Determining the Abstractness of Determiners

Charles Yang[a]

[a]University of Pennsylvania

Virginia Valian[b]

[b]Hunter College and CUNY Graduate Center

Corresponding author: Charles Yang, charles.yang@ling.upenn.edu, Department of Linguistics, University of Pennsylvania, Philadelphia, PA 19104 USA

Virginia Valian, virginia.valian@hunter.cuny.edu, Department of Psychology, Hunter College – CUNY, 695 Park Avenue, New York, NY 10065 USA

CRediT statement:
Charles Yang: Conceptualization, Methodology, Software, Validation, Formal Analysis, Investigation, Data Curation, Writing – Original Draft, Writing – Review & Editing, Visualization.
Virginia Valian: Conceptualization, Methodology, Validation, Formal Analysis, Investigation, Writing – Original Draft, Writing – Review & Editing, Visualization.

Abstract

Via four empirical studies and one mathematical analysis we demonstrate that children's early grammars represent the abstract category determiner and use instances of that category, specifically *a* and *the*, productively, with the same nouns. We use the Manchester corpus of 12 children, visited 34 times over a one-year time period starting around age 1;10. Analyses 1 and 2 replicate previous empirical and formal accounts showing that both children and adults use the determiner category productively. Analyses 3 and 4 replicate a sampling-based method (Pine, Freudenthal, Krajewski, & Gobet 2013, *Cognition*) and identify several paradoxical results, including a *reductio ad absurdum*, with that method. The mathematical formulation in Analysis 5 reveals the design flaws of the sampling method and shows that all results, including the paradoxical ones, are predictable, thus strengthening the conclusion that children have productive knowledge of determiners. We discuss the implications of our studies for conclusions about children's early syntactic knowledge and suggest how empirical and methodological studies of determiners can inform future research on language acquisition.

Keywords: grammatical development; modeling language acquisition; syntactic categories; corpus analysis methods; language productivity

Highlights:

- We demonstrate that children have the abstract syntactic category of determiner from the onset of combinatorial speech.
- We demonstrate that certain sampling methods intended to remove bias instead create bias, leading to a *reductio ad absurdum* that adults with sparse output lack the determiner category.
- We demonstrate the need for rigorous mathematical analysis of quantitative methods in the study of language acquisition.

1. Introduction

Two little words – *a* and *the* – have been the focus of debates about the abstractness of children's early grammatical representations. On one hand are researchers who claim that categories are abstract and genuinely syntactic as soon as children's speech can be systematically analyzed (Valian, 1986; Valian, Solt, & Stewart, 2009; Yang, 2013). On the other hand are investigators who claim that children's apparent early mastery of syntactic categories is instead an example of lexically-based formulae (Pine, Freudenthal, Krajewski, & Gobet, 2013; Pine & Lieven, 1997; Pine & Martindale, 1996). According to these theorists, children say "a ball" not because they understand that the class of determiners (including members like *a*) takes the class of Nouns (such as *ball*) as their complement, but because they have heard "a ball" multiple times.

Determiners are a particularly good focus of inquiry. First, unlike nouns, which potentially could be learned via ostension, determiners do not have referents. They have semantic content, but not content that can be readily pointed to. Second, at some point in development, as all participants in the discussion agree, children's syntactic categories, including determiners, are genuinely formal and syntactic. The end state is clear. The question is how early the child demonstrates that knowledge. Third, determiners are the thin edge of the wedge. If determiners are abstract from the outset of combinatorial speech, so are certain other categories such as nouns, which are entailed by determiners. It is impossible to have an abstract representation of determiners without also having an abstract representation of nouns.

If evidence shows that children represent determiners abstractly very early – say, at the onset of combinatorial speech, around age 2 for most children – that demonstration puts constraints on possible learning theories: those theories must yield abstract categories at least as soon as connected speech is present. For that reason, early acquisition provides support to theories that propose early, perhaps innate, availability of categories (e.g., Valian et al., 2009; Yang, 2013). Early acquisition is a challenge to theories that require more prolonged input and learning, such as empiricist approaches (e.g., Pine et al., 2013) in which the child starts with low-level details and successively generalizes from the exemplars she has stored until she eventually creates an abstract category (e.g., Abbot-Smith & Tomasello, 2006; Ibbotson & Tomasello, 2009).

1.1. Overlap as evidence of syntactic productivity

What, then, will count as evidence for syntactic categories? One approach is to use standard linguistic diagnostics and see whether children's "language" patterns as expected if they in fact have categories, such as by substituting the word *it* for a noun phrase (Valian, 1986). Valian reported that children passed standard diagnostics for determiner, noun, noun phrase, preposition, and prepositional phrase. A criticism of the diagnostic approach is that children might pass some diagnostic tests with a single correct example (Pine & Lieven, 1997; Pine & Martindale, 1996).

Another approach is to see if children and their parents (henceforth, parents and adults are referred to as mothers) are equally flexible in how they combine determiners and nouns. The overlap test (Pine & Lieven, 1997; Pine & Martindale, 1996) examines a subset of determiners, the articles *a* and *the*, and their use before singular nouns. If both children and their mothers use *a* and *the* with their noun types to the same extent, children are credited with the determiner category. For example, if the child only says "a ball", while the mother says both "a ball" and "the ball", that suggests that children do not have an equivalence class of determiners, but merely use frequently encountered combinations, akin to formulae. Pine and colleagues reported that children showed less overlap than their mothers and hence did not have the category determiner and claimed that two-year-olds failed the test, showing significantly less overlap than their mothers did.

In this paper we use the overlap test, while noting some of its conceptual limitations. For example, restricting determiners to *a* and *the* risks losing information about the breadth of children's knowledge: the equivalence class of determiners is much larger. If the goal is to establish whether children's grammars represent the category determiner, then looking at a subset of children's determiners is arbitrary and limits how much generalization it is possible to observe. Nor is it obvious that children and mothers *should* show the same degree of overlap. Children should show some overlap but one cannot argue from *less* overlap to the absence of a category without other ancillary premises, as we have discussed elsewhere (Valian et al., 2009; Yang, 2013).

Early implementations had methodological issues, such as including nouns that were only used only once (Pine & Martindale, 1996), thus making overlap impossible (Valian et al., 2009), or defining developmental periods so that the second period included the first (Pine & Martindale, 1996), so that both children and mothers appeared to increase overlap over time (Valian et al., 2009), or using small samples (Pine & Lieven, 1997), some of which were not audiotaped.

Later implementations that corrected for methodological problems found that there was no difference between children and their mothers. In six different analyses, Valian et al. (2009) showed that, once proper controls were included, there was no difference between children and their mothers. They used the Valian cross-sectional corpus consisting of 21 children ranging in age from 1;10 to 2;6 and in MLU from 1.53 to 4.38. Sample size – the number of times a noun was used with a determiner – was critical: for both children and mothers, to the same degree, overlap was a function of the opportunity to observe overlap. Apparent increases in overlap in both child and mother could be understood as due to an increase in sample size and attendant opportunity to observe overlap.

In a variety of other tests, Valian et al. (2009) showed that a) when the range of children's determiners was not restricted to *a* and *the*, but included all determiners, children showed massive overlap with no change as a function of MLU, b) when children and mothers were compared on exactly the same Det-N (determiner-noun) pairs, overlap was even greater, again with no difference between children and mothers, c) children's and mothers' formulaicity was identical, and d) children made negligibly few syntactic errors involving determiners.

Yang (2013) developed a statistical benchmark for productivity by considering the effects on overlap of two factors. The first is Zipf's Law (1949), according to which many word/noun types will appear sparsely in any linguistic sample, providing few opportunities to be used with both determiners, which necessarily lead to low overlap values. The second factor is the tendency for nouns to be used more with one article than the other. For example, *the bathroom* appears more frequently than *a bathroom* whereas *a bath* is more common than *the bath.* We refer to this imbalance in favor of a particular determiner as determiner bias, which we describe in detail in Section 3. Multiple occurrences of a noun may be paired with the favored determiner exclusively, again reducing the overlap score. Taking both statistical facts into account, Yang (2011, 2013) provided a method for calculating the expected overlap values of the nouns in a corpus under the assumption that their combination with the two determiners is fully productive and thus interchangeable. Quantitative analyses show that even for children under age 2, the average overlap value of nouns is statistically indistinguishable from what could be expected from a fully productive rule. Similarly to Valian et al.'s (2009) finding that overlap depended on the number of times a noun was used with a determiner, Yang found that overlap was predicted by the token/type ratio of nouns, i.e., the average number of times nouns are used in a sample of determiner-noun combinations.

Metrics of different sorts, all aimed at measuring productivity and overlap, with different groups of children in English, and in several other languages, have almost uniformly confirmed the presence of overlap across a corpus (e.g. Goldin-Meadow & Yang, 2017, home sign; Joo & Yoo, 2018; Meylan, Frank, & Levy, 2013, English; Meylan, Frank, Roy, & Levy, 2017, English; Silvey & Christodoulopoulos, 2016, English; Szagun & Schramm, 2019, German). The findings in children's production are matched by findings in comprehension (for sample findings, see Shi & Melançon, 2010; Kedar, Casasola, & Lust, 2006; Kedar, Casasola, Lust, & Parmet, 2017; for reviews, see Dye, Kedar, & Lust, 2019; Valian, 2009; Valian, 2013).

The combined production and comprehension data would thus seem to lead to two conclusions: very young children and mothers overall do not differ in their productive use of Det-N combinations; children understand that determiners are an equivalence class. The failure to detect productivity in some studies is due to small sample sizes and inadequate acknowledgment of the extent to which nouns "prefer" one or another determiner.

1.2. Current study

One study, however, which examines the Manchester corpus, is an outlier (Pine et al., 2013). The Manchester corpus includes twelve children recorded 34 times each (with a few exceptions), starting around age 1;10 and ending a year later. Pine et al. developed a sampling-based method, which we describe in detail later, to compare children's and mothers' overlap; the procedures were intended to fairly compare children and mothers. These authors also introduced a specific way of partitioning the children's longitudinal data into five developmental phases. They report that the Manchester children show less overlap than do their mothers, especially at very young ages.

We argue that two restrictions in Pine et al. (2013)'s sampling method introduce significant biases in the estimation of both children's and mothers' overlap values and thus incorrectly represent the syntactic productivity of each group. The methods lead to paradoxical results: for instance, they will assess some adults to be less grammatically productive than other adults, and, indeed, less productive than themselves under certain conditions.

In the current paper, we present five analyses to investigate the productivity of determiners in the Manchester corpus. The code, data, and results for all analyses are available on Github at https://github.com/charles-yang-upenn/Determiner-Productivity.

Analyses 1 and 2 are full replications of the findings of Valian et al. (2009) and Yang (2013), respectively: when all nouns are used in overlap calculation, mothers and children both show full productivity – including at the earliest developmental stage in the case of children. Analysis 3 replicates as closely as possible the main findings of Pine et al. (2013). Analysis 4 tests the method by comparing two adults with each other, and by comparing subsets of maternal data with the full data set. If the sampling method is distortion-free, two different adults should show comparable rates of overlap, and a mother's subset should be comparable to her entire corpus. Analysis 5 provides a formal analysis of the biases introduced by two questionable restrictions in Pine et al.'s methods.

For this and all subsequent analyses reported in this paper, we extracted Det-N pairs for each child and each mother (a total of 24 samples, 12 children and 12 mothers) in the Manchester Corpus. The data extraction method follows that in Pine et al. (2013). We used the mor-tier within CHILDES to identify Det-N pairs consisting of *a/an* or *the* with a singular noun immediately following the determiner, which was the method of data extraction in previous work (e.g., Yang, 2013). We also include cases where a word intervenes between the determiner and noun, following Pine et al. (2013). Note that the latter method introduces errors, such as coding 'the…home' as a Det-N pair in 'all the way home'. Nevertheless, we adopted Pine et al.'s method in order to meet one of our principal goals, namely, to replicate their results. Our statistical results do not change significantly when we exclude nonadjacent Det-N pairs.

2. Analysis 1 – Replication of Valian, Solt, and Stewart, 2009

2.1. Introduction

One of the principal results of Valian et al. (2009) was that overlap increased, equally for child and mother, as a function of how often a noun was used with a determiner. When overlap was stratified by number of opportunities for overlap, the correlation between overlap and opportunity for overlap was .80 for children and .83 for mothers. That analysis examined all determiners, not just *a/an* and *the*. We replicate that analysis here, using only *a/an* and *the*.

2.2. Method

For each sample, overlap is computed as the percentage of noun types used with both determiners divided by the total number of noun types used with either determiner (or both determiners). We computed both raw overlap values and arc-sine-transformed values (because overlap is a proportion) and found the same results with each. For Tables 1 and 2 below we report the raw data, which include, for child and mother, the number of noun types, the number of noun tokens, the tokens/type ratio, the degree of determiner bias, the predicted overlap, and the observed overlap.

2.3. Results and discussion

As Fig. 1 shows, we successfully replicated the findings of Valian et al. (2009)'s Analysis 3, which plotted overlap as a function of the number of the number of times a noun type appeared with a determiner. Child overlap was significantly correlated with the number of times a noun appeared with a determiner, ranging from 2-8 or more times (r=.96, p = .001), as was mother overlap (r=.98, p < .001). There are no differences between children and mothers (paired two-tailed t(7)= .96, p = 0.37). Both children and mothers show more overlap, to the same extent, as the number of noun types used with a determiner increases.

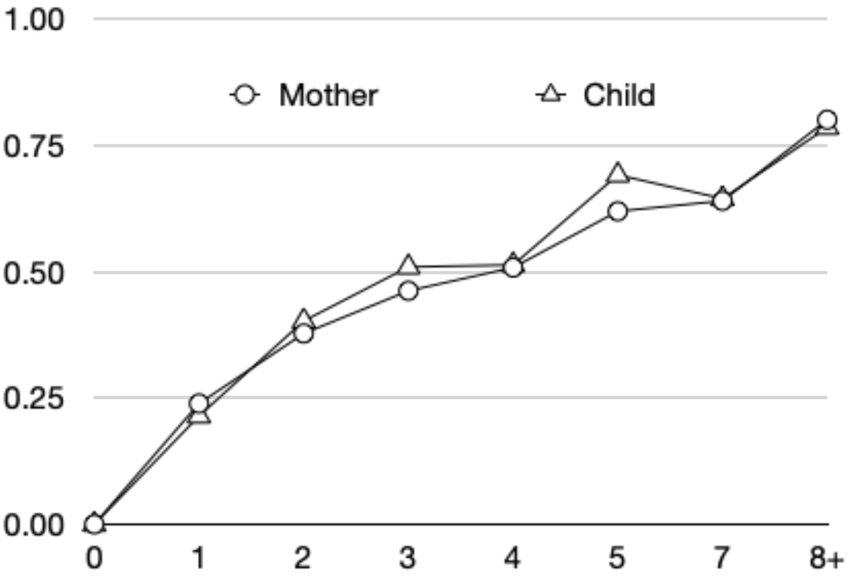

Figure 1
Determiner overlap for children and mothers as a function of opportunity for overlap

The correlational results are robust and the correlations are even higher than those reported by Valian et al. (2009). Two differences between the earlier analysis and this one may be relevant. Here we include only *a* and *the*, while Valian et al. included all determiners. Here we observed 12 longitudinal children, while Valian et al.'s sample consisted of 21 cross-sectional children.

Stratifying the data solves two methodological problems. By disaggregating the data one can use all the data, increasing the robustness and reliability of the findings. Disaggregating the data also avoids the possibly improper deflation of maternal overlap that Pine et al. (2013) allude to, due to the fact that mothers, more often than children, produce many nouns only once or twice with a determiner. Our results with the Manchester data set replicate the results of Valian et al. (2009): observed overlap is a function of opportunity to observe overlap. Children are productive as early as we can measure.

3. Analysis 2 – Replication of Yang (2013)

3.1. Introduction

Analysis 2 uses the Manchester corpus to replicate the results of Yang (2013). We predict that children and their mothers will show an effect of the tokens/type ratio – the higher the ratio the more overlap both children and their mothers will show. We also predict that both children and mothers will have a determiner bias. We test the effect of those two variables by comparing children's and mothers' observed overlap with their expected overlap under the assumption of full productivity.

3.2. Method

Yang (2013) calculates the expected value of overlap for a sample of Det-N combinations, under the assumption that the determiner and nouns, as syntactic categories, combine productively and fully interchangeably. Let *S* be the size of the sample, the token frequency sum of all Det-N pairs, and *N* be the number of unique noun types. We refer the reader to that paper for details of the mathematical formulation but will highlight two conceptual points central to that study.

First, because noun frequencies can be approximated by Zipf's Law, the *r*th most frequent noun will have the expected frequency of S/rH_N, where H_N is the harmonic number $1+1/2+1/3+\ldots+1/N$. While this statistical property may not strictly hold in smaller corpora such as those in language acquisition studies (e.g., McCauley & Christiansen 2014, Pine et al. 2013), the accuracy of Zipf's Law has been extensively documented in numerous large-scale corpus studies (e.g., Baroni, 2009), and thus, we argue, serves as a better proxy for "true" word frequencies in any corpus than the actual empirical frequencies. It follows that only nouns with a rank lower than S/H_N, out of the N nouns, can be expected to have overlap, under the assumption that the determiner and nouns combine productively. Given the slow growth of H_N (approximately lnN), the overlap value of a sample will be strongly correlated with S/N, the token/type ratio of the nouns in the corpus. As shown in Valian et al. (2009) and Analysis 1, the overlap values of mothers and children do not differ in samples stratified by frequency.

Second, Yang (2013) takes into account the fact that different nouns favor one determiner over the other – there is a determiner *bias* (*B*). We quantify the bias toward one or another determiner for each speaker. For each noun in a corpus, we count and compare the number of times the noun appears with *a/an* and the number of times it occurs with *the*. The ratio of the larger number of determiners to the total number is the measure of determiner bias. Thus, if a noun has 10 uses of determiners, 5 of each type, the determiner bias is .50. If all 10 uses are a single determiner type, the value would be 1. Determiner bias values can range from .50 to 1.0. The closer the value is to 1.0, the larger the bias. The determiner bias is summed over all nouns for a speaker and then divided by the total number of times that the speaker used those nouns to provide a quantitative measure of determiner bias. For the purpose of overlap calculation, it does not matter which determiner is favored, only that one of them is. Yang (2013) documented this for the one-million-word Brown corpus: the 6,000 annotated singular nouns have an average bias value of 0.79. In the much larger Corpus of Contemporary American English (COCA), the overall determiner bias is 0.82, even if we only include the 19,480 singular common noun types used at least twice: only 1,380 (7%) of these have no bias—a balanced use of *the* and *a.* (We thank Martin Chodorow for the COCA analysis.)

The three variables, i.e., the number of noun tokens (*S*), the number of Noun types (*N*), and determiner bias (B), were used to create predicted overlap values for each child and each mother (see Yang, 2013 for the relevant equations). We compared the predicted overlap values with the actual observed overlap values for each child and mother.

In addition to computing overall overlap values from all files separately for child and mother, we computed overlap values by phase. Phase was defined as in Pine et al. (2013); see Analysis 3 for details. Here we focus on Phase 1, which represents the earliest longitudinal data starting at age 2;0, the stage during which children produced no fewer than 50 unique Det-N pairs. The children's Det-N combinations were extracted and the overlap values for the twelve corpora were computed and compared to the predicted values, again using the three variables of *S*, *N*, and *B*.

3.3. Results and discussion

3.3.1. *Overall results and discussion*

As shown in Table 1, we replicated the findings of Yang (2013). Children's observed overlap values were almost identical to their predicted values (.30 observed vs .28 predicted; paired *t*-test *t*(11), *p*=0.47), as were mothers' values (.34 observed vs .36 predicted, paired *t*-test *t*(11), *p*=0.25). Thus, both children's and mothers' language is consistent with the hypothesis that they have a rule that combines determiners and nouns fully productively (Yang, 2013).

The two groups were almost identical in determiner bias values (.82 for children and .81 for mothers, paired *t*-test *t*(11), *p*=0.26), confirming the result that, on average, children are not more lexically specific than their mothers in terms of determiner-noun combinations. This is not surprising. As the examples of *bathroom* and *bath* illustrate, the determiner bias is unlikely to be rooted in one's linguistic knowledge, but instead reflects the routines and circumstances of life.

	Child				Mother			
Name	Tokens/ Types	Det Bias	Predicted	Observed	Tokens/ Types	Det Bias	Predicted	Observed
Anne	4.06	0.83	**0.24**	**0.33**	9.18	0.83	**0.41**	**0.40**
Aran	5.22	0.79	**0.32**	**0.32**	8.62	0.80	**0.41**	**0.34**
Becky	4.36	0.84	**0.24**	**0.33**	6.42	0.82	**0.33**	**0.35**
Carl	11.60	0.76	**0.59**	**0.46**	8.57	0.78	**0.46**	**0.40**
Dom	3.09	0.84	**0.20**	**0.22**	8.55	0.79	**0.45**	**0.30**
Gail	3.01	0.87	**0.16**	**0.21**	4.77	0.83	**0.24**	**0.30**
Joel	3.51	0.87	**0.18**	**0.24**	5.04	0.84	**0.25**	**0.27**
John	5.78	0.78	**0.36**	**0.33**	5.67	0.79	**0.32**	**0.36**
Liz	4.78	0.85	**0.25**	**0.29**	5.74	0.83	**0.29**	**0.31**
Nicole	3.30	0.85	**0.19**	**0.26**	5.77	0.82	**0.30**	**0.31**
Ruth	4.13	0.79	**0.28**	**0.26**	7.71	0.80	**0.39**	**0.35**
Warren	6.81	0.76	**0.41**	**0.37**	7.91	0.79	**0.41**	**0.35**
Mean	4.97	0.82	**0.28**	**0.30**	7.00	0.81	**0.36**	**0.34**

Table 1
Tokens/types, determiner bias score, and predicted and observed raw overlap values for children and mothers in the Manchester corpus

Note. Tokens/types is the number of tokens per type. Det Bias is the overall extent to which a particular determiner was preferred. Predicted is the overlap predicted on the basis of the values of the tokens/type ratio and the determiner bias (Yang, 2013). Observed is the observed overlap.

Children had significantly lower overlap than their mothers (average 0.30 vs. 0.34, paired $t(11) = -2.81$, $p < 0.02$). While such discrepancies have been taken as evidence that children's language is not as productive than their mothers, they are directly accounted for by the two variables of interest—token/type ratio and determiner bias. Because the determiner bias values do not differ between the two groups, the token/type ratio predicts overlap values nearly perfectly. The correlation between the token/type ratios and (empirical) overlap values is .91 ($p < 0.001$) for children and 0.65 for mothers ($p<0.03$). The difference between children and mother's overlap values is the result of the difference between their token/type ratios (average 4.97 vs. 7.00, paired $t(11)=-3.04$, $p=0.01$)

3.3.2. Phase 1 results and discussion

Because the Manchester corpus spans roughly a year, an overlap measure across the entire span may disguise developmental trends and an early period when children are not productive. Thus, we examined children at the earliest phase; again, the definition of phase will be discussed in Analysis 3 below. The results are given in Table 2.

Name	Tokens/ Types	Det Bias	Predicted	Observed
Anne	1.73	0.97	0.04	0.04
Aran	3.16	0.84	0.24	0.19
Becky	2.08	0.94	0.09	0.12
Carl	2.48	0.90	0.15	0.16
Dom	1.42	0.91	0.08	0.13
Gail	1.60	0.97	0.04	0.05
Joel	1.80	0.95	0.06	0.07
John	2.52	0.87	0.17	0.19
Liz	2.07	0.96	0.06	0.09
Nicole	1.74	0.96	0.05	0.08
Ruth	2.32	0.81	0.21	0.13
Warren	3.36	0.87	0.23	0.18
Mean	**2.19**	**0.91**	**0.12**	**0.12**

Table 2
Empirical and expected overlap values for children in the first developmental phase
Note. See Table 1 for definitions of column headings.

As would be expected, given the relatively small number of noun types and tokens, the predicted and observed overlap values are considerably lower than those in Table 1, only .12. As was the case for the overall data, there is again no difference between the predicted and observed values of overlap (paired *t*-test *t*(11)=0.007, *p*=0.99). The token/type ratio (average 2.19) correlates highly and positively with overlap values (*r*(12)=0.80, *p* < 0.01), and determiner bias correlates highly and negatively with overlap values (*r*(12) = -0.81, *p* = 0.001, similar to the overall data.

The determiner bias values in Table 2, for phase 1, are considerably higher than those from the children's entire corpus in Table 1 (paired t-test, t(11) = 9.1383, p<0.001). This may give the misleading impression that very early child language is lexically specific. As we discuss in Analysis 5, the higher determiner bias is due to Pine et al.'s method of demarcating developmental phases. And, as Table 2 shows, children demonstrate full productivity once their determiner biases are factored into the overlap calculations (Yang 2013).

3.4. Conclusions – Analyses 1 and 2

Analyses 1 and 2 confirm the conclusions of Valian et al. (2009) on 21 children from the Valian corpus and from Yang (2013) on six other children in the CHILDES corpus. Once one controls for the opportunities of usage, child and mother overlap measures are indistinguishable. Even at the beginning of combinatorial speech, children appear to have a fully productive determiner-noun rule and to represent determiners and nouns abstractly.

4. Analysis 3 – Replication of Pine et al. (2013)

4.1. Introduction

Analyses 1 and 2 demonstrate the dependence of overlap on opportunity to observe overlap. Having replicated our prior results with the Manchester corpus, we now consider what happens when we compare the Manchester children and their mothers using Pine et al.'s (2013) method.

4.2. Method used by Pine et al. (2013) and replicated here

Pine et al. (2013) introduced several restrictions in an attempt to provide a controlled comparison between children and mothers. We briefly discuss these in turn, highlighting the design choices that differ from previous studies of overlap calculation.

First, Pine et al. (2013) used only nouns that were used with both *a* and *the* by the twelve mothers collectively. This restriction is intended to ensure that the nouns used for overlap calculation in fact show overlap in the input data, presumably providing children with the opportunity to learn the interchangeability of the two determiners. Note, however, that this restriction entails, among other things, that if a child used a noun with both determiners—and thus has overlap—but the child's mother used it with only one determiner, then the noun would be excluded, depriving the child an opportunity to display the productivity of their determiners.

Second, the children's data are subdivided into longitudinal phases that are intended to capture potential developmental changes in the grammar. Phases are defined as follows. Starting from the beginning of each child corpus, Det-N pairs are collected until there are 50 unique pairs: for example, *a ball* and *the ball* would each be considered a unique pair. In phase 1, then, this requirement means that the child sample would contain at least 25 and at most 50 unique nouns for overlap calculation. Collecting continues until the end of the recording session in which the criterion is reached. This forms the end of phase 1. Phase 2 starts at the endpoint of phase 1 and terminates at the age (transcript file) by which the child has produced 100 unique determiner-noun pairs from the beginning of the corpus. Phases 3, 4, and 5 are defined similarly, with 150, 200, and 250 unique determiner-noun pairs.

Third, for each noun used to compare children and their mothers, Pine et al. (2013) require that the noun be used by the child—but not necessarily by the mother—at least twice in that phase. The exclusion of singletons guarantees that the child has the opportunity to show overlap for each noun included in the sample. The restriction reduces the number of nouns used in overlap calculation. In phase 1, for instance, an average of only19 nouns were used for children's overlap calculation.

The fourth and most important feature of Pine et al.'s (2013) method is requiring that the child and mother have the same noun types and the same number of noun tokens when they are being compared. The larger sample is reduced to the size of the smaller sample in the following way (see Yang, 2011). For each noun, one draws uniformly, and with replacement, from the larger sample to match the number of occurrences in the smaller sample. In most cases, the mother's sample is considerably larger than the child's sample. That occurs partly because mothers generally talk more than children but also because the child's sample is from a specific developmental phase whereas the mother's sample is collected over her entire corpus.

An example illustrates the controlled sampling procedure. Say the noun *car* is used twice with a determiner {*the car*, *the car*} by the child in a particular phase. Say the mother, by contrast, has used Det+*car* 20 times over her entire corpus; she has 20 tokens of *the car* and *a car* in some mixture. In order to equalize the number of opportunities for overlap, Pine et al.'s (2013) method draws uniformly (with replacement) two tokens from the mother — thus matching the child sample size — from the set of 20 and calculates the overlap value in the sample. This process is completed for all nouns eligible for overlap calculation 100 times, and the average overlap value is recorded. The overlap value for *car* in the smaller (child) sample is calculated empirically: zero in the present example. If the number of times a noun is used is already equal in the child's and mother's sample, the overlap values are calculated empirically for both without the need for the controlled sampling process.

Pine et al. (2013)'s sampling method generates stochastic estimates of overlap values when comparing the corpora of a child and their mother. As noted above, when comparing the overlap values for each noun, the sampling method always matches the lower of the frequencies of that noun in the two corpora. Generally, the child's corpus will have a lower noun frequency: the mother's corpus is down-sampled to match the child. But on occasion the mother's corpus has the lower noun frequency, in which case the child's corpus is down-sampled to match the mother. Because of the vagaries of sampling, the overlap values for both the child and the mother from this sampling scheme will be somewhat different for each simulation. We report the overlap values averaged over 100 simulations; although we do not have access to the original code of Pine et al. (2013), their results would also differ somewhat from those reported in their paper if additional simulations were conducted.

4.2. *Results and discussion*

	Current study			Pine et al. 2013		
Phase	Token/Type	Child overlap	Mother overlap	Token/Type	Child overlap	Mother overlap
1	4.03 (2.58-6.60)	.31 (.15-.50)	.43 (.24-.58)	4.46 (2.96-6.96)	.34 (.13-.53)	.49 (.25-.63)
2	3.79 (2.53-5.97)	.35 (.18-.61)	.48 (.33-.62)	4.12 (2.70-8.14)	.34 (.13-.70)	.48 (.34-.65)
3	4.07 (2.67-7.70)	.34 (.22-.55)	.46 (.31-.56)	3.84 (2.42-6.77)	.31 (.06-.53)	.45 (.31-.60)
4	3.47 (2.77-5.44)	.33 (.12-.55)	.47 (.34-.61)	3.59 (2.68-5.03)	.30 (.14-.47)	.46 (.35-.57)
5	3.78 (2.57-5.51)	.28 (.06-.43)	.44 (.29-.54)	3.60 (2.14-5.74)	.28 (.06-.47)	.46 (.30-.62)

Table 3
Mean (and range) child and controlled mother overlap scores by phase

4.2.1. Overall comparison with Pine et al. (2013)

Table 3 compares the values of the current study with those of Pine et al. (2013) for each phase. As shown in Table 3, our values approximate those of Pine et al. for child values at each phase, once slightly lower (phase 1), three times slightly higher (phases 2-4), and once exactly the same (phase 5). Our values for the mothers also approximate Pine et al.'s, except for phase 1. In phase 1, we show an overlap value of .43 and Pine et al. report .49; that is the biggest discrepancy between our calculations and Pine et al.'s. At phase 2 we report the same value; for phases 3 and 4 we report very slightly higher values; for phase 5 we report a slightly lower value. In general, our data show a smaller range of values for all measures, including tokens per type, than do Pine et al.'s. As noted earlier, the results from Pine et al. would be subject to a certain level of variation during the nature of the sampling scheme.

Children displayed significantly lower overlap values than did their mothers for each phase via paired *t*-test: Phase 1, *t*(11) = -3.34, *p* = .007; phase 2 *t*(11) = -3.82, *p* = .003; phase 3 *t*(11) = -6.12, *p* < .001; phase 4 *t*(11) = -5.39, *p* < .001; phase 5 *t*(11) = -8.20, *p* < .001. We thus replicate one finding from Pine et al. (2013).

The question is why the difference between child and mother occurs, in seeming contradiction to the conclusions of Analyses 1 and 2. We hypothesize that the problem is that the sampling method is biased, a hypothesis we test in Analyses 4 and 5.

4.2.2. Lack of increase in overlap as development proceeds due to small sample size

Note a curious feature of the results in Table 3: overlap does not increase by phase. One would expect children to develop syntactic categories by age 3, even if, around age 2, they begin with low-level formulae. In discussing their results, Pine et al. (2013) first take their data at face value, stating (pp 354-355), "The implication is that children's use of the determiners *a/an* and *the* is significantly less flexible than that of their mothers, and that this difference in flexibility persists until relatively late in development (i.e., until most of the children have entered Brown's Stage III)." On that interpretation, children have little abstract knowledge of determiners over the entire one-year period.

To test that possibility, we treated the entire one-year period as a single phase, while obeying the other restrictions of Pine et al.'s (2013) procedures. The results are shown in Table 4. With the larger samples for the children, there is no difference between children and their mothers (*t*(11) = -.97, *ns*). Indeed, half the children have larger overlap values than their mothers.

Name	Tokens/ Types	Child Overlap	Mother Overlap
Anne	6.01	0.54	0.47
Aran	8.60	0.58	0.57
Becky	6.66	0.54	0.53
Carl	12.82	0.66	0.63
Dom	5.42	0.46	0.51
Gail	4.67	0.40	0.46
Joel	5.78	0.45	0.51
John	8.54	0.53	0.63
Liz	6.81	0.46	0.55
Nicole	5.34	0.52	0.48
Ruth	6.92	0.48	0.51
Warren	10.64	0.62	0.58

Mean	**7.35**	**0.52**	**0.54**

Table 4
Child and mother overlap scores over the entire longitudinal period

Pine et al. (2013, pp 355-356) carried out a somewhat similar analysis, with similar results, but interpreted the results differently. They divided their data into two phases instead of five. The first phase was identical to phase 1 as already described; the second phase consisted of all remaining files – but only using nouns that were present in the first phase. The overlap values show a significant increase, which the authors attribute to more "flexibility" over the one-year period as the original nouns became less lexically bound.

But their reasoning fails: the comparison they make is guaranteed to show an increase in overlap. If noun types are held constant and restricted to those that occurred in the first phase, the second phase, which is a much larger corpus, will guarantee a higher token/type ratio and hence a greater opportunity to detect overlap. It is not surprising, then, that children's overlap in Pine et al.'s (2013) new second phase is 0.50, very similar to our overlap values of 0.52 for the complete data set, of which the first phase constitutes a very small portion. Pine et al.'s claim of greater flexibility is simply an artifact of sample size. Statistically, the facts could not have been otherwise. Nouns that a child has used frequently are nouns that have more opportunity for overlap and are thus more likely to show overlap. Pine et al. have made our point.

5. Analysis 4 – Adult-adult comparisons

5.1. Introduction

Analysis 3 closely replicated the results of Pine et al. (2013), following their sampling methods. Children's overlap values were significantly lower than their mothers' during each of the five developmental phases (Table 3), but were the same as their mothers' when the whole corpus was considered (Table 4). We now address more fully the reason for that difference, exploring the possibility that Pine et al.'s sampling methods introduce biases that generally, although not always, underestimate the productivity of the smaller corpus in a comparison. Since the child's sample in each phase is compared with the mother's *entire* sample, the child will almost always have a reduced sample compared to the mother.

Analysis 4 uses Pine et al.'s (2013) sampling methods to compare three pairs of adult speakers, each of whom presumably has productive knowledge of the determiner category. In each pair, one speaker has a considerably larger corpus than the other. A valid measure of productivity should return similar overlap scores for the two speakers, since their linguistic knowledge is not in question. Indeed, any significant deviation from identity between the samples would signal a defective method.

5.2. *Method*

Our first test uses two corpora from CHILDES: the MacWhinney corpus (combining the transcripts of the two target children Mark and Ross) and the Bloom corpus from the target child Peter. In the MacWhinney corpus, we compare the parents, Mary and Brian, labeling Mary as the "child" and Brian as the "mother" because Mary's sample (n=25K) is much smaller than

Brian's (n=285K). That difference in output, a 10:1 ratio, is similar to the mother/child ratio in many of the Manchester child-mother pairs in the phase-based comparisons.

The Bloom corpus for Peter records interspersed conversations among Peter, Peter's mother, and two investigators Patsy and Lois. We make two comparisons: in each case the mother is labeled as the child and the two investigators, Patsy and Lois, are labeled as the mother. Patsy produced about 61,000 words and Lois produced about 47,000 words, while the mother produced only some 15,000 words.

We chose the MacWhinney and Bloom corpora because they had a large enough sample for the adult labeled "child" that we could divide them into five phases à la Pine et al. (2013), which we did for each dyad, using the same procedures described above. We used the "child" nouns per phase and the "mother's" aggregate nouns, sampling in the same way as described above, averaging the results from 100 simulations.

Table 5
Adult-adult comparisons, with "child" data from the adult with the smaller corpus, using Pine et al's (2013) sampling method

5.3. Results and discussion

5.3.1. Comparison between adult speakers

As Table 5 shows, the "children's" overlap scores are significantly lower than the "mothers'" across the three dyads (paired *t*-test $t(14)=-4.47$, $p < 0.001$). Each "child" phase represents a subset of the child's overall data, created using Pine et al.'s (2013) procedures, while the "mother's" data are drawn from the entire corpus. Each "child" phase, then, consist of significantly less data than the "mother's" phase. As Table 5 shows, the smaller corpus is almost always disfavored in overlap comparison but yields comparable and even slightly higher scores in 2 of the 15 phases. If we took these results at face value, we would conclude that these adult speakers with a smaller sample do not have the category determiner.

5.3.2. Comparison of mothers and themselves

To further test the validity of Pine et al.'s methods, we also compare the Manchester mothers against *themselves*, as shown in Table 6, in two ways.

First, we divide the mother's corpus into longitudinal "phases" and evaluate the overlap scores for each phase against the mothers' own data drawn from their entire corpus. Thus, each "child" phase, by design, is a strict subset of the mother's own data ("mother"). The overlap values from the five phases are averaged for each mother. The phase-based overlap values are significantly lower than the estimation from the full corpus (mean 0.33 vs. 0.42, paired *t*-test $t(11)=-5.08$, $p < 0.001$).

Phase	Mary "child"	Brian "mother"	Peter's mother "child"	Patsy "mother"	Peter's mother "child"	Lois "mother"
1	0.19	0.41	0.18	0.56	0.22	0.51
2	0.00	0.59	0.48	0.51	0.47	0.52
3	0.29	0.49	0.16	0.52	0.20	0.54
4	0.50	0.52	0.31	0.58	0.40	0.59
5	0.17	0.40	0.69	0.62	0.64	0.63
Mean	0.23	0.48	0.36	0.56	0.39	0.56

Second, we take a random 50% sample of the mother's data and evaluate those data against their full corpus. Here, despite the smaller corpus size, the 50% sample results in a higher overlap value than the full corpus, for every mother (mean 0.56 vs. 0.51, paired *t*-test $t(11)=9.67$, $p < 0.001$). In addition, for every "child", the overlap values are higher for the 50% sample than for the phase data. We investigate the reasons for these results in Analysis 5.

Finally, note that the data for the mother's full corpus differ in overlap value, depending on whether the comparison is with their data by phase or is with a random 50% of their data. That is because the mother's value itself changes depending on what nouns the "child" uses. Given Pine et al.'s (2013) sampling method, even the full corpus is not full. For every mother, the "full" corpus is larger for the 50% sample than for the phase sample.

Child	Mean of 5 Phases	Full Corpus	Random 50%	Full Corpus
Anne	0.26	0.35	0.59	0.52
Aran	0.42	0.38	0.53	0.51
Becky	0.38	0.44	0.57	0.48
Carl	0.32	0.44	0.59	0.54
Dom	0.45	0.46	0.54	0.50

Gail	0.27	0.38	0.52	0.46
Joel	0.28	0.36	0.51	0.46
John	0.34	0.50	0.60	0.56
Liz	0.28	0.34	0.48	0.45
Nicole	0.33	0.44	0.56	0.51
Ruth	0.25	0.42	0.59	0.53
Warren	0.39	0.50	0.63	0.57
Mean	**0.33**	**0.42**	**0.56**	**0.51**

Table 6
Manchester mothers' overlap values, calculated with different subsets of their data and their full corpus

5.3.3. Conclusion

The cross-adult comparisons in Table 5 and the within-adult comparisons in Table 6 show that something is seriously awry with the sampling method of Pine et al. (2013). A procedure designed to create well-matched samples has assessed two fully competent adult native speakers (Mary and Peter’s mother (from two different samples) as less knowledgeable and productive than three other fully competent adult native speakers (Brian, Lois, and Patsy). Even more telling, the Manchester mothers, whose use of determiners is productive, as confirmed statistically in Analysis 2, are assessed to be less productive on “longitudinally” delimited phases, as if their grammars were becoming more "flexible" in the same way that two-year-olds' grammars were considered to be by Pine et al. (2013). These results constitute a *reduction ad absurdum* of Pine et al.'s sampling method.

The cross- and within-adult results are largely but not entirely attributable to sample size differences. In a few cross-adult cases and in one within-adult case, the smaller sample has a higher overlap value than the larger sample. This is also true for the child-mother comparisons in Pine et al.’s (2013) original data and in our replications in Analysis 3. As the range of the overlap values show in Table 3, even when the child-by-phase corpus is smaller than the mother’s full corpus, children show higher overlap values than mothers in several comparisons. In the within-mother comparison with 50% of the mother's corpus, the pattern is fully reversed, and the smaller sample has higher overlap values for every mother, even though the 50% sample is uniformly higher than the data by phase.

It seems clear, then, Pine et al.’s (2013) methods produce inconsistent and even paradoxical results. In Analysis 5, we provide a formal analysis in order to understand the source of the inconsistencies.

6. Analysis 5 – A formal analysis of Pine et al. (2013)

6.1. Introduction

Analysis 5 explores the key feature of Pine et al.’s (2013) method, the reduction of a large corpus to match a small corpus, via sampling, while holding the nouns under comparison constant. For convenience, we will assume that the large corpus is from the mother and the small corpus is from the child. That is almost always the case for the nouns under comparison in the current empirical study.

We put forward two claims. *First*, Pine et al.'s (2013) method of sampling down from the larger corpus to match the size of the small corpus will almost always result in an underestimation of overlap in the larger (i.e., mother's) corpus. *Second*, the child's overlap, already reduced because their smaller corpus size reduces the opportunity to detect overlap, is further underestimated by the way in which Pine et al. delineate developmental phases. In general, the underestimation of the mother is overwhelmed by the underestimation of the child, albeit inconsistently. After establishing the two claims empirically, we undertake a formal analysis to show that once the two biases are taken into consideration, we can account for the results from Analyses 3 and 4, including the paradoxical ones. Throughout we take as a given what we demonstrated in Analyses 1 and 2 – that children's knowledge of determiners is abstract and fully productive. We also assume that adults' knowledge is similarly abstract and productive.

6.2. Matched sampling: Biases against adults

We start with the following point, which has been made before:

(1) Sampling from a corpus almost always reduces the overlap value of that corpus.

Yang (2011, 2013) tested an item-based learning model much like Pine et al.'s (2013), based on a well-known usage-based learning proposal (Tomasello, 2000). The model samples from Det-N pairs in mothers' speech in order to match children's corpus size. The model consistently produces overlap values lower than children's empirical values, shown in our Analysis 2. The reason is as follows. First, if a word in a corpus has no overlap, then no samples from it can ever produce overlap. Second, if a word in the corpus does have overlap, i.e., it is used with both determiners, there is always a non-zero probability that only one of the two determiners is drawn exclusively from the sample, no matter how large the sample is.

Consider a concrete example. Suppose a noun appears in one corpus 10 times, 7 times with *the* and 3 times with *a*. The overlap value is 1. Suppose the same noun appears in another, smaller, corpus only 4 times. Using Pine et al.'s (2013) procedure, we create a matched sample by drawing from the larger corpus 4 times with replacement. With a probability of 7/10, a Det-N combination with *the* will be drawn, and with a probability of 3/10, a Det-N combination with *a* will be drawn. Thus, the probability of drawing a sample that contains overlap is $1- (7/10)^4 - (3/10)^4 = 0.752$, a significant reduction from 1 in the original sample.

The probability that the actual overlap value will be reduced by sampling is greater than zero even if the number of times the sample is drawn from is greater than the original corpus size. That is, if we draw 20 times from the 10-item sample (with a 7-3 split), the probability of yielding overlap is$1-(7/10)^{20}-(3/10)^{20}$, which is almost 1, but still less than 1. Thus, Pine et al.'s method will, at least slightly, underestimate the overlap value of every noun from the mother's sample, irrespective of the size of the child sample.

6.3. Phase delimitation: Biases against children

Recall that Analysis 2 introduces the determiner bias *B,* which characterizes the (im)balance of Det-N combinations according to which a noun prefer one or the other determiners: A higher bias value makes overlap less likely. Analysis 2 shows that, in general, the average bias value is approximately 0.8, and similar for both children and mothers across their entire corpora.

An important point, when considering bias, is that although the *average* bias value is roughly the same across speakers, it is not the same for *every* noun for *every* speaker for *every* time period. For example, parents probably use *the baby* more often than *a baby*, since they are referring to their (single) baby. In contrast, *a baby* would make up a larger proportion in corpora collected in a busy maternity ward where there are many babies. For the noun *baby*, then, a parent would be evaluated as less productive than a medical specialist — through no fault of their grammar. Relatedly, the bias value for any specific noun can vary from moment to moment for the same speaker. A doctor talking about their own baby at home will refer to *the baby*, but in the hospital will refer both to *a baby* and *the baby*. The same doctor would be assessed as less productive if measured at home instead of at the hospital. We are interested in using overlap to ascertain whether a speaker represents a category, not in tracking how that overlap varies from moment to moment.

The fluctuation of the bias value, which affects overlap, is exacerbated by Pine et al.'s (2013) delimitation of developmental phases, which correspond to longitudinal recording sessions in the Manchester corpus. Within each session, nouns, especially those used frequently enough to be included for overlap calculation, can and do significantly deviate from their average bias across the full corpus, producing markedly higher bias values in some sessions.

To take a concrete example of the "clumpiness" of Det-N usage, consider the most frequently used noun by the Manchester children – *car*. Carl produced *car* 389 times, with a bias value of 0.67 across his full corpus. Of Carl's 33 sessions, four sessions alone account for 117 tokens. The densest session took place at 2;8, which has a strong transportation theme: 50 uses of *car* were recorded, of which 45 were *the car* — which Carl was insistent on getting into the garage — yielding a bias value of 0.9, considerably higher than his corpus average.

Such clumpy distributions are typical. Table 7 shows the bias value for all twelve children's most frequent noun across the full corpus and also for each of the five developmental phases. The clumpiness of determiner-noun usage significantly inflates the bias value for each developmental phase. Of the 45 phases in which the noun qualifies for overlap calculation, 38 have a higher bias than the corpus average. Overall, the bias value is inflated by 0.12 under Pine et al.'s (2013) division into phases, and that results in depressing children's overlap measures, a point to which we return later.

Child	Noun	Frequency	Phase 1	Phase 2	Phase 3	Phase 4	Phase 5	Overall
Anne	bit	66	**1.00**	NA	**1.00**	NA	**1.00**	0.94
Aran	man	80	**0.88**	**1.00**	0.50	**0.80**	NA	0.79
Becky	baby	55	**1.00**	0.75	**1.00**	**1.00**	NA	0.76
Carl	car	389	**1.00**	**1.00**	**0.89**	0.65	**0.81**	0.67
Dom	train	47	**0.80**	**0.75**	**1.00**	**0.70**	**0.80**	0.51
Gail	bit	58	**1.00**	NA	0.88	**1.00**	**1.00**	0.90
Joel	bit	48	NA	NA	0.80	**1.00**	**1.00**	0.92
John	box	93	NA	NA	**0.93**	**1.00**	**1.00**	0.88
Liz	bit	56	NA	**1.00**	NA	**1.00**	**1.00**	0.89
Nicole	minute	35	NA	NA	NA	NA	**1.00**	0.97
Ruth	baby	141	**0.64**	**0.67**	**0.72**	**0.62**	0.50	0.57
Warren	car	206	**0.81**	**0.71**	**1.00**	**1.00**	0.57	0.67

Table 7
Bias values of children's most frequent noun, with values higher than the corpus average bolded

Taken together, Pine et al.'s (2013) method is a contest between two inaccurate estimators, one that reduces mothers' overlap and one that reduces children's overlap. The outcome of the contest depends on the extent of the inaccuracies. In general, the child's corpus is disfavored more heavily than the mother's because the child's much smaller sample size per phase produces more opportunity for clumps. As the child corpus increases, through the inclusion of more recording sessions, the clumpiness, and thus the bias against children, is reduced. The bias against mothers remains: it is possible for a small corpus to match or even overtake the larger corpus in overlap values. As Table 4 shows, when children's full longitudinal corpus is pitted against the mothers, the overlap values are no longer different, even though children still produced considerably less data than the mothers.

6.4. Formal analysis of biases in Pine et al. (2013)

6.4.1. Comparisons of small and large corpora

Table 8 reports several key statistics of the samples under comparison using Pine et al.'s (2013) sampling method. The four comparisons are: children's phase one data against their mothers' full corpus, children's full corpus data against their mothers' full data (Table 4), mothers' phase one data against their own full data (Table 6), and a random 50% sample of the mothers' data against their own full data (Table 6). Recall the apparently paradoxical findings from these comparisons. Children's phase-delimited corpora—we use phase one here as an example—show lower overlap than mothers' full corpus (Table 3, i.e., Pine et al.'s main result), but children's full corpus shows comparable overlap to mothers' full corpus (Table 4). Mothers' phased-delimited corpora—again, we use phase one here as an example—show lower overlap than their own full corpus (Table 6) but a random 50% sample of the mother's corpus shows higher overlap than their own full corpus (Table 6).

In each case, there is a small corpus against a large corpus. Not all nouns in these corpora are included for overlap comparison as Pine et al.'s (2013) methods, described in Analysis 3, place various restrictions such that only a subset of nouns in these corpora are used. Table 8 gives the average determiner bias and frequencies of the nouns included for comparison in the small corpus and those values in the large corpus. The frequency measures report how often the usable Ns occurred in the smaller sample and the how often they occurred (before being sampled down) in the larger sample.

Note that the bias and frequency values in Table 8 are different from those in Table 2 where, following Yang (2013), all nouns are included in overlap comparison. The average noun frequency for both child phase one corpora (4.09) and child full corpora (8.29) in Table 8 are also slightly higher than those in Tables 3 (4.03) and 4 (7.35). In Tables 3 and 4, the frequency of a noun used in overlap comparisons is the smaller value of its frequency in the child sample and its frequency in the mother's sample, following the sample size matching design of Pine et al. (2013). Table 8 reports the noun frequency in the child sample: it is generally lower than that in the mother's sample but occasionally higher. Note also that average noun frequency in the large corpus (e.g., mother full corpus) decreases when the small corpus is larger, in the order of child phase one, mother phase one, and mother random 50% corpus. As the small corpus gets larger, more noun types will be included, and their average frequency will necessarily drop.

As expected, the average frequencies of the nouns in the small corpus are much lower than those in the larger corpus; the latter will be sampled down to match the former under Pine et al.'s (2013) methods. As expected, given our earlier discussion, the mean determiner biases are significantly higher in the small corpus than in the larger ones in all four sets of comparisons: child phase one against mother full corpus (paired *t*-test: *t*(11)=5.84, *p*<0.001), child full corpus against mother full corpus (paired *t*-test: *t*(11)=6.29, *p*<0.001), mother phase one against mother full corpus (paired *t*-test: *t*(11)=9.38, *p*<0.001), mother random 50% against mother full corpus (paired *t*-test: *t*(11)=11, *p*<0.001). (For simplicity, we report the mean of the individual dyad results, which can be found on the Github repository.) The difference is much reduced, though still significant, when we compare the children's full corpus and mothers' random 50% corpus with the mother's full corpus.

	Small corpus		Large corpus	
Corpora (Small ~ Large)	Avg bias	Avg N frequency	Avg bias	Avg N frequency
Child phase one ~ Mother full	0.88	4.09	0.78	41.77
Mother phase one ~ Mother full	0.87	3.51	0.78	38.06
Child full ~ Mother full	0.83	8.29	0.78	19.34
Mother random 50% ~ Mother full	0.82	7.58	0.80	14.75

Table 8
Comparison of determiner bias and frequency of the nouns in the small and large corpora that are used for overlap comparison with Pine et al. (2013)'s method

6.4.2. Formal analysis of Pine et al.'s (2013) sampling method

We now provide a formal analysis to show that when a typical noun under comparison has a determiner bias and frequency like the values presented in Table 8, all results that follow Pine et al.'s (2013) methods, including the paradoxical ones summarized as the beginning of this section, can be qualitatively accounted for under the hypothesis that both children and mothers have fully productive knowledge of determiners. In other words, the results are paradoxical only because of the flaws introduced by Pine et al., and in fact provide evidence for full productivity.

Consider the expected overlap value of a noun under the conditions imposed by Pine et al.'s (2013) sampling method. Suppose its frequency is *f* in the small corpus, and *F* in the large corpus, and its determiner bias is *b* in the small corpus and *B* in the large corpus. As shown in Table 8, we assume *b>B* and *f<F*: the noun is more biased and less frequent in the small corpus. Let *o(b, f)* be the expected overlap value obtained from the smaller corpus, and *O(B, F)* be the expected overlap value from the large corpus.

The value *o(b, f)* can be calculated directly:

$$o(b,f) = 1 - b^f - (1-b)^f$$

which is simply one minus the probability of the noun being paired with one of the determiners exclusively on all f trials, again under the assumption that Det-N combination is fully productive and thus statistically independent.

The calculation of *O(B, F)* is slightly more complicated because a sampling scheme à la Pine et al. (2013) is introduced to match f from F. Specifically,

$$O(B,F) = \sum_{i=0}^{F} \Pr(i;F,B)\left[1 - (\frac{i}{F})^f - (1-\frac{i}{F})^f\right] \text{ where } \Pr(i;F,B) = \binom{F}{i} B^i (1-B)^{F-i}$$

Under full productivity, the term *Pr(i; F, B)* specifies the binomial probability of the noun being used in the *(i, F-i)* mixture where one determiner is used i times and the other (F-i) times. The term in the square bracket above is the probability of drawing from that mixture f times (with replacement), matching the smaller corpus frequency *f*, while resulting in overlap (i.e., not drawing one of the two determiners exclusively on all f trials). It is clear that *O(B, F)* depends on *f* from the small corpus as well as *F* and *B* in the large corpus.

By using the values of *f, b, F*, and *B* from the four sets of small vs. large corpus comparisons in Table 8, we calculate the values of *o(b, f)* and *O(B, F)* seen in Table 9. For simplicity, we present only the average of the expected overlap values for the four comparisons, each of which contains 12 dyads; the full data can be found on the Github repository.

Note that the expected overlap values in Table 9 are those of a (single) noun with the average bias value and average frequency in the two corpora under comparison: the results thus can be thought of as those of a typical or average noun. As such, they are different from the results in analyses (e.g., Table 3, 4, and 6) that use Pine et al.'s methods, which are the average overlap values of all nouns.

Table 9 very clearly illustrates statement (1), that sampling from a corpus almost always reduces the overlap value of that corpus. For instance, as shown in Table 8, when compared against the child phase one corpus, the average noun frequency in the mother's full corpus is 41.77; when compared against the child full corpus, in contrast, the average noun frequency in the mother 's full corpus is 19.34. With a bias value of 0.78, such a noun is virtually guaranteed to have an overlap. But because its frequency in the child corpus is much lower (i.e., 4.09 in phase one and 8.29 in full; Table 8), its overlap value in the frequency matched sample is significantly reduced, to 0.60 and 0.77 respectively. Furthermore, the frequency reduction from the mother's corpus to the child's corpus is much more dramatic for the phase one data (41.77 to 4.09) than for the full data (19.34 to 8.29): this results in a more substantial overlap reduction (from essentially 1 to 0.60) in the former comparison than the latter (from essentially 1 to 0.77), even though the latter starts with a lower frequency than the former (41.77 vs. 19.34). Subtleties of this type are difficult to uncover without formal mathematical analysis.

Corpora (Small ~ Large)	*o(b, f)*	*O(B, F)*
Child phase one ~ Mother full	0.40	0.60
Mother phase one ~ Mother full	0.38	0.55
Child full ~ Mother full	0.74	0.77
Mother random 50% ~ Mother full	0.76	0.73

Table 9
Expected overlap values of nouns in the small and large corpus, *o(b, f)* and *O(B, F)* with average determiner bias and frequency taken from Table 8.

Four observations are noteworthy. *First*, the child's small phase one-delimited corpus has a smaller overlap value than the size-matched sample from the mother's larger corpus (row 2, Table 9; paired *t*-test, *t*(11)= -5.7, *p*<0.001), replicating Analysis 3 (Table 3), Pine et al.'s (2013) main result. *Second*, the same holds for the comparison of the mother's phase one corpus against their own full corpus (row 3, Table 9; paired *t*-test, *t*(11)=-7.4, *p*<0.001), replicating Analysis 4, and demonstrating the flaw in Pine et al.'s sampling method, which produces the paradox in which some adults are assessed as less productive than other adults (e.g., Mary vs. Brian MacWhinney, Table 5) and a mother is analyzed as less productive than herself (part of Table 6).

Third, we explain why the large corpus overlap advantage diminishes as its size advantage over the small corpus is reduced. Even though the small corpus is still considerably smaller, the disparity between the determiner bias in the two samples is reduced sufficiently that the smaller corpus yields comparable or even higher overlap values than the larger corpus. The expected overlap values of nouns in the child full corpus are not significantly different from those in the mother's full corpus (paired *t*-test: *t*(11)=-2.03, *p*=0.06), replicating the finding in Table 4. Fourth, and in contrast, the expected overlap values of nouns in the random 50% sample of the mother's corpus are in fact significantly higher than those in their own full corpus (paired *t*-test: *t*(11)=12.03, *p*<0.001), replicating the finding in (part of) Table 6.

6.5. Summary

To summarize, *all* of the statistical results we report that use Pine et al.'s (2013) method are fully accounted for by the formal analysis developed here. Their method introduces two major biases which lead to apparently paradoxical results. We have shown that those paradoxical results are in fact *predicted*—under the hypothesis that both children and mothers have fully productive knowledge of determiners: the calculations of *o(b, f)* and *O(B, F)* demonstrate that.

7. General discussion

7.1. Replication of Valian et al. (2009) and Yang (2013): children are productive

Our data make several facts clear: children productively use the determiners *a* and *the* with their nouns very early in acquisition – essentially as soon as children begin producing a reasonable sample of multiword combinations. In Analyses 1 and 2, we successfully replicated our own previous findings (Valian et al., 2009; Yang, 2013) and found that the extent to which children used both *a* and *the* with their nouns was a direct function of how many tokens children produced for their nouns. In particular, the more times a child used a noun with a determiner, the higher the overlap. Exactly the same pattern holds for mothers.

7.2. Replication of Pine et al. (2013): even with fully mature speakers – a flawed method

Using their restrictions and sampling methods, we replicated Pine et al.'s (2013) findings (Analysis 3). At each of five successive phases, children showed less overlap than their mothers. We demonstrated, however, that even fully mature speakers, whose knowledge of the category determiner is not in doubt, will look exactly like children if we apply Pine et al.’s sampling methods. Analysis 4 was a *reductio ad absurdum*, showing that when adults with small samples are compared with adults with large samples, and divided into "phases", they show less overlap. Even a mother compared with herself looks unproductive if only 10% of her data are used. The sampling methods Pine et al. recommended are fatally flawed.

7.3. Mathematical analysis of Pine et al.'s (2013) sampling method

In Analysis 5, we identify two main sources of biases in the methods and demonstrate that once Pine et al.’s methods are formalized, the full range of statistical results is derivable from the hypothesis that children freely combine determiners and nouns. One source of bias concerns the larger sample: whenever it is reduced, the actual overlap values will also be reduced, thus underestimating the overlap in the larger sample. The other main source of bias concerns the smaller sample: small samples, particularly those that are arrived at through dividing the data into small longitudinal groups, are likely, for reasons related to the choice of conversation, to show "clumpiness" of Det-N pairings, with a concomitant reduction in overlap. The smaller sample is also, simply in virtue of being very small, less likely to show the overlap that a large sample will show.

7.4. Comprehension studies

Comprehension studies provide converging evidence with production studies. The conclusions we draw from production data are consistent with findings from comprehension studies. To take one example, French 14-month-olds can use determiners to categorize nonsense words as nouns or verbs, suggesting a determiner equivalence class (Shi & Melançon, 2010). To take another example, English-speaking 18-month-olds are more successful in identifying a referent if *the*, rather than other small words, is used before a noun (Kedar et al., 2006; Kedar et al., 2017). In addition, Valian (1986) found that the six two-year-olds she observed never sequenced determiners, but did sequence determiners and adjectives, and occasionally sequenced adjectives. Children did not simply have a "pre-noun" category in which determiners and adjectives were lumped together, but had a separate determiner category.

7.5. Methodological implications

We suggest that our comprehensive analysis of Pine et al. (2013) is instructive for the future quantitative study of child language. Recall that our earlier work was in part a reaction to methods in previous research (e.g., Pine & Lieven 1997; Pine & Martindale, 1996; Tomasello, 1992) that proposed overlap and other distributional measures of productivity. Those measures were taken at face value to draw conclusions about the underlying grammar. For example, if children's overlap values are lower than their mothers', they lack determiner knowledge.

When, however, a proper baseline is established, either by stratifying the data and comparing children with mothers (Valian et al., 2009), or by comparing children's observed values with a mathematically derived benchmark of productivity (Yang, 2011, 2013), the (low) overlap values in fact support the opposite conclusion of full productivity.

Pine et al. (2013) introduced a new and complex sampling scheme for overlap assessment in order to avoid problems they saw with previous analyses. Informally justifying their methods, they drew conclusions without testing the validity of their scheme on other samples of clearly competent speakers. While the analytical approach taken in our Analysis 5 may be beyond the scope of their study, the adult-to-adult comparisons in our Analysis 4 are purely empirical and could have been deployed as a sanity check, since the status of the adult’s grammar is not in doubt. The biases in their methods would have been revealed and their conclusion that children lack the category determiner would have been shown to be unfounded.

As computational and statistical tools for linguistic data analysis become more complex, it is even more important to develop a proper understanding of them before accepting the conclusions they produce. An analytical approach such as Analysis 5, which can yield a complete understanding of the method, is always preferable. Failing that, sanity checks against empirical cases with known outcomes, such as our Analysis 4, should be pursued exhaustively to detect unforeseen design flaws.

7.6. Final words

One reason for paying so much attention to children's determiner knowledge is, as we remarked in the introduction, the implications for the rest of the child's grammar. Production and comprehension tests converge in showing that, at least by age 2, children acquiring English represent determiners as a syntactic class and treat its members equivalently. If there is a period when determiners are *not* abstract, it is before children are combining words. We can thus constrain acquisition theories by requiring that any model, whether nativist or empiricist, yield abstract knowledge of the category determiner no later than age 2.

Note further that the grammatical knowledge of determiners is language specific at least in part: the Det-N rule is specific to English, and there are languages that do not have the counterpart of *a* and *the*. The acquisition of these rules thus must make use of language-specific data; see Yang (2016) for a mathematical model in which such rules are learned on the basis of child-directed input.

Determiners are the thin edge of the wedge. Children's early abstract knowledge of determiners suggests that other syntactic categories may also be represented abstractly and formally in children's early grammars: the empirical and methodological advances made with determiners should prove fruitful in future investigations.

Acknowledgments: We acknowledge with thanks Edward Wadsworth's contribution to a preliminary study of the current project, and the feedback from audiences at the 2014 Boston University Conference on Language Development and the 2015 Linguistic Society of America annual meeting. We benefited from comments from and discussions with Martin Chodorow and Kyle Gorman.

Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

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.

Ambridge, B. (2017). Syntactic categories in child language acquisition: Innate, induced, or illusory? In H. Cohen & C. Lefebvre (Eds.), *Handbook of categorization in cognitive science,* 2nd Ed., pp. 567-580. Amsterdam, NL: Elsevier.

Baroni, M. (2009). Distributions in text. In A. Lüdeling & M. Kytö (Eds.) *Corpus linguistics: An international handbook* (pp. 803-821). Berlin: Mouton de Gruyter.

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 and Biobehavioral Reviews*, *81*(Part B),150-157.

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.

Ibbotson, P., & Tomasello, M. (2009). Prototype constructions in early language acquisition. *Language and Cognition*, *1*(1), 59-85.

Kedar, Y., Casasola, M., & Lust, B. (2006). Getting there faster: 18-and 24-month-old infants' use of function words to determine reference. *Child Development*, *77*(2), 325-338.

Kedar, Y., Casasola, M., Lust, B., & Parmet, Y. (2017). Little words, big impact: Determiners begin to bootstrap reference by 12 months. *Language Learning and Development*, *13*(3), 317-334.

Joo, K. J., & Yoo, I. W. (2018). Early articles in English child language: Impostors or overt instantiations of a nominal functional category?. *Language Acquisition*, *25*(2), 224-230.

McCauley, S. M., & Christiansen, M. H. (2014, January). Reappraising lexical specificity in children's early syntactic combinations. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, *36*(36), 1000-1005.

Meylan, S., Frank, M., & Levy, R. (2013, January). Modeling the development of determiner productivity in children's early speech. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *35*(35), 3031-3037.

Meylan, S. C., Frank, M. C., Roy, B. C., & Levy, R. (2017). The emergence of an abstract grammatical category in children's early speech. *Psychological Science*, *28*(2), 181-192.

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.

Pine, J. M., & Lieven, E. V. (1997). Slot and frame patterns and the development of the determiner category. *Applied Psycholinguistics*, *18*(2), 123-138.

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.

Shi, R., & Melançon, A. (2010). Syntactic categorization in French-learning infants. *Infancy*, *15*(5), 517-533.

Silvey, C. & Christodoulopoulos, C. (2016). Children's production of determiners as a test case for innate syntactic categories. In S. Roberts, C. Cuskley, L. McCrohon, L. Barceló-Coblijn, O. Fehér, & T. Verhoef (Eds.), *The evolution of language: Proceedings of the 11th international conference* (EVOLANGX11).

Szagun, G., & Schramm, S. A. (2019). Lexically driven or early structure building? Constructing an early grammar in German child language. *First Language*, *39*(1), 61-79.

Tomasello, M. (1992). *First verbs: A case study of early grammatical development*. Cambridge, MA: Harvard University Press.

Tomasello, M. (2000). Do young children have adult syntactic competence? *Cognition*, *74*(3),209-253.

Valian, V. (1986). Syntactic categories in the speech of young children. *Developmental Psychology*, *22*, 562-579.

Valian, V. (2009). Abstract linguistic representations and innateness: The development of determiners. In W. Lewis, S. Karimi, H. Harley, & S. Farrar (Eds.), *Language: Theory and practice: Papers in honor of D. Terence Langendoen* (pp 189-206). Amsterdam: John Benjamins.

Valian, V. (2013). Determiners: An empirical argument for innateness. In M. Sanz, I. Laka, & M. Tanenhaus. (Eds.). *Language down the garden path: The cognitive and biological basis for linguistic structure* (Chapter 14, pp 272-279). New York: Oxford University Press.

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.

Yang, C. (2011). A statistical test for grammar. In *Proceedings of the 2nd workshop on Cognitive Modeling and Computational Linguistics* (pp. 30-38). Association for Computational Linguistics.

Yang, C. (2013). Ontogeny and phylogeny of language. *Proceedings of the National Academy of Sciences*, *110*, 6324-6327.

Yang, C. (2016). *The price of linguistic productivity: How children learn to break rules of language.* Cambridge, MA: MIT Press.

Zipf, G. K. (1949). *Human behavior and the principle of least effort: An introduction to human ecology*. Cambridge, MA: Addison-Wesley.