general and which are language-specific during language development. They can also provide data concerning the (in)consistency of differences in language development due to specific internal and external factors. The present cross-linguistic study focused on the possible effects of internal and external factors on lexical and syntactic skills across three European languages of different typology. Specifically, the aim of this study was to analyze the effects of gender, birth order and maternal/parental education on the lexical ability and word combination skills of 2-year-old children speaking Croatian, Estonian or Finnish, based on the short form of the CDI.

Our data indicate that although the three samples showed differences in birth order and maternal/paternal education level reflective of the local demographics, parents in all three groups indicated that their children used an average of 100 i.e. 103 words. This data stands for expressive vocabulary list i.e. those that parents more reliably assess than use of receptive vocabulary (Tomasello and Mervis, 1994). Similar proportions of children in all three samples were judged by their parents as able to produce multi-word utterances. These results suggest that, independently of the language, children around 2 years old begin to face morphological complexity and begin to combine words. Our results are consistent with the finding that children around the age of 2 progressively improve their syntactic ability by adding new grammatical structures, they complexify the relationships among those structures and they lengthen their utterances (Fenson et al., 2007; Stolt et al, 2007; 2009).

Although the short version of the CDI is a screening method which provides limited information about early lexical ability, our results confirm the previously reported relationship

between expressive lexical ability and syntactic development (Maital et., 2000; Stolt et al., 2009; Thordardottir, Weismer & Evans, 2002). Namely, our analysis shows that with high certainty syntactic development can be predicted by lexical development, and even more parents themselves recognize that once a child attains a lexicon with sufficient words, he or she also attempts to combine them. Consistent with this, we found that across all three languages, 2-year-old children who combined words had vocabularies approximately four times larger than those who did not. This provides additional, cross-linguistic evidence of the lexical-syntactic relationship, such that lexical development can predict syntactic development.

The present cross-linguistic study suggests that gender significantly affects lexical ability as well as word combination ability: in all three languages, girls showed a larger expressive vocabulary than boys, and they were more likely to combine words into clauses. These results are consistent with those from a study of Portuguese-speaking children (Frota et al., 2016) and from a cross-linguistic study (Eriksson et al., 2011). In Eriksson's et al. cross-linguistic study ten genetically and typological different languages were included (Austrian-German, Basque, Croatian, Danish, Estonian, French, Galician, Slovenian, Spanish and Swedish) and in all of them the factor gender has been confirmed as the robust factors that do not change between language communities. All these findings suggest that biological factors such as gender may influence early lexical and syntactical skills more than environmental factors. This biological determination may in turn lead to a more stimulating language environment for girls, which consolidates and increases their advantage over boys. Studies have suggested that mothers and fathers adopt a different language style with their daughters than with their sons, expressing emotions more frequently and varying their language more often with daughters (Adams, Kuebli & Boyle 1995).

Mothers tend to speak more and employ more supportive speech with their daughters than with their sons (Leaper, Anderson & Sanders, 1998). In light of the impact of social skills it is plausible that from an early age, girls in Croatian, Estonian and Finnish settings have more opportunities to engage in social situations that provide richer and more varied language input (Junefelt & Tulviste, 1997; Tulviste, Mizera & De Geer, 2004).

Our results do not support studies that found a significant effect of birth order and maternal/paternal education level on the language development of 2-year-old children (Brooks & Kempe, 2012; Zambrana, Ystrom & Pons, 2012). Our negative findings may be explained, at least in part, by the possibility that these environmental factors do not influence the particular language measures that we assessed or do not influence children in the age range of our sample. For example, Schults, Tulviste & Konstabel (2012) found difference in expressive vocabulary between first and second born children just between 8 and 16 months but not later.

We conclude from the present cross-linguistic data that gender, but not birth order or maternal/paternal education level, significantly influences lexical ability and word combination skills around 2 years of age. Our analysis identifies gender as a robust factor in early language development, independently of language typology.

Nevertheless, our findings should be interpreted with caution because of several limitations. First, we included only parents with secondary and tertiary education levels, so it is unclear whether our results can be generalized to households where parents have a primary education level. Second, the CDI provides only limited information about early lexical and syntactic skills. For example, it cannot elucidate how children apply morpho-syntactic rules to combine words. In addition, estimating total vocabulary size from the original English CDI is highly

non-linear: when toddlers know 90% of the words on the CDI, their vocabulary is likely to be about three times larger (Mayor & Plunkett, 2011). Therefore, future studies should apply other instruments for estimating total vocabulary size across languages. Third, the present study did not examine all internal or external factors that contribute to early lexical and syntactic development. Future studies should consider other variables, such as the quantity and quality of language input provided by the mother, father and older siblings. This is especially important since parents may assess linguistic skills of their sons differently from those of their daughters. Future studies should also investigate differences in children's lexicons with respect to typological differences across languages, as well as the role of each word type in the production of early sentence structures.

7. Conclusion

The present study was motivated by contradictory findings of previous studies regarding the contribution of various internal and external factors to language development. In addition, previous research has focused either on a single language, usually English, or it has taken a cross-linguistic perspective but focused on only one factor. The present study integrated data from three languages to examine the influence of one internal factor (gender) and two external factors (birth order and maternal/paternal level of education) on early lexical and syntactic skills.

This study, based on the short form of MacArthur-Bates Communicative Developmental Inventories, found that parents in all three language groups gave similar estimates for how many words their 2-year-old children knew and whether or not their children could combine those words. Our results support the idea that children's ability to combine words depends on their

lexical ability: when children's vocabulary achieves a critical size, they begin to produce on a syntactic level.

The present study found that in each of the language groups, gender but not birth order or parental education level was associated with 2-year-old children's number of words and ability to combine words. This gender bias likely reflects differences in biology, exposure to language environments and approach to verbal situation. Our findings suggest that, independently of language typology, internal factors may affect language development more strongly than external factors around 2 years of age.

References

Adams, S., Kuebli, J., Boyle, P. A. (1995). Gender differences in parent-child conversations about past emotions: A longitudinal investigation. *Sex Roles*, 33(5-6), 303-323.

Barić, E., Lončarić, M., Malić, D., Pavešić, S., Peti, M., Zečević, V., Znika, M. (1997.) Hrvatska gramatika [The Croatian grammar]. Školska knjiga. Zagreb.

Basit, T. N., Hughes, A., Iqbal, Z., Cooper, J. (2015). The influence of socio-economic status and ethnicity on speech and language development. *International Journal of Early Years Education*, 23, 115-133. <https://doi.org/10.1080/09669760.2014.973838>.

Berglund, E., Eriksson, M., Westerlund, M. (2005). Communicative skills in relation to gender, birth order, childcare and socioeconomic status in 18-month-old children. *Scandinavian Journal of Psychology*, 46, 485-491. doi: 10.1111/j.1467-9450.2005.00480.x

Berman, R. (2014) Cross-linguistic comparisons in child language research. *Journal of Child Language*, 41(1), 26-37. 10.1017/S0305000914000208

Bornstein, M. H., Leach, D. B., Haynes, M. (2004) Vocabulary competence in first- and secondborn siblings of the same chronological age. *Journal of Child Language*, 31(4), 855-873. doi: 10.1017/S0305000904006518

Brooks, P. J., Kempe, V. (2012). *Language Development*. Chichester: John Wiley & Sons.

Caselli, M. C., Bates, E., Casadio, P., Fenson, J., Fenson, L., Sanderl, L., Weir, J. (1995). A cross-linguistic study of early lexical development. *Cognitive Development*, 10(2), 159–199. [https://doi.org/10.1016/0885-2014\(95\)90008-X](https://doi.org/10.1016/0885-2014(95)90008-X)

Caselli, M. C., Casadio, P., Bates, E. (1999). A comparison of the transition from first words to grammar in English and Italian. *Journal of Child Language*, 26(1), 69-111. 10.1017/s0305000998003687.

Croatian Bureau of Statistics. Statistical Report. https://www.dzs.hr/Hrv_Eng/publication/2016/SI-1583.pdf Accessed 13 November 2019.

Devescovi A., Caselli M. C., Marchione D., Pasqualetti P., Reilly J., Bates E. (2005) A crosslinguistic study of the relationship between grammar and lexical development. *Journal of Child Language*, 32(4), 759-786. 10.1017/s0305000905007105

Eriksson, M. (2006). Sex differences in language development as a topic for cross-cultural comparisons. In M. Eriksson, (Eds.), *Proceedings from the First European Network Meeting on the Communicative Development Inventories* (103-114), Gävle: University of Gävle.

Eriksson, M., Marschik, P. B., Tulviste, T., Almgren, M., Perez Pereira, M., Wehberg, S., Marjanovič-Umek, Lj., Gayraud, F., Kovačević, M., Gallego, C. (2011). Differences between girls and boys in emerging language skills: Evidence from 10 language communities. *British Journal of Developmental Psychology*, 30, 326-343. doi: 10.1111/j.2044-835X.2011.02042.x.

http://www.estonica.org/en/Society/Population/Family_and_birth_of_children/ Accessed 13 November 2019.

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J., Tomasello, M., Mervis, C. B., Stiles, J. (1994) Variability in Early Communicative Development. *Monographs of the Society for Research in Child Development*, 59, 1-185.

Fenson, L., Pethick, S., Renda, C., Cox, J. L., Dale, P. S., Reznick, J. S. (2000). Short-form versions of the MacArthur Communicative Development Inventories. *Applied Psycholinguistics*, 21, 95-116. <https://doi.org/10.1017/S0142716400001053>.

Fenson, L., Marchman, V. A., Thal, D. J., Dale, P. S., Reznick, J. S., Bates, E. (2007). MacArthur–Bates Communicative Development Inventories: User’s guide and technical manual (2nd ed.). Baltimore, MD: Brookes Publishing.

Frota, S., Butler, J., Correia, S., Severino, C., Vicente, S., Vigarrio, M. (2016). Infant communicative development assessed with the European Portuguese MacArthur-Bates Communicative Development Inventories short forms. *First Language*, 36, 525-545. doi: 10.1177/0142723716648867.

Hoff, E. (2003). The Specificity of Environmental Influence: Socioeconomic Status Affects Early Vocabulary Development Via Maternal Speech. *Child Development*, 74, 1368-1378. doi: 10.1111/1467-8624.00612.

Junefelt, K., Tulviste, T. (1997) Regulation and Praise in American, Estonian, and Swedish Mother-Child Interaction. *Mind Culture and Activity*, 4(1), 24-33. 10.1207/s15327884mca0401_3

Kovačević, M., Kuvač Kraljević, J., Cepanec, M. (2006). Sex differences in lexical and grammatical development in Croatian. In M. Eriksson (Ed.), *Proceedings from the First European Network Meeting on the Communicative Development Inventories* (5-16), Gävle: University of Gävle.

Kovačević, M., Jelaska, Z., Kuvač Kraljević, J., Cepanec, M. (2007). *Komunikacijske razvojne ljestvice (KORALJE)*. [The Croatian versions of the MacArthur-Bates Communicative Development Inventories]. Jastrebarsko: Naklada Slap.

Kovačević, M., Palmović, M., Hržica, G. (2009) The Acquisition of Case, Number and Gender in Croatian. Stephany, Ursula, Voeikova, Maria (Eds) *Development of Nominal Inflection in First Language Acquisition: A Cross-Linguistic Perspective*. (153-177). Berlin: Mouton De Gruyter.

Kuvač Kraljević, J., Cepanec, M., Kovačević, M. (in press). *Kratka verzija komunikacijske razvojne ljestvice - kratke KORALJE*. [The Croatian short form versions of the MacArthur-Bates Communicative Development Inventories]. Jastrebarsko: Naklada Slap.

Leaper, C., Anderson, K. J., Sanders, P. (1998) Moderators of gender effects on parents' talk to their children: A meta-analysis. *Developmental Psychology*, 34(1), 3-27. <http://dx.doi.org/10.1037/0012-1649.34.1.3>.

Le Normand, M. T., Parisse, C., Cohen, H. (2008). Lexical diversity and productivity in French preschoolers: developmental, gender and sociocultural factors. *Clinical Linguistics & Phonetics*, 22, 47-58. doi: 10.1080/02699200701669945.

Letts, C., Edwards, S., Sinkas, I., Schaefer, B., Gibbons, W. (2013). Socio-economic status and language acquisition: children's performance on the new Reynell Developmental Language Scales. *International Journal of Language and Communication Disorders*, 48, 131-143. doi: 10.1111/1460-6984.12004.

Luijk, M. P. C. M., Linting, M., Henrichs, J., Herba, C. M., Verhage, M. L., Schenk, J. J., Arends, L. R., Raat, H., Jaddoe, V. W. V., Hofman, A., Verhulst, F. C., Tiemeier, H., Van IJzendoorn, M. H. (2015). Hours in non-parental childcare are related to language development in a longitudinal cohort study. *Child Care Health and Development*, 41(6), 1188-1198. doi:10.1111/cch.12238.

Maital, S. L., Dromi, E., Sagi, A., Bornstein, M. H. (2000). The Hebrew Communicative Development Inventory: Language specific properties and cross-linguistic generalizations. *Journal of Child Language*, 27 (1), 43-67. doi: 10.1017/S0305000999004006.

Mayor, J., Plunkett, K. (2011) A statistical estimate of infant and toddler vocabulary size from CDI analysis. *Developmental Science*, 14(4), 769-785.

Oshima-Takane, Y., Goodz, E., Deverensky, J. L. (1996). Birth order effects on early language development: Do secondborn children learn from overheard speech? *Child Development*, 67(2), 621-634. <https://doi.org/10.2307/1131836>.

Pace, A., Luo, R., Hirsh-Pasek, K., Michnick Golinkoff, R. (2016). Identifying Pathways Between Socioeconomic Status and Language Development. *The Annual Review of Linguistics*, 3, 285-308. doi: 10.1146/annurev-linguistics-011516-034226.

Schults, A., Tulviste, T., Konstabel, K. (2012). Early vocabulary and gestures in Estonian children. *Journal of Child Language*, 39, 64 - 686. doi: 10.1017/S0305000911000225.

Silva, C., Cadime, I., Ribeiro, I., Santos, S., Santos, A. L., Viana, F. L. (2017). Parents' reports of lexical and grammatical aspects of toddlers' language in European Portuguese: Developmental trends, age and gender differences. *First language*, 37 (3), 267-284. <https://doi.org/10.1177/0142723716689274>

Statista <https://www.statista.com/> accessed 21 July 2020.

Statistics Finland. National Statistical Service in Finland.

https://www.stat.fi/til/perh/2015/02/perh_2015_02_2016-11-25_kat_001_en.html Accessed 13 November 2019.

Stolt, S., Klippi, A., Launonen, K., Munck, P., Lehtonen, L., Lapinleimu, H., Haataja, L. & the PIPARI study group (2007). Size and composition of the lexicon in prematurely born very-low-birth-weight and full-term Finnish children at two years of age. *Journal of Child Language*, 34, 283-310. doi: 10.1017/s0305000906007902.

Stolt, S., Haataja, L., Lapinleimu, H., Lehtonen, L. (2008). Early lexical development of Finnish children - a longitudinal study. *First Language*, 28, 259-279. <https://doi.org/10.1177/0142723708091051>

Stolt, S., Haataja, L., Lapinleimu, H., Lehtonen, L. (2009). Associations between lexicon and grammar at the end of the second year. *Journal of Child Language*, 36 (4), 779-806. doi:10.1017/S0305000908009161.

Stolt, S., Savini, S., Guarini, A., Caselli, M., Matomäki, J., Lapinleimu, H., Haataja, L., Lehtonen, L., Alessandroni, R., Faldella, G. & Sansavini, A. (2017) Does the native language influence lexical composition in very preterm children at two years of age? A cross-linguistic comparison study of Italian and Finnish children. *First language*, 1-27. DOI: 10.1177/0142723717698006

Stolt, S. (2018). Early lexicon and the development that precedes it and the development that follows - a developmental view to early lexicon. In A. Bar-On, D. Ravid (Eds.), *Handbook of communication disorders*. (91-110). Berlin: De Gruyter Mouton.

Stolt, S., Vehkavuori, S. (2018). Sanaseula. MacArthur-Bates Communicative Development Inventories -arviointimenetelmän lyhyt, suomalainen versio [Sanaseula. The Finnish short form

versions of the MacArthur-Bates Communicative Development Inventories]. Jyväskylä: Niilo Mäki Instituutti.

Szagun, G., Stumper, B., Schramm, S. A. (2009). A normative study of early language development in German using an adaptation of the long and short version toddler CDI. www.gieselaszagun.com/en/Szagun_et_al_FRAKIS_2009.pdf [16.04.2017.]

Thordardottir, E. T., Weismer, S. E., Evans, J. L. (2002). Continuity in lexical and morphological development in Icelandic and English-speaking 2-year-olds. *First language*, 22, 3-28. <https://doi.org/10.1177/014272370202206401>.

Toivainen, J. (1997). The acquisition of Finnish. In Dan Slobin (ed.), *The cross-linguistic study of language acquisition*, Vol. 4, 87–182. Mahwah, NJ: Lawrence Erlbaum.

Tomasello, M., & Mervis, C. B. (1994). The instrument is great, but measuring comprehension is still a problem. *Monographs of the Society for Research in Child Development*, 59, 174–179.

Tse, S. K., Chan, C., Kwong, S. M., Li, H. (2002). Sex differences in syntactic development: Evidence from Cantonese-speaking preschoolers in Hong Kong. *International Journal of Behavioral Development*, 26(6), 509-517. <https://doi.org/10.1080/01650250143000463>.

Tulviste, T., Mizera, L., De Geer, B. (2004). Expressing communicative intents in Estonian, Finnish, and Swedish mother-adolescent interactions. *Journal of Child Language*, 31 (4), 801–819.

Tulviste, T. (2006). Variation in vocabulary development among Estonian children as a function of child's gender, birth order, child-care, and parental education. In M. Eriksson (Ed.), *Proceedings from the First European Network Meeting on the Communicative Development Inventories* (16-22), Gävle: University of Gävle.

Tulviste T., Schults, A. (2019). Parental reports of communicative development at the age of 36 months: The Estonian CDI-III. *First Language*, 40(1), 64-83. <https://doi.org/10.1177/0142723719887313>.

Urm, A., Tulviste, T. (2016). Sources of individual variation in Estonian toddlers' expressive vocabulary. *First Language*, 36(6), 580-600. <https://doi.org/10.1177/0142723716673951>.

Urm, A., Tulviste, T. (in press). Toddlers' communicative skills as assessed by the short form version of the Estonian CDI-II. Submitted.

van Houdt C. A., van Wassenaer-Leemhuis A. G., Oosterlaan J., van Kaam A. H., Aarnoudse-Moens C. S. H. (2019) Developmental outcomes of very preterm children with high parental education level. *Early Human Development*, 133, 11-17.

Vasilyeva, M., Waterfall, H., Huttenlocher, J. (2008). Emergence of syntax: commonalities and differences across children. *Developmental science*, 11(1), 84-97. doi: 10.1111/j.1467-7687.2007.00656.x.

Westerlund, M., Lagerberg, D. (2008). Expressive vocabulary in 18-month-old children in relation to demographic factors, mother and child characteristics, communication style and shared reading. *Child: care, health and development*, 34(2), 257-266. doi: 10.1111/j.1365-2214.2007.00801.x

Zambrana, I. M., Ystrom, E., Pons, F. (2012). Impact of gender, maternal education, and birth order on the development of language comprehension: a longitudinal study from 18 to 36 months of age. *Journal of Developmental and Behavioral Pediatrics*, 33(2), 146-55. doi: 10.1097/DBP.0b013e31823d4f83.

Zauche, L. H., Thul, T. A., Mahoney, A. E. D., Stapel-Wax, J. L. (2016). Influence of language nutrition on children's language and cognitive development: An integrated review. *Early Childhood Research Quarterly*, 36, 318-333. <https://doi.org/10.1016/j.ecresq.2016.01.015>.