Early vocabulary and grammar development in Albanian-speaking children: An MB-CDI adaptation study

Enkeleida Kapia
University Ludwig Maximilian, Munich, Germany
Albanian Academy of Sciences, Tirana, Albania

Shanley E. M. Allen
University of Kaiserslautern-Landau, Germany

Doruntinë Zogaj
University of Saarbrucken, Germany

Abstract: This is the first exploratory study of a large sample of Albanian-speaking children and their communicative development, based on the Albanian adaptation of three sections of the MacArthur–Bates Communicative Development Inventory II Words and Sentences: The vocabulary checklist (Part I Section A), How children use words (Part I Section B), and Word Endings (Part II Section A). Parental report data were collected from 112 children between the ages of 13 and 36 months. Correlation analyses for early vocabulary were conducted. Developmental trajectories of children were compared based on the demographic characteristics of sex and parental education. Results show that lexical growth for Albanian is comparable to that reported for other languages, and neither sex nor maternal education level correlate with vocabulary size. However, fathers' educational level correlates with production of early morphological markers. We conclude that the Albanian CDI is a useful tool for the assessment of and research on the language development of Albanian-speaking children. Future directions include testing larger numbers of children from different dialect and socioeconomic backgrounds.

Keywords: MB-CDI, early lexical and grammar development, Albanian

Corresponding author: Enkeleida Kapia, Institute for Phonetics and Speech Processing, LMU, Schellingstrasse 3, 80799, Munich, Germany. Email: enkeleida.kapia@phonetik.uni-muenchen.de.

Orcid ID: https://orcid.org/0000-0002-9676-3509

Citation: Kapia, E., Allen, S. E. M., & Zogaj, D. (2024). Early vocabulary and grammar development in Albanian-speaking children: An MB-CDI adaptation study. *Language Development Research, 4*(1), 175–206.http://doi.org/10.34842/wve8-yn80

Introduction

To date very little research exists on the early language development of Albanian-speaking infants and toddlers. This limits our understanding of how Albanian is learned and what the relationship is between various aspects of early language development. Information about Albanian-speaking children's language development from an early age is valuable not only for the purpose of discovering milestones and developmental steps for children with typical development, but also for comparing the process of language development in Albanian with that of other languages with the same or different typologies. Furthermore, such information is essential for identifying children who have developmental delays or other atypical language profiles such as autism and aphasia, as well as for creating and evaluating effective intervention strategies for language disorders, and for distinguishing linguistic and cognitive delays.

The present study is the first to provide a detailed exploratory examination of the early acquisition of vocabulary and grammar in Albanian. It is based on data from 112 Albanian-speaking infants and toddlers aged 13 to 36 months using the Albanian adaptation of three sections of the MacArthur-Bates Communicative Development Inventories II Words and Sentences: The vocabulary checklist (Part I Section A), How children use words (Part I Section B), and Word Endings (Part II Section A). To this end, this study describes the characteristics of early vocabulary in child Albanian including its size, its composition and some language-internal and language-external factors that may be associated with it.

Early vocabulary and grammar development in child language

Results from studies of numerous languages using their own adaptations of the MB-CDI indicate very similar routes and speeds in early vocabulary and grammar development across languages among infants and toddlers (Bates et al., 1994; Bleses et al., 2008; Devescovi et al., 2005; Frank et al., 2021; Gendler-Shalev & Dromi, 2022; Stolt et al., 2009). Despite large individual differences (Bates et al., 1994; Bleses et al., 2008; Devescovi et al., 2005; Fernald et al., 2001), most children around the world are found to produce their first words between 1;0 and 1;8 years of age, experiencing a vocabulary spurt soon after 1;6 (Bates & Goodman, 2001; Fernald et al., 2001).

Early vocabulary consists of items that denote concrete things in children's immediate environment, such as body parts, kitchen tools, food and drinks, toys, routine activities, etc. (Caselli et al., 1999; Karmiloff & Karmiloff-Smith, 2001). These open-class or content words, which show up in high numbers during the first years of life, have been argued to aid the later emergence of closed-class words such as prepositions and articles (Bates & Goodman, 2001). The later acquisition of closed-class words has also been linked to them being phonologically shorter and less emphasized in speech,

which makes them harder to be noticed by children from early on (Morgan et al., 1996).

In addition, extensive data from children crosslinguistically has shown that the size of early vocabulary is closely linked to the development of grammar. For example, for two-word utterances to begin appearing in children's speech, children need to have already acquired around 50-100 words (Bates & Goodman, 2001; Marchman & Bates, 1994). A recent study looking at data from nearly 8000 English-speaking infants and toddlers has suggested that lexical and syntactic ability begins to grow roughly in the 20- to 24-month age range; however, variability is high and this ability does not stabilize until children reach 30 months or so (Day & Elison, 2022). This study reveals a distinct lag effect, where children's ability to use syntax is almost always lower than their vocabulary skills. However, although a relationship between early vocabulary size and later morphological and syntactic development is evident across languages, this relationship is not identical across languages, reflecting differences in language structure as well as cultural differences in language use. For example, Thordardottir et al. (2002) show that Icelandic-speaking children need to have acquired a larger vocabulary than English-speaking children before they begin to use plural inflections on nouns and past tense inflections on verbs, linking this observation with the more complex inflectional system in Icelandic as compared to English. Thus, it is important that more studies investigate the timing of the relationship of early grammar to early lexicon, especially from languages with a complex grammatical system such as Albanian.

And lastly, long standing work has also shown that several demographic factors including sex and parental education play a role in early language development. On average, it has been reported that boys produce fewer words at a given age than girls (Bates et al., 1994; Bouchard et al., 2009; Day & Elison, 2022; Hulle et al., 2004), and children of more highly educated mothers (Bates et al., 1994) – with mixed findings with respect to fathers (Bates et al., 1994; Pancsofar & Vernon-Feagans, 2006) – produce more words.

The main objective of the present study is to build on this literature in order to establish the characteristics of early vocabulary in Albanian-speaking infants and toddlers from 13 – 36 months of age, as reported by the Albanian MB-CDI. In Section 2 we present general information about the Albanian language and its early development in infants and toddlers, as well as detail the development of the Albanian MB-CDI form.

Albanian language and cultural context

Albanian language

Albanian is a language of the Indo-European family with 6-7 million speakers (J. S.

Klein et al., 2017; Rusakov, 2017) who live mostly in the Republic of Albania, the Republic of Kosovo, Montenegro and the Republic of North Macedonia, as well as in Albanian-speaking minority communities in Italy, Greece, Croatia and Ukraine. In this article, however, we focus on Albanian as spoken in Albania. Albanian is widely accepted to form a branch of its own within the Indo-European language family (Bopp, 1855; Çabej, 1976; Pedersen, 1897); no evidence to date relates Albanian to any other language within this family (Demiraj, 2018). Albanian has traditionally been described as comprising two main dialects: Gheg, spoken in northern and central Albania, and Tosk, spoken in the south of the country (e.g. Desnickaja, 1976; Gjinari, 1988; Hahn, 2013). The Shkumbin River located in the center of the country forms the approximate boundary between the two dialect regions. A third variety, the Standard variety, is also present alongside the two dialects; it was institutionalized in 1972 via a National Congress of Orthography and was based mostly on the Tosk dialect (Ismajli, 1998). Standard Albanian, which will be the focus of this article, is characterized by a relatively free word order (Rushi, 1983) and has a fairly complex fusional morphology (Agalliu et al., 2002). For example, it has noun declensions in 5 different cases, which are also marked for gender, number, and definiteness, as well as verbs which are marked for mood, tense, person, and number (Agalliu et al., 2002).

Albanian early language development

Although children's language acquisition has been studied extensively in many languages, research on the acquisition of Albanian has been almost completely absent from this literature. Exceptions are just a few studies published in recent years (Cenko, 2017; Cenko & Budwig, 2007; Kapia, 2010, 2014; Shashaj, 1996), and a couple of recent Master's theses (Dule, 2023; Jia, 2023; Sehitaj, 2015; Zogaj, 2021). These studies provide a preliminary view of children's speech at an early age, with quantitative and qualitative descriptions focusing on the lexicon, syntax and, more broadly, grammatical constructions.

The earliest study that focused on children's language development was motivated by the educational policy demands during the communist regime (Shashaj, 1996). This policy sought ways to understand how children acquired language in relation to general education, with the purpose of influencing their language habits more towards the Standard norm. Shashaj (1996) investigated the speech of three children, of which two produced their first words before 8 months old, and the third child after 12 months. According to this study, which does not report any details about its methodology, Albanian-speaking children at age 1 typically produce around 20 words while children at age 3 produce around 1000 words. These words are not the same for every child, but depend on the environment in which they grow up including their home environs, playmates, kindergarten, and caregivers (Shashaj, 1996).

A small number of studies have also investigated the acquisition of syntax and pragmatics, and provide preliminary reports about children's ability to combine words together into phrases or initial short sentences. Shashaj (1996) reports that children begin to be able to form more complex phrases between the second and third year of life. Studies by Cenko and Budwig (2007) as well as Kapia and colleagues (Kapia, 2010, 2014) have focused on different language structures and different stages of acquisition across age groups in monolingual and bilingual children. Kapia (2010, 2014), for instance, found that Albanian-speaking children perform at an adult-like level with clitic doubling of dative object nouns at around age 2-3 years, but show difficulty with clitic doubling of accusative object nouns since these require differentiation between old and new information – a pattern which seems to suggest a delay in their development of pragmatics, but not of syntax. In another study, Cenko & Budwig (2007) found that most 2-year-old Albanian-speaking children are able to use at least one verb with the correct morphological marking in both the transitive and the unaccusative form. This demonstrates their flexibility in verb construction use, in contrast to the findings for English-speaking 2-year-old children. This flexibility was suggested to be linked to the rich morphological markings on the verb that emphasize the differences between transitive and unaccusative constructions, consistent with similar findings for other morphologically rich languages (Cenko & Budwig, 2007). Apart from these studies, our knowledge about lexical and morphosyntactic development in Albanian is very limited, which is why more detailed and larger scale studies are needed at this point. In the present study, we aim to fill some of this gap by focusing on early vocabulary and grammar development, their relation to each other, and some of the language-external factors that may influence both of these components of early language. We achieve this by investigating early language development using the Albanian MB-CDI tool.

In terms of vocabulary learning, we do not have any reason to believe that Albanian-speaking children generally learn words any differently than children from other language contexts, apart from perhaps learning certain words earlier or later depending on the cultural context. Nonetheless, determining the trajectory of vocabulary developmental for Albanian-speaking children specifically is crucial for clinical and educational reasons given that assessing their vocabulary level using norms taken from other languages is likely to lead to inaccurate results. In relation to how children use words (operationalized by 'displaced events' in the CDI), to the extent that these are conceptual, we also do not think that Albanian-speaking children will show different developmental tracks than children learning other languages. Where we do expect a difference, however, is the rate of acquisition of morphology based on the fact that Albanian has quite a complex morphological system. For instance, verb forms are often realized not just by adding a suffix or one or two particles to them, but also sometimes by metathesis or stem change altogether, as is the case for *ha* 'to eat' vs *hëngra* 'ate' vs *pata ngrënë* 'had eaten' or *shoh* 'to see' vs *shihja* 'saw' vs *pata parë* 'had seen'. We predict that learning of grammatical morphemes will take longer and occur

later in life than for children with simpler morphological systems such as English.

Social context

Considering that parental gender and education has been reported to be linked to early lexical and grammatical development (see Section 1.1), a secondary aim of this study is to investigate whether mother's and father's educational levels play a role in early child language (vocabulary and grammar) development. Thus, we outline here some relevant details about the role of women and men in Albanian society. Although Albania is now on the rise from one of the poorest countries in Europe to a middle-income country, there are still gaps in both education levels and labor market opportunities between women and men. The World Bank reports that, unlike in most patriarchal societies, more women than men in Albania receive a postsecondary education, with the gap between genders being around 25% (World Bank, 2018). However, when it comes to the labor market, women's participation in the labor force has dropped drastically from 78% in 1989 to 46% in 2005, finally reaching 50% participation in 2013 (INSTAT, 2017). This is likely due to the Albanian society holding onto strong patriarchal values that place women of reproductive age outside the labor market, with few opportunities for retraining and qualification (Young, 2018).

Development of the Albanian MB-CDI form

Version 1

The present adaptation of the CDI for Albanian was developed by the first author in collaboration with Enila Cenko (University of New York Tirana), a researcher of early child development, in close consultation with the second author and Nancy Budwig (Clark University). They received authorization for the Albanian adaptation project from the CDI Advisory Board in 2008, and had regular cooperation with the European Network on Communicative Development Inventories including researchers from all over Europe developing CDIs for various European languages. The adaptation was completed in 2010.

The starting point in constructing the Albanian CDI was the American English MB-CDI *Words and Sentences* which consists of two parts (Bates et al. 1994; Fenson et al., 2007). Part I, *Words Children Use,* focuses on the child's use of words and is split into sections A and B. Section A, *Vocabulary Checklist,* is a checklist of 680 words divided into 22 semantic categories. Section B, *How Children Use Words,* asks whether or not the child has started using words to talk about displaced events such as events in the future or past, or objects not present in the context.

Part II, *Sentences and Grammar,* deals with early grammar and is divided into five sections. Section A, *Word Endings I,* tests the child's use of word endings such as -ed, -ing

etc. Section B, *Words Forms,* deals with word forms, nouns and verbs. Section C, *Word Endings II,* checks the proper use and errors when using word endings. Section D, *Examples,* asks for the three longest sentences the child has said recently. Finally, Section E, *Complexity,* deals with the complexity of the child's morphosyntactic skills, for example, does the child say "two foot" or "two feet"?

Our adaptation of the Albanian CDI only includes Part I Sections A & B and Part II Section A. Adaptations of the remaining sections are ongoing.

Part I Section A: Lista e fjalëve (Vocabulary Checklist)

For the adaptation of Part I Section A, we began by translating all the words of the American English version (all 22 semantic categories) from English to Albanian. We then had a person unfamiliar with the project translate those items back from Albanian to English, in order to ensure that the back-translation resulted in the same items as originally intended. All words at this step of the process were back-translated as intended.

Some sections, however, required adaptation due to differences between English and Albanian morphosyntax. In the action word category, for example, listing verbs in their root form as in English would not be appropriate because Standard Albanian (the dialect used here) lacks infinitives. Instead, verbs were listed in typical citation form used in Albanian dictionaries: first person singular form using indicative mood, active voice and present tense, and preceded by a first person pronoun (e.g., *(unë) punoj* '(I) work', *(unë) shkruaj* '(I) write'). Differently from English, in Section II Part A (Word Endings I), we inquired whether children have knowledge of the subjective form of verbs which are formed with particles *për të* + verb. These forms denote an action to be completed and are the quasi-analogue of the infinitive which only exists in the Gheg dialect, but not in the Standard variety tested here (Cipo, 1949). Furthermore, adjectives in Albanian are obligatorily inflected for gender. Thus, the words in the descriptive word category included the relevant gender variations (e.g., *i bukur / e bukur* 'beautiful.M / beautiful.F', *jeshil / jeshile* 'green.M / green.F). Additionally, words in the pronoun category were changed to fit the Albanian pronoun system which agrees in gender with the noun it refers to, so gender variations were included when necessary (e.g., *i imi / e imja* 'mine.M / mine.F', *këta / këto* 'these.M / these.F').

Once the list was finalized, we tested it with two focus groups comprised of caregivers of children aged 1-3 years. Focus Group I, consisting of 15 caregivers (7 had 8 or fewer years of education) received the full vocabulary checklist. Their task was to mark the words that their children produced and comprehended. Focus Group II, with another set of 15 caregivers (6 had 8 or fewer years of education), received a blank form with just the semantic categories of the American English MB-CDI. Their task was to write

down as many words as they could remember that their children produced and comprehended within these categories and add any other words that were not captured by the categories on the form. After the two focus groups, we combined the responses from the 30 participants into one large list of words. For each word, we determined the number of children out of 30 that produced that word, and rank-ordered the words in terms of the frequency of occurrence. Several words on the original list of the American English MB-CDI were excluded such as *hamburger, peanut butter, babysitter, backyard, soda,* and *pancake*. Other words relevant to Albanian culture and culinary practices were added in, such as *çiçi* 'peepee', *dum dum* 'small van', *kola* 'coca cola', *petulla* 'a type of fried dough', *byrek* 'a type of savory pie', and *gjizë* 'a type of cottage cheese'. The final result was the first complete draft of Part I Section A, *Lista e fjalëve* (Vocabulary Checklist), of the Albanian CDI.

Part I Section B: Si i përdorin fjalët fëmijët (How Children Use Words)

Part I Section B comprised five questions that check whether children refer to displaced events. Two of these questions focus on concepts of time, asking whether the child talks about events in the past (e.g., the child visited the beach last week, later he/she mentions sea, sun, sand) or in the future (e.g., if they are going to visit grandmother, the child says *nëna* 'grandmother'). Three further questions focus on displaced objects and people, asking whether the child talks about objects or people that are not present (e.g., asking for the father while he is at work), whether the child understands requests for objects or people that are not present (e.g., when asked "where is the ball?", s/he can get the ball from another room), and if the child can point to an object that belongs to a person that is not present (e.g., the child can point to mother's shoes and says "mother"). For each question, caregivers selected either 'never', 'sometimes', or 'often' as their answer. We discussed each of these questions with the caregivers in Focus Groups I and II. These consultations resulted in no changes in Part I Section B of the Albanian CDI from the American English CDI, as all the categories of time and displaced events were deemed appropriate and necessary.

Part II Section A: Mbaresat e fjalëve I (Word Endings I)

Finally, Part II Section A is constructed in the same way as in the American CDI, and also benefitted from discussions with caregivers from both Focus Groups mentioned earlier. Our goal in adapting this section was to include elements of grammar that are relevant to the rich morphology of noun and verbal systems in Albanian. Consultations with caregivers revealed similar elements of grammar as relevant for this part of the Albanian CDI as for the American English CDI: plurality in nouns and several different verb tenses. This section contains five questions, one each about whether the child uses the forms in (1-5), whose morpheme-by-morpheme glosses follow the Leipzig glossing conventions (Croft, 2002; Lehmann, 1982) .

(1) Regular plural forms
vajzë *vajza*
girl.F.NOM.SG girl.F.NOM.PL

(2) Verbs in simple past tense
ha *hëngra*
eat.IND.PRS.1SG eat.IND.PST.1SG

(3) Verbs in present continuous tense
jam duke ngrënë
am -ing eat.PRS.PTCP

(4) Verbs in indicative form
për të ngrënë
for to eat.PRS.PTCP

(5) Verbs in future tense
do të ha
will to eat.PRS.IND

Questions are structured as in (6).

(6) *A e keni dëgjuar fëmijën tuaj të përdori emra në shumës, si për shembull: libër-libra, vajzë-vajza, mollë-mollë?*
'Have you ever heard your child use nouns in plural, as in the examples book-books, girl-girls, apple-apples?'

For each question, caregivers selected either 'never', 'sometimes', or 'often' as their answer.

In sum, Version 1 of the Albanian CDI form resulted in having two parts: Part I Section A *Lista e fjalëve* (Vocabulary Checklist) and Section B *Fjalë të tjera që përdorin fëmijët* (Words Children Use), as well as Part II Section A *Mbaresat e fjalëve* (Word Endings).

Version 2

Version 1 of the Albanian CDI form was tested in a pilot study with a group of 40 parents and grandparents (it is quite common in Albania that children grow up in extended households where they receive input from grandparents and parents at the same time). The results of this pilot were used to develop Version 2.

Words from Part I Section A were rank-ordered in terms of frequency of occurrence.

Words that never occurred were removed from the CDI, for example *dëshiroj* 'wish'. Table 1 compares the number of words per category of Part I Section A of Version 2 of the Albanian CDI with the American English, Danish and Norwegian forms (because these were the CDI forms for which we could readily find word lists). As seen from this table, the difference between these forms in the number of words for each semantic category is relatively small. One aspect to highlight here is that the biggest differences are in action words, descriptive words and prepositions and locations. The reasons for these differences are many, but in the case of action words, for example, some words in the American CDI such as *skate* or *ride* are irrelevant in the Albanian context, as there are not many opportunities for skating in the warm Albanian Mediterranean climate or for riding since not many families own cars and neither do children ride ponies or bikes when they are little, as is pretty standard in some Western cultures. Another example in this section that seemed inappropriate to focus groups at the time of the adaptation was the action word *hate;* no one imagined children this young using the word *hate* 'urrej' in their daily speech. Other examples of the American CDI action word list that ended up being deleted from the Albanian counterpart were examples that were not expressed via one word but a group of words, such as *cuddle* 'rri gushe-gushe/përqafuar dhe duke u përkëdhelur'. These expressions were generally avoided unless they denoted some really basic activity in children's lives such as *pee* 'bëj çiçin'.

Table 1: ***Comparison of the categories and number of items in the vocabulary lists of Albanian, Danish, Norwegian and American MB-CDIs***

	Albanian	American English	Danish	Norwegian
1. Sound effects & animal sounds	11	12	12	12
2. Animals (real or toy)	42	43	43	44
3. Vehicles (real or toy)	15	14	14	14
4. Toys	12	18	18	18
5. Food and drink	70	68	68	68
6. Clothing	34	28	30	30
7. Body parts	24	27	28	27
8. Small household items	70	50	50	50
9. Furniture and rooms	-[1]	33	33	34
10. Outside things	31	31	31	31
11. Places to go	18	22	22	22
12. People	31	29	40	36

[1] *Small household items* and *Furniture and rooms* are combined in the Albanian CDI in one semantic category labelled *Things and rooms around the house*.

13. Games and routines	32	25	27	27
14. Action words	83	103	103	108
15. Descriptive words	54	63	63	62
16. Words about time	15	12	15	16
17. Pronouns	26	25	31	31
18. Question words	7	7	7	7
19. Prepositions and locations	29	26	41	41
20. Quantifiers and articles	14	17	21	22
21. Auxiliary verbs	26	21	21	22
22. Connecting words	6	7	6	9
Total Vocabulary	**650**	**680**	**725**	**731**

Part I Section B and Part II Section A remained the same as in the previous version since they functioned as expected and no changes were necessary. This version - Version 2 - was used in three unpublished MA theses (Jia, 2023; Sehitaj, 2015; Zogaj, 2021), the latter two of which comprise the data for the present study.

Methods

Participants and materials

A total of 112 Albanian-speaking infants and toddlers aged 13 to 36 months old were recruited for this study. Not much is known about Albanian vocabulary acquisition. Thus, in line with the protocol followed during the earlier phases of the development of the American English CDI inventories (Fenson et al., 2007), we wanted to ensure that the form we chose to adopt first showed steady developmental regularity in all of its components not just for the suggested age window (16-30 months), but also for a few months below (3 months) and above (6 months). As a result, we expected to see numerous floor and ceiling effects. This way we could firmly say that the decision to cut off the age range for the norming study to 16-30 months was dictated by the results of our preliminary study, not just by following verbatim the American CDI. All participants lived in Albania. Data from two participants were excluded from the analysis because their caregivers did not complete some part of the CDI. Additionally, data from one participant was excluded from the analysis due to a very low vocabulary size – more than 1.5 SD below the mean of his/her age group. This yielded a final participant pool of 109 children with an age range from 13 to 36 months (M = 26.3, SD = 6.2, 55 females and 54 males). All children had normal birth weight, no serious illnesses, and no developmental delays. Children were recruited from five different cities in Albania to ensure a certain degree of generalizability, as this was the first study of this kind: Tirana, Fier, Vlorë, Elbasan, and Pogradec. Note that almost all these cities come from the Tosk-speaking areas, the dialect which also forms the base for the Standard variety. We purposefully did not recruit from the Gheg-speaking areas other

than Tirana (which is the capital and where almost half of the population lives). We believe that a separate CDI form is needed for the deep Gheg-speaking areas of northern Albania as well as Kosovo due to the lexical, and sometimes grammatical, differences between the two varieties. Children were also recruited from homes with different parental education levels (see Table 3).

In checking the distribution of our data, we noticed two instances of skewness. First, the majority of our participants fall within the upper age ranges of 25-36 months, as shown in Table 2 and in the density plot in Figure 1. Second, the majority of the parents of our participants have a relatively high level of education - either postsecondary (for mothers) or postsecondary or secondary (for fathers) – as shown in Table 3 and Figure 2. We had very few participants whose parents have completed only primary school. The skewness of education level may have occurred for two reasons. One reason is that, as a former communist state, Albania has a tradition of granting access to higher education to all, making university degrees common among the population. In recent years, the multiplication of private universities has further increased the graduation rate at the Master's level; in fact, Albania is reported to be the country with the highest number of private universities in Southeast Europe. A second reason is that parents with higher education may be more willing to take part in research studies like this – a general trend also noticed in other CDI studies (deMayo et al., 2021).

Thus, given that our sample does not have a normal distribution, and also because of our inability to make evidence-based hypotheses about children's performance on the CDI due to the absence of knowledge about Albanian language acquisition, our study will be exploratory and descriptive in nature. We perform statistical testing solely as an auxiliary tool to explore the description of the trends we observe.

Table 2. ***Number and sex of participants, divided in 3-month intervals***

Age	Female	Male
13-15	4	4
16-18	1	6
19-21	5	4
22-24	6	8
25-27	13	6
28-30	12	13
31-33	6	7
34-36	8	6

Table 3. ***Education level of participants' parents***

Completed Education	Mother	Father
Primary School	6 (5.5%)	6 (5.5%)
Secondary School	29 (26.6%)	48 (44.0%)
Post-Secondary Education (Bachelor's degree or higher)	74 (67.9%)	55 (50.5%)

Data for the study were provided by a primary caregiver of the child, either a parent or grandparent (as noted in Section 2.3, grandparents often live with the family and/or babysit children during the day). Prior to data collection, permission to conduct the study was obtained from the school boards of the corresponding municipalities. Children were recruited in two ways. First, preschool teachers distributed the CDI test to interested parents of the children that were enrolled in their preschools. Second, researchers involved in the study handed out the CDI test to interested relatives, neighbours, friends, etc., who had children corresponding to the ages relevant for the study. In both cases, parents provided written consent for their child's participation in the study. Data were collected over the period 2013-2015 using Version 2 of the Albanian CDI, as detailed above.

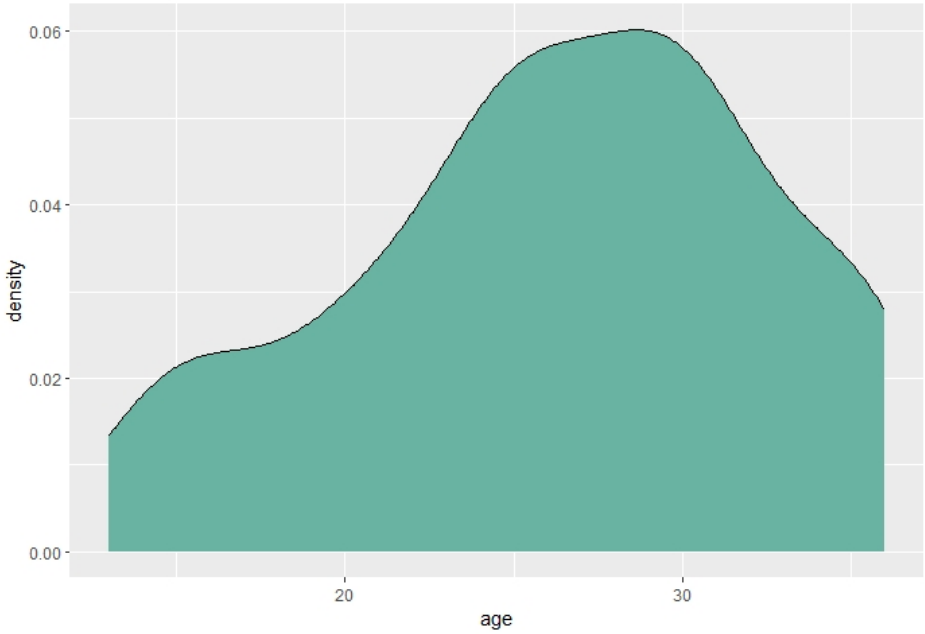

Figure 1. ***Density plot showing the distribution of the sample by variable age in months***

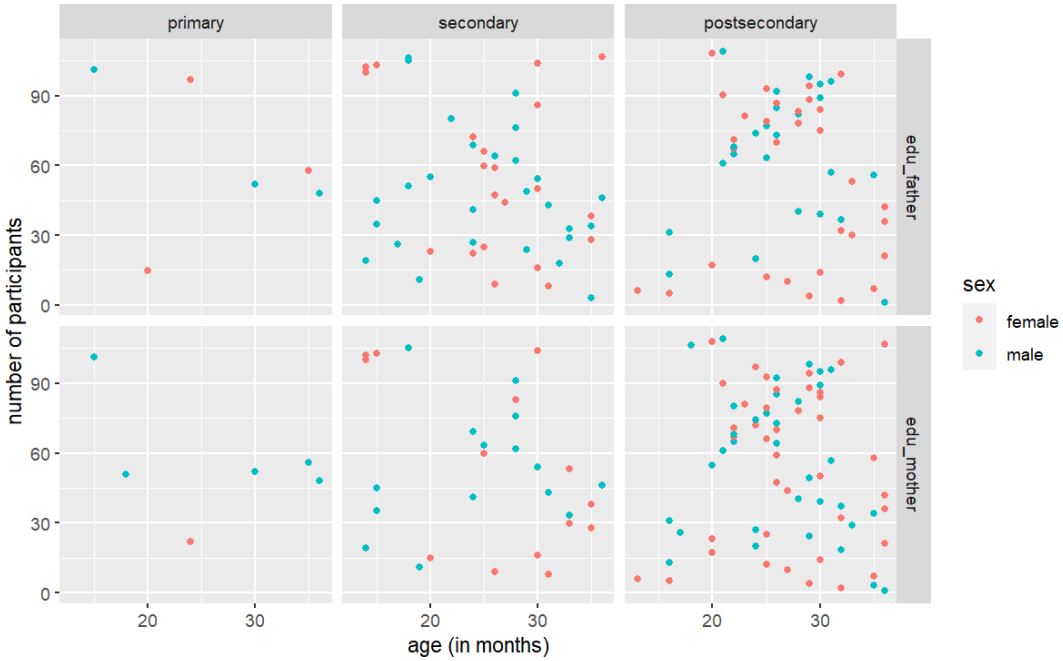

Figure 2. ***Distribution of our participants per demographic variables of sex and level of parental education, i.e. maternal education (edu_mother) and paternal education (edu_father)***

Analysis

For ease of reference in the analysis, we labelled Part I Section A as 'vocabulary size', Part I Section B as 'displaced events', and Part II Section A as 'morphology'. We also divided the words in the vocabulary list from Part I Section A into two categories: open class words (nouns, verbs, adjectives, adverbs) and closed class words (prepositions, exclamations, articles, pronouns). The vocabulary size was calculated as one point per word reported as used by the child, with a maximum score of 650. The scoring for displaced events and morphology was calculated as one point per item that the caregiver reported used either 'sometimes' or 'often'; items reported as used 'never' were scored as 0 points. Thus, the maximum score was 5 for each of these two parts, since each part contained 5 questions.

Given that some of the variables are bound within certain ranges, as was clear from the distribution plots presented above, we used Spearman correlation analyses to explore two types of relationships: a) the relationship between age and vocabulary size, and between age and grammatical development (i.e., score for displaced events + morphology); and b) the relationship between vocabulary size and other aspects of language development, such as number of open class words, number of closed class words, score for displaced events, and score for morphology. Lastly, we explored the

possible effects of demographic factors (children's sex, maternal education and paternal education) om language development (vocabulary size, displaced events, morphology) using a beta regression model (Ferrari & Cribari-Neto, 2004). This is a form of regression used when the response variable – for example, total vocabulary size – takes values within (0,1), and is assumed to take a beta distribution. The values of the response variables were beta-transformed as suggested in Smithson & Verkuilen (2006)[2]. Importantly, however, due to the unusual shape of the sample distribution and its bias towards older children and parents with higher educational levels, these analyses are used only as an auxiliary tool in the descriptions of the trends that we observe in the data.

Results

Age and acquisition of vocabulary, displaced events and morphology

Most frequent words

As a first step, a descriptive analysis showed that several semantic categories are included in the most frequently produced words in the Albanian CDI. These include the following (percentage of all children who produced the word is shown in parentheses).

a) kinship terms: *babi* 'father' (97.2%), *mami* 'mother' (90.8%), *teta* 'aunt' (83.5%), *gjyshi* 'grandfather' (79.8%), and the child's own name (69.7%)
b) terms used in the contexts of greeting and parting: *alo* 'answering phone call' (88.9%), *jo* 'no' (84.4%), *po* 'yes' (74.3%)
c) animals and their sounds: *macja* 'cat' (77.9%), *qeni* 'dog' (76.1%), *lopa* 'cow' (73.4%), *pula* 'chicken' (72.5%), *bee* 'baa' (82.6%), *ciu ciu* 'tweet-tweet' (81.7%), *ham ham* 'woof-woof' (84.5%)
d) food items: *buka* 'bread' (80.7%), *banane* 'banana' (77.1 %), *biskota* 'cookies' (74.3%)
e) toys: *topi* 'ball' (77.9%), *lapsi* 'pen' (73.4%), *lodra* 'toys' (73 %)

On the other hand, words that were used less frequently and only by older children (from 21 months old) were as follows: *i/e ngathët* 'clumsy' (26.4%), *i/e përgjumur* 'sleepy' (31,8%), *i/e varfër* 'poor' (33.6%), *pasnesër* 'day after tomorrow' (29.1%), *i/e tyre* 'theirs' (27.3%), *i yni* 'ours' (30.1%), and also two traditional desserts *kadaif* 'type of pastry dessert' (32.7%), *mualebia* 'type of baby food' (31.8%). As we can see, words that are learned latest are mostly adjectives, adverbs, and pronouns. The low frequency counts of words such as *kadaif* and *mualebia* in this study show the fast evolving nature of post-communist Albanian society; these words were deemed as important for this list by two different focus groups in the early phases of the Albanian CDI, but appear

[2] We thank one of the anonymous reviewers for this helpful suggestion which allows a more appropriate analysis for a sample with a distribution such as ours.

to have gone "out of fashion" now. It is very likely that we will exclude them from the norming phase of the CDI test.

Relationship between age and vocabulary

After a first descriptive analysis, we then assessed the relationship between children's age and their vocabulary size. A Spearman's correlation analysis for monotonic relationships between two variables revealed a moderately strong and statistically significant correlation (rho 0.40, $p < 0.01$). As can be seen in Figure 3, the total number of words produced by children increases as a function of age. However, two trends deserve mention here: a) data points are much sparser for children younger than 25 months of age, and b) a large number of children appear to reach ceiling, i.e. produce all the words in the CDI list, around 25 months old.

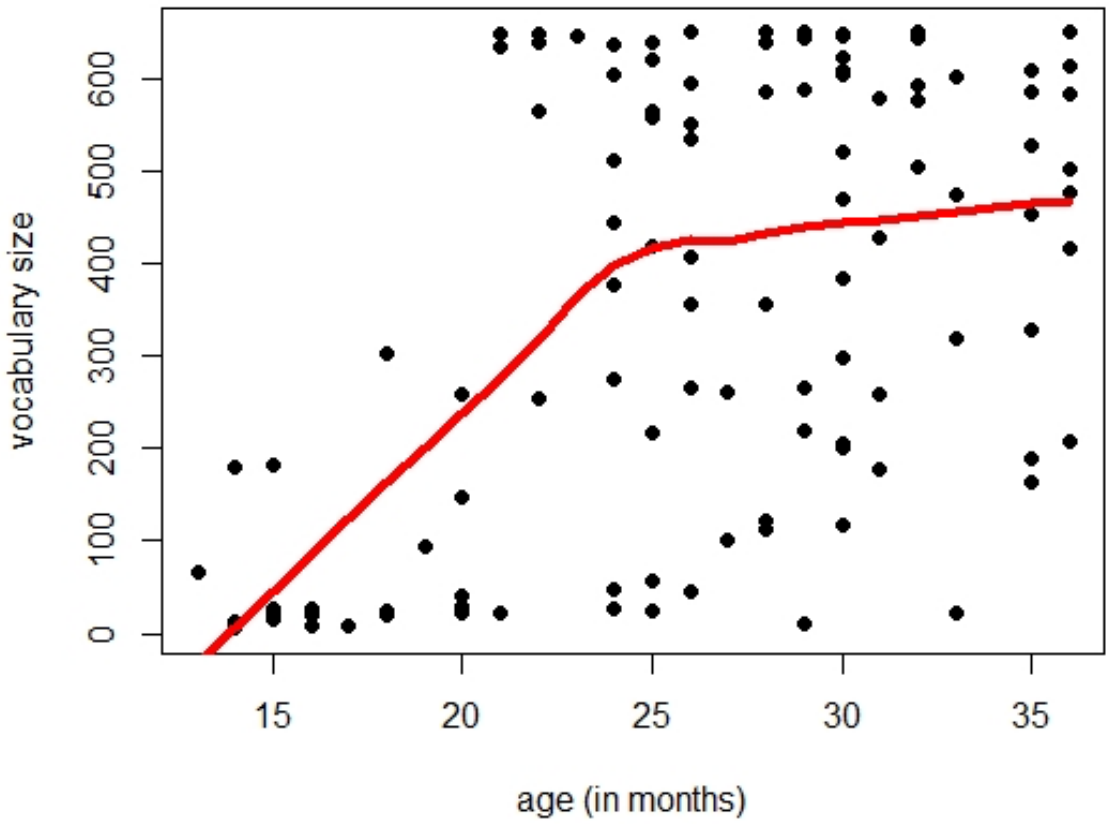

Figure 3. ***Number of words produced as a function of age (in months)***

Relationship between age and displaced events

We also looked at 'displaced events' – at how children performed with regards to whether they a) talk about events that happened in the past, b) talk about events that will happen in the future, c) talk about objects or people that are not present, d) understand requests for objects or people that are not present, and e) point/talk to an object that belongs to a person that is not present. Each participants' score varied depending on the number of displaced events they showed evidence of having conceptualized; for example, a score of 5 means that the participant showed evidence of having conceptualized all five of them. A Spearman's correlation analysis revealed a moderately strong and statistically significant correlation (rho 0.41, $p < 0.01$). Again, we note the high number of participants that are either at ceiling or at floor.

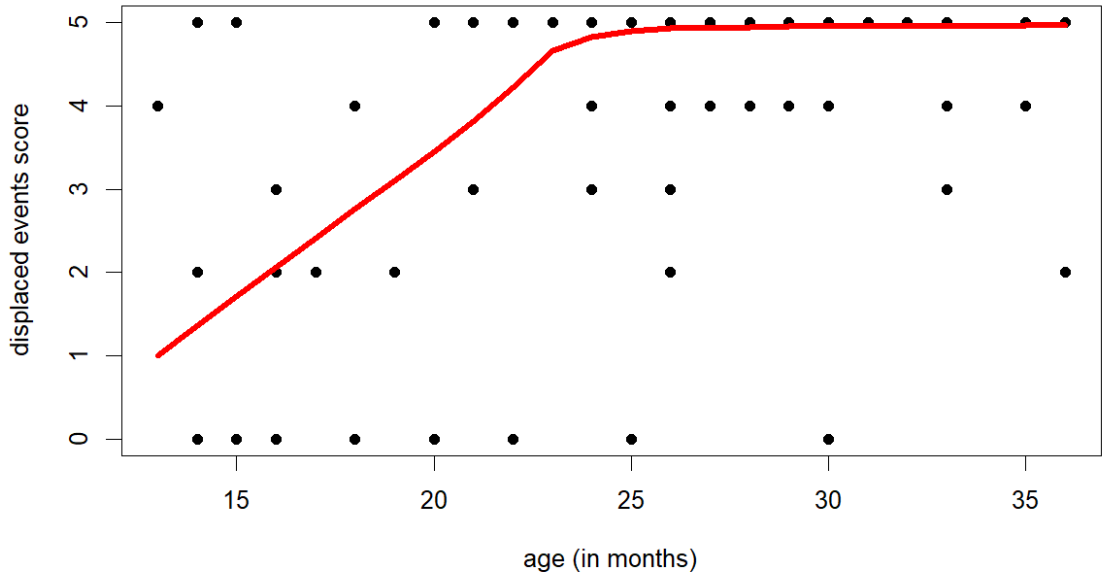

Figure 4. ***Displaced events score produced as a function of age (in months)***

Relationship between age and morphological knowledge

In the same way, we also analysed the development of morphology – whether children produce 1) regular plural forms (*vajzë-vajza* 'girl-girls'); 2) verbs in simple past tense (*hëngra* 'ate'); 3) verbs in present continuous tense (*jam duke ngrënë* '(I) am eating'); 4) verbs in the subjunctive form (*për të ngrënë* 'to eat'); and 5) verbs in future tense (*do të ha* '(I) will eat'). Again here, each participants' score varied depending on the number of morphological forms they showed evidence of having produced; for example, a score of 5 means that the participant showed evidence of having produced all five of them. As Figure 5 shows, children's abilities with these morphological forms

increases as they age. A correlation analysis revealed a moderate and statistically significant correlation (rho 0.53, $p < 0.01$). As with displaced events, a high number of participants are either at ceiling or floor in terms of their morphology as assessed in this section of the CDI.

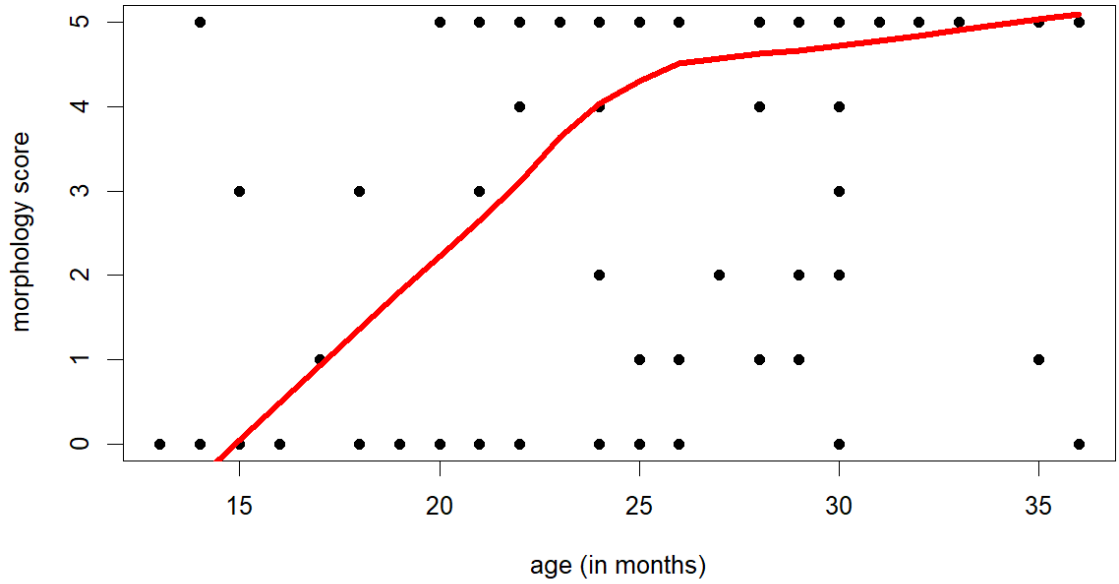

Figure 5. ***Morphology score as a function of age (in months)***

Relationship between vocabulary size and other aspects of language development

Next, we examined the relationship between the overall vocabulary size and production of other language-related phenomena, such as open class words and closed class words, displaced events and morphology. Following Bates et al (1994), Figure 6 portrays developmental trends in vocabulary composition for both open class and closed class words as a function of vocabulary size. Both open class and closed class words increase with age and there is a clear linear relationship between both of them and vocabulary size. Correlation analyses reveal strong correlations for both open class words (rho 0.99, $p < 0.01$) and closed class words (rho 0.97, $p < 0.01$).

We also found a moderately strong significant correlation between vocabulary size and the expression of displaced events (rho 0.57, $p < 0.01$), as well as between vocabulary size and the use of morphology (rho 0.64, $p < 0.01$), using Spearman's correlation analysis for monotonic relationships. As can be seen in Figure 7, the total number of

grammatical concepts realized through morphology increases as a function of vocabulary, but there is noticeable variability as shown by many data points in the upper end of the scale, revealing a ceiling effect.

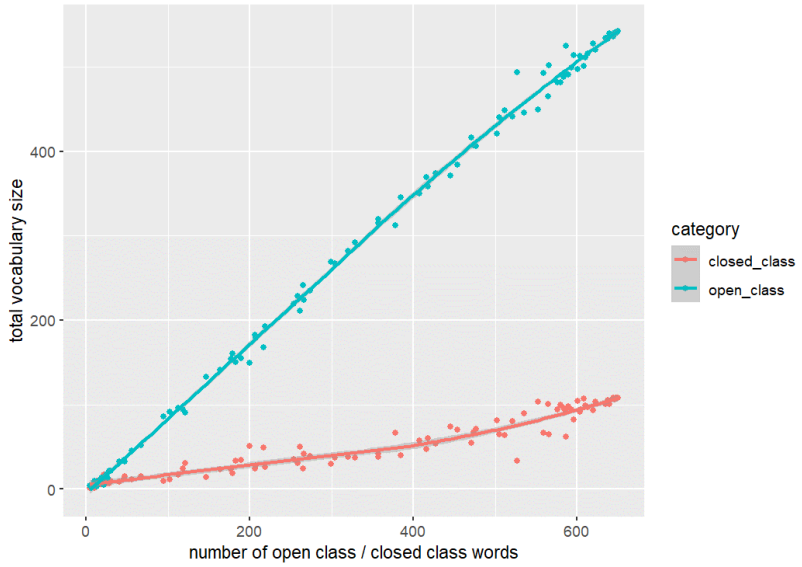

Figure 6. ***Number of open class and closed class words as a function of vocabulary size***

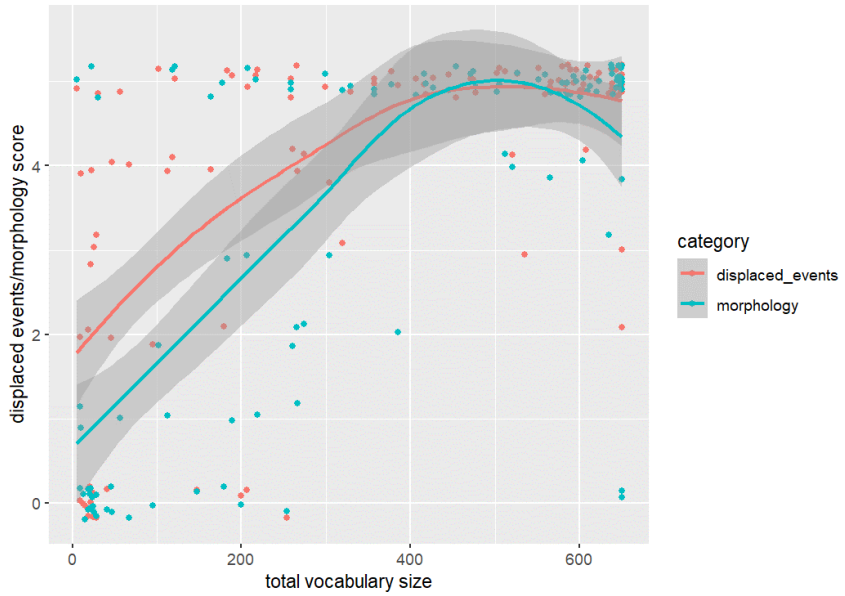

Figure 7. ***Relationship between vocabulary size and frequency of mention of displaced events (blue) and morphological inflections (red)***

The role of demographic variables

Finally, we investigated the effect of three demographic variables – children's sex, maternal education, and paternal education – on vocabulary size, displaced events and morphology. We fitted three beta regression models via the *betareg* function (Ferrari & Cribari-Neto, 2004) in R (R Core Team, 2023) to predict the total number of words, the displaced events score, and the morphology score based on each of the three variables, employing the standard logit link in *betareg*. The data is visualized in Figure 8 for sex, in Figure 9 for maternal education and in Figure 10 for paternal education.

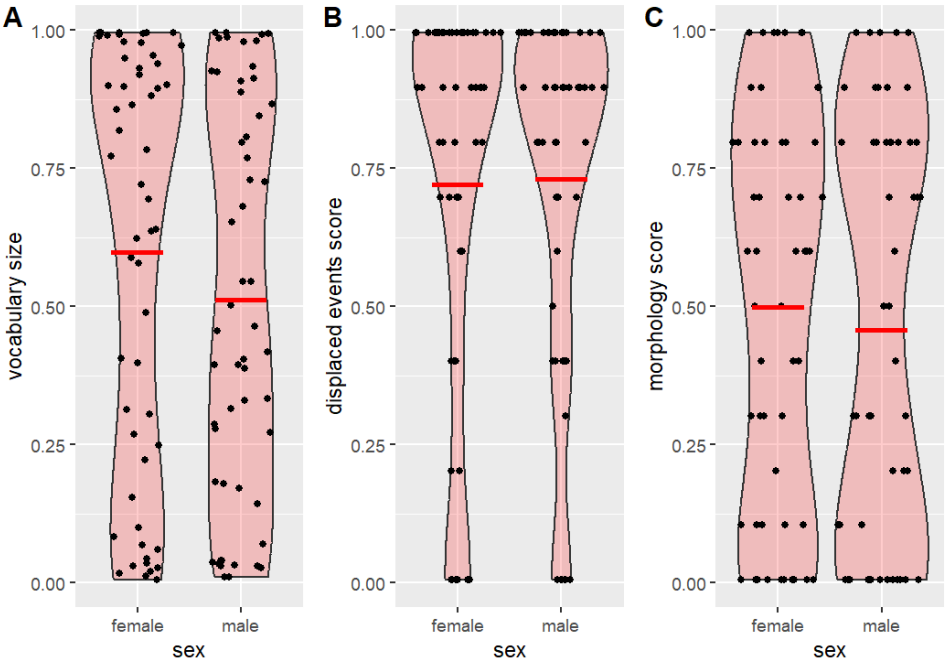

Figure 8. ***The relationship between children's sex and vocabulary size (plot A), displaced events score (plot B), and morphology score (plot C)***

In general, our observations, aided by the beta analysis, showed that none of the three demographic variables tested, i.e. sex, maternal education and paternal education had an effect on the total number of words that children learned over their first three years of life, or on the displaced event score or morphology score. However, an effect of paternal education was found on children's morphology score, as shown in the output of this analysis in Appendix 1. Children of fathers that had postsecondary education degrees were more likely to have higher morphology scores than were children of other fathers. However, these results should be considered with caution given that the distribution of our data is skewed towards older children that have parents with postsecondary degrees.

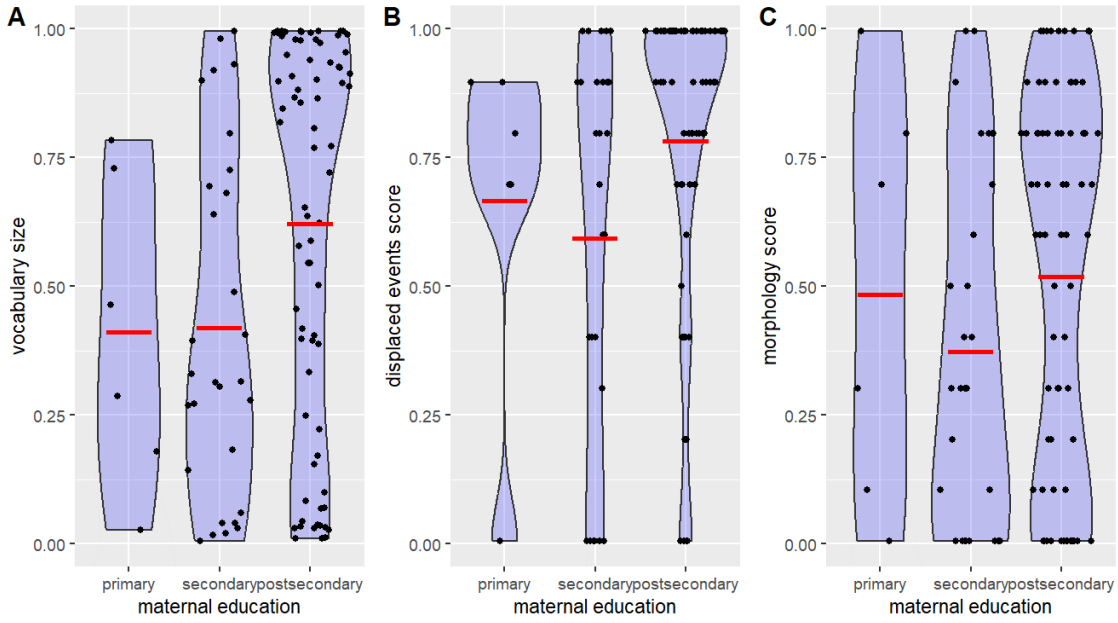

Figure 9. ***The relationship between maternal education and vocabulary size (plot A), displaced event score (plot B), and morphology score (plot C)***

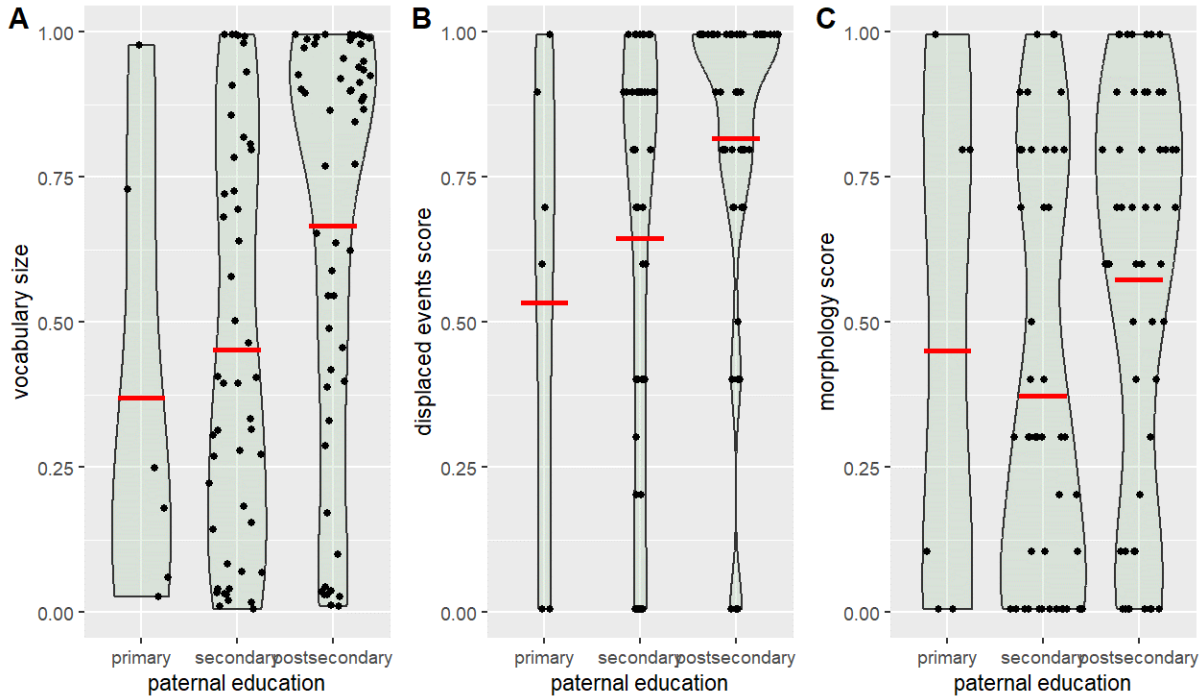

Figure 10. ***The relationship between paternal education and early expressive vocabulary (plot A), displacement score (plot B), and morphology score (plot C)***

Discussion and Conclusion

This study is the first study investigating the characteristics of early vocabulary and grammar acquisition (i.e. displaced events and morphology) in Albanian-speaking infants and toddlers using data from the first adaptation of the Albanian CDI. It serves as a first step in the development of this instrument for Albanian, a largely understudied language of the Indo-European family. In addition, it provides the first description of lexical and grammatical development of a large group of Albanian-speaking children, at least compared to what has been previously reported in the literature. Generally speaking, we observed language development trends similar to other reported languages in that vocabulary size, conceptualization of displaced events, and use of morphology to indicate plurality and tense all increase as children grow older. However, given that our sample distribution was skewed towards older children and children whose parents had postsecondary degrees, as well as to children living in urban areas, our results should be seen as mainly explorative in nature to guide specific hypotheses that will be tested during our forthcoming norming study of the Albanian CDI with a more representative distribution of children.

In measuring early vocabulary growth, we found that the correlation between age and vocabulary size reported here for Albanian appears similar to that reported for other languages at a general level (Frank et al., 2017). However, the clear levelling out in growth around 25 months as illustrated in Figure 3 seems to differ somewhat from other languages, and overall, the Albanian-speaking children seem to master a higher number of words than children in many other languages (see Bleses et al., 2008, p. 641). We believe that two reasons may explain this result. First, the distribution of our sample is largely skewed towards older children, as shown in the density plot in Figure 1, and it is likely that these older children know more words in general than the younger children. Note that 27 of the 109 children are older than 30 months of age, which is the recommended upper age limit for CDI II (the version adapted and tested here). The ceiling effect in these children's performance validates the CDI recommendations for 30 months being the upper age limit for CDI II. Second, our sample is also biased towards children whose parents have postsecondary degrees. This high level of education may indeed lead to higher vocabulary levels, or some other factor might be at play (e.g., these parents may have a tendency to claim that their children are doing well by checking off all the words in the CDI checklist). We plan to explore the relationship between children's vocabulary development and parental education level further by having a wider distribution of parental education in future studies.

Another trend noticeable in our data is that during the 2nd year of life, children's vocabularies increase dramatically. Although the number of words for 24-month-old Albanian-speaking children seems quite high, previous work has suggested that toddlers of the same age across different languages show substantial differences when it comes to the number of words in the expressive vocabulary (e.g Bates et al., 1988;

Fernald et al., 2001). The observed difference between the 19- to 21-month old group and the 22- to 24-month old group might also indicate the occurrence of two stages in the development of vocabulary, thus matching previous findings reported in the literature (e.g Bates & Goodman, 2001; Gendler-Shalev & Dromi, 2022). The stages that we discern here roughly match those reported for English, for instance, for which a second phase of vocabulary development begins around 2;0 to 2;6 (Day & Elison, 2022). For Albanian-speaking children, it seems that it is around age 2;0 that a new phase of vocabulary learning begins. These phases have been linked to neuro-maturational changes that impact the route and the speed with which children can acquire new vocabulary (Gendler-Shalev & Dromi, 2021).

Children's vocabulary patterns revealed that the overall number of open-class words (nouns, verbs, and adjectives) was higher than that of closed-class words (prepositions, pronouns, adverbs, interjections, etc.) from the very start (see Figure 6). However, a closer look at children's first 100 words below the age of 20 months showed only 5 verbs alongside 77 nouns and 13 interjections, 2 adverbs, 2 pronouns, and 1 conjunction. These results generally match the findings reported for other languages (Conboy & Thal, 2006; Day & Elison, 2022; Marjanovič-Umek et al., 2011), which establish that toddlers tend to produce more open-class or content words such as nouns, verbs and adjectives earlier than closed-class words such as prepositions, determiners and pronouns.

A second research question focuses on the relationship between children's vocabulary development and whether they refer to displaced events (things that are absent in space or time) or morphology knowledge use. We found that the use of both of these correlated strongly with vocabulary size. On a first look, it appears that the more words children produced, the more knowledge of morphology they used in their speech. These results resonate with those found in other languages in which infants' and toddlers' grammatical development is closely linked to vocabulary size (e.g Bates et al., 1988, 1994; Bleses et al., 2008; Caselli et al., 1999; Day & Elison, 2022; Devescovi et al., 2005; Jackson-Maldonado et al., 1993; Stolt et al., 2009, 2009). But it is not clear if this relationship between early vocabulary and displaced events/morphology is linear or not. It may also be the case that both the lexicon and displaced events/morphology might actually be developing synchronously in the early years of life, along the lines of the proposal put forth in Dixon and Marchman (2007). Indeed, the scores for displaced events, for example, associate positively with vocabulary size, converging with the idea that these close-timing synchronies can be interpreted as evidence that lexicon is not necessarily learned earlier, but is "part and parcel of the child's transition to grammatical language" (Anisfeld et al., 1998).

Our study also investigated the role of language-external factors in early vocabulary development, specifically sex and parental education. The results revealed no effect of sex in children's early vocabulary, displaced events and morphology, which is not

coherent with previous work showing girls outperforming boys at this age (Eriksson et al., 2012; Simonsen et al., 2014). However, another aspect of previous CDI work that has found partial support in our study is the idea that children's early language differs based on their parents' education level (Day & Elison, 2022; Fernald & Marchman, 2012; Hoff & Ribot, 2015). More specifically, we found that paternal educational levels correlated with children's morphology score. The more educated fathers were, the more knowledge of morphology was reported in children's early language. Interestingly, this effect is also reflected in the fact that the word *babi* 'father' is also at the top of Albanian-speaking children's first 100 words, followed by *mami* 'mom' in third place. This role of fathers could be linked to more modern trends in recent years with changes in family structure across the globe and the changing role of men in these structures, wherein more fathers play an active role in children's development (Pancsofar & Vernon-Feagans, 2006). It is not clear, however, that this is the case for Albanian society to date. Studies have also previously shown that in collectivist cultures like those in Asia or South America, fathers indirectly influence children through their effects on the mother-child relationship, providing resources that promote learning and language, and directly through their interactions with children (Tamis-LeMonda et al., 2008). In addition, fathers with more education are typically able to provide more resources and learning opportunities to their children compared to fathers with less education (Cabrera & Peters, 2000). Whether this is true of the Albanian context is still unclear and deserves further investigation, especially given the fact that our sample distribution was somewhat biased with respect to the education levels of caregivers. The patriarchal nature of Albanian society should also be considered in understanding the patterns that we have uncovered here.

A limitation of this study is that it did not fully capture the diverse nature of the Albanian society, with highly educated and urban parents being over-represented in our sample. Future steps should aim to test children that were under-represented in this study, such as those from less urban areas, those brought up by parents that have low educational attainment, as well as those from different dialectal backgrounds in order to better understand lexical and grammatical development of early language in Albanian-speaking children. Procedures for norming of the CDI in the future will be better informed if we have tackled some of these populations beforehand and compared their development to the developmental trajectories reported here.

In sum, this study examined empirical parental report data from the Albanian version of the MacArthur-Bates Communicative Development Inventory (Albanian CDI) concerning the vocabulary and grammatical development of Albanian-speaking children aged 13-36 months. The adaptation reported here covered three sections of the original CDI II Words and Sentences: The vocabulary checklist (Part I Section A), How children use words (Part I Section B), and Word Endings (Part II Section A). The findings outlined here provide insight into patterns of vocabulary growth and the development of communicative competence in Albanian-speaking children from 1 to 3

years old. We have shown that an adaptation of the MB-CDI for Albanian - a yet unstudied language/culture – is able to produce data comparable to those found for other languages using the same instrument, and thus conclude that this adaptation is successful and promising so far. A full adaptation of the entire test is necessary to develop the tool for the assessment of and research on the language development of Albanian-speaking children. Further data collection, instrument improvement, and the full development of Albanian MB-CDI will contribute to shedding more light on the patterns of communicative development in a language and a cultural context that has been largely understudied.

Appendix 1

Table 1. Summary of the beta regression analysis with *morphology* score as dependent variable and *gender, maternal education* and *paternal education* as independent

```
Call:
betareg(formula = morf_beta ~ gender + edma + edfa, data = df2)

Standardized weighted residuals 2:
    Min      1Q  Median      3Q     Max
-1.7720 -0.8010  0.1090  0.5689  1.9643

Coefficients (mean model with logit link):
                Estimate Std. Error z value Pr(>|z|)
(Intercept)      0.19306    0.20576   0.938   0.3481
gendermale      -0.14089    0.25023  -0.563   0.5734
edmaelementary   0.38062    0.63029   0.604   0.5459
edmahigh school -0.08982    0.31714  -0.283   0.7770
edfaelementary  -0.59109    0.62979  -0.939   0.3480
edfahigh school -0.61218    0.29280  -2.091   0.0365 *

Phi coefficients (precision model with identity link):
      Estimate Std. Error z value Pr(>|z|)
(phi)   0.9365     0.1007   9.302   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Type of estimator: ML (maximum likelihood)
Log-likelihood: 47.42 on 7 Df
Pseudo R-squared: 0.07335
Number of iterations: 13 (BFGS) + 2 (Fisher scoring)
```

References

Agalliu, F., Demiraj, S., Domi, M., & Instituti i Gjuhësisë dhe i Letërsisë (Akademia e Shkencave e RSH) (Eds.). (2002). *Gramatika e gjuhës shqipe*. Botimi i Akademisë së Shkencave.

Anisfeld, M., Rosenberg, E. S., Hoberman, M. J., & Gasparini, D. (1998). Lexical acceleration coincides with the onset of combinatorial speech. *First Language, 18*(53), 165–184. https://doi.org/10.1177/014272379801805303

Bates, E., Bretherton, I., & Snyder, L. S. (1988). *From first words to grammar: Individual differences and dissociable mechanisms*. Cambridge University Press.

Bates, E., & Goodman, J. C. (2001). On the inseparability of grammar and the lexicon: Evidence from acquisition. In *Language development: The essential readings.* (pp. 134–162). Blackwell Publishing.

Bates, E., Marchman, V., Thal, D., Fenson, L., Dale, P., Reznick, J. S., Reilly, J., & Hartung, J. (1994). Developmental and stylistic variation in the composition of early vocabulary. *Journal of Child Language, 21*(1), 85–123. https://doi.org/10.1017/S0305000900008680

Bleses, D., Vach, W., Slott, M., Wehberg, S., Thomsen, P., Madsen, T. O., & Basbøll, H. (2008). Early vocabulary development in Danish and other languages: A CDI-based comparison. *Journal of Child Language, 35*(3), 619–650. https://doi.org/10.1017/S0305000908008714

Bopp, F. (1855). *Über das Albanesische in seinen verwandtschaftlichen Beziehungen*. Dümmler.

Bouchard, C., Trudeau, N., Sutton, A., Boudreault, M.-C., & Deneault, J. (2009). Gender differences in language development in French Canadian children between 8 and 30 months of age. *Applied Psycholinguistics, 30*(4), 685–707. https://doi.org/10.1017/S0142716409990075

Çabej, E. (1976). *Studime gjuhësore, III*. Academy of Sciences.Tirana.

Cabrera, N., & Peters, H. E. (2000). Public Policies and Father Involvement. *Marriage & Family Review, 29*(4), 295–314. https://doi.org/10.1300/J002v29n04_04

Caselli, 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. https://doi.org/10.1017/S0305000998003687

Cenko, E. (2017). The Early Acquisition of Verb Constructions in Albanian: Evidence from Children's Verb Use in Experimental Contexts. *Academic Journal of Interdisciplinary Studies, 6*(1), 87–96. https://doi.org/10.5901/ajis.2017.v6n1p87

Cenko, E., & Budwig, N. (2007). The acquisition of early verb constructions in Albanian: A first look at transitives and intransitives. *A Supplement to the Proceedings of the 31st Boston University Conference on Language, 31*.

Cipo, K. (1949). *Gramatika e gjuhës shqipe*. Instituti i Shkencave. Tirana.

Conboy, B. T., & Thal, D. J. (2006). Ties Between the Lexicon and Grammar: Cross-Sectional and Longitudinal Studies of Bilingual Toddlers. *Child Development, 77*(3), 712–735. https://doi.org/10.1111/j.1467-8624.2006.00899.x

Croft, W. (2002). *Typology and Universals* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511840579

Day, T. K. M., & Elison, J. T. (2022). A broadened estimate of syntactic and lexical ability from the MB-CDI. *Journal of Child Language, 49*(3), 615–632. https://doi.org/10.1017/S0305000921000283

deMayo, B. E., Kellier, D., Braginsky, M., Bergmann, C., Hendriks, C., Rowland, C., Frank, M., & Marchman, V. (2021). *Web-CDI: A system for online administration of the MacArthur-Bates Communicative Development Inventories*. https://doi.org/10.34842/KR8E-W591

Demiraj, B. (2018). The evolution of Albanian. In J. Klein, B. Joseph, & M. Fritz (Eds.), *Handbook of Comparative and Historical Indo-European Linguistics* (pp. 1812–1815). De Gruyter. https://doi.org/10.1515/9783110542431-021

Desnickaja, A. V. (1976). Çështje të dialektologjisë historike të gjuhës shqipe. *Jehona, 4*, 305–310.

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. https://doi.org/10.1017/S0305000905007105

Dixon, J. A., & Marchman, V. A. (2007). Grammar and the Lexicon: Developmental Ordering in Language Acquisition. *Child Development, 78*(1), 190–212. https://doi.org/10.1111/j.1467-8624.2007.00992.x

Dule, X. S. (2023). MA Thesis. University of Kaiserslautern.Germany.

Eriksson, M., Marschik, P. B., Tulviste, T., Almgren, M., Pérez Pereira, M., Wehberg, S., Marjanovič-Umek, L., Gayraud, F., Kovacevic, M., & Gallego, C. (2012). Differences between girls and boys in emerging language skills: Evidence from 10 language communities. *British Journal of Developmental Psychology, 30*(2), 326–343. https://doi.org/10.1111/j.2044-835X.2011.02042.x

Fenson, L., Marchman, V., Thal, D., Dale, P., Reznick, J. S., & Bates, E. (2007). *MacArthur-Bates Communicative Development Inventories (CDI) Words and Sentences.* Brookes Publishing.

Fernald, A., & Marchman, V. A. (2012). Individual Differences in Lexical Processing at 18 Months Predict Vocabulary Growth in Typically Developing and Late-Talking Toddlers: Lexical Processing and Vocabulary Growth. *Child Development, 83*(1), 203–222. https://doi.org/10.1111/j.1467-8624.2011.01692.x

Fernald, A., Swingley, D., & Pinto, J. P. (2001). When Half a Word Is Enough: Infants Can Recognize Spoken Words Using Partial Phonetic Information. *Child Development, 72*(4), 1003–1015. https://doi.org/10.1111/1467-8624.00331

Ferrari, S., & Cribari-Neto, F. (2004). Beta Regression for Modelling Rates and Proportions. *Journal of Applied Statistics, 31*(7), 799–815. https://doi.org/10.1080/0266476042000214501

Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2017). Wordbank: An open repository for developmental vocabulary data. *Journal of Child Language, 44*(3), 677–694. https://doi.org/10.1017/S0305000916000209

Gendler-Shalev, H., & Dromi, E. (2022). The Hebrew Web Communicative Development Inventory (MB-CDI): Lexical Development Growth Curves. *Journal of Child Language, 49*(3), 486–502. https://doi.org/10.1017/S0305000921000179

Gjinari, J. (1988). *Dialektologjia shqiptare* (revised and expanded). Akademia e Shkencave e Shqipërisë.Tirana.

Hahn, J. G. (2013). *Studime shqiptare. Origjinali gjermanisht: Albanesiche Studien. Wien: Friedrich Mauke, 1854.* Akademia e Shkencave e Shqipërisë.Tirana.

Hoff, E., & Ribot, K. M. (2015). Language Development: Influence of Socio-Economic Status. In *International Encyclopedia of the Social & Behavioral Sciences* (pp. 324–328). Elsevier. https://doi.org/10.1016/B978-0-08-097086-8.23132-2

Hulle, C. A. V., Goldsmith, H. H., & Lemery, K. S. (2004). Genetic, Environmental, and Gender Effects on Individual Differences in Toddler Expressive Language. *Journal of Speech, Language, and Hearing Research*, *47*(4), 904–912. https://doi.org/10.1044/1092-4388(2004/067)

INSTAT. (2017). *Burrat dhe gratë në Shqipëri—Men and Women in Albania*. Albanian National Institute of Statistics. http://www.instat.gov.al/media/2316/burrat_dhe_grat__ne_shqiperi_2017_libri.pdf

Ismajli, R. (1998). *"Në gjuhë" dhe "për gjuhën" (rrjedha të planifikimit të shqipes në Kosovë (1945-1968)*. Dukagjini.

Jackson-Maldonado, D., Thal, D., Marchman, V., Bates, E., & Gutierrez-Clellen, V. (1993). Early lexical development in Spanish-speaking infants and toddlers. *Journal of Child Language*, *20*(3), 523–549. https://doi.org/10.1017/S0305000900008461

Jia, X. (2023). *Early lexical acquisition of Albanian: An online: CDI-based study*. MA Thesis. University Ludwig-Maximilian, Germany.

Kapia, E. (2010). *The Role of Syntax and Pragmatics in the Structure and Acquisition ofClitic Doubling in Albanian*.PhD Dissertation. Boston University. Boston, MA, USA.

Kapia, E. (2014). Acquisition of Dative and Accusative Clitic Doubling in Albanian: A Syntactic-Pragmatic Approach. In *Developments in the Acquisition of Clitics*. Cambridge Scholars Publishing.

Karmiloff, K., & Karmiloff-Smith, A. (2001). *Pathways to language: From fetus to adolescent.* (pp. ix, 256). Harvard University Press.

Klein, J. S., Joseph, B. D., & Fritz, M. (Eds.). (2017). *Handbook of comparative and historical Indo-European linguistics*. De Gruyter Mouton.

Lehmann, C. (1982). Directions For Interlinear Morphemic Translations. *Folia Linguistica*, *16*(1–4). https://doi.org/10.1515/flin.1982.16.1-4.199

Marchman, V. A., & Bates, E. (1994). Continuity in lexical and morphological development: A test of the critical mass hypothesis. *Journal of Child Language, 21*(2), 339–366. https://doi.org/10.1017/S0305000900009302

Marjanovič-Umek, L., Fekonja, U., Podlesek, A., & Kranjc, S. (2011). Assessing toddler language competence: Agreement of parents' and preschool teachers' assessments. *European Early Childhood Education Research Journal*, *19*(1), 21–43. https://doi.org/10.1080/1350293X.2011.548957

Morgan, J. L., Shi, R., & Allopenna, P. (1996). Perceptual bases of rudimentary grammatical categories: Toward a broader conceptualization of bootstrapping. In *Signal to syntax: Bootstrapping from speech to grammar in early acquisition.* (pp. 263–283). Lawrence Erlbaum Associates, Inc.

Pancsofar, N., & Vernon-Feagans, L. (2006). Mother and father language input to young children: Contributions to later language development. *Journal of Applied Developmental Psychology*, *27*(6), 571–587. https://doi.org/10.1016/j.appdev.2006.08.003
Pedersen, H. (1897). Die albanesichen l-Laute. *KZ*, *33*, 535–551.

R Core Team. (2023). *R: A language and environment for statistical computing* [Computer software]. https://www.R-project.org.

Rusakov, A. (2017). Albanian. In M. Kapović (Ed.), *The Indo-European Languages* (2nd ed., pp. 552–608). Routledge.

Rushi, T. (1983). Rreth rendit të gjymtyrëve në fjalinë dëftore. [De l'ordre des termes dans la phrase énonciative]. *Studime Filologjike*, *1*, 71–85.

Sehitaj, G. (2015). *Zhvillimi i fjalorit dhe i gramatikës tek fëmijet shqipfolës nga mosha 1-3 vjeç*. MA Thesis. Akademia e Studimeve Albanologjike, Insituti i Gjuhësisë dhe Letërsisë.

Shashaj, A. (1996). *Të folurit e fëmijëve të moshës parashkollore dhe puna për zhvillimin e tij*. Shtëpia botuese Onufri.

Simonsen, H. G., Kristoffersen, K. E., Bleses, D., Wehberg, S., & Jørgensen, R. N. (2014). The Norwegian Communicative Development Inventories: Reliability, main developmental trends and gender differences. *First Language*, *34*(1), 3–23. https://doi.org/10.1177/0142723713510997

Smithson, M., & Verkuilen, J. (2006). A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. *Psychological Methods, 11*(1), 54–71. https://doi.org/10.1037/1082-989X.11.1.54

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

Tamis-LeMonda, C. S., Adolph, K. E., Lobo, S. A., Karasik, L. B., Ishak, S., & Dimitropoulou, K. A. (2008). When infants take mothers' advice: 18-month-olds integrate perceptual and social information to guide motor action. *Developmental Psychology,*

44(3), 734–746. https://doi.org/10.1037/0012-1649.44.3.734

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*(1), 3–28. https://doi.org/10.1177/014272370202206401

World Bank. (2018). *Education Statistics: Country at a Glance Albania*. https://datatopics.worldbank.org/education/country/albania

Young, E. (2018). The Borgen Project. *10 Facts about Girls' Education in Albania*. https://borgenproject.org/tag/girls-education-in-albania/

Zogaj, D. (2021). *Design and analysis of the MacArthur Bates Communicative Development Inventory for Albanian*. MA Thesis. University of Kaiserslautern, Germany.

Data, Code and Materials Availability Statement

Data, code, and materials are available at https://osf.io/9fzyt/

Ethics statement

Ethics approval was obtained from the ethics committee of the Academy of Albanological Sciences in Tirana, Albania. Caregivers of the children who participated in this study gave informed written consent for their participation.

Authorship and Contributorship Statement

All authors approved the final version of the manuscript and agree to be accountable for all aspects of the work in ensuring the accuracy or integrity of any part of the work presented here. EK conceived of the study. EK designed the experiment. EK, SA, DZ performed the analyses. EK and SA wrote the manuscript.

Acknowledgements

The first author was supported by a DFG grant (project number 499336853) for this study. We would like to thank Gjyste Sehitaj who did prior work in her Master's thesis leading up to the present paper. This study has also benefitted tremendously from the comments and feedback of two anonymous reviewers and the LDR statistics reviewer, as well as statistics discussions with Stefano Coretta at University of Edinburgh and Michele Gubian at the Institute of Phonetics and Speech Processing at LMU Munich. In addition, we are grateful to all the parents that filled out the questionnaires and participated in this study without any financial compensation.

License