Do young children have adult-like syntactic categories? Zipf's law and the case of the determiner

Julian M. Pine [a,*], Daniel Freudenthal [a], Grzegorz Krajewski [b], Fernand Gobet [a]

[a] Institute of Psychology, Health and Society, University of Liverpool, UK
[b] Faculty of Psychology, University of Warsaw, Poland

ARTICLE INFO

ABSTRACT

Generativist models of grammatical development assume that children have adult-like grammatical categories from the earliest observable stages, whereas constructivist models assume that children's early categories are more limited in scope. In the present paper, we test these assumptions with respect to one particular syntactic category, the determiner. This is done by comparing controlled measures of overlap in the set of nouns with which children and their caregivers use different instances of the determiner category in their spontaneous speech. In a series of studies, we show, first, that it is important to control for both sample size and vocabulary range when comparing child and adult overlap measures; second, that, once the appropriate controls have been applied, there is significantly less overlap in the nouns with which young children use the determiners *a/an* and *the* in their speech than in the nouns with which their caregivers use these same determiners; and, third, that the level of (controlled) overlap in the nouns that the children use with the determiners *a/an* and *the* increases significantly over the course of development. The implication is that children do not have an adult-like determiner category during the earliest observable stages, and that their knowledge of the determiner category only gradually approximates that of adults as a function of their linguistic experience.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

A central issue in the field of first language acquisition is the nature of children's early grammatical categories. On the one hand, many generativist researchers argue that children have adult-like syntactic categories from the earliest observable stages (e.g. Pinker, 1984; Valian, 1986, 1991; Bloom, 1990; Wexler, 1994, 1998). On the other, many constructivist researchers argue that children's early grammatical categories are more limited in scope, and only gradually approximate those of adults as a function of some form of data-driven learning (Bowerman, 1973; Braine, 1976, 1992; Maratsos, 1990; Maratsos & Chalkley,

1980; Schlesinger, 1982, 1988; Pine, Lieven, & Rowland, 1998; Tomasello, 1992, 2000). These positions are obviously very different in principle. However, in practice, they are more difficult to distinguish than they might at first appear. In the present paper, we seek to differentiate and test them with respect to one particular syntactic category, the determiner, which has been the focus of some debate in the literature (Pine & Lieven, 1997; Pine & Martindale, 1996; Valian, 1986; Valian, Solt, & Stewart, 2009; Yang, 2010). We do this by comparing controlled measures of overlap in the set of nouns with which children and their caregivers use different instances of the determiner category in the Manchester Corpus (Theakston, Lieven, Pine, & Rowland, 2001). In a first study, we show why it is important to control for differences in vocabulary range as well as sample size when comparing child and adult overlap measures. In a second study, we show that, once the appropriate controls have been applied, there is signif-

* Corresponding author. Address: School of Psychology, University of Liverpool, Eleanor Rathbone Building, Bedford Street South, Liverpool L69 7ZA, UK. Tel.: +44 151 794 1113.

E-mail address: Julian.Pine@Liverpool.ac.uk (J.M. Pine).

icantly less overlap in the nouns with which children use the determiners *a/an* and *the* in their speech than in the nouns with which their caregivers use these determiners. In a third study, we show that the level of (controlled) overlap in the nouns that the children use with the determiners *a/an* and *the* increases significantly across two developmental phases. The implication is that children do not have an adult-like determiner category during the earliest observable stages, and that their knowledge of the determiner category only gradually approximates that of adults as a function of their linguistic experience.

1.1. Generativist and constructivist models of syntactic development

A central assumption of generativist models of syntactic development is that children have adult-like grammatical categories from the earliest observable stages. For example, according to Pinker's (1984) semantic bootstrapping model, the child's early categories are the result of a process whereby innate knowledge of syntactic structure, together with innate linking rules, is used to classify words in the input language into traditional word class categories such as determiner, adjective and noun. These categories may initially include a smaller number of items than those of adults, because of the child's more limited vocabulary. However, the categories themselves do not change over the course of development, since they are defined from the beginning in terms of their place within an adult-like grammatical system. For example, for Pinker the category of determiner is defined in terms of the phrase structure rules in which it participates (roughly as the set of words that have been parsed as determiners using the rule: NP → det (Adj) Noun).

This view of the nature of children's early syntactic categories has two important implications for children's early multi-word speech. The first is that the way in which children combine the words in their vocabularies should be essentially adult-like from the beginning. Thus, although children may omit words from obligatory contexts for performance reasons (e.g. Bloom, 1990), those words that they do produce should pattern correctly with respect to each other in the child's speech, and should combine as productively in the child's speech as they do in the adult's speech. That is to say, once the child has categorised a set of words as determiners and a set of words as nouns, she should be able to combine the words in these categories as productively as adults using the rule: NP → det (Adj) Noun (see Braine, 1976, 1988, for a similar argument with respect to the categories of NPsubject and VP).

The second implication is that there should be no change in the productivity of children's use of the words in their early vocabularies. Thus, although children's language will inevitably increase in productivity as vocabulary size increases, there should be no change in the productivity with which children combine the words that they are using in the early stages, since their knowledge of the syntactic properties of these words does not change over the course of development. That is to say, since the child's determiner and noun categories are defined in terms of the rule: NP → det (Adj) N from the beginning, there is no reason why the child's ability to combine the

determiners and nouns in her early vocabulary should increase with development.

These predictions can be directly contrasted with those of constructivist models of grammatical development. The central assumption of constructivist models is that children's early grammatical categories are more limited in scope than those of adults, and only gradually approximate the categories of the adult grammar as a function of some form of data-driven learning. For example, Tomasello (1992) argues that children construct syntactic categories through a process of functionally based distributional analysis, which involves analogising across words on the basis of their semantic and distributional similarity. This view is sometimes taken to predict that children's early word combinations will be completely unanalysed. For example, Yang (2010) rejects constructivist models of determiner development on the basis that children's early determiner use is more productive than one would expect if all of the determiner + noun combinations in their speech were rote-learned sequences. In fact, however, most constructivist models assume at least some level of productivity from very early in development. For example, Tomasello (1992) argues that, during the early stages, children have a productive Noun/Object word category, but do not generalise across verbs or other predicate structures. Thus, constructivist models do not predict that children's early knowledge will be completely lacking in productivity, but rather that it will be less productive than that of adults.

This prediction can be broken down into two more specific predictions. The first is that the way in which children combine the words in their early vocabularies will initially be more restricted than the way in which adults combine those same words. Thus, even when one focuses only on the words that children produce in their early combinations, these words should combine less flexibly with each other than the same words combine in the speech of adults. The second is that the flexibility with which children combine their early words should increase as the scope of their categories increases. Thus, even when one focuses only on the words that children produce in their early combinations, there should be a significant increase in the flexibility with which children use these words with respect to each other over the course of development.

To summarise, the critical difference between generativist and constructivist models of grammatical development is that generativist theories predict that children's early speech will pattern in adult-like ways and constructivist models predict that children's speech will be less productive than that of adults even when one has controlled for differences in lexical knowledge. In the present paper, we test these predictions with respect to one particular syntactic category, the determiner. The reason for focusing on the determiner category is that it has recently been claimed that the distribution of determiners in early child English provides evidence that children have an adult-like determiner category from very early in development (Valian et al., 2009; Yang, 2010). The aim of the present paper is to test this claim, and, in so doing, to address the more general question of whether children have adult-like syntactic categories from the earliest observable stages.

1.2. Syntactic categories in the speech of young children

The strongest empirical argument for attributing adult-like syntactic categories to young children is made by Valian (1986). In a now classic study, Valian examined speech samples from 6 children ranging in age from 2;0 to 2;5 and in Mean Length of Utterance (MLU) from 2.93 to 4.14 for evidence of 6 syntactic categories: determiner, adjective, noun, noun phrase, preposition and prepositional phrase. Valian's method involved evaluating children's use of instances of particular categories against criteria based on syntactic diagnostics. For example, children were credited with a syntactic determiner category provided that they:

(1) Only generated correctly ordered strings (i.e. if present in an NP, a determiner had to appear pre-adjective, pre-noun or pre-both).
(2) Did not produce determiners alone as the sole content of an utterance (e.g. ∗a, ∗the, ∗my).
(3) Did not produce two or more determiners in sequence (e.g. ∗kick the my ball, ∗that's a her car).

Valian found that all 6 of the children passed these criteria, and indeed the criteria for each of the other five categories (except for the lowest MLU child, whose performance was borderline on adjectives and prepositional phrases). In a slightly more stringent analysis, Ihns and Leonard (1988) replicated these findings for the determiner and NP categories on data from Brown's (1973) subject Adam, obtained through the Child Language Data Exchange System (CHILDES, MacWhinney, 2000).

Valian's (1986) results have been taken as evidence that children have adult-like syntactic categories from the earliest observable stages. However, in a subsequent analysis, focusing specifically on the determiner category, Pine and Martindale (1996) pointed out that Valian's criteria are actually rather lax and could be passed by children with a relatively limited repertoire of rote-learned phrases (e.g. Kick the ball, In a minute) or lexically specific formulae (e.g. That's a + X, On the + X). Note that Pine and Martindale's point is a methodological rather than an empirical one. That is to say, the point is not that there is strong evidence that children's early determiner use reflects knowledge of rote-learned sequences and lexically specific formulae rather than an adult-like determiner category, but rather that Valian's criteria are not stringent enough to rule out such an interpretation. The basic problem is that showing that children's use of different instances of a particular category conform to the adult grammar is not the same as showing that the child has an adult-like category that includes these different instances, since the child may simply have learned to use different instances of the category separately as part of less abstract representations. For example, the child may have learned that the indefinite article a/an can be produced after That's and before an object name in an object-labelling construction (e.g. That's a + Object Name), and that the definite article the can be produced after On and before an object name in a location-specifying construction (On the + Object Name). This kind of knowledge would allow the child to behave as if she had an abstract determiner category, even

if she were completely insensitive to the fact that the definite and indefinite articles belonged to the same syntactic category.

In view of this problem, Pine and Martindale (1996) proposed an alternative way of distinguishing between generativist syntactic and constructivist limited scope accounts of category development. This was to look at the extent to which children showed overlap in the nouns and predicates with which they produced different determiners (specifically the indefinite article a/an and the definite article the, which were the most frequent determiners in both their own and Valian's data). The rationale was that, if the child had a category that included these two determiners, as opposed to separate knowledge about how each of these lexical items patterned, then any knowledge that the child acquired about one member of the category (e.g. the indefinite article a) should immediately become available for use with another member of the category (e.g. the definite article the). This should result in a relatively high degree of overlap in the contexts in which the child used different determiners (equivalent to that shown by an adult control). On the other hand, if the child's knowledge was more limited in scope, this should result in a relatively low level of overlap in the contexts in which the child used different determiners (significantly lower than that shown by an adult control).

Note that what Pine and Martindale are advocating here is not the use of overlap as an absolute criterion for attributing adult-like knowledge, but the use of overlap as a relative measure of productivity that can be applied to both children and adults, and hence used to investigate whether children's language use is less productive than one would expect if they had adult-like categories. The distinction between overlap as an absolute criterion and overlap as a relative measure is an important one, since it is perfectly possible for a child to show reasonably high levels of overlap in the nouns with which different determiners are used in the absence of any knowledge of the relation between different instances of the category. For example, in line with our earlier discussion of Valian's criteria, a child with the limited scope formulae That's a + X (used to point out objects in the environment) and On the + X (used to specify the location of objects in the environment) would be likely to show reasonably high levels of overlap in the nouns produced with a and the, simply because these formulae are likely to take similar sets of nouns as slot fillers. However, such a child would still be expected to show less overlap than an adult control. The implication is that overlap is best viewed not as a criterion for attributing knowledge to children, but as a measure of the flexibility with which children use different instances of the same putative category. The critical question is therefore not whether children show low levels of overlap in their speech, but whether they show significantly lower levels of overlap than adult controls.

Pine and Martindale (1996) applied this kind of approach to data from 7 children and their caregivers, by calculating the proportion of nouns and predicates used with either a/an or the that were also used with both a/an and the, controlling both for sample size (in terms of number of multi-word utterances) and vocabulary range (by only including in the

analysis nouns and predicates that occurred with *a/an* and *the* in the children's data as a whole). They found a significant difference in the noun and predicate overlap shown by the children and their caregivers at Time 1 (when the children ranged in age from 1;1 to 2;4 and in MLU from 2.20 to 3.40), and a significant difference in the predicate overlap shown by children and their caregivers at Time 2 (when the children ranged in age from 2;1 to 2;6 and in MLU from 2.33 to 3.90), though the difference in noun overlap was no longer significant at this point (*p* = .109, two-tailed). They therefore concluded that the data were more consistent with a limited scope than a syntactic account.

1.3. Critiques of Pine and Martindale (1996)

Pine and Martindale's (1996) findings appear to show that children's early use of the determiners *a/an* and *the* is less productive than one would expect if they had an adult-like determiner category. However, this conclusion has recently been challenged by Valian et al. (2009). Valian et al. accept the logic of Pine and Martindale's overlap measure, but argue that Pine and Martindale's results underestimate children's knowledge of the determiner category because they are based on relatively small samples of determiner + noun sequences, and include nouns that only occur once with a determiner, and hence on which the child could not possibly show overlap. Yang (2010) extends this argument by pointing out that linguistic distributions tend to obey Zipf's law (Zipf, 1949), according to which relatively few words are used with any great frequency and most words are used very rarely, with many occurring only once in even large samples of text. As Yang shows, one of the consequences of this fact is that the level of overlap in the lexical contexts in which two instances of a category occur tends to be low even in adult speech. This is because most lexical contexts (e.g. nouns) are so rare in the data, that the chances of them occurring with more than one instance of another category (e.g. the determiner) are extremely low. This problem is exacerbated by the fact that most nouns are more likely to occur with one determiner than another. For example, English speakers are more likely to say *a bath* than *the bath* but more likely to say *the bathroom* than *a bathroom,* although all four of these sequences are, of course, perfectly grammatical. The implication is that the low level of overlap in children's speech is much less significant than Pine and Martindale (1996) assume, and consistent with the claim that young children do have an adult-like determiner category.

Valian et al. (2009) test this interpretation of the data in two ways. First, they compute child and adult overlap scores based on Pine and Martindale's original formula, and show that both sets of scores are very low, and that there is no significant difference between them. Second, they compute child and adult overlap scores based on a wider range of determiners, and show that, in this case, the overlap scores are much higher, but that there is still no significant difference between the child and adult measures. These findings are taken as evidence that Pine and Martindale's original measures underestimated children's knowledge and that, once one corrects this problem, young children do show adult-like levels of overlap.

Yang adopts a different approach, and compares observed and expected overlap scores for 6 corpora of child speech taken from the CHILDES database (MacWhinney, 2000), and a large corpus of adult speech (the Brown corpus, Kucera & Francis, 1967), where expected overlap scores are calculated on the assumption that both the Nouns and the Determiners in these corpora conform to a Zipfian distribution. Yang shows, first, that the level of overlap in the Brown corpus is relatively low (25.2%), and, second, that there is no significant difference between the observed and expected values in the child and adult corpora, which are almost perfectly correlated (R^2 = .97) with a slope that is close to 1 (slope = 1.08). These findings are taken as evidence that the low overlap scores reported in previous research are simply a reflection of the Zipfian distribution of Nouns and Determiners in naturalistic speech samples, and that young children do have an adult-like Determiner category after all.

1.4. Problems with Valian et al's. (2009) and Yang's (2010) analyses

Valian et al.'s and Yang's analyses appear to show that Pine and Martindale's original results are due to a sampling artefact, and that children's low overlap scores can be explained entirely in terms of the Zipfian properties of naturalistic speech. In fact, however, there are issues with both analyses, which raise serious doubts about this interpretation of the data.

In the case of Valian et al.'s analysis, there are two problems. The first is that, in the only analysis that uses Pine and Martindale's original measure, Valian et al. fail to control the identity of the nouns entering into the analysis. This is problematic since, as Yang has shown, a major determinant of the probability of overlap is the proportion of nouns that occur in the data with very low frequency. On the assumption that nouns show a Zipfian distribution, this proportion is likely to be higher in speakers with larger noun vocabularies (e.g. adults) than speakers with smaller noun vocabularies (e.g. children). Valian et al.'s failure to control noun identity is therefore likely to underestimate the level of overlap shown by adults on the nouns that their children produce, and hence to mask the kind of caregiver–child differences reported by Pine and Martindale.

The second problem is that Valian et al.'s decision to increase the amount of data under consideration by expanding the range of determiners on which overlap measures are based is flawed since it has the effect not only of increasing sample size, but also of considerably reducing the sensitivity of the overlap measure. Thus, whereas Pine and Martindale's measure only credits overlap when a given noun is used with both of the determiners *a/an* and *the*, Valian et al.'s measure credits overlap when a given noun is used with any two of a much larger number of different determiners (e.g. *a/an, the, some, my, one, another*). Valian et al.'s measure therefore results in much higher overlap scores in both children and their caregivers than Pine and Martindale's measure, but it does so by making overlap much easier to achieve, and hence the kind of differences reported by Pine and Martindale much more difficult to detect.

Yang's analysis is subject to similar problems. First, like Valian et al.'s adult measures, Yang's adult measure is not comparable to Pine and Martindale's measures, since it is not based on a controlled set of nouns, but on an adult corpus that includes between 5 and 16 times as many different nouns as any of the child corpora being analysed. On the assumption that these nouns conform to a Zipfian distribution, the proportion of nouns that occur with low frequency in the adult corpus is likely to be considerably higher than the proportion of nouns that occur with low frequency in the children's corpora, which means that Yang's adult measures are likely to seriously underestimate the level of overlap shown by adults on those nouns that the children themselves produce. The implication is that this uncontrolled overlap measure tells us very little about how we should interpret the level of overlap in children's speech.

Second, Yang's child measures are based on corpora that span very long periods of development (Mean = 27.3 months, Range = 9–48 months). While these corpora have the advantage that they provide relatively large datasets for Yang's mathematical analysis, they also include data from periods that are far too late in development to be relevant to the question at hand. For example, by the end of the period of analysis, 4 of the 6 children analysed are more than 12 months older than any of the children in Pine and Martindale's study, and 2 are as old as 5;1, an age at which no current theory would predict a difference between Yang's observed and expected values. This is a major problem for Yang's analysis because it means that, although potentially more reliable than Pine and Martindale's analysis, it is also considerably less sensitive. The implication is that, although Yang's results may have important methodological implications for the field (in the sense that they identify an important confound that needs to be eliminated in future research), they do not rule out the possibility that children show significantly lower levels of overlap than expected during the early stages, which simply cannot be detected in an analysis with such a wide developmental window.

To summarise, while both Valian et al.'s and Yang's analyses clearly demonstrate the need to take account of sampling issues when considering the level of overlap in children's speech, neither shows that the relatively low level of overlap in young children's speech can be explained in terms of sampling issues alone, and hence provides any real support for the claim that young children have an adult-like determiner category.[1] The aim of the present paper is therefore to take a fresh look at this claim by compar-ing controlled measures of noun overlap both between children and their caregivers, and between the same children at different points in development. In Study 1, we explore the effects of differences in vocabulary range on adult overlap measures by comparing overlap measures based on determiner + noun sequences that occur in the child's speech with overlap measures based on determiner + noun sequences that do not occur in the child's speech. In Study 2, we look for differences in noun overlap between children and their caregivers by comparing measures of noun overlap controlled both for the identity of the relevant nouns and the number of determiner + noun sequences in which they occur. In Study 3, we look for developmental differences in noun overlap by comparing controlled overlap measures based on two separate developmental stages. In all three studies we use a measure of overlap in the nouns used with the indefinite article *a/an* and the definite article *the*. This measure is used partly to ensure comparability with previous studies, but mainly because it is much more sensitive than Valian et al.'s measure based on multiple determiners, and hence more likely to reveal differences in the flexibility of children and adults' determiner use, should they exist.

2. General method

All of the studies that follow use the same basic method. This involved automatically searching CHAT-formatted transcripts (i.e. transcripts formatted according to the conventions of the CHILDES database) for determiner + noun pairs. Determiner + noun pairs were identified by focusing on the mor-line (in which words are coded for their syntactic class) and extracting instances of *a/an* and *the* and the nouns that follow them, either directly or with one word intervening between the determiner and the noun. This approach was used to analyse both the adult and the child speech.

2.1. Corpora

All analyses were conducted on the Manchester Corpus (Theakston et al., 2001), which is available in the CHILDES database (MacWhinney, 2000). This corpus consists of 34 h of data for each of 12 English-speaking children from the United Kingdom, collected over a period of 12 months between the ages of approximately 2;0 and 3;0. Each hour of data consists of 30 min of structured play and 30 min of unstructured play recorded in the child's home environment.

2.2. Coding procedure

In order to restrict the analysis to nouns that are grammatical with both the definite and indefinite article (i.e. singular count nouns), the combined maternal data were first searched for singular nouns used with both *a/an* and *the*, and all subsequent analyses were restricted to just this set of nouns (N = 1053). Instances of determiner + noun pairs including either *a/an* or *the* were identified in the child and adult data and used to calculate overlap scores

---

[1] In a more recent study, Wang (2012) addresses this issue by comparing the extent to which the overlap observed in children's and adult's speech deviates from that predicted on the basis of the relative frequency with which particular nouns occur with *a* and *the* in the adult corpus as a whole, and the frequency with which those nouns occur in particular child and adult samples. Wang reports no significant difference between child and adult deviation scores in either English or German. However, since the amount of overlap predicted for the children tends to be much lower than the amount of overlap predicted for the adults, it is not at all clear that scores that measure the extent to which observed overlap deviates from expected overlap are equivalent in the two cases, and can hence be meaningfully compared. In the present study, we avoid this problem by comparing child and adult overlap directly on an equivalent number of instances of each of a shared set of nouns.

on a child-by-child or adult-by-adult basis. All overlap scores were calculated as the proportion of relevant nouns used with *a/an* or *the* that occurred with both *a/an* and *the*. In some of the analyses, overlap scores were calculated in a way that controlled for sample size. This was done by randomly sampling (with replacement[2]) a fixed number of determiner + noun tokens from the relevant pool of items. In other analyses, overlap scores were calculated in a way that controlled for both the identity of the nouns entering into the overlap measure and the frequency with which those nouns occurred with either *a/an* or *the*. This was done by identifying the nouns used with either *a/an* or *the* in both of the relevant datasets (e.g. Child and Caregiver or Child during Phase 1 and Child during Phase 2) and randomly sampling determiner + noun tokens from the sample with the larger number of determiner + noun tokens for that particular noun. This procedure was carried out for every noun that occurred at least twice in both samples. The resulting measures were thus based on exactly the same set of nouns and exactly the same number of determiner + noun tokens for each noun in the set, and only included nouns that occurred at least twice with a determiner, and hence for which there was some chance that overlap would occur. Since these measures are necessarily proportions that are bounded by 0 and 1, all statistical analyses were conducted on arcsin-transformed data to ensure that they met parametric testing assumptions.

3. Study 1: Effects of differences in vocabulary range on overlap measures

As both Valian et al. (2009) and Yang (2010) point out, a critical determinant of the likelihood of any particular noun occurring with both *a/an* and *the* in a particular dataset is the frequency with which that noun occurs with either *a/an* or *the* in that dataset. For example, all other things being equal, the probability of observing overlap when $N = 2$ is $1 - 2/2^2 = .50$, the probability of observing overlap when $N = 3$ is $1 - 2/2^3 = .75$ and so on. Yang (2010) goes on to demonstrate that this fact interacts with the Zipfian distribution of nouns and determiners in naturalistic speech corpora to result in low overlap on all but the most frequent nouns. For example, consider the Zipfian distribution plotted in Fig. 1, where the frequency of any particular item is equal to the frequency of the highest ranked item divided by the rank of the item in question. It is clear from Fig. 1. that, in such a distribution, frequency decreases rapidly as rank order increases such that a large proportion of the items in the distribution (.50 in this case) have an expected frequency of less than two, making it impossible to observe overlap with respect to these particular items. The implication is that overlap scores are likely to be relatively low in any naturalistic corpus, regardless of the nature of the underlying grammar, since many of the

items will simply not occur frequently enough for overlap to be observed.

This argument is clearly correct as far as it goes. However, what is perhaps less obvious is that it also has potentially important implications for the way in which overlap measures should be compared in speakers with different vocabulary sizes (e.g. children and their caregivers). This is due to the fact that, as Yang (2010) points out, Zipfian distributions tend to result in low overlap because of the proportion of the nouns in these distributions that occur very infrequently (i.e. because such distributions have long tails), and one of the consequences of Zipf's law is that the proportion of nouns in the distribution that occur very infrequently increases with vocabulary size (i.e. Zipfian distributions based on larger vocabularies have longer tails than Zipfian distributions based on smaller vocabularies, and so a higher proportion of the vocabulary items are drawn from the tail). For example, consider the distribution plotted in Fig. 2.

This distribution is simply a truncated version of the distribution plotted in Fig. 1, from which the 10 lowest frequency items have been removed to control for the fact that they have yet to be learned by the child [3]. The fact that this distribution is based on fewer vocabulary items than the distribution plotted in Fig. 1. does not make it any less Zipfian. However, it does mean that the average frequency of the items in the second distribution is greater than the average frequency of the items in the first distribution (5.9 versus 3.6). The chances of observing overlap in the second case are therefore higher than the chances of observing overlap in the first case. The implication is that measures based on larger vocabularies (which include higher proportions of low frequency items) are likely to underestimate the level of overlap relative to measures based on smaller vocabularies (which include lower proportions of low frequency items); or, to put it another way, that measures that are not restricted to the same set of (relatively high frequency) nouns used by young language-learning children (such as those reported by Yang for the Brown corpus) are likely to underestimate the level of overlap shown by adults on this more restricted set of nouns, and hence to obscure any differences in productivity that exist between adults and young children.

In Study 1 we investigate this possibility by comparing adult overlap measures based on nouns that do occur in the child data and adult overlap measures based on nouns that do not occur in the child data. The prediction is that measures based on nouns used by both adults and children will be significantly higher than measures based on nouns used only by adults. It is also predicted that adult measures based on nouns used by both adults and children will decrease with development (i.e. as the child's vocabulary

---

[2] The decision to sample with replacement reflects the fact that sampling without replacement has the potential to bias overlap scores by increasing the chances of sampling previously unsampled items as overall sample size decreases. We are grateful to an anonymous reviewer for pointing this out to us.

[3] For ease of exposition we have kept the frequency of the highest frequency item constant across the two Figures. An alternative approach would be to present Zipfian distributions based on the same overall sample size, in which case the frequency of the highest frequency item would increase from 20 to 24, resulting in a further increase in the difference in the average frequency of the items across the two distributions. In either case, it is clear that the average frequency with which vocabulary items occur (and hence the chances of observing overlap on those items) increases as vocabulary size decreases.
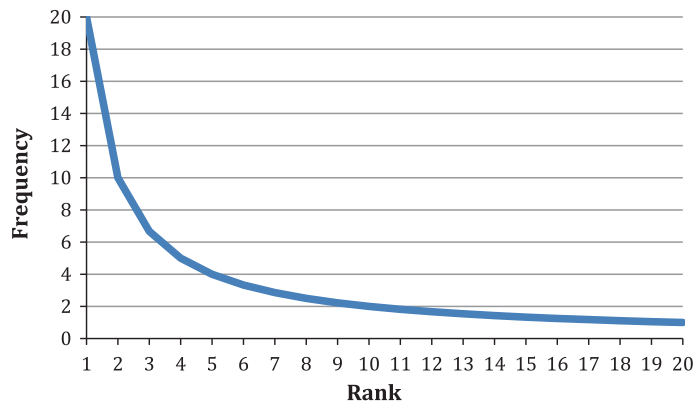
Fig. 1. Zipfian distribution for 20 items, where highest frequency item occurs 20 times.
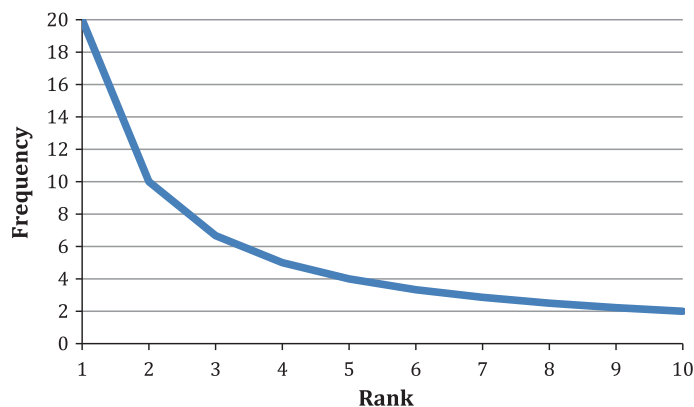
Fig. 2. Zipfian distribution for 10 items, where highest frequency item occurs 20 times.

incorporates more and more nouns from the tail of the distribution). If the results confirm these predictions, they will show that the Zipfian distribution of lexical items in naturalistic speech interacts with differences in vocabulary range in a way that is likely to mask differences in overlap between speakers with different vocabulary sizes. They will thus validate our earlier critique of Valian et al. and Yang's analyses and show that, when comparing overlap scores, it is necessary to control the identity of the nouns on which they are based.

3.1. Method

The aim of this study was to explore the effect of differences in vocabulary range on noun overlap scores by comparing adult overlap scores based on nouns used by both adults and their children and nouns used only by adults. In order to do this, three sub-corpora were extracted from the data of each of the children in the Manchester corpus, where the first consisted of the data from transcripts 1 to 10, the second consisted of the data from transcripts 11 to 20 and the third consisted of the data from transcripts 21 to 30. Each of these sub-corpora was searched for nouns that occurred with either the indefinite or the definite article. The corpus of adult speech directed at each child was

then searched for nouns used with either the indefinite article or the definite article and overlap scores were calculated separately for those nouns that occurred with *a/an* or *the* in each of the child sub-corpora and those that did not. In a further analysis, scores for the first sub-corpus were calculated based on samples of different numbers of determiner + noun pairs. In this case, the samples were obtained by randomly sampling from the relevant pool of determiner + noun pairs in the adult data (i.e. those determiner + noun pairs used by the child in the first sub-corpus and those determiner + noun pairs not used by the child in the first sub-corpus). For each condition/sample size, 10 different samples were drawn and the results were averaged across samples.

3.2. Results

Table 1 presents overlap scores for each of the children's caregivers for each of the three sub-corpora for determiner + noun pairs used by both caregiver and child and determiner + noun pairs used only by the caregiver. Also presented is the number and mean rank frequency of the nouns over which these measures have been computed. It is clear from Table 1 that overlap scores based on determiner + noun pairs used by both caregiver and

Table 1

Caregiver overlap scores for nouns used by the child and nouns not used by the child in three developmental phases.

	Nouns used by child			Nouns not used by child		
	Phase 1	Phase 2	Phase 3	Phase 1	Phase 2	Phase 3
Anne	0.73	0.67	0.65	0.44	0.38	0.40
Aran	0.75	0.73	0.71	0.44	0.41	0.41
Becky	0.64	0.64	0.62	0.37	0.29	0.31
Carl	0.75	0.65	0.66	0.34	0.34	0.28
Dominic	0.79	0.74	0.64	0.36	0.34	0.32
Gail	0.61	0.62	0.64	0.37	0.34	0.32
Joel	0.64	0.62	0.53	0.32	0.27	0.27
John	0.75	0.67	0.64	0.35	0.38	0.35
Liz	0.70	0.63	0.58	0.33	0.26	0.27
Nicole	0.70	0.70	0.63	0.35	0.32	0.31
Ruth	1.00	0.81	0.69	0.44	0.41	0.37
Warren	0.70	0.66	0.68	0.41	0.37	0.33
Mean Noun Overlap	0.73	0.68	0.64	0.38	0.34	0.33
Mean Noun Types	80.33	135.00	161.50	420.92	366.25	339.75
Mean Rank Frequency	159.98	192.11	199.44	387.75	408.64	422.66

child are m[...] [...]er-
miner + n[...] [...]
sub-co[...] [...]
creas[...] [...]
rank[...] [...]
the [...]

[...] [...]
data[...] [...]
facto[...] [...]
duce[...] [...]
menta[...] [...]
significa[...] [...]
$p < .001$[...] [...]ns
produ[...] [...]ot pro-
duce[...] [...] [...]ain effect of
Developmental Stage ($F_{2,22} = 9.69$, $p = .008$, $\eta_p^2 = .468$),
where overlap decreased with developmental stage
(Mean = .55 for Stage 1, Mean = .51 for Stage 2 and
Mean = .48 for Stage 3), and no significant interaction be-
tween these factors ($F_{2,22} = 2.43$, $p = .141$, $\eta_p^2 = .181$).

One interesting thing to note about these results it that
the significant decrease in overlap scores with develop-
mental stage occurs despite the fact that there is a dra-
matic increase in sample size for determiner + noun pairs
used by both caregiver and child across the three sub-cor-
pora. The obvious explanation for this pattern of results is
that, consistent with Zipf's law, despite the increase in
sample size, the average frequency with which particular
determiner + noun pairs occur decreases steadily as more
nouns are included from the tail of the distribution.
Changes in average frequency can also explain the de-
crease in overlap for nouns that are only used by the care-
giver across the three segments. This effect, which is also
consistent with Zipf's law [...]flects the fact th[...]
miner + noun pairs that nev[...]
to be drawn from even furth[...]
tribution than those deter[...]
but only relatively late [...]

It is clear from the a[...]
the nouns on which ove[...]
precisely the location of t[...]

quency distribution) is a critical factor in determining the
size of overlap scores. However, it is also the case that, in
practice, overlap scores tend to be based on much smaller
sample sizes than those reported in Table 1. It is therefore
possible that the scores presented in Table 1 may exagger-
ate the difference between overlap scores based on nouns
used by both the caregiver and child and nouns used by
only the caregiver in smaller samples (where all of the
nouns are likely to be relatively high frequency items). In
order to investigate this possibility, a further analysis
was performed on the data from Segment 1, which in-
volved randomly sampling 100, 200 and 500 instances of
determiner + noun pairs from those used by both the care-
giver and child and those used only by the caregiver for
each of the 12 corpora. For each condition/sample size,
10 different samples were drawn and the results were
averaged across samples.

The results of this analysis are presented in Table 2,
from which it can be seen that overlap scores tend to in-
crease with sample size, as does the difference between
overlap scores based on nouns used by both caregiver
and child and nouns used only by the caregiver. However,
it is also clear that there are substantial differences be-
tween the two types of overlap measures at all three sam-
ple sizes.

This pattern of results was confirmed by submitting the
data to a 2 × 3 repeated measures ANOVA, where the first
factor was Noun Type (Produced by the child, Not pro-
duced by the child) and the second factor was Sample Size
(100, 200 or 500). This analysis revealed a significant main
effect of Noun Type ($F_{1,11} = 15.80$, $p = .002$, $\eta_p^2 = .590$),
where overlap was higher for nouns produced by the child
(Mean = 0.41) than nouns not produced by the child
(Mean = .12), a significant main effect of Sample Size
[...]12, $p < .001$, $\eta_p^2 = .859$), where overlap increased
[...]ize (Mean = .18 for 100, Mean = .25 for 200
[...]r 500), and a significant interaction be-
[...]nd Sample Size ($F_{2,22} = 13.94$, $p < .001$,
[...] analysis using pair-wise comparisons
[...]e was a significant effect of Noun Type
[...]s of Sample Size (all $ps < .005$).

Table 2
Caregiver overlap scores for nouns used by the child and nouns not used by the child in samples of 100, 200 and 500 tokens.

	Nouns used by child			Nouns not used by child		
	100	200	500	100	200	500
Anne	0.19	0.29	0.47	0.07	0.09	0.17
Aran	0.18	0.29	0.45	0.05	0.08	0.16
Becky	0.23	0.33	0.48	0.06	0.10	0.18
Carl	0.19	0.25	0.42	0.09	0.14	0.21
Dominic	0.54	0.59	0.68	0.08	0.13	0.21
Gail	0.20	0.32	0.41	0.06	0.10	0.15
Joel	0.25	0.34	0.46	0.07	0.09	0.16
John	0.19	0.28	0.45	0.06	0.10	0.19
Liz	0.19	0.37	0.50	0.06	0.10	0.18
Nicole	0.30	0.40	0.54	0.05	0.09	0.16
Ruth	0.80	1.00	1.00	0.07	0.11	0.23
Warren	0.23	0.31	0.50	0.06	0.10	0.19
Mean Overlap	0.29	0.40	0.53	0.07	0.10	0.18

It is clear from the above analysis that, even in relatively small samples, the identity of the nouns on which overlap measures are based is a critical factor in determining the size of overlap scores. The implication is that the Zipfian distribution of lexical items in naturalistic speech interacts with differences in vocabulary range in such a way that it is likely to mask differences in overlap between young children and their caregivers. It is therefore necessary not only to control the sample size on which overlap measures are based, but also the identity of the nouns on which they are based, before comparing overlap scores in children and their caregivers.

confound = shared nouns

4. Study 2: comparing overlap scores in children and their caregivers

The results of Study 1 show that comparing overlap in children and their caregivers without controlling for both sample size and vocabulary range has the potential to mask important differences between children and adults. However, none of the studies in the previous literature have controlled satisfactorily for both of these confounds. Thus, Pine and Martindale (1996) do not control adequately for differences in sample size, whereas Valian et al. (2009) do not control adequately for differences in sample size or vocabulary range,[4] and Yang presents low uncontrolled measures of overlap in adults as if they undermined the idea of using overlap measures in analyses of children's speech. The implication is that the question of whether young children use determiners as flexibly as their caregivers, and hence show evidence of having an adult-like determiner category, is still very much an empirical issue.

In Study 2 we investigate this issue by comparing child overlap measures over five different developmental phases with adult overlap measures controlled both for sample size and vocabulary range. The rationale for comparing child and adult measures is that, if young children have an adult-like determiner category from the earliest observable stages, there should be no difference in the flexibility with which young children and their parents use different instances of the determiner category, at least with respect to those nouns that occur with determiners in both the child and the parent's speech. On the other hand, if children's early knowledge of the determiner category is less abstract than that of adults, there should be a significant difference in the flexibility with which young children and their parents use different instances of the determiner category.

The approach adopted in this study involves controlling sample size and vocabulary range by comparing child and caregiver measures based on exactly the same set of nouns and exactly the same number of instances of each of these nouns. Note that this approach has the advantage that it not only eliminates the confounds identified in Study 1, but also controls for a number of other differences between children and their caregivers, including differences in vocabulary size (since each pair of adult and child measures is based on exactly the same set of nouns), and differences in determiner provision and MLU (since each pair of measures is based on exactly the same number of determiner + noun combinations). It therefore allows us to conduct a strong test of the claim that, once one has controlled for differences in children's and adults' lexical knowledge and performance capabilities, there will be no difference in the flexibility with which children and adults use determiners in their early speech.

4.1. Method

The aim of Study 2 was to test the prediction that there would be a significant difference in the flexibility with which young children and their caregivers used different instances of the determiner category by comparing controlled overlap measures for children and their caregivers. This was done by identifying 5 different developmental phases in each child's corpus, calculating overlap scores based on the data for each child for each developmental phase, and then calculating 5 controlled overlap scores for each caregiver, one for each of the child overlap scores. Because there was considerable variation in rate of development across the 12 children, the different developmen-

---

[4] In fact, Valian et al. (2009) do control for sample size and vocabulary range in their later analyses. However, these analyses use a measure that focuses on overlap across a range of different determiners, and so are problematic for other reasons.

Table 3
Mean lengths of utterance (MLUs) for children during Phases 1–5.

	Phase 1	Phase 2	Phase 3	Phase 4	Phase 5
Anne	1.87	2.48	2.87	2.85	
Aran	1.68	2.15	2.29	2.50	
Becky	1.47	1.98	2.36	2.54	
Carl	2.13	2.17	2.03	2.1	
Dominic	1.80	2.38	2.97	2.6	
Gail	2.02	2.47	2.84	2.6	
Joel	1.56	2.02	2.26	2.48	
John	2.04	1.88	2.07	2.09	
Liz	1.78	2.25	2.71	2.73	
Nicole	1.30	1.73	2.00	2.14	2.
Ruth	1.68	2.06	2.20	2.70	2.90
Warren	2.04	2.24	2.64	2.98	3.06
Mean	1.78	2.15	2.44	2.54	2.62

tal phases were defined in terms of the number of different nouns with which each child used the definite and/or the indefinite article in their speech. Thus, Phase 1 was defined as the period from the beginning of the corpus to the point at which the child had used 50 nouns with one of the determiners *a/an* or *the;* Phase 2 was defined as the period from this point to the point at which the child had used 100 different nouns with *a/an* or *the*, and so on. The required numbers of different nouns for Phases 3, 4 and 5 were 150, 200 and 250, respectively. MLUs for the data for each child in each phase are presented in Table 3. It can be seen from Table 3 that Phase 1 roughly corresponds to Brown's Stage I (MLU Range 1.0–2.0); Phase 2 and 3 roughly correspond to Brown's Stage II (MLU Range 2.0–2.5); and Phases 4 and 5 roughly correspond to Brown's Stage III (MLU Range 2.5–3.00) (Brown, 1973).

Controlled overlap scores were calculated as follows. For each noun entering into the child measure, an equivalent number of instances of that noun in combination with *a/an* or *the* was drawn randomly with replacement from the total number of instances in the caregiver's data. The sample for each noun was then analysed to determine whether overlap occurred, and a controlled overlap measure was calculated by dividing the number of nouns for which overlap was observed by the total number of nouns considered. Like the child measures, the controlled overlap measures were thus measures of the proportion of nouns that occurred with both *a* and *the* in the controlled samples. In order to take account of random variation, the sam-

pling procedure was repeated 100 times, resulting in 100 sets of controlled overlap scores for each developmental [phase]. All of the analyses that follow are based on averages

get token frequency

... an child and caregiver overlap ... developmental phases. Also pre-... umber of tokens per noun type that ... measures. Note that for each develop-... both the identity and the number of instances of each noun are the same for both child and caregiver. However, the identity and number of instances of each noun are ... t develop-... condu...

randomly draw from adult data, a small proportion as "child" and the rest as "adult" data

tio... dev... scor... .06 and ... range ... and ... for Phase 5. How ... ...e to realise that this variation is at least partly due to variation in sample size and vocabulary range, which is controlled within each dyad, but is not controlled across the children or their caregivers.

The relation between variation in overlap scores and sampling considerations was investigated by correlating both the child and caregiver measures with the average number of tokens per noun type that contributed to the scores for each caregiver–child pair. This analysis revealed significant correlations for the child measures for Phases 1–4 (all $rs > .62$, all $dfs = 10$, all $ps < .05$), and a marginally significant correlation for Phase 5 ($r = .50$, $df = 10$, $p = .098$), and significant correlations for the caregiver measures for Phases 2, 3 and 4 (all $rs > .82$, all $dfs = 10$ all $ps < .002$), and marginally significant correlations for Phases 1 and 5 (both $rs > .49$, both $dfs = 10$, both $ps < .10$).

It is clear from these results that the level of overlap in both children's and adults' speech is strongly influenced by the number of instances of each particular noun type in the data. This finding is consistent with Valian et al.'s (2009) and Yang's (2010) critiques of Pine and Martindale (1996), and confirms the need to control the number of times that each noun occurs with *a/an* or *the* before comparing overlap scores. However, the fact that the level of overlap in children's and adults' speech is highly sensitive to sampling considerations does not rule out the possibility that children's determiner use is also significantly less flexible than that of their caregivers. Moreover, it is clear from Table 4. that there are substantial differences in the level of overlap shown by children and their caregivers, with children scoring between 14 and 18 points lower than their caregivers during all 5 developmental phases. These differences were analysed using paired sample *t*-tests, which revealed significant effects at all 5 developmental phases (all $ts > 3.95$, all $dfs = 11$, all $ps < .003$). The implication is that children's use of the determiners *a/an* and *the* is significantly less flexible than that of their caregivers, and that

Table 4
Mean child and controlled caregiver overlap scores for Phases 1–5.

	Tokens per noun type (range)	Child overlap (range)	Caregiver overlap (range)
Phase 1	4.46 (2.96–6.96)	.34 (.13–.53)	.49 (.25–.63)
Phase 2	4.12 (2.70–8.14)	.34 (.13–.70)	.48 (.34–.65)
Phase 3	3.84 (2.42–6.77)	.31 (.06–.53)	.45 (.31–.60)
Phase 4	3.59 (2.68–5.03)	.30 (.14–.57)	.46 (.35–.57)
Phase 5	3.60 (2.14–5.74)	.28 (.06–.47)	.46 (.30–.62)

this difference in flexibility persists until relatively late in development (i.e. until most of the children have entered Brown's Stage III).

When taken as a whole, these results show that, although the level of overlap in both adults' and children's speech is strongly determined by sampling considerations, there are nevertheless significant differences in the flexibility with which young children and their caregivers use the determiners *a/an* and *the*. These differences are apparent early in development. However, they also seem to persist until at least Brown's Stage III. They therefore count against the view that children have an adult-like determiner category from the beginning, and are consistent with the view that children's knowledge of the determiner category only gradually approximates that of adults.

5. Study 3: comparing overlap scores in children at two points in development

The results of Study 2 show that, contrary to the claims of Valian et al. (2009) and Yang (2010), there are differences in the flexibility with which children and their caregivers use the determiners *a/an* and *the* in their speech. They also suggest that these differences are not restricted to the earliest observable stages. However, since the child and caregiver measures reported in this Study were not controlled across developmental phase, it is not possible to compare them directly in order to look for developmental changes in the flexibility of children's determiner use. The aim of Study 3 was to overcome this problem by explicitly comparing overlap measures controlled for sample size and vocabulary range in two different developmental phases.

The rationale for comparing child measures over different periods of development is that, if young children have an adult-like determiner category from the earliest observable stages, there should be no change in the flexibility with which they use different instances of the determiner category, at least with respect to those nouns that are used with determiners during both developmental phases. On the other hand, if children's early knowledge of the determiner category is less abstract than that of adults, there should be a significant increase in the flexibility with which children use different instances of the determiner category over the course of development.

5.1. Method

The aim of Study 3 was to test the prediction that there would be a significant increase in the flexibility with which children used different instances of the determiner category by comparing controlled overlap measures based on two different developmental phases. Phase 1 was defined much in the same way as in Study 2 (i.e. as the period from the beginning of the study to the point at which the child had used 50 different nouns with either *a/an* or *the*). However, in order to maximise the power of the analysis, Phase 2 was defined as the period from this point to the last taping session in each child's corpus. Overlap measures for Phase 1 and Phase 2 were obtained by identifying those

nouns that occurred at least twice with either *a/an* or *the* in both segments of the child's and the caregiver's data, and randomly sampling determiner + noun tokens from the segments with the larger number of determiner + noun tokens for each of these nouns in exactly the same was as in Study 2. This allowed us to conduct a very tightly controlled test of the hypothesis that there would be a significant increase in the flexibility of children's use of *a/an* and *the* by computing overlap scores for both the child and the caregiver during Phase 1 and Phase 2 based on exactly the same set of nouns and exactly the same number of instances of each of these nouns. Again the scores presented are averaged across 100 runs of the sampling procedure.

5.2. Results

Table 5 presents controlled overlap scores for each child and his or her caregiver during Phases 1 and 2. It is clear from Table 5, that, although there is no difference in overlap for the caregivers between Phases 1 and 2 (Mean = .51 in both cases), there is a substantial increase in overlap for the children (Mean = .37 versus Mean = .50), who show substantially lower levels of overlap than their adult controls during Phase 1, but very similar levels of overlap during Phase 2.

This pattern of results was confirmed by submitting the data to a 2×2 repeated measures ANOVA, where the first factor was Participant (Child, Caregiver) and the second factor was Developmental Phase (Phase 1, Phase 2). This analysis revealed a significant main effect of Participant ($F_{1,11} = 13.75$, $p = .003$, $\eta_p^2 = .556$), a significant main effect of Developmental Phase ($F_{1,11} = 8.67$, $p = .013$, $\eta_p^2 = .441$), and a significant interaction between Participant and Developmental Phase ($F_{1,11} = 4.97$, $p = .048$, $\eta_p^2 = .311$). Post hoc analysis using pairwise comparisons confirmed that the children showed significantly less overlap than their caregivers during Phase 1 ($p = .007$), and a significant increase in overlap between Phases 1 and 2 ($p = .019$). There was no significant difference in overlap scores for the caregivers during Phases 1 and 2 ($p = .834$), or for the children and their caregivers during Phase 2 ($p = .926$).

The results of this analysis not only confirm that there are significant differences in the flexibility with which young children and their caregivers use the determiners *a/an* and *the* in their use of these determiners, but also show that the child/caregiver comparison conducted here is rather different from the child/caregiver comparisons conducted in Study 2. This is because the Study 2 comparisons are based on all of the nouns used by the child during the relevant developmental phase, whereas the present comparison is based only on nouns used by the child during both Phase 1 and Phase

Table 5
Controlled child and caregiver overlap scores for Phases 1 and 2 for nouns that occurred with *a/an* or *the* during both developmental phases.

	Tokens per noun type	Child overlap during Phase 1	Child overlap during Phase 2	Caregiver overlap during Phase 1	Caregiver overlap during Phase 2
Anne	2.95	0.28	0.31	0.25	0.29
Aran	8.07	0.59	0.69	0.65	0.62
Becky	3.50	0.33	0.46	0.55	0.52
Carl	5.04	0.57	0.64	0.68	0.66
Dominic	4.09	0.52	0.57	0.59	0.67
Gail	2.54	0.15	0.56	0.45	0.52
Joel	2.50	0.27	0.29	0.34	0.43
John	3.67	0.47	0.68	0.68	0.57
Liz	2.89	0.16	0.32	0.44	0.37
Nicole	3.67	0.12	0.51	0.48	0.40
Ruth	4.29	0.53	0.41	0.50	0.52
Warren	6.50	0.39	0.61	0.50	0.56
Mean	4.14	0.37	0.50	0.51	0.51

2. The present results therefore suggest that, by Phase 2, there is no difference in the flexibility with which children and adults use *a/an* and *the* with the nouns with which they used *a/an* and *the* during Phase 1. However, they do not rule out the possibility that children are using *a/an* and *the* less flexibly than adults with those nouns that did not appear with determiners in the Phase 1 data. Indeed, when combined with the results of Study 2, they suggest that this is precisely what is happening. The obvious explanation of this pattern of results is therefore that the flexibility of children's use of *a/an* and *the* only gradually approximates the adult level, with children achieving adult-like performance earlier for nouns used with determiners from the beginning (which tend to be more frequent) than for nouns that are not combined with determiners until later in development.

6. Discussion

The aim of the present paper was to differentiate and test the predictions of generativist and constructivist models of children's early multi-word speech, by focusing on one particular syntactic category, the determiner, and analysing the extent to which children used different instances of this category interchangeably in their speech. This analysis involved comparing controlled measures of noun overlap both between children and their caregivers and between the same children at different points in development.

In a first study, we investigated the implications of Zipf's law for the use of overlap measures by comparing overlap scores based on nouns used by both the child and the caregiver with overlap scores based on nouns used only by the adult. The results of this study showed that, although overlap measures were sensitive to sample size, they were also highly sensitive to the identity of the lexical items over which they were computed, even when sample size was controlled. Thus, caregivers were much more likely to show overlap on those nouns that occurred with determiners in their children's speech than they were to show overlap on those nouns that did not, even in small samples. This finding is not particularly surprising since, as both Valian et al. and Yang point out, overlap scores are sensitive to the number of times that a particular noun occurs with a determiner in the relevant speech sample, and nouns that only occur in the adult's speech are much more likely to be from the tail of the Zipfian frequency distribution than nouns that occur in both the child's and the adult's speech. However, it does suggest that it is necessary to control for vocabulary range as well sample size when comparing child and adult overlap measures, since uncontrolled overlap scores in adults (such as those reported by Valian et al. and Yang) are likely to underestimate the level of overlap shown by adults with respect to the (high frequency) nouns that occur in children's early speech, and hence hide important differences in the flexibility with which adults and children use determiners with these particular nouns.

In a second study, we investigated whether there were significant differences in the overlap shown by children and their caregivers, once sample size and vocabulary range had been controlled. The results of this study revealed large and significant differences between children and their caregivers for all of 5 different developmental phases. These results are important for two reasons. First, they show that the level of overlap shown by adults in samples carefully matched with those of their children is actually rather high (averaging between .45 and .49 across the different developmental phases). They thus undermine the claim that low overlap scores in young children's speech are a Zipfian artefact, and underline the need to control for sample size and vocabulary range when comparing the level of overlap in adults' and children's speech. Second, they show that, once the appropriate controls have been made, there are differences in the flexibility with which children and their caregivers use the determiners *a/an* and *the* in their speech, which persist until at least Brown's Stage III. They therefore count against the view that children have an adult-like determiner category from the beginning, and are consistent with the view that children's knowledge of the privileges of occurrence of particular determiners is initially more lexically restricted than that of adults in the sense that it is embedded in particular constructions that take particular sets of nouns as arguments.

In a third study, we investigated the issue of developmental change directly by asking whether there was a significant increase in the level of overlap shown by young children with respect to a fixed set of nouns across two different developmental phases. Note that this analysis provides a particularly strong test of the claim that children's knowledge is changing over the course of development, since in addition to comparing each child's behaviour with that of an adult control, it compares each child's behaviour with control behaviour from the same child with respect to the same nouns during a later phase of development. The results of this study revealed a significant increase in the level of overlap shown by the child between Phases 1 and 2, together with a significant advantage for the adult over the child during Phase 1, but not during Phase 2. These results confirm that there are significant differences in the flexibility with which young children and their caregivers use the determiners a/an and the early in development. More importantly, they provide strong evidence that children's use of these determiners becomes significantly more flexible over the course of development, and hence only gradually approximates that of adults.

These results have a number of implications for the field as a whole. First of all, they underline the need to control for sampling effects when investigating the scope of children's early ... Yang (2010) ... particularly ... Pine, Freudenthal, ... lli, ... with the ... production data as evidence that children's early categories are more limited in scope than those of adults. In fact, however, the lexical specificity of children's early multi-word speech is difficult to interpret in isolation, since the frequency distribution of words in naturalistic speech is such that even adult speech tends to look somewhat lexically specific, particularly in small samples. Thus, one of the consequences of Zipf's law is that many of the words that occur in both adult and child speech occur with very low frequency. This fact, together with the fact that most words are more likely to occur in some of the contexts in which they are grammatical than in others, inevitably reduces the chances of particular words occurring in more than one syntactic context, and hence tends to give both child and adult speech samples a lexically specific look. The implication is that the lexical specificity of children's speech can only be properly interpreted with reference to some index of how lexically specific one would expect the child's speech to be if she had adult-like knowledge.

One way of deriving such an index is to use a mathematical model such as that described in Yang (2010). However, one of the disadvantages of this kind of approach is that it requires the modeller to have a very good understanding of the distributional properties of the corpora under investigation. For example, Yang's model is heavily reliant on the assumption that the nouns in the corpora being analysed conform to a Zipfian distribution. However, while this assumption may be true of very large datasets, there is no guarantee that it will be true of the kind of child language corpora used in Yang's analyses. Moreover, care-

ful consideration of these corpora suggests that it actually reflects a rather idealised view of the data. For example, Fig. 3. shows the frequency statistics of the 10 most frequent nouns that occur with a/an or the in the 6 child corpora that Yang analyses together with their expected frequencies assuming a Zipfian distribution.[5]

It can be seen from Fig. 3 that, although noun frequency does decrease substantially b... and the 10th most f... dramatically as ... assumptions, with... significantly flatter ... dfs = 9, all ps < .000... the implica... overlap measures reported by Yang for these corpora may underestimate the level that would be expected given the actual distribution of nouns in the data, raising doubts about the validity of his conclusions.

An alternative way of deriving an index of expected overlap is to impose the same sampling restrictions that obtain for the target child on a sample of maternal speech, or speech from the same child at a later point in development. This approach, which is the one used in the present paper, is not without its limitations. For example, because it involves controlling for sampling considerations within dyads rather than across the entire sample, it generates measures that are not directly comparable across children. However, because it controls for sampling considerations directly, it has the advantage that it allows one to conduct targeted analyses of data from relatively early in development without making potentially unwarranted assumptions about the distributional properties of the corpora being analysed. The results of the present study suggest that this kind of approach may be a powerful way of testing the predictions of generativist and constructivist models of early multi-word speech, particularly on the kind of child language corpora that are currently available.

A second implication of the present study is that the way in which the kind of sampling effects identified by Valian et al. (2009) and Yang (2010) interact with the distributional properties of naturalistic speech at different points in development is actually rather more complex than one might assume. Thus, although it might be tempting to assume that child and adult measures are directly comparable, at least once one has controlled for sample size, it is clear from our results that measures of lexical specificity are sensitive not only to differences in sample size, but also to differences in the identity of the lexical items over which they are computed. One obvious explanation for this phenomenon is that a key factor in determining whether a high proportion of vocabulary items occur in more than one context is the proportion of vocabulary items that occur with reasonably high frequency in the sample (or the average frequency of all the relevant vocabulary items). This variable is obviously related to sample size. However, it is also related to the average rank of the relevant vocabulary items in the Zipfian frequency distribution, which tends to increase with vocabulary size and hence to increase the level of lexical specificity in more mature speakers. The implication is that lexical specificity scores can only be meaningfully compared if they have been matched both for sample size and vocabulary range.
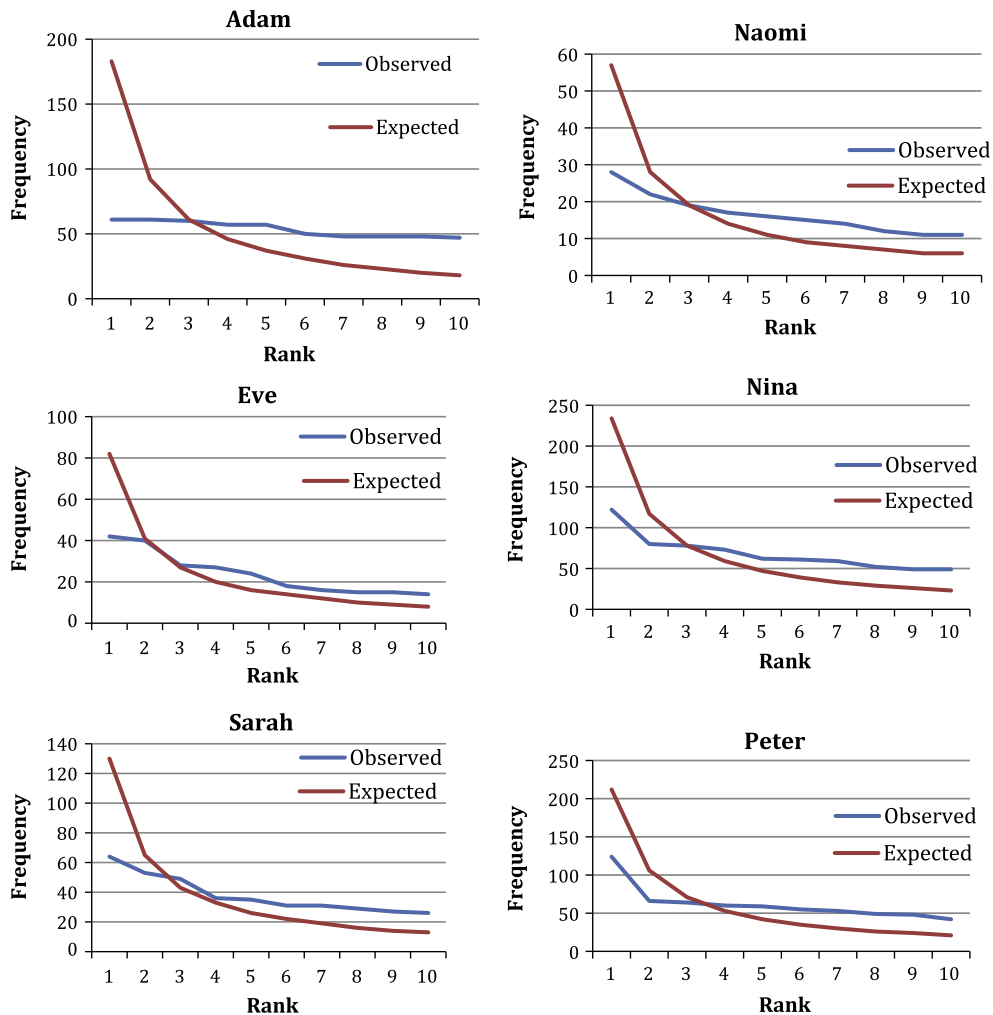
Fig. 3. Observed frequency of the 10 most frequent nouns with *a/an* or *the* in 6 child corpora and their expected frequency assuming a Zipfian distribution.

One way of doing this is to derive control measures by sampling an equivalent number of instances of each item from a control sample, as in the present study. An obvious advantage of this approach is that it controls directly not only for differences in sample size and vocabulary range, but also for differences in the likelihood that particular words will occur in different contexts in the adult language. For example, it controls for the fact that some nouns are likely to occur with both *a/an* and *the* even in relatively small samples, since they tend to occur with *a/an* and *the* with approximately equal frequency in the adult language, whereas others are unlikely to occur with both *a/an* and *the* even in large samples, since they tend to occur much more often with one determiner than they do with the other. The present approach controls directly for these

kinds of item effects because it samples instances of particular items from data with a realistic frequency distribution (i.e. data which incorporates the same item effects).

Finally, our results suggest that, although many previous analyses may not have controlled adequately for sampling effects, it would be a mistake to dismiss the apparent lexical specificity of children's early speech as a sampling artefact (see Aguado-Orea, 2004 and Krajewski, Lieven, & Theakston, 2012 for similar conclusions with respect to the lexical specificity of children's early knowledge of morphology). This is because, although sampling considerations inevitably make both children's and adults' speech look more lexically-specific than it actually is, they also interact with differences in vocabulary range to obscure differences between children and adults, and between children at different points in development. Thus, while the speech of both children and adults has a lexically specific look about it, children's use of the high frequency items that dominate their early speech is less productive than that of adults and becomes more productive over time.

---

[5] These expected frequencies were obtained by summing the observed scores for Nouns 1 through 10 for each child, and dividing the total for each child by 2.93 (which is $1 + 1/2 + \cdots 1/9 + 1/10$) to arrive at the expected score for Noun 1. This score was then used to obtain the expected scores for Nouns 2 through 10, assuming a Zipfian distribution based on the same overall N.

The most straightforward interpretation of this pattern of results is that it reflects a gradual increase in the abstractness of children's syntactic representations. According to this view, the kind of effects found in the present study map more or less directly onto differences in the abstractness of children's and adults' categories and the abstractness of children's categories at different points in development. These differences reflect the fact that children's knowledge of the determiner category is initially embedded in lexically-specific frames, and becomes progressively more abstract as children generalise across these frames, and hence learn to use particular determiners in a wider range of contexts (see Taelman, Durieux, and Gillis (2009) for a similar account of the pattern of determiner development in Dutch).

There are, however, at least two possible alternatives to this interpretation of the data. The first of these is that, rather than reflecting a gradual increase in the abstractness of the children's syntactic representations, the pattern of results actually reflect some difference between child and adult speech that is not controlled in the present analysis. According to this view, the kind of effects found in the present study confound potential differences in the abstractness of children's and adults' categories, with potential differences in the way that instances of these categories are used in naturalistic corpora. For example, it is possible that differences in child and adult overlap scores reflect differences in the range of contexts in which children and adults use the nouns that they produce, and hence in the probability that they will use these nouns with both the definite and indefinite article. Since it is impossible to control for all of the potential differences between adults and children, this kind of explanation cannot be ruled out. However, while it does provide a reasonably plausible explanation of the difference between child and adult overlap measures, it is somewhat less plausible as an explanation of the difference between child scores at different points in development (at least within the rather narrow developmental period examined in this study).

A second alternative possibility is that the lexical specificity of children's early determiner use is not a reflection of the scope of the child's determiner category per se, but of inferences drawn by the child about how instances of that category should be used given the patterning of the input data. According to this view, although children have adult-like syntactic categories from the beginning, they need to establish which particular processes are productive in the language being learned, on the basis of the patterns that they encounter in their input. Lexically specific effects therefore arise when an adult-like generalisation is not licensed sufficiently strongly by the input data, and disappear as the child encounters instances of the category in a wider range of contexts (see Conwell, O'Donnell, and Snedeker (2011) for such an account of differences in the range of arguments with which young children use the prepositional and the double object dative).

Since this kind of account makes very similar predictions about the nature of children's spontaneous utterances to an account that takes lexical specificity at face value, distinguishing empirically between these two alternatives is likely to require the use of different methods from those presented here. We therefore leave it as a question for future research. What is clear from the results of the present study, however, is that young children's use of the determiners *a/an* and *the* is less flexible than that of adults, and becomes more flexible over the course of development. These findings are certainly open to more than one interpretation. However, they provide strong evidence against the claim that the lexical specificity of children's early language is a Zipfian artefact, and are at least consistent with the view that children's knowledge of the determiner category is less abstract than that of adults, and becomes progressively more abstract over the course of development.

References

Aguado-Orea, J.J. (2004). *The acquisition of morpho-syntax in Spanish: Implications for current theories of development*. Unpublished doctoral thesis. University of Nottingham.

Bloom, P. (1990). Subjectless sentences in child language. *Linguistic Inquiry, 21*, 491–504.

Bowerman, M. (1973). Structural relationships in children's utterances: Syntactic or semantic? In T. E. Moore (Ed.), *Cognitive development and the acquisition of language* (pp. 197–213). New York: Academic Press.

Braine, M. D. S. (1976). Children's first word combinations. *Monographs of the Society for Research in Child Development, 41*(1).

Braine, M. D. S. (1988). Review of 'Language learnability and language development' by S. Pinker. *Journal of Child Language, 15*, 189–219.

Braine, M. D. S. (1992). How much innate structure is needed to bootstrap into syntax. *Cognition, 45*, 77–100.

Brown, R. (1973). *A first language: The early stages*. Cambridge, MA: Harvard University Press.

Conwell, E., O'Donnell, T. J., & Snedeker, J. (2011). Frozen chunks and generalized representations: The case of the English dative alternation. In N. Danis, K. Mesh, & H. Sung (Eds.), *Proceedings of the 35th Boston University conference on language development* (pp. 132–144). Somerville, MA: Cascadilla Press.

Ihns, M., & Leonard, L. B. (1988). Syntactic categories in early child language: Some additional data. *Journal of Child Language, 15*, 673–678.

Krajewski, G., Lieven, E. V. M., & Theakston, A. L. (2012). Productivity of a Polish child's inflectional morphology. *Morphology, 22*, 9–34.

Kucera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English*. Providence: Brown University Press.

MacWhinney, B. (2000). *The CHILDES project: Tools for analysing talk* (3rd ed.). Mahwah, NJ: Erlbaum.

Maratsos, M. (1990). Are actions to verbs as objects are to nouns? On the differential semantic bases of form, class, category. *Linguistics, 28*, 1351–1379.

Maratsos, M., & Chalkley, M. (1980). The internal language of children's syntax: The ontogenesis and representation of syntactic categories. In K. E. Nelson (Ed.). *Children's language* (Vol. 2, pp. 127–214). New York: Gardner Press.

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

Pine, J. M., Lieven, E. V. M., & Rowland, C. F. (1998). Comparing different models of the development of the English verb category. *Linguistics, 36*, 807–830.

Pine, J. M., & Martindale, H. (1996). Syntactic categories in the speech of young children: The case of the determiner. *Journal of Child Language, 23*, 369–395.

Pinker, S. (1984). *Language learnability and language development*. Cambridge, MA: Harvard University Press.

Pizzuto, E., & Caselli, M. C. (1992). The acquisition of Italian morphology: Implications for models of language development. *Journal of Child Language, 19*, 491–557.

Schlesinger, I. M. (1982). *Steps to language: Toward a theory of native language acquisition*. Hillsdale, NJ: Erlbaum.

Schlesinger, I. M. (1988). The origin of relational categories. In Y. Levy, I. M. Schlesinger, & M. D. S. Braine (Eds.), *Categories and processes in language acquisition* (pp. 121–178). Hillsdale, NJ: Erlbaum.

Taelman, H., Durieux, G., & Gillis, S. (2009). Fillers as signs of distributional learning. *Journal of Child Language, 36*, 323–353.

Theakston, A. L., Lieven, E. V. M., Pine, J. M., & Rowland, C. F. (2001). The role of performance limitations in the acquisition of Verb-Argument structure: An alternative account. *Journal of Child Language, 28*, 127–152.

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

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

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

Valian, V. (1991). Syntactic subjects in the early speech of American and Italian children. *Cognition, 40*, 21–81.

Valian, V., Solt, S., & Stewart, J. (2009). Abstract categories or limited-scope formulae? The case of children's determiners. *Journal of Child Language, 36*, 743–778.

Wang, H. (2012). *Acquisition of functional categories*. Unpublished doctoral thesis. University of Southern California.

Wexler, K. (1998). Very early parameter setting and the unique checking constraint: A new explanation of the optional infinitive stage. *Lingua, 106*, 23–79.

Wexler, K. (1994). Optional infinitives, head movement and the economy of derivation in child grammar. In N. Hornstein & D. Lightfoot (Eds.), *Verb movement* (pp. 305–350). Cambridge: Cambridge University Press.

Yang, C. (2010). *Who's afraid of George Kingsley Zipf?* Unpublished manuscript.

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