Children's Early Acquisition of Syntactic Category: A Corpus-Based Analysis of English Determiner-Noun Combinational Flexibility

Yumeng Lin¹ & Songqi Li¹

¹ School of Philosophy, Psychology and Language Science, University of Edinburgh, Edinburgh, UK

Correspondence: Yumeng Lin, Unit3, Building No.9, Qianniwasanqu, Fengtai District, Beijing City, China. 100071, China. E-mail: linyumengElsa@126.com

Received: August 30, 2022	Accepted: October 13, 2022	Online Published: October 24, 2022
doi:10.5539/ijel.v12n6p78	URL: https://doi.org/10.5539/ijel.v	v12n6p78

Abstract

While the generativist account posits that an abstract specification of syntactic categories is innate and children show adult-like performance from an early stage, the constructivist account postulates that children's early acquisition of grammatical categories is item-based and reflects limited rules later. The present study tests these assumptions in a specific category, the English determiners. More specifically, we took the controlled measures of overlap (e.g., the use of definite article *the* and indefinite articles *a/an* before the same noun type) in 16 children and their mothers' spontaneous speech as an indicator of determiner-noun combinational flexibility. A series of three studies were conducted, in which we strictly controlled the impact of differences between children and adults in lexical knowledge. In Study 1 and Study 2, we find that children's use of determiners shows a significant difference from adults but this difference disappeared later. Furthermore, Study 3 investigates the influence of external environment with birth order and family's social class as factors and emphasizes that the input factor is worthy of further investigation in future studies. These findings are consistent with one of the constructivist claims, namely that children's early acquisition of determiners is not category-based and their flexibility in using determiners gradually approximates that of adults with development.

Keywords: child syntactic categories, English determiners, grammatical development, input factor, lexical specificity

1. Introduction

1.1 Children's Early Grammar Development

The development of children's syntactic knowledge has long been controversial and one of the core debates focuses on the nature of the early development of grammatical categories. In order to generate adult-like grammar, children not only need to acquire the syntactic category that adult speakers are claimed to possess, but also need to figure out the relationship between categories to arrive at the correct word orders for grammatical sentences (Abbot-Smith, Lieven, & Tomasello, 2004; Akhtar & Tomasello, 1997; Dye et al., 2018; Pinker, 1984). The dominant theoretical accounts of the syntactic category development can be generally categorized as the generativist approach (Carnie, 2006; Guasti, 2004; Radford, 2004; Valian et al., 2009; Valian et al., 2014) and the constructivist approach (Croft, 2001; Croft & Cruse, 2004; Tomasello, 2003). Generativist researchers assume that at least a certain level of linguistic categories and principles of utterance formation are innate, and some even claim that children possess adult-like grammar categories from early on, while constructivist researchers argue that even though they may have the potential to acquire language at birth, they do not develop their grammar innately but rather gradually construct it based on the exposure to input. One of the keys to distinguishing these contrasting theoretical approaches is the productivity and flexibility in using syntactic knowledge at children's early stage of development, in which generativist approaches predict no variation in the productivity of the words that children use to combine within their early stage, whereas constructivist models propose that, as children's grammatical categories are limited, their less productive early knowledge will approximate adult's grammar through data-driven learning.

A number of studies have examined the acquisition of a special syntactic category that is at the core of early grammar, the determiner (Pine & Lieven, 1997; Pine et al., 2013; Valian et al., 2009; Yang, 2013). Since in many languages including English, German, and French, nominal determiners like definite and indefinite articles are obligatory for marking the status of noun referents, the acquisition of determiners is regarded as a central process

of grammar development and the mastery of discourse functions (Bassano, 2015, p. 25). The emergence of using determiners with nouns rather than bare nouns is regarded as a crucial developmental step in children's speech production, indicating the development of grammatical properties of the noun category (Chomsky, 1965; Berwick & Chomsky, 2016, p. 9). There is evidence that children master the basic semantic and morphological distinctions in their determiner system at the age of three or four (Bassano et al., 2008; Bassano et al., 2013; Montrul, 2011). Moreover, some scholars also believe that this ability to show combinational use of determiners with nouns in context is also the hallmark of language evolution, as children's use of a finite inventory of linguistic units to form an unbounded range of meaningful expressions seems to show that they have already acquired a combinational grammar (Goldin-Meadow & Yang, 2017; Hurford, 2012). Nevertheless, there are observations that young children display an optional omission of determiners in obligatory contexts (e.g., saying want cat instead of I want a/the cat), and have limited and formulaic combinatorial flexibility in their early determiner-noun combinations (Tomasello, 2003, pp. 57–58). Empirical cross-linguistic data on children's acquisition of determiners provided evidence of linguistic influences including prosodic and morphological properties (Demuth, 2001, 2011; Lleó, 2001; Kupisch et al., 2009; Guasti et al., 2008; Taelman et al., 2009), lexical-semantic and discourse-pragmatic properties (Bassano et al., 2008; Bassano et al., 2013; Guasti et al., 2008; Kupisch et al., 2009; Ruhlig & Bittner, 2013), and other sociocognitive components such as input effects (Trueswell & Gleitman, 2007; van Dijk & van Geert, 2013) and general cognitive abilities (De Cat, 2011; Skarabela, 2007). Since syntactic knowledge does not exist in a vacuum, various components can interact with each other in children's early grammar development. The influences of these factors and the relations between them on the emergence and development of determiners across languages seem to be complex, and there is still much remaining to be studied given previous mixed results (Guasti et al., 2008; Kupisch, 2007). Except that children may have sensitivity to grammatical morphemes in their input language from an early age, much less is agreed upon about the nature of the determiner optionality in children's early speech, which leaves more questions in the process of the syntactic development to be understood (Gerken & McIntosh, 1993; Hallé et al., 2008; Höhle, et al., 2004; Shi & Melancon, 2010).

1.2 Syntactic Categories in Child Language Acquisition

The much-debated issue is whether children's acquisition of syntactic categories is innate. Opposing views have been raised mainly by two rival accounts, namely the constructivist account and the generativist account. The generativist account's position is that children are born with the innate knowledge of linguistic categories such as verbs and nouns and they learn the details about how such categories apply in their target language (cf. Chomsky, 1957; Valian, 2014). The constructivist account, on the other hand, posits that children start by observing details (cf. Braine, 1963) and gradually create abstract structures (e.g., Tomasello, 2003; Abbot-Smith & Tomasello, 2006). Despite their distinction on the innateness of children's syntactic categories, however, both accounts postulate a certain level of abstractness at the end-state of adult grammar. While previous studies on child syntactic categories can also be divided into the following three possibilities. They vary in their assumptions of whether children arrive at adult-like grammar eventually as well as the explanation of how children obtain this knowledge.

1.2.1 Innate Syntactic Category

The first possibility is that children are born with innate syntactic categories and they need to learn how those abstract categories behave in their target language, which falls under the category of nativism (Valian, 2009, p. 744). Nativists posit that syntactic categories in adults' knowledge can be best represented by basic syntactic categories (e.g., NOUN, VERB). Although this claim has been challenged by opponents, who point out that languages have great diversity in terms of their syntactic categories (e.g., Haspelmath, 2007), nativists put forth the idea that those categories mainly serve as a "toolkit" (Jackendoff, 2002, p. 263) and language users may select "tools" from it. To address the concern of how children acquire the knowledge of such adult-like syntactic categories, three main hypotheses under the nativist account were developed.

The semantic bootstrapping hypothesis (Pinker, 1982, 1984, 1987) assumes that, in addition to syntactic categories such as NOUN and VERB, children also have the innate knowledge of the mapping between semantic basis and syntactic types and functions (see example 1). Admittedly, semantic categories such as action or person are observable in the real world, which would be possible for children to acquire. However, the main issue with this hypothesis is that semantic types do not always link to a unique syntactic category (Benedict, 1979) and vice versa. For instance, actions can be denoted not only by verbs (e.g., rise, fall) but they can also be expressed through prepositions (e.g., up, down).

(1) Alice made a cake.

(Step 1: use the linking rules to posit categories)

name	action	?	thing
NOUN	VERB	?	NOUN

(Step 2: assign the thematic roles and syntactic functions)

agent	action	?	patient
subject	verb		object

The second hypothesis based on the innateness of syntactic categories is the prosodic bootstrapping hypothesis (e.g., Christophe, Millotte, Bernal, & Lidz's, 2008). Its assumption is that phonological and prosodic boundaries provide cues for distinguishing syntactic boundaries. According to Jusczyk (2002, pp. 16–18), children first split clauses into phrases (e.g., [The boy] [is running]) based on their prosodic boundaries and then assign the functional content within each phrase into the relevant categories (e.g., *boy* to NOUN, *running* to VERB). Empirical evidence (Jusczyk, Culter, & Redanz, 1993) suggests that even infants as young as four to five months old show sensitivity to prosodic boundaries, which coincide with syntactic phrase boundaries. Nevertheless, this hypothesis remains problematic since prosodic cues may not be practical without acquiring basic language structure, not to mention that language structure varies in different languages.

The third hypothesis in support of the innateness of syntactic categories is the frequent frames hypothesis (Mintz, 2003, p. 112). It is proposed that children first form distributional clusters based on distributional learning and then assign the syntactic labels (e.g., NOUN, VERB) to the clusters defined by distributional learning. Note that, distinct from the distributional learning approach (see the next section for details of this approach), this hypothesis assumes an innate mechanism that guides the linking between syntactic categories and distributionally defined clusters. However, this hypothesis has yet to be examined in terms of how categories in addition to NOUN and VERB (e.g., DETERMINER) can be linked to distributional categories. Taken together, more attempts are required from nativist accounts to explain how children recognize all syntactic categories.

1.2.2 Induced Syntactic Category

The second possible explanation for child syntactic categories is that they are induced from the input on the basis of distributional learning (Tomasello, 1992, pp. 26–29). For instance, the distributional clustering procedure leads learners to the realization that words after *the* and *a* or *an* are highly likely to be NOUNs and learners are believed to use this distribution information to aid their real-time sentence processing.

Notice that both generativist accounts and constructivist accounts postulate a certain level of abstractness at the end-state of adult grammar. The main distinction between those two accounts lies in how the distributionally defined clusters are utilized. While the generativist accounts posit that those clusters are directly linked to innate categories with the help of an innate linking mechanism (Yang, 2008, p. 206), the constructivists' assumption is that they are directly used in language processing and production. Although distributional analysis has been widely adopted in both generativist studies and constructivist studies, its focus lies on form instead of function or meaning (Ambridge, 2017). Consequently, it lacks explanatory force in terms of how children may utilize the grammatical functions and semantic properties of words.

1.2.3 Illusory Syntactic Category

The illusory category is the third possible explanation for syntactic categories that children develop. According to this proposal, language acquisition is based on storing exemplars (e.g., phrases, sentences) as well as their meanings (Abbot-Smith & Tomasello, 2006, p. 276). In comparison with the induced category through which learners eventually develop abstractions that replace schemas they have memorized in early stages, developing an illusory category indicates that learners merely make real-time analogical comparisons to their stored exemplars without forming free-standing categories such as NOUN or VERB.

It is also posited that a certain level of exemplar representation also exists in adult grammar. For example, adults may show high sensitivity to the distributional properties of specific verbs. Despite the fact that verbs can be followed by various types of complements, language users tend to have preferences for a certain type of complements over others (Trueswell, Tanenhaus, & Kello, 1993). For instance, both the verb *remember* and *suspect* can be used with a noun phrase (NP) complement or a sentence complement as shown in (2). However, *remembered* is more often followed by an NP, whereas the complement of *suspected* is more frequently a sentence. Note that the limited amount of empirical evidence for this approach has yet to explain how the generalization across exemplars is achieved.

(2) a. Elsa remembered the fact....

b. Elsa suspected the fact....

1.3 Previous Studies on Child Determiner Acquisition

A fundamental crux of understanding children's language representations and how they may change through development is their early language productivity and abstraction. This issue has been controversial, and researchers adopting generativist models (e.g., Valian et al., 2013) and constructivist models (e.g., Pine et al., 2013) seek to focus on the existence of adult-like syntactic categories in young children's determiner system. As some researchers oppose constructive models by arguing that the distribution of English determiners in young children supports the generativist claim that children have an adult-like determiner category from early on (Valian et al., 2009; Yang, 2010; Yang, 2013), studies from constructivist researchers point out that these empirical arguments are invalid due to the inappropriate analyses and interpretation of the data (Pine, 2013; Lieven, 2014). The main debate has been over the methodology of how to reliably measure the use of determiners with different nouns in child speakers. Previous findings on children's ability to use a specific category conform to adult grammar are regarded as evidence supporting the adult-like category pattern in children's early stage (Valian, 1986; Ihns & Leonard, 1988). However, this kind of production criteria is too lax in a way that it fundamentally failed to rule out the possibility that children lack adult-like categories, because they may only acquire the knowledge of using different instances of the grammatical category separately in a less abstract and more limited way. On the surface, this knowledge enables children to behave as though they possess adult-like abstract determiner categories, yet they could be completely unaware that the definite and indefinite articles share the same syntactic category. As an alternative, Pine and Martindale (1996, pp. 370–380) advocate a way to measure the extent to which children exhibit overlaps in nouns and predicates with which they combined with different determiners. Instead of measuring children's syntactic knowledge, this relative criterion focuses on measuring the flexibility of children using various instances of the same putative category and comparing it to test whether they are significantly less productive than the adult control (Pine et al., 2013). The data of Pine and Martindale (1996) seem to indicate that children's early determiner use is more in line with an explanation of limited scope as they are less productive than the level if they acquire adult-like determiner categories.

The logic of this overlap measurement is accepted by various scholars, yet the conclusion has received various challenges claiming that the original observations of children's low level of determiner overlap are caused by a sampling artefact in terms of the Zipfian properties in natural linguistic distribution (Valian et al., 2009; Yang, 2013). Based on the result that they computed on a wider range of determiners, Valian et al. (2009) argue that young children do have adult-like levels of overlap and that Pine underestimated children's knowledge of the determiner category because of samples based on a relatively small number of noun sequences, which takes into account of those nouns that only combined once with a determiner as well. Additionally, Yang (2013) maintains that, according to Zipf's law (Chao & Zipf, 1949), relatively few words are used with any high frequency and most words are used very rarely, with many occurring only once in even large samples of text. Yang rejects constructive models and argues that the low overlap scores are sampling artefacts, as the analysis of comparing the overlap scores in child and adult corpus did not show a significant difference in his study when calculated on the assumption.

In response to these critiques, Pine et al. (2013) note that there are flaws in both of these studies, in which Valian's method masked the differences between children and caregivers since it neglected the control of the identity of nouns and increased the sample size in a way that it considerably reduced the sensitivity of the overlap measurement; whereas Yang's corpora comparations were unreliable due to the fact that the nouns in the analysis were uncontrolled with a significantly larger number of nouns in the adult corpus than the child corpus, and the time span in the child corpora included late development up to 48 months which could be irrelevant to the discussion of early syntactic development. After controlling the overlap measurements, Pine and his colleagues carried out a series of corpus analysis studies comparing the performance of children and their caregivers, combining different instances of the determiner category with controlled noun pools in the Manchester Corpus (Pine et al., 2013). Their first study result found that the overlap comparison is highly sensitive to the vocabulary range, and once they control the variables of the sample size and vocabulary range in their following studies, the less flexible performance in children compared to adults is consistent with the assumption that young children's determiner system is less abstract, proving that children's lexical specificity is not a Zipfian artefact and their knowledge of determiner category becomes increasingly abstract over time. Despite both Valian et al. and Yang's analyses confirming the importance of considering sampling factors when assessing the level of overlap in the child corpus, neither of their claims proved that the relatively low level of overlap in young children's speech can be explained purely by sampling issues, and therefore Pine noted that they did not provide any solid evidence to support the claim that children have an adult-like determiner category. The consensus in these studies calculating overlap scores is that Zipf's Law is generally acknowledged to play a crucial role in controlling the frequency distribution, as it improves the measurement of language productivity. Specifically, the lexical items should be used by both the child and the caretaker so that it is possible to randomly reduce the adult vocabulary range to the same size as that of the child (Pineet al., 2013; Lieven, 2014; Rajewski et al., 2012). Once these relevant factors are controlled, young children should demonstrate systematically less productivity in generating the determiner and noun combinations by comparison with their caretakers.

In view of the findings from Pine's analysis, the present study attempts to take their implications into the consideration to further assess children's syntactic category development in different theoretical frameworks, with even younger participants and other potential input factors. Firstly, in terms of the lexical specificity analysis in child-caregiver corpus comparison, data of children's early speech production should be properly interpreted with reference to an index of how lexically specific the speech would be assuming the child had adult-like knowledge. The sampling effect problems are that many vocabularies have low frequency in both child and adult speech since the distribution of words in naturalistic speech obeys Zipf's law and preferable syntactic contexts for words to be grammatical deduce their appearances in other contexts. Thus, in order to derive an index for expected overlap, one approach is to apply the same restricted sampling standards to the targeted corpora for comparisons which enables the analyses to be concerned less about the distributional properties of the language corpora in early development. The comparison can be a synchronic child-mother pair or a diachronic speech comparison of the same child at different points of development. As Pine et al. (2013) suggest, this is a promising way to test the theoretical model that enjoys benefits that the computational modeling approach does not have. The present study adopts this method for the overlap index and extended their research from only child-maternal speech comparison to diachronic comparison in several syntactic developmental stages. Secondly, the measurement of lexical specificity scores needs to be controlled for their sample size as well as the identity of the lexical items in comparison. The present study samples instances of lexical items from data that incorporate the same item effects to ensure our analyses are meaningful comparisons. By sampling an equal number of instances of each item from a fixed control sample, the sampling issues mentioned before, namely that including the item effects of some nouns having different likelihood to be combined with different determiners in adult language, should be solved. Thirdly, the nature of this observed phenomenon has several possible interpretations which require further investigation. When adopting a lexically specific inspection of the controlled speech, the items with a higher frequency that dominate children's early production are less productive than that of their caregivers, and children's flexibility and productivity grow as the frequency of the noun item increases. Since Pine's observation opens to various theoretical accounts, explanations of this finding remain to be explored. The present research evaluates the theoretical explanations discussing whether the syntactic categories are innate, induced, or illusory, and offers insights into the syntactic development mechanism by examining input factors additionally.

1.4 The Present Study

The present study aims to examine children's early syntactic category development by probing into their flexibility in using English determiners *a*/*an*/*the*, as compared with adults (in our case, the mothers). Their flexibility in using determiners is measured by the overlap score, which is calculated by dividing the number of determiner-noun pairs showing overlap (e.g., the noun tokens are modified both by *a*/*an* and by *the*) by the total number of determiner-noun pairs in the selected corpora. After having strictly controlled the sample, three sets of analyses were conducted on the selected corpora. In Study 1, we investigate whether children's categorization (if any) of determiners is adult-like in early developmental stages. To that endeavor, we compare children's and mothers' overlap measures across development and test whether the difference between child and mother (if any) would vary as child MLU increases. In Study 2, rather than looking at overlap measures of children according to their language ability represented by MLU, we take time as the indicator and look for differences among the same group of children and mothers in two test points. Furthermore, a follow-up analysis in Study 3 is included to delve into whether other input-related variables (e.g., birth order, social class) could have an impact on children's overlap measure.

In these studies, predictions vary under the three possible theoretical explanations about child syntactic categories as introduced above. Assuming that children are equipped with innate categories from the first place, we would expect to find their overlap measures as high as their mothers' and it should remain the case throughout their development or, in our case, two different test points. Considering that generativists emphasize more on the innateness of the category, input quality and quantity should not make much difference. If, on the other hand, children's generalization of determiners is induced gradually based on distributionally defined

clusters, they are predicted to show less flexibility in using determiners (as compared with their mothers) but their flexibility increases gradually over time. We may also observe that, in early stages, children only use a certain article rather than others (e.g., *a/an* instead of *the* or vice versa) since they store exemplars through rote-learning. A third possibility is that children may not form abstract syntactic categories at all, but they analogically compare their input with stored exemplars and such exemplars may exist for a long time even when learners grow up. In that case, we expect to find that, similar to the last prediction, children may show less overlap (compared with their mothers) in using determiners since they are still storing exemplars and their performance will improve at a later time when they can make judgments based on their stored exemplars. Another prediction following this hypothesis is that, regardless of whether those learners form abstract syntactic categories eventually, they may still show better performance or higher sensitivity when encountering exemplars that they have stored at the initial language acquisition stages. As for the input influence, both the induced and illusory perspective value it to have related impacts as item-based and exemplar-based ways of language acquisition require children to store information from language exposures. Despite its significance in distinguishing whether syntactic categories are induced (usage-based) or illusory (exemplar-based), finding such evidence would require more data from studies on adults or longitudinal studies and therefore may only be partially addressed in the current study due to the limitation of the scope.

2. General Method

All the studies adhere to the same basic method for corpus analysis using the CLAN program (MacWhinney, 2000). This involves searching transcripts in CHAT format for determiner plus noun sequences in the corpora. Determiners plus nouns pairs can be identified through the mor-line. Instances of a/an and the followed by nouns either directly or with intervening words like adjectives in between are extracted for analysis.

2.1 Corpora

All analyses were based on the Howe Corpus (Howe, 1981), downloaded from the Child Language Data Exchange System (CHILDES) (MacWhinney, 2000). The corpus includes transcripts from 16 Scottish child-mother pairs (seven girls and nine boys) during a session they played with toys at home. Participants lived in Glasgow, Scotland, and were randomly selected from a small university and nearby villages. Social class was divided into the "middle class" (with fathers of 13 children having professional or managerial occupations) and the "working class" (with fathers of 11 children having skilled or semiskilled manual occupations). Data of 40 minutes duration were collected at two time points. More specifically, children aged 1:6 to 1:8 (mean 1:7) in the first test, and 1:11 to 2:1 (mean 2:0) in the second test. In each videotaped session, children had 20 minutes to play with their own toys and 20 minutes in which they played with a special set of toys presented in a specific order.

2.2 Analysis Control

In order to control not only the sample size but also the identity of nouns that reflects the frequency of nouns combined with either one of the articles, samples were controlled through three criteria in each individual corpus. All criteria must be met before assessing the existence of overlap in the controlled samples. Overlap scores were then calculated by dividing the number of nouns showing overlap by the total number of nouns that met the controlling criteria.

Note that, in all of the analyses, overlap scores were calculated under strict control of sampling size. This was achieved by random sampling (with replacement) determiner + noun tokens from the mother's corpus in a way that, for each noun type (e.g., the noun *man*), the mother's corpus consisted of the same number of tokens as the child's. For instance, if a child's corpus only contains two tokens containing the noun *man* modified by a determiner, we would randomly select two determiner + *man* tokens from the relevant mother's corpus regardless of how many determiner + *man* exist in the mother's corpus. We took the process of random sampling as a necessary step because, according to the Zipfian distribution, nouns occurring with higher frequency (tokens) would theoretically show a higher probability to be modified by both indefinite articles and definite articles, therefore more easily showing overlap (Pine et al., 2013). After the process of random sampling, the final overlap scores for each mother were then calculated by averaging their overlap scores in 100 times of random sampling.

Criterion 1: Nouns must be grammatical with both a/an and the (i.e., singular countable nouns). As in Pine et al.'s (2013) critique of Valian et al.'s (2009) controlled measures, considering overlaps of all determiners (e.g., some, my, another, etc.) could reduce the sensitivity of overlap measures and make overlap much easier. To avoid the seemingly high overlap value, this research focused on the analysis of indefinite determiner a/an and definite determiner the to ensure the sensitivity of overlap measurements to reflect flexibility.

Criterion 2: Nouns must appear at least twice in both the child's and the mother's speech. On the assumption that noun distribution obeys Zipf's law (Chao & Zipf, 1949), the proportion of nouns occurring with low frequency is likely to be higher in adults' corpus than in children's corpus because adults own larger noun vocabulary size. To avoid undermining the overlap scores of adults on nouns that their children produce and exclude the impact of differences in lexical knowledge, it is important to control the identity of nouns within each child-mother pair.

Criterion 3: Nouns must appear at least twice with either *a/an* or *the* in both the child's and the mother's speech. As Valian et al. (2009) pointed out, the overlap is, by definition, impossible in a noun occurring only once with a determiner. Nouns modified by determiners should appear at least twice so that overlap can possibly occur.

These three criteria enable the research to compare measures of noun overlap after controlling the sample size and identity of relevant nouns as well as the frequency of these nouns combined with either a/an or the. The location of these nouns in the Zipfian frequency distribution is a crucial factor in determining the size of overlap scores because lexical items in naturalistic speech interact with differences in vocabulary range in a way that those nouns with low frequency will have a higher probability of overlap in adults than in children with a smaller noun vocabulary range. Therefore, it is likely to mask differences in the overlap between young children and their mothers without proper control of the noun identity. The current research addressed the issue by directly comparing overlap scores between child and adult pairs on an equivalent number of instances of a shared set of nouns that contained the same fixed set of nouns and the same number of a/an/the + noun tokens for each noun in the pool.

3. Study 1: Comparison of Overlap Scores in Children and Their Mothers

Study 1 aims to investigate whether children have an adult-like determiner category in early development by comparing child overlap measures with adult overlap measures. If children do have an adult-like determiner category, they should show noticeable overlap to an extent that is close to their mother's performance and it shall remain the same regardless of children's MLU values. Otherwise, differences should be noticed in children's flexibility in using determiners compared with their mothers, which will later decrease as children's MLU increases.

3.1 Method

Analyses in Study 1 were achieved first by comparing children's overlap measures with their mother's. Considering that in Study 1 we would like to know how children's performance can vary as their language ability (represented by their MLU) varies regardless of the time point they were tested, recordings at two test points from the same child were treated as two distinct samples in Study 1. 18 out of 32 samples were excluded for failing to meet the controlling criteria (see Appendix A for controlling results). Overlap scores after random sampling, which was repeated 100 times, were then calculated for each child and the mother. In addition to delving into the differences between children and their mother (child overlap score subtracted by mother overlap score), we also probed into whether such differences (if any) would vary as children's language ability (represented by their MLU) varies.

3.2 Results

Table 1 shows child and mother overlap scores of the 14 controlled samples. Also presented is the child's MLU and the average number of tokens per noun type for each child-mother pair. Note that, due to the control of sample size, the number of noun types and noun tokens were the same within each child-mother pair, therefore resulting in the same ratio for their tokens per noun type.

Sample	MLU	Tokens Per Noun Type	Overlap S	core	MOT-CHI Overlap Difference
			Child	Mother	
Graham1	1.09	2.00	0.00	0.40	0.40
Barry1	1.30	4.33	0.00	0.18	0.18
Wayne1	1.33	6.00	0.00	0.57	0.57
Wayne2	1.39	4.25	0.75	0.43	-0.32
Oliver1	1.53	2.00	0.00	0.27	0.27
Philip2	1.54	3.33	0.67	0.00	-0.67
Eileen2	1.57	2.33	0.00	0.31	0.32
Nicola2	1.67	2.00	0.33	0.00	-0.33
Yvonne2	1.69	6.00	0.00	0.69	0.69
Yvonne1	1.72	6.00	0.00	0.00	0.00
Kevin2	1.87	2.00	0.00	0.00	0.00
Sally2	1.98	3.33	0.50	0.36	-0.14
Oliver2	2.04	6.00	0.43	0.46	0.03
Richard2	2.13	3.00	0.75	0.00	-0.75

Table 1. Overlap scores of child and mother in the 14 controlled same

Before investigating the difference between mother overlap score and child overlap score, a correlation test was conducted to find out whether previous sampling consideration was effective. This was achieved by correlating both children's and mothers' overlap scores with the average number of tokens per noun type. Analysis indicated a marginal correlation for mothers' overlap measures (r = .47, df = 12, p = .09 < .10), suggesting that mother's overlap scores are positively correlated with tokens per noun type to a marginal extent. In line with Pine et al.'s (2013) results in which marginal to significant correlation was found for mothers, our result also indicated that controlling the determiner + noun tokens in the mother's corpus was a necessary controlling measure, without which mother's overlap scores would in principle be higher than children's if they say more determiner + noun tokens than children do. However, no significant correlation was found for children (r = .05, df = 12, p = .87). Although this non-significance was different from the marginal to significant correlation found in Pine et al. (2013), our result is actually reasonable considering the possibility that children may not have a determiner system but rather stick with a certain combination (e.g., *a cat*, *a man*, *a tiger*), in which case children's use of determiners remains the same pattern regardless of how many determiner + noun tokens they repeated (e.g., in the corpus of Barry in test 1, he used *a* rather than *the* even though he repeated the noun *man* for three times, *tiger* for five times, and *cat* for five times).

With the effectiveness of the sampling consideration confirmed, further analyses were carried out to probe into the difference between children and mothers. It is noticeable from the descriptive data in Table 1 that variation exists in child and mother overlap since the difference between them is not zero in most cases. These patterns of results were verified by paired sample t-test, which did not find significant differences between children and mothers (t = -.15, df = 13, p = 0.88). Although the overall result of the t-test suggests no significant difference between mother and child in the flexibility in using determiners, it does not rule out the possibility that differences between the two groups may exist at first but diminish later as children's MLU increases, thus showing no significant differences in total.

To further examine the possibility that the difference between groups diminishes with the increase of children's language ability (symbolized by MLU), linear regression analysis was run to test the relationship of children's MLU and overlap difference between each child and mother (calculated by subtracting child overlap score from mother overlap score), with tokens per noun type as a main factor as well. As is shown in Figure 1 and Table 2, we found a marginal relationship between children's MLU and overlap difference (p = .095 < .10) whereas the fixed effect of tokens per noun type showed no significant nor marginal relationship with overlap difference (p = .226). It indicates that, as children's language ability grows, their differences with their mother decreases.

Table 2. Relationship between children's MLU and overlap difference

	Estimate	SE	t	р	
(Intercept)	.778	.630	1.236	.242	
Children's MLU	660	.361	-1.826	.095	
Tokens per noun type	.084	.065	1.284	.226	

Note. Formula: Overlap difference ~ MLU + Tokens per noun type.



Figure 1. Scatterplot of child MLU and mother-child overlap difference

To summarize, although we did not find a significant difference between the child overlap measure and the mother overlap measure when we looked at all the data overall, we found evidence supporting the possibility that their difference may start to be large but later diminishes gradually as children's MLU increases.

4. Study2: Comparison of Overlap Scores in Children at Two Time Points

While the previous analysis in Study 1 was based on children's MLU which is relevant to children's grammatical abilities, the goal of Study 2 is to investigate child-mother differences from a different perspective by comparing controlled overlap measures at two time points. If children have an adult-like determiner category from early on, no change in children's overlap measure is predicted. However, if children's early knowledge of determiners is unlike adults', we expect to find children's performance to increase over time.

4.1 Method

Whereas Study 1 looked at how children's performance varies as MLU increases, Study 2 focused on children's variation in performance in the two tests recorded within a five-month interval. The controlling criteria were the same as mentioned in the general method section. While in Study 1 data from the same child at two test points were treated as two separate samples each represented by the child's MLU at the corresponding test (thus resulting in 14 valid child-mother pairs of samples out of 32 samples), data in Study 2 treated data from test 1 and test 2 as an independent variable, which led to 16 child-mother samples overall (see Appendix B for the overlap scores of all 16 samples at two test points). Out of the 16 samples, five child-mother samples were eliminated since no applicable data was left in both test 1 and test 2 after strict selecting criteria. As a result, 11 child-mother pairs of samples were included in the analysis. Note that in the eleven samples, some overlap scores were represented by "NA" because there was no utterance left in the individual's corpus in that specific test. For the convenience of calculation, the NAs were replaced with zero in the following analyses.

4.2 Results

Table 3 presents the controlled overlap scores for each of the 11 child and mother pairs during Test 1 and 2. According to Table 3, children's average performances were lower than mothers' in Test 1 (mean = .00 versus mean = .15) but surpassed mothers' score in Test 2 (mean = .31 versus mean = .21). It can also be observed in Figures 2 and 3.

This pattern was confirmed by a 2x2 repeated measures ANOVA, which revealed a non-significant main effect of participant (p = .86), a marginally significant main effect of test (p = .09), and a significant interaction between participant and test (p = .04). Post hoc analysis using pairwise comparisons confirmed that the children showed marginally significantly less overlap than mothers during Test 1 (p = .059). We found significant increase of children's overlap across Test 1 and 2 (p = .009), and marginally significant increase between mothers' overlap scores in Test 1 and 2 (p = .09). There was no significant difference between children and mothers during Test 2 (p = .92). Taken together, the results suggest that, compared with mothers, children showed less flexibility in Test 1, but their performance significantly increased in Test 2 to an extent that no difference with their mothers was observed in Test 2.

These results not only confirm significant differences in the flexibility of children and adults in using determiners but also indicate that children's use of determiners becomes significantly more flexible with development over time and even surpassed the level of their mothers. Note that mothers' average overlap score increased from Test 1 to Test 2, it does not necessarily mean that mothers also improved their flexibility in using determiners. This is because, as most studies (from both constructivists and generativists) recognized, the grammatical ability of adults is already at a stable stage. Rather, it could be explained by the assumption that overlap scores tend to increase with sample size, so does the difference between overlap scores based on nouns used by both the mother and the child. Furthermore, there is a considerable amount of evidence that child-directed speech (CDS) has the adaptive phenomenon, in which adults tend to adapt their language to the child's ability using simplified vocabulary and syntax (Huttenlocher et al., 2007; Van Dijk & Van Geert, 2011). Although there is the possibility that those caretakers adapted their linguistic ability as their children develop, since our study controlled the sample size as well as the identity of nouns and only focused on the limited linguistic area of English determiner, the relevant potential influence should have been controlled.

	Child overlap in Test 1	Mother overlap in Test 1	Child overlap in Test 2	Mother overlap in Test 2
Barry	0.00	0.18	NA	NA
Eileen	NA	NA	0.00	0.32
Graham	0.00	0.40	NA	NA
Kevin	NA	NA	0.00	0.00
Nicola	NA	NA	0.33	0.00
Oliver	0.00	0.27	0.43	0.46
Philip	NA	NA	0.68	0.00
Richard	NA	NA	0.75	0.00
Sally	NA	NA	0.50	0.36
Wayne	0.00	0.57	0.75	0.43
Yvonne	0.00	0.19	0.00	0.69
Mean	0.00	0.15	0.31	0.21

Table 3. Overlap scores of child and mother in test 1 and test 2

Note. 'NA' refers to 'Not Applicable'. It indicates that there was no utterance left in the individual's corpus in that specific test after having controlled the identity and frequency of nouns and the vocabulary range. 'NA' was counted as 0 for calculation, but it was represented this way for readers to understand what happened.



Figure 2. The line chart of child and mother overlap in test 1 and test 2



Figure 3. The box plot of child and mother overlap in test 1 and test 2

5. Study 3: Investigation of Other Input Variables

As children need to process the input in order to learn from it, non-syntactic factors also interact with syntactic development. Previous cross-linguistic evidence has shown that external environmental factors such as input quantity and quality and parent's educational level as well as socioeconomic status (SES) can modulate children's language development (Hurtado et al., 2008; Oller & Eilers, 2002; Weisleder & Fernald, 2013). It is important to realize that social classes and communicative partners may affect variability in input quality and quantity, thus influencing the rate of acquisition. For instance, former studies have suggested that the amount of speech that mothers address impacts vocabulary growth differently in middle-class children and in working-class children (Huttenlocher et al., 1991; Hart & Risley, 1998). Children's interaction with their younger peers might also influence their language acquisition (Stiegliz et al., 2013). Accordingly, Study 3 investigated whether input variation in social class or birth order influences children's overlap scores.

5.1 Methods

Analyses in Study 3 were achieved by comparing children's overlap measures across social class and birth order. The corpora selection followed the criteria mentioned in the general method section. Out of all 32 samples, 14 samples that did not contain "NA" value entered the analysis in Study 3.

5.2 Results

The analysis was achieved by a Factorial Mixed ANOVA, which factors MLU, and the other factors are social class (middle or working) and birth order (0 as non-first-born, 1 as first-born). It suggests that there was no significant main effect of MLU, and neither birth order (p = .32) nor social class (p = .69) had a significant effect on overlap scores. Nevertheless, there was a marginally significant interaction between social class and birth order (p = .081). As shown in Figure 4, it seems that the effect of birth order has different impacts on the overlap score in children from different classes, in which first-born children in the middle class have lower overlap scores than non-first-born children, while the pattern is the opposite in working-class families. However, since the social environment has complex impacts on input, the interaction of this social class factor and birth order factor on overlap score can be interpreted as a reflection of the complexity of social contexts, in which the class and birth order differences may not be directly informative on the quality and quantity of their language environments.



Figure 4. The line chart of child overlap with different birth order and social class

6. Discussion

Present analyses aim to distinguish and test the predictions of three main assumptions of child syntactic category, namely that child syntactic category is innate (category-based), induced (usage-based), or illusory (exemplar-based). More specifically, we tested these assumptions on a specific syntactic category, English determiners. To probe into whether an adult-like determiner category exists in children's early developmental stages, we compared child and mother overlap across language development (measured by children's MLU) and across two time points (Test 1 and Test 2). The vocabulary range and the frequency and identity of nouns occurring with either *a/an* or *the* have been strictly controlled.

Analyses in Study 1 investigated whether there were differences between children's flexibility in using determiners and their mothers' (in our case, mothers') and how the difference (if any) would vary as children's language ability (measured by MLU) increases. The results of this study showed that, although overlap measures did not vary between groups overall, the difference between each pair of mother and child decreased as child MLU increased. This trend provides evidence for the assumption under the constructivist account that, as children's language ability develops, they are gradually catching up with adults' flexibility in using determiners.

However, the growingly smaller difference between children and their mothers is only half of the picture. We would also like to know whether the trend also exists when looking at their differences across the time dimension. To that endeavor, we investigated whether there was a significant increase in children's overlap scores from Test 1 to Test 2 and whether their performance differed from their mothers at two test points within a five-month interval. The results of this study indicated marginally significant increase in overlap scores for both children and adults. Additionally, the results captured a marginally significant difference between child flexibility and mother flexibility in Test 1 which later diminished in Test 2. Taken together, results from Study 1 and Study 2 provide evidence that children do not have an adult-like syntactic category in early development, and that their flexibility in using determines increases over development and gradually approximates their adults' flexibility.

In Study 3, we investigated the implication of language input on children's overlap scores by analyzing the effects of birth order and social class. Although the result of this study only revealed a marginally significant interaction between these factors, the absence of any significant main effects could be due to the limitation of controlling measures and sample size. Previous cross-linguistic research has shown that external environmental factors involving the frequency and quality of input relate to the acquisition of vocabulary or language development and, in particular, the number and type of children's communicative partners including adults and young peers can be responsible. The social-economic status factor could partially account for the variability in input quantity and caregiver-child interaction since better SES can enhance the degree of engagement that adult adopts in communicative interaction with their children and increase input quantity and quality which facilitate children's language development (Brown, 2001). Nevertheless, relevant elements such as the form of input including child-directed speech (CDS), whether children overheard (Stiegliz et al., 2013; Weisner & Gallimore, 1977), and the structure and quantity of the input can be hard to control. As a result, it is likely that the information on social class and birth order in this corpus is not sufficient enough to measure and reflect the real input situation. Additionally, due to the highly restricted feature of the definite and indefinite articles in the English determiner

system, the differences in overlap flexibility in early child development may be subtle. During early stages, children predominantly use indefinite DPs, and their high rates of optionally omitting English determiners from obligatory contexts subsides to a minimum of around 36 months old (Abu-Akel et al., 2004). As the average age of our child participants in both tests are 19 months and 24 months, the fact that they are still at an early stage of being susceptible to making omission errors with limited overlap scores may mask the developmental differences caused by input factors.

In terms of the input influence on determiner acquisition, past studies focusing on CDS influences on the emergence of determiners obtained mixed results. The influence of the input factor on the development of the determiner system has been controversial, with various studies focusing on bidirectional influences between CDS and the gradual development of children's determiners (Bassano et al., 2011; Huttenlocher et al., 2007; van Dijk & van Geert, 2013; van Geert & Steenbeek, 2005). Nevertheless, due to the variation in the analysis of different relevant indicators, the results are inconsistent and thus the impact of such a factor is still unclear. There is a need for further research with more sophisticated approaches to examine the relationships between these factors in children's acquisition of determiners and syntactic development. Since from theoretical perspectives of syntactic development, the role of input plays different roles in generativist accounts versus constructivist accounts, the discussion of relevant factors also facilitates revealing the mechanism of syntactic development. Those who advocate nativist accounts of language acquisition emphasize children's innate ability for language acquisition and their maturational constraints, and this innate schedule in language learning renders them to regard input quality and quantity as being less important (Chomsky, 1981, 1995; Valian, 2014). Despite the claim from nativist researchers that children could identify and assign instances of innate category in the language input, their explanations including bootstrapping and distributional learning (Lleó, 2001; Mintz, 2003; Ruhlig & Bittner, 2013) still involve certain input influences as children need to have input observations for a successful allocation and deduction of syntactic category, and yet the input factor seems to be less focused in their discussion. Instead, some generativists even propose that children's evolved biological capacity for language learning generates combinatorial productivity without external linguistic input (Goldin-Meadow & Yang, 2017). They advocate that their statistical results provided evidence that, instead of memorizing specific word combinations from caregiver speech, children have a productive grammar that follows abstract rules including the determiner-noun combination in which they combine words independently (Yang, 2013).

In contrast, input plays an essential role for children to eventually arrive at the abstractness of syntactic categories in constructivist theories. The traditional constructivist account assumes that children have an initial item-based learning mechanism in their beginning stages, where they memorize chunks available in the input as holophrases, and then break up memorized chunks into less lexically specific slot-and-frame schemas. The adult-like abstract representation with syntactic categories may only be gradually reached after passing the third developmental stage where specific lexical items are abstracted from stored schematization and analogy on the basis of enough language-by-language input (Croft, 2001; Croft & Cruse, 2004). Thus, this process is characterized as item-based and data-driven, since the higher the frequency of a string encountered the input, the better a child would exhibit in his linguistic performance on an equivalent stored utterance. McCauley and Christiansen (2014) evaluated the abilities of two computational models on children's flexibility in determiner-noun combination, and their analysis showed that the item-based model (McCauley & Christiansen, 2011), in which it gradually creates an inventory of building blocks, outperforms the class-based model in which the built-in syntactic categories like determiner and noun are combined independently (Yang, 2013). They claim that the simulation of the developmentally motivated item-based learning model improved with exposure to more linguistic input, and it successfully captures the actual determiner-noun combinations in the dense child corpus. In summary, although the current result on input factors may not be informative, it can be regarded as a probe into revealing the environmental factor in children's early grammatical development, which appeals for more research to provide evidence for the theoretical debates on the mechanism of syntactic development.

7. Conclusion

Given our results on English determiner and noun combinations, there are several insightful findings for the competing theories in discussing children's development of syntactic categories. First, regarding the research question of children's early syntactic development, we argue that even when the lexical and performance ability of children and adult speakers are controlled, children do not have adult-like performance in their early grammatical development. This is based on the result that the flexibility difference only gradually decreases as child MLU increases and the existence of the significant overlap difference between children and their mothers in the chronological study. Since children's overlap data follows a pattern where their less flexible ability to use determiners gradually catches up with adult flexibility, these studies provide evidence resonating with the

constructivist theory assuming that children's less flexible productivity progressively approximates adult linguistic performance in later developmental stages. As the analyses in this research indicate that the low overlap score in the early syntactic development is not a Zipfian artefact, we interpret this catching-up phenomenon as an indicator that, instead of having adult-like syntactic knowledge at the beginning as nativist theories assert, children's syntactic category is progressively becoming abstract through data-driven learning rather than innateness. Note that this discussion focuses more on the result of whether children have an adult-like syntactic category at early stages, thus being insufficient to unravel the detailed process of how they arrived at such abstractions and generalizations (if any).

In order to further understand how the acquisition mechanism work and also consider the third possibility of exemplar-based accounts' claim that syntactic categories are illusory rather than induced, this research additionally evaluate the corpus data and found combinational patterns in children who already exhibited the ability to make noun-determiner combinations. In addition to omission errors of determiner in an obligatory context like "Back the car." "It tiger.", there are ungrammatical combinations like "a back" in inappropriate contexts that appeared in several children's corpora. With closer examination, instead of acquiring this combination from the mother's speech, all of these "a back" utterances were initiated by the child in child-mother pairs. According to the exemplar-based proposal that learners are simply storing the original sentence strings with detailed probabilistic information, it seems unlikely that this combination of the indefinite article would be followed with the noun "back" referring to a space. As this account proposes that the surface form of young children's language productions is generalized across their stored multiword linguistic exemplars, in which there are fine-grained distributional properties with frequency information, this kind of rare combination should not easily emerge by itself. Moreover, the optionality of using determiners also poses challenges to the exemplar-based account. In view of the child-mother corpus, there is no adults' use of inappropriate bare nouns, yet the variability of the child's omission of determiners is observed. The key difference between the exemplar-based account and a traditional constructivist account (usage-based) is whether there is a free-standing abstraction of categories like NOUN and DETERMINER in the end state of syntactic development. Regarding the similar part of item-based processing in both the induced and illusory syntactic categories, there has been computational and psycholinguistic evidence suggesting that, despite children's increasingly flexible and abstract grammatical categories, item-based processes continue to play an important part throughout development (Arnon & Snider, 2010; Croft, 2001; Goldberg, 2006; McCauley & Christiansen, 2011). To conclude, our results of the gradual increase of flexibility in young children's determiner use can be taken as evidence that their knowledge of syntactic categories is not innate since they were not adult-like from the beginning and their flexibility did not remain the same but rather it increases with the development of their language ability. In addition, the findings on input impact as well as combinational errors in child corpus provide new perspectives for future research to answer.

The present results yield implications for the investigation of child syntactic development. To begin with, they verified the effectiveness of the sampling considerations proposed in Pine et al. (2013). That is to control the frequency and identity of nouns within each child-mother pair and calculate the average overlap score based on results of random sampling (with replacement) repeated 100 times. This finding is necessary because, theoretically, mothers are more likely to show overlap if they produce more tokens for a specific noun as compared with their children. Therefore, controlling both the noun type and the number of determiner + noun tokens can eliminate the potential bias and create an equal probability for the child and the mother to show overlap. The results suggest that, as found in Pine et al. (2013), the mother's overlap score increases as the tokens per noun type increases. However, unlike in Pine et al. (2013), where child overlap score was also positively correlated with tokens per noun type, no significant nor marginal correlation was found between them in the current study. Our finding is actually reasonable in that if children do not have an adultlike syntactic system but rather they build their understanding of determiners on schemas or exemplars (e.g., if they just say "a cat" rather than "a cat" or "the door" rather than "a door"), the number of noun tokens would not impact their overlap possibility in the first place. The implication is that the control of the noun identity and noun frequency is necessary so that the theoretical possibility for overlap to occur would be the same for the child and the mother, leading us to see the actual overlap without the bias caused by the higher noun frequency in either the mother or the child corpora.

A second implication is that the population involved in this study is younger (age range 1:6 to 1:8 in Test 1 and 1:11 to 2:1 in Test 2) than in previous studies on children's early syntactic categories. For instance, children aged from 1:8.22 to 2:0.25 when starting the one-year data collection in Pine et al. (2013), in which significant differences were found among children and caregivers in Phase 1 but not in Phase 2, suggesting that children have caught up with their caregivers over time. In addition to confirming the similar trend that children's

flexibility approximates that of adults, our findings put forth the idea that this catching-up time may occur even earlier (before children reached the age range 1:11 to 2:1) than previous studies suggested.

On top of the new evidence on when children catch up with adults from the dimension of time (distinguished by Test 1 and Test 2 which were five months apart), this study also yields credible results from another perspective, namely language development (represented by child MLU). Note that although Pine et al. (2013, p. 354) reported significant differences among all five development phases (with Phase 1, Phase 2 & 3, Phase 4 & 5 each roughly corresponding to Brown's Stage I, Brown's Stage II, and Brown's Stage III) (Brown, 1973), it was unclear whether the difference varies for children across different developmental stages. However, the present study bridges the gap with the discovery that the difference between mother overlap and child overlap gradually decreases as the child's language ability develops (measured by the child's MLU).

There are, however, some limitations in the current study. Due to strict controlling criteria, the amount of valid data entering the analyses was not sufficient enough. As a result, the overlap scores of some children and mothers were represented with NA, which is less informative. Additionally, those NAs were replaced by zero in the calculation in Study 2, which may open the result to several possible interpretations. Recall that NA in the overlap score column does not necessarily mean that the individual has zero flexibility in using determiners. Rather, the individual might have used the D + N combination several times which might show overlap but this data was not included in the analysis since this noun did not appear in both the mother and the child corpora. It should be noted that the replacement of NA with zero may decrease the absolute value of the average overlap scores for both the mother and her child. NAs occur for a child-mother pair at the same time owing to the controlling criteria and differences between child-mother pairs where overlaps scores were marked as NA will be taken as zero. In this way, the overlap difference between groups in Test 2 may be lower than the actual situation considering that children and mothers may show a certain level of overlap. However, the replacement of NA with zero should not heavily bias the data since there were only two child-mother pairs in which NA exists out of eleven pairs. The overall trend that children approximate adults should remain similar, whereas the exact timepoint and to what extent children catch up with adults requires future investigation with larger sample and higher data collection frequencies (e.g., more tests evenly separated in the age range of 1:6 to 2:1).

Acknowledgments

It should be noted that the authors contributed equally to the study. The authors would like to thank Yaoyuan Fan for suggestions on the coding for random sampling.

References

- Abbot-Smith, K., Chang, F., Rowland, C., Ferguson, H., & Pine, J. (2017). Do two and three year old children use an incremental first-NP-as-agent bias to process active transitive and passive sentences?: A permutation analysis. *PloS One*, *12*(10), e0186129. https://doi.org/10.1371/journal.pone.0186129
- Abbot-Smith, K., & Tomasello, M. (2006). Exemplar-learning and schematization in a usage-based account of syntactic acquisition. *The Linguistic Review*, 23(3), 275–290. https://doi.org/10.1515/TLR.2006.011
- Abu-Akel, A., Bailey, A. L., & Thum, Y. M. (2004). Describing the acquisition of determiners in English: A growth modeling approach. *Journal of Psycholinguistic Research*, 33(5), 407–424. https://doi.org/10.1023/B:JOPR.0000039548.35396.c2
- Akhtar, N., & Tomasello, M. (1997). Young children's productivity with word order and verb morphology. *Developmental Psychology*, 33(6), 952–965. https://doi.org/10.1037/0012-1649.33.6.952
- Ambridge, B. (2017). Chapter 22. Syntactic Categories in Child Language Acquisition. In H. Cohen & C. Lefebvre (Eds.), *Handbook of Categorization in Cognitive Science* (2nd ed., pp. 567–580). https://doi.org/10.1016/b978-0-08-101107-2.00023-3
- Bassano, D. (2015). The acquisition of nominal determiners. *The Acquisition of Reference*, 15, 25–49. https://doi.org/10.1075/tilar.15.02bas
- Bassano, D., Maillochon, I., Korecky-Kröll, K., Van, D. M., Laaha, S., Dressler, W. U., & Van G., P. (2011). A comparative and dynamic approach to the development of determiner use in three children acquiring different languages. *First Language*, 31(3), 253–279. https://doi.org/10.1177/0142723710393102
- Bassano, D., Maillochon, I., & Mottet, S. (2008). Noun grammaticalization and determiner use in French children's speech: A gradual development with prosodic and lexical influences. *Journal of Child Language*, 35(2), 403–438. https://doi.org/10.1017/S0305000907008586
- Bassanol, D., Korecky-Krollz, K., Maillochon, I., & Dressler, W. U. (2013). The acquisition of nominal

determiners in French and German. Grammaticalization and First Language Acquisition: Crosslinguistic Perspectives, 50, 37. https://doi.org/10.1075/bct.50.02bas

- Benedict, H. (1979). Early lexical development: Comprehension and production. *Journal of Child Language*, 6(2), 183–200. https://doi.org/10.1017/S0305000900002245
- Berwick, R. C., & Chomsky, N. (2016). Why only us: Language and evolution (p. 224). MA: MIT Press. https://doi.org/10.7551/mitpress/9780262034241.001.0001
- Braine, M. D. S. (1963). The ontogeny of English phrase structure: the first phase. *Language*, 39(1), 1–13. https://doi.org/10.2307/410757
- Brown, P. (2001). Learning to talk about motion UP and DOWN in Tzeltal: is there a language-specific bias for verb learning? In *Language acquisition and conceptual development* (pp. 512–543). Cambridge University Press. https://doi.org/10.1017/CBO9780511620669.019
- Carnie, A. (2006). Guglielmo Cinque & Richard S. Kayne (eds.), The Oxford handbook of comparative syntax. Oxford: Oxford University Press, 2005. Pp. xii+ 977. *Journal of Linguistics*, 42(2), 444–452. https://doi.org/10.1017/S0022226706224095
- Chao, Y. R., & Zipf, G. K. (1949). Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology. *Language*, *26*, 394. https://doi.org/10.2307/409735
- Chomsky, N. (1957). Syntactic Structures. The Hague: Mouton. https://doi.org/10.1515/9783112316009
- Chomsky, N. (1965). Aspects of the Theory of Syntax. MIT Press. https://doi.org/10.21236/AD0616323
- Christophe, A., Millotte, S., Bernal, S., & Lidz, J. (2008). Bootstrapping lexical and syntactic acquisition. Language and Speech, 51(1-2), 61-75. https://doi.org/10.1177/00238309080510010501
- Croft, W. (2001). *Radical construction grammar: Syntactic theory in typological perspective*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198299554.001.0001
- Croft, W., & Cruse, D. A. (2004). Cognitive linguistics. Cambridge University Press. https://doi.org/10.1017/CBO9780511803864
- Demuth, K. (2001). Prosodic constraints on morphological development. *Language Acquisition and Language Disorders*, 24, 3–22. https://doi.org/10.1075/lald.24.02dem
- Demuth, K. (2011). The acquisition of phonology. In J. Goldsmith, J. Riggle & A. Yu (Eds.), *The handbook of phonological theory* (pp. 29–79). Malden, MA: Blackwell. https://doi.org/10.1002/9781444343069.ch17
- Dye, C., Kedar, Y., & Lust, B. (2019). From lexical to functional categories: New foundations for the study of language development. *First Language*, *39*(1), 9–32. https://doi.org/10.1177/0142723718809175
- Gerken, L., & McIntosh, B. J. (1993). Interplay of function morphemes and prosody in early language. *Developmental Psychology*, 29(3), 448. https://doi.org/10.1037/0012-1649.29.3.448
- Goldberg, A. E. (2006). *Constructions at work: The nature of generalization in language*. New York: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199268511.001.0001
- Goldin-Meadow, S., & Yang, C. (2017). Statistical evidence that a child can create a combinatorial linguistic system without external linguistic input: Implications for language evolution. *Neuroscience & Biobehavioral Reviews*, 81, 150–157. https://doi.org/10.1016/j.neubiorev.2016.12.016
- Guasti, M. T. (2017). Language acquisition: The growth of grammar. MIT press. https://doi.org/10.1093/oxfordhb/9780199573776.013.12
- Guasti, M. T., Gavarró, A., De Lange, J., & Caprin, C. (2008). Article omission across child languages. *Language* Acquisition, 15(2), 89–119. https://doi.org/10.1080/10489220801937346
- Hallé, P. A., Durand, C., & de Boysson-Bardies, B. (2008). Do 11-month-old French infants process articles? Language and Speech, 51(1-2), 23-44. https://doi.org/10.1177/00238309080510010301
- Haspelmath, M. (2007). Pre-established categories don't exist: Consequences for language description and typology. *Linguistic Typology*, *11*(1), 119–132. https://doi.org/10.1515/LINGTY.2007.011
- Höhle, B., Weissenborn, J., Kiefer, D., Schulz, A., & Schmitz, M. (2004). Functional elements in infants' speech processing: The role of determiners in the syntactic categorization of lexical elements. *Infancy*, 5(3), 341–353. https://doi.org/10.1207/s15327078in0503_5
- Howe, C. (1981). Acquiring language in a conversational context. New York: Academic Press.

- Hurford, J. R., & Hurford, J. R. (2012). *The origins of grammar: Language in the light of evolution II* (Vol. 2). Oxford University Press.
- Hurtado, N., Marchman, V. A., & Fernald, A. (2008). Does input influence uptake? Links between maternal talk, processing speed and vocabulary size in Spanish learning children. *Developmental Science*, *11*(6), F31–F39. https://doi.org/10.1111/j.1467-7687.2008.00768.x
- Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M., & Lyons, T. (1991). Early vocabulary growth: relation to language input and gender. *Developmental Psychology*, 27(2), 236. https://doi.org/10.1037/0012-1649.27.2.236
- Huttenlocher, J., Vasilyeva, M., Waterfall, H. R., Vevea, J. L., & Hedges, L. V. (2007). The varieties of speech to young children. *Developmental Psychology*, 43(5), 1062. https://doi.org/10.1037/0012-1649.43.5.1062
- Ihns, M., & Leonard, L. B. (1988). Syntactic categories in early child language: Some additional data. *Journal of Child Language*, 15(3), 673–678. https://doi.org/10.1017/S0305000900012630
- Jackendoff, R. (2002). *Foundations of language: Brain, meaning, grammar, evolution*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198270126.001.0001
- Jusczyk, P. W. (2002). How Infants Adapt Speech-Processing Capacities to Native-Language Structure. *Current Directions in Psychological Science*, 11(1), 15–18. https://doi.org/10.1111/1467-8721.00159
- Jusczyk, P. W., Cutler, A., & Redanz, N. J. (1993). Infants' Preference for the Predominant Stress Patterns of English Words. *Child Development*, 64(3), 675–687. https://doi.org/10.2307/1131210
- Krajewski, G., Lieven, E. V., & Theakston, A. L. (2012). Productivity of a Polish child's inflectional noun morphology: a naturalistic study. *Morphology*, 22(1), 9–34. https://doi.org/10.1007/s11525-011-9199-0
- Kupisch, T. (2007). Testing the effects of frequency on the rate of learning: determiner use in early French, German, and Italian. In *Frequency effects in language acquisition-defining the limits of frequency as an explanatory concept* (pp. 83–113). https://doi.org/10.1515/9783110977905.83
- Kupisch, T., Anderssen, M., Bohnacker, U., & Snape, N. (2009). Little words: Their history, phonology, syntax, semantics, pragmatics and acquisition (pp. 223–236). Article acquisition in English, German, Norwegian and Swedish. Proceedings of The Georgetown University Round Table, GURT 2007.
- Lieven, E. (2014). First language development: a usage-based perspective on past and current research. *Journal of Child Language*, 41(S1), 48–63. https://doi.org/10.1017/S0305000914000282
- Lleó, C. (2001). The interface of phonology and syntax: The emergence of the article in the early acquisition of Spanish and German. *Language Acquisition and Language Disorders*, 24, 23–44. https://doi.org/10.1075/lald.24.03lle
- MacWhinney, B. (2000). The CHILDES project: Tools for analysing talk (3rd ed.). Mahwah, NJ: Erlbaum.
- McCauley, S. M., & Christiansen, M. H. (2011). Learning simple statistics for language comprehension and production: The CAPPUCCINO model. In L. Carlson, C. Hölscher & T. Shipley (Eds.), *Proceedings of the* 33rd Annual Conference of the Cognitive Science Society (pp. 1619–1624). Austin, TX: Cognitive Science Society.
- McCauley, S. M., & Christiansen, M. H. (2014). Reappraising lexical specificity in children's early syntactic combinations. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 36, No. 36).
- Mintz, T. H. (2003). Frequent frames as a cue for grammatical categories in child directed speech. *Cognition*, 90(1), 91–117. https://doi.org/10.1016/S0010-0277(03)00140-9
- Montrul, S. (2011). Multiple interfaces and incomplete acquisition. *Lingua*, 121(4), 591–604. https://doi.org/10.1016/j.lingua.2010.05.006
- Oller, D. K., & Eilers, R. E. (Eds.). (2002). *Language and literacy in bilingual children* (Vol. 2). Multilingual Matters. https://doi.org/10.21832/9781853595721
- Panagos, R. J. (1998). Meaningful Differences in the Everyday Experience of Young American Children. *Remedial and Special Education*, 19(6), 371. https://doi.org/10.1177/074193259801900609
- Pine, J. M., Freudenthal, D., Krajewski, G., & Gobet, F. (2013). Do young children have adult-like syntactic categories? Zipf's law and the case of the determiner. *Cognition*, 127(3), 345–360. https://doi.org/10.1016/j.cognition.2013.02.006

- Pine, J. M., & Lieven, E. V. (1997). Slot and frame patterns and the development of the determiner category. *Applied Psycholinguistics*, 18(2), 123–138. https://doi.org/10.1017/S0142716400009930
- Pine, J. M., & Martindale, H. (1996). Syntactic categories in the speech of young children: The case of the determiner. *Journal of Child Language*, 23(2), 369–395. https://doi.org/10.1017/S0305000900008849
- Pinker, S. (1982). A theory of the acquisition of lexical interpretive grammars. In J. Bresnan (Ed.), *The Mental Representation of Grammatical Relations* (pp. 655–726). Cambridge, MA: MIT Press.
- Pinker, S. (1984). Language learnability and language development. Cambridge, MA: Harvard University Press.
- Pinker, S. (1987). The bootstrapping problem in language acquisition. In M. W. Brian (Ed.), *Mechanisms of Language Acquisition* (pp. 339–441). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Radford, A. (2004). *English Syntax: An Introduction*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511841675
- Ruhlig, N., & Bittner, D. (2013). Introduction: Emergence and applications of the Lexical Bootstrapping Hypothesis. In *Lexical Bootstrapping* (pp. 1–10). De Gruyter Mouton. https://doi.org/10.1515/9783110308693.1
- Shi, R., & Melançon, A. (2010). Syntactic categorization in French learning infants. *Infancy*, 15(5), 517–533. https://doi.org/10.1111/j.1532-7078.2009.00022.x
- Stieglitz, J., Gurven, M., Kaplan, H., & Hooper, P. L. (2013). Household task delegation among high-fertility forager-horticulturalists of lowland Bolivia. *Current Anthropology*, 54(2), 232–241. https://doi.org/10.1086/669708
- Taelman, H., Durieux, G., & Gillis, S. (2009). Fillers as signs of distributional learning. *Journal of Child Language*, 36(2), 323–353. https://doi.org/10.1017/S0305000908008982
- Tomasello, M. (1992). *First verbs: A case study of early grammatical development*. Cambridge, MA: Cambridge University Press. https://doi.org/10.1017/CBO9780511527678
- Tomasello, M. (2003). *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Cambridge, MA: Harvard University Press.
- Trueswell, J. C., Tanenhaus, M. K., & Kello, C. (1993). Verb-specific constraints in sentence processing: Separating effects of lexical preference from garden-paths. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19, 528–553. https://doi.org/10.1037/0278-7393.19.3.528
- Valian, V. (1986). Syntactic categories in the speech of young children. Developmental Psychology, 22(4), 562. https://doi.org/10.1037/0012-1649.22.4.562
- Valian, V. (2014). Arguing about innateness. *Journal of Child Language*, 41(S1), 78–92. https://doi.org/10.1017/S0305000914000336
- Valian, V., Solt, S., & Stewart, J. (2009). Abstract categories or limited-scope formulae? The case of children's determiners. *Journal of Child Language*, 36(4), 743–778. https://doi.org/10.1017/S0305000908009082
- Van D., M., & Van G., P. (2011). Exploring patterns of adaptation in child-directed speech during the process of early grammaticalization in child language. *Language, Interaction and Acquisition*, 2(1), 61–81. https://doi.org/10.1075/lia.2.1.03dij
- Van G., P., & Steenbeek, H. (2005). The dynamics of scaffolding. *New Ideas in Psychology*, 23(3), 115–128. https://doi.org/10.1016/j.newideapsych.2006.05.003
- Weisleder, A., & Fernald, A. (2013). Talking to children matters: Early language experience strengthens processing and builds vocabulary. *Psychological Science*, 24(11), 2143–2152. https://doi.org/10.1177/0956797613488145
- Weisner, T. S., Gallimore, R., Bacon, M. K., Barry III, H., Bell, C., Novaes, S. C., ... Williams, T. R. (1977). My brother's keeper: Child and sibling caretaking [and comments and reply]. *Current Anthropology*, 18(2), 169–190. https://doi.org/10.1086/201883
- Yang, C. (2008). The great number crunch. *Journal of Linguistics*, 44, 205–228. https://doi.org/10.1017/S0022226707004999
- Yang, C. (2013). Ontogeny and phylogeny of language. *Proceedings of the National Academy of Sciences*, *110*(16), 6324–6327. https://doi.org/10.1073/pnas.1216803110

Appendix A Selection of 32 Samples

Sample	Nouns that meet Criterion 1 and 2	Nouns that meet	Taken into the controlled
		Criterion 3	sample or not
Barry 1	baby bear cat drink hat man monkey nose push tiger	cat tiger man	Yes
Barry 2	bed	NA	No
Eileen 1	NA	NA	No
Eileen 2	baby bear bed car hat man shop tail tellie train	baby shop train	Yes
Faye 1	chair girl top	NA	No
Faye 2	book box car door	NA	No
Graham 1	box dress elephant fence lion lorry tiger	lorry	Yes
Graham 2	baby bear book box clock doll giraffe helicopter horse mouse owl post	NA	No
	teddy van zoo		
Ian 1	car color elephant garage house penguin place	NA	No
Ian 2	breakfast car cup doll door garage gate	NA	No
Jason 1	book cock teddy	NA	No
Jason 2	bear boat box car cheek clock doll door giraffe letter pocket shop sock	NA	No
	teddy train		
Kevin 1	baby bed cake car clock door duck face hat lid man sleep teddy tip tractor	NA	No
Kevin 2	baby basket day door man panda play shop towel	basket	Yes
Lucy 1	baby drink	NA	No
Lucy 2	baby bottle cup egg fork hat spoon	NA	No
Melanie 1	NA	NA	No
Melanie 2	box girl lady play	NA	No
Nicola 1	baby ball dog man nose	NA	No
Nicola 2	apple baby bed bird boat box brush chair dickle dinner dog drawing	dickle doge	Yes
	flower garden girl hat lady letter nose pencil post pot prickle sleep tiger	flower	
Oliver 1	baby camel chair door elephant house kangaroo llama lorry man monkey	door	Yes
	penguin tortoise		
Oliver 2	baby bed blanket car crash cup door engine field hole horse leg man	baby car door	Yes
	monkey morning night ride sleep spoon wheel	field man	
		monkey spoon	
Philip 1	dog man	NA	No
Philip 2	baby book bus car cow dog letter	baby car dog	Yes
Richard 1	elephant	NA	No
Richard 2	bang car cup elephant hole horse kangaroo lorry play top train	elephant lorry	Yes
		train	
Sally 1	bird book bottle box chicken church doll hat lorry milk plane pot sleep	NA	No
Sally 2	animal ball bear bench book brick camel camera car cat cup doll elephant	camel camera	Yes
j.	hippo horse house man mouse play top tower	car cup man	
	II	tower	
Wavne 1	bag ball banana car door man show	ball car door	Yes
Wavne 2	baby ball bath book box car coat door drawer lorry man shoe spoon tune	ball bath box	Yes
· · · · j · · · =	,	man	
Yvonne 1	baby bag ball bed box cat man picture rabbit teddy	man	Yes
Yvonne 2	animal body car cat deer doll eye garden knee man picture	cat man	Yes

Note. 'Barry 2' in 'Sample' column = the second recording of Barry; "Nouns that meet Criterion 1 and 2" = nouns that are countable and occurred twice or more than twice in both the mother and the child speech; 'Nouns that meet Criterion 3' = Nouns that occurred with a determiner twice or more than twice; 'Yes' = there are nouns left in the second selection and the sample can remain in the controlled sample whereas; 'No' = no nouns in this sample meet all of Criterion 1, 2, and 3; 'NA' = no nouns in this sample meets the current criterion/criteria.

	Child overlap in Test 1	Child overlap in Test 2	Mother overlap in Test 1	Mother overlap in Test 2
Barry	0.00	NA	0.18	NA
Eileen	NA	0.00	NA	0.32
Faye	NA	NA	NA	NA
Graham	0.00	NA	0.40	NA
Ian	NA	NA	NA	NA
Jason	NA	NA	NA	NA
Kevin	NA	0.00	NA	0.00
Lucy	NA	NA	NA	NA
Melanie	NA	NA	NA	NA
Nicola	NA	0.33	NA	0.00
Oliver	0.00	0.43	0.27	0.46
Philip	NA	0.67	NA	0.00
Richard	NA	0.75	NA	0.00
Sally	NA	0.50	NA	0.36
Wayne	0.00	0.75	0.57	0.43
Yvonne	0.00	0.00	0.19	0.69
Mean	0.00	0.21	0.10	0.14

Appendix B Complete Data of Controlled Child and Mother Overlap Scores for Test 1 and 2

Note. 'NA' refers to 'Not Applicable'. It indicates that there was no utterance left in the individual's corpus in that specific test after having controlled the frequency of nouns and the vocabulary range. 'NA' was counted as 0 for calculation, but it was represented this way for the readers to understand what happened in that corpus.

Copyrights

Copyright for this article is retained by the author, with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/4.0/).