J. Child Lang. **36** (2009), 743–778. © 2009 Cambridge University Press
doi:10.1017/S0305000908009082 Printed in the United Kingdom

Abstract categories or limited-scope formulae? The case of children's determiners*

VIRGINIA VALIAN
Hunter College and CUNY Graduate Center

STEPHANIE SOLT AND JOHN STEWART
CUNY Graduate Center

(*Received* 14 *November* 2005. *Revised* 28 *January* 2008. *First published online* 5 *January* 2009)

ABSTRACT

Six tests of the spontaneous speech of twenty-one English-speaking children (1;10 to 2;8; MLUs 1·53 to 4·38) demonstrate the presence of the syntactic category determiner from the start of combinatorial speech, supporting nativist accounts. Children use multiple determiners before a noun to the same extent as their mothers (1) when only *a* and *the* or (2) all determiners are analyzed, or (3) when children and mothers are matched on determiner and noun types and determiner+noun tokens. (4) Overlap increases as opportunity for overlap increases: children use multiple determiners with more than 50% of nouns used at least twice with a determiner and with 80% of nouns used at least six times with a determiner. (5) Formulae play a limited role in low-MLU children's determiner usage, INCREASING with MLU. (6) Less than 1% of determiner uses are errors. Prior results showing no overlap are likely a sampling artifact.

How do children's earliest syntactic representations develop? We contrast two possible approaches to children's development of syntactic categories. We focus on categories because even radically incompatible approaches agree that humans' mental grammars are eventually expressed in terms of categories.

[*] This work was supported in part by an award from the National Science Foundation to Hunter College (SBE-0123609). Earlier versions of the results were presented at the Architectures and Mechanisms of Language Processing (AMLaP) conference in Aix-en-Provence in September 2004 and at the Boston University Conference on Language Development in November 2006. We thank A. Geogo, M. Lesnick, T. Lesnick, B. Marroquín, D. Sette and C. Theodorou for their help in hand-coding and hand-analyzing computer outputs. We thank our reviewers for their detailed, thoughtful and constructive comments, which led to new data analyses. Address for correspondence: Virginia Valian, Department of Psychology, Hunter College – CUNY, 695 Park Avenue, New York, NY, 10065, USA. e-mail: virginia.valian@hunter.cuny.edu

On one approach, which we will argue against, the child advances from details to abstract structure (Pine & Lieven, 1997; Pine & Martindale, 1996; Tomasello, 2000, 2003). The child's initial representations are lexically specific formulae; her knowledge is limited to individual words and item-based templates. After a period of item-specific learning, the child generalizes from commonalities in her representations to create syntactic categories. The child's task is creation: she must create categories from lower-level representations. This approach is sometimes called constructivism (e.g. Tomasello, 2003). Since constructivism attributes no innate syntactic content to the child, we refer to it as a form of syntactic empiricism.

On the alternate approach, which we will argue for, the child begins with an abstract specification of syntactic categories and must learn the details of how those categories behave in her target language. For two reasons, learning will not be instantaneous. First, the child has to learn what phonetic forms correspond to members of the category. Second, category members do not behave identically. To take determiners as an example, *the* can precede any noun, but *a* can only precede singular count nouns, and *this* can only precede singular count nouns and mass nouns. The three words pattern similarly in that they all precede adjectives and nouns, cannot be sequenced, cannot stand alone as the sole content of an utterance and so on. But the particular nouns that they can be in construction with vary. The child's task here is mapping: she must map the categories she possesses onto the words that she hears. Since, on our approach, the child begins with the abstract category, it is a form of nativism.

We propose that the child is not instantly successful in categorizing every word she hears or produces, but that she has access to all the categories – lexical and functional – from the earliest stages of combinatorial speech. In associating individual words with an abstract category, the child uses a type of pattern learning based on distributional regularities, especially frequent regularities, in the speech she hears. Thus, our scenario incorporates a learning mechanism that constructivists might also accept to solve some acquisition problems. The difference is that in our proposal the mechanism is enriched by a stock of innate syntactic categories available from the beginning of acquisition. The child does not develop non-syntactic local combinatorial rules she will later abandon, though she may use some formulaic speech to cover gaps.

The present paper concentrates on the accessibility of the functional category determiner[1] in children's early speech, as partial support of the

[1] In recent generative syntactic theory (e.g. Abney 1987; Zamparelli 2000), the words that we classify here as determiners are not taken to be members of a single syntactic category, but rather located variously in the head or specifier positions of a series of

claim that all categories – both lexical and functional – are accessible at the outset of combinatorial speech. Determiners are our focus because, as functional categories, they do not participate directly in assigning semantic roles or relations and convey less semantic information than do lexical categories like noun and verb. They are good candidates for a category that, if it had to be learned, would be learned late. Nevertheless, there is a range of data – from spontaneous speech, elicited imitation tasks, comprehension tests and infant preferences – to suggest that children's earliest grammars do contain the abstract determiner category.

On the basis of distributional regularities in the speech of six two-year-olds, Valian (1986) argued that two-year-olds' grammars contained the category determiner, as well as the categories adjective, noun, noun phrase, preposition and prepositional phrase. In particular, she argued that children distinguished between determiners and adjectives, sequencing the latter but not the former, using the latter but not the former as the sole content of an utterance, placing the latter but not the former as the final word of an utterance, and always using the correct word order in determiner–adjective–noun sequences.

A subsequent longitudinal investigation of the speech of a child aged 2;3 at the beginning of observations corroborated Valian's analysis, finding that early determiners were distributed across a variety of nouns and showed no semantic localization (Ihns & Leonard, 1988). A later longitudinal study of the spontaneous speech of seventeen English-speaking children, beginning at 1;6, similarly found that children used determiners from the onset of combinatorial speech and made very few errors other than omissions (Abu-Akel, Bailey & Thum, 2004). Data from different laboratories thus converge on suggesting early representation of the determiner category in English-speaking children.

In an analysis of the spontaneous speech of a child learning Swedish, Bohnacker (1997) similarly found evidence of very early use of determiners. In Swedish, determiners pattern differently from the way they do in English. Definite determiners can be bound morphemes, in which case they exist as noun suffixes, or unbound morphemes, in which case they occur before the noun. Indefinite determiners are always prenominal and unbound. In one case, there is a dual definite determiner, where an unbound morpheme appears before the noun and a determiner-like bound morpheme is also suffixed to the noun. Bohnacker found that the child she observed produced fully grammatical determiner phrases from the start of

functional projections in the extended noun phrase (the determiner phrase or DP). We abstract away from this level of detail because it is not central to the present discussion. The question here is whether the child possesses any sort of abstract knowledge of functional categories such as determiner, not the precise structure of the functional domain.

the observations at 1;8. Over the course of the study, up to 2;1, about 2·9% of the child's determiners were error-free double definite constructs, a usage rate even greater than the adult's 1·5%. The fact that a noun suffixed by a definite determiner never occurred in an indefinite context, and that the same noun appeared elsewhere without the suffix, suggests that the child did not simply reproduce the suffixed form as an unanalyzed unit. Rather, by 2;1, Bohnacker's child appeared to have sophisticated knowledge of the Swedish determiner system, including the syntax of double-definiteness. Cross-linguistic data suggest variation in the extent to which children use determiners early in acquisition, with no single property (e.g. frequency in the input, language typology) responsible for the variability (Kupisch, 2007).

Elicited imitation data have likewise suggested that very young children both attend to and understand determiners. Two-year-olds are more likely to repeat an English noun if it is preceded by an English determiner than if it is preceded by a short nonsense word (Gerken, Landau & Remez, 1990: high MLU children, Experiment 1; low MLU children, Experiment 2). A child who hears 'Pete push-o na car', for example, is less likely to repeat *car* than a child who hears 'Pete pushes the car', suggesting that real determiners help young learners segment their speech and identify nouns. Children are also more likely to repeat nonsense words that occupy the determiner slot than to repeat English determiners, indicating that even very young children distinguish between English and nonsense determiners that have the same prosodic and segmental characteristics.

Comprehension data also demonstrate an early sensitivity to the role of determiners in distinguishing between proper nouns and common nouns. Girls at 1;5 are able to choose one or another animal or doll depending on whether a determiner is used before a nonsense word to guide the child's choice (Katz, Baker & Macnamara, 1974). In a later, better-controlled, experiment, children aged 2;7 used the presence or absence of a determiner to guide their choice of a stuffed animal or toy from a selection of two animals and two block-like toys (Gelman & Taylor, 1984). When an animal is introduced as 'a zav', for example, and the child is asked to put a zav in a basket, she chooses either the animal originally named or the other similar animal to a roughly equal degree. But when the animal is introduced as 'zav', the child primarily chooses that specific animal to put into a basket. Whether the child would behave similarly with any determiner (e.g. *the*, *this* or even a nonsense determiner) is not known, but children at least know that the absence of a determiner in such a context signals a proper name.

Finally, a head-turning experiment with German-speaking infants aged 1;2–1;4 found that when children had familiarization trials with a German determiner (the indefinite article *ein*) followed by a nonsense word (*glamm* or *pronk*), they subsequently showed a novelty effect, listening more to

passages where the nonsense word was in a verb context rather than a noun context featuring a different determiner (e.g. the definite article *das* or the demonstrative *dieses*). When the nonsense word was preceded by a pronoun (*sie*) in familiarization trials, the infants showed no preference (Höhle, Weissenborn, Kiefer, Schulz & Schmitz, 2004). Thus, by 1;2, German children appear sensitive to the cue provided by the indefinite determiner *ein* – that a noun will follow – and further appear able to use the presence of various members of the determiner category (possibly in conjunction with other information) to distinguish noun contexts from non-noun contexts. Since there were some prosodic differences in the two types of test sentences, it is possible that prosody was an effective cue. Somewhat similar results have been reported by Waxman & Booth (2001).

In total, previous research provides a broad range of converging data in favor of the claim that children's early grammars include the abstract category determiner. Taken individually, none of the experiments is definitive, but as a whole these studies show that two-year-olds use determiners in their spontaneous speech, use them to guide selection of an object and treat them as members of a class. While this research does not pinpoint a minimum age or MLU at which the determiner category is conclusively present in children's grammars, neither does it identify an age or MLU at which it is absent.

Although the data suggest that children's earliest grammars contain at least one functional category (namely, determiner), a critic might argue that the data allow for an alternate explanation. Rather than supporting children's abstract category knowledge, it could be suggested, the experimental results to date might merely demonstrate sensitivity to the distributional properties of certain highly frequent lexical items (e.g. English *the* and *a*(*n*); German *ein*) which in the adult grammar are categorized as determiners. We agree that additional evidence would clarify this issue.

Opponents of nativist syntactic accounts of young children's grammars have proposed that relatively concrete item-based knowledge could result in a form of linguistic productivity, without the presence of abstract categories or grammatical rules. On this account, the grammatical knowledge underlying young children's very early speech is not syntactic categories but rather item-specific formulae or templates. Among studies proposing item-specific formulae rather than the attribution of abstract category knowledge to young children, Pine & Lieven (1997) and Pine & Martindale (1996) specifically criticize as too lax Valian's (1986) criteria for establishing the presence of the determiner category in young children's grammars. They observe that a child who had a very limited repertoire of rote-learned phrases or lexically specific formulae such as 'where's the X' or 'that's a X' could satisfy Valian's criteria. They thus propose a further test to distinguish between the nativist's 'syntactic' position and the constructivist's

'limited scope' account, namely, an overlap test. They reason that if a child genuinely had the category determiner, and not simply words that adults would label as determiners, then any knowledge acquired about one member of the category (e.g. the indefinite article *a*) should immediately become available to other members of the category (e.g. the definite article *the*). This should result in a high degree of overlap in the contexts in which a child uses different determiners. For example, a child who hears 'a ball' should be able to produce both 'a ball' and 'the ball'. If, on the other hand, children show little or no overlap, that would argue in favor of an item-based approach.

More recent work extends the item-based approach to an account which focuses on the role of frequent exemplars and usage, a combination of prototype and examplar accounts (Abbot-Smith & Tomasello, 2006; Tomasello, 2003). Rooted in recent cognitive-functional theories of grammar (e.g. Goldberg, 1995), which hold that adult grammatical linguistic competence consists of mastery of the full inventory of constructions of the language, such accounts take children's earliest representations to consist of specific utterances the child has encountered. Such an approach would make predictions similar to those of Pine and his colleagues (Pine & Lieven, 1997; Pine & Martindale, 1996).

According to exemplar approaches, as the child encounters more exemplars in the target language, her initial concrete representations become more abstract: the child extracts smaller units from utterance-level wholes and identifies patterns across utterances. The development of abstract categories and constructions relies on general cognitive and social skills, including the ability to identify patterns and to interpret the communicative intentions of others. Exemplar theories of natural language concepts have several problems (Murphy, 2004), however, and it is not clear from current sketches whether exemplar models can be successfully applied to syntax acquisition or natural language concepts.

Although we cannot evaluate here constructivist accounts in toto, we can examine how well they apply to the syntactic category of determiner. We agree with the spirit of Pine & Martindale's (1996) reasoning: ceteris paribus, a child who says 'a ball' and 'the horse' should also say 'the ball' and 'a horse'. There are, however, some conditions that have to be fulfilled in order for the overlap reasoning to be valid. One is that the child produces nouns often enough with a determiner to allow an adequate assessment of overlap (see Tomasello & Stahl (2004) for a similar point about the need for adequate sampling in order to detect rare but stable events). If the child produces determiners with only a small fraction of the nouns she uses, the chance that overlap will be observed is low, especially if the total sample of speech analyzed is small. In particular, if in a given sample of speech, a child uses each of her nouns only one time with a determiner, overlap will

trivially be measured at zero, regardless of the type of syntactic structure underlying her production.

Another precondition is that the child produce enough different types of determiners to allow an adequate assessment. In the limiting case, if the child only produces one determiner type, there will again necessarily be no measured overlap, but that finding would similarly yield no insight into the nature of the child's grammar. The child could have the category determiner but only be able to identify one member of it.

If those conditions were met, however, and overlap was not observed, that would constitute some evidence that the child was not operating with categories but with sequences that were not syntactically analyzed. We say 'some evidence' rather than 'knock-down-drag-out evidence' because it is possible that there were no contexts in the sample of speech under consideration that demanded a determiner other than the one the child typically used.

Pine & Martindale (1996) analyzed spontaneous speech data from seven children learning English, looking at the extent to which multiple determiners occurred before nouns, and with verbs or other predicates. We concentrate on the results for overlap before nouns, because predictions are clearer there and because overlap before nouns is the focus of Pine & Lieven (1997). We further focus only on the most rigorous analysis that Pine & Martindale conducted, one which examined nouns that occurred at least once in a child's speech with either *a* OR *the*.

Pine & Martindale (1996) found that, at the first taping period, when the children ranged in age from 1;11 to 2;4 and in MLU from 2·2 to 3·4, the extent of overlap in children's use of the determiners *a* and *the* before nouns was significantly lower than that observed in the speech of adults (usually the mother but sometimes the experimenter) to them. Overlap for children was roughly 16% and overlap for adults 30% (Pine & Martindale, 1996: Table 11). Children's early determiner usage, they concluded, is better described by a limited scope account than a syntactic account.

Two months later, when the children's ages ranged from 2;1 to 2;7 and their MLUs ranged from 2·33 to 3·90, they were audiotaped again. The children's speech from the second taping was combined with the speech from the first. Now there was no significant difference in the amount of overlap between children and adults – roughly 33% for children and 44% for adults (Pine & Martindale, 1996: Table 11). Although Pine & Martindale do not comment on the change from Taping 1 to Taping 1+2, by their reasoning the children they observed must have developed the determiner category in two months, when all of them were less than three years old. An alternate and more plausible interpretation would be that the children had the category at Taping 1 but did not reveal it because of the relatively small sample of speech that was analyzed.

Pine & Lieven (1997) examined the spontaneous speech of eleven children over a two year period, from ages 1;0 to 3;0. The focus again was *a* and *the*, this time in the first 400 of the children's utterance types (as distinguished from tokens). Roughly half the data came from parents' diaries or notes by observers during home visits; the other half came from audiotapes (made after children produced more than 20 word types). Of the eleven children, five showed no overlap at all; these children produced relatively few noun types with determiners (averaging 7, ranging from 4 to 9). The six children who did show overlap had relatively more noun types used with determiners (averaging 11, ranging from 6 to 15); three of these six showed overlap that was significantly different from zero. Pine & Lieven, based on the overall low level of measured overlap and the prevalence of lexically specific 'frames' such as 'want a X' in the speech of the children they studied, concluded that the children's determiner usage was best described as resulting from limited-scope formulae rather than true syntactic knowledge. We suggest instead that the small sample of speech analyzed explains why a child who showed overlap on 2 of the 6 nouns he produced with a determiner was nevertheless not found to demonstrate significant overlap.

The data from the studies challenging the determiner category are thus not definitive as they stand, even on the assumption that the methods were unproblematic. But there are also two methodological issues. First, both analyses focused exclusively on the determiners *a* and *the*. The rationale was that, as Valian (1986) had noted, *a* and *the* are the most frequently occurring determiners in young children's speech, accounting in her sample for 72% of the children's determiner tokens. But although that is true for children's determiner usage as a whole, any individual child may use few tokens of either *a* or *the*. For example, in Pine & Martindale's (1996) study, one child used *a* only 15 times out of 1019 utterances, and a second used it only 18 times out of 983 utterances. Those two children showed the lowest levels of noun overlap; we argue that such low overlap is due in part to the likelihood that those two determiners under-represented their total use of determiners. In order to ensure that the full range of a given child's determiner usage is being reflected, it is preferable to perform at least two analyses, one which examines *a* and *the* and one which examines all of a child's determiners.

Secondly, and more significantly, neither Pine & Lieven (1997) nor Pine & Martindale (1996) took into account how often each given noun occurred with ANY determiner. To detect overlap of determiners before a particular noun, the noun must obviously appear at least twice with a determiner in a given child's corpus. If a noun occurs only once with any determiner, overlap is impossible by definition. Yet Pine and his colleagues do not specify that their analyses were restricted to nouns that occurred at least twice with a determiner.

Furthermore, even twice may not be enough. Suppose that a hypothetical child has the category determiner and uses *a* and *the* with equal frequency across all nouns. If a given noun is used twice in a particular sample of the child's speech, there is only a 50% chance that the two determiners will be different and that overlap will be detected. The likelihood increases to 75% if the noun is used 3 times and to 88% if it is used 4 times. Apparent lack of overlap could thus be an artifact of lack of opportunity rather than genuine lack of category knowledge.

We agree with Pine & Martindale (1996) and Pine & Lieven (1997) that overlap, if properly assessed, is a useful measure for distinguishing between a syntactic and a limited-scope account of children's early determiner usage. It is, however, only a necessary condition. If a child produced word salad, she would pass the overlap test but fail to show knowledge of determiners. A second necessary condition is thus that the child exhibit knowledge of the syntactic restrictions on the placement of determiners. Finally, the child should show little dependence on formulae. Some formulae are to be expected, since adults use them as well, but we predict that formulae will make a minimal contribution to the child's productions with determiners.

In the present study, we examine spontaneous speech from the twenty-one children studied in Valian (1991) and from their mothers; we implement several versions of the overlap test that avoid the problems in the earlier works, and supplement this with two complementary tests for the presence of the determiner category. Altogether, we conduct six analyses.

Analysis 1 examines overlap of *a* and *the* before nouns, in a partial replication of the work of Pine and colleagues. *Analysis 2* expands the overlap calculation to include all determiners, not just *a* and *the*.

Analysis 3 assesses overlap for children and mothers as a function of how many times each noun occurred with a determiner. That allows us to plot degree of overlap as a function of opportunity. If, no matter how many times a child or mother uses a noun with a determiner, and if, no matter how many determiners the child or mother has in her repertoire, she shows no overlap, that will constitute good evidence that she is using fixed phrases rather than categories. If, however, both children and mothers show more overlap as there is more opportunity for overlap, that will constitute good evidence for the presence of the abstract determiner category.

Analysis 4 matches children and their mothers on the number of determiner types used, the number of noun types used and the number of determiner + noun pairs used, and calculates overlap in the resulting data-set. An individual who uses more different determiner types is likely to show more overlap, and an individual who uses more noun types is likely to show less overlap. By matching, misleading implications based on such differences can be avoided.

Analysis 5 examines the children's combinations specifically for formulae, with the aim of verifying that overlap in children's determiner usage is not largely driven by their occurrence in fixed formulae or frames.

Analysis 6 examines children's distribution errors. If children do not pass Valian's (1986) criteria for determiner category membership, they do not have the category determiner. A child who shows overlap might know only that *a* and *the* often precede nouns; the overlap criterion alone is itself too lax.

We predict that children will show overlap to the same extent as their mothers in all versions of the overlap analysis, including the one in which children and mothers are matched. We further predict that children's overlap – and mothers' – will be accounted for by measures of the opportunity for overlap, namely, the proportion of nouns that an individual uses more than once with a determiner and by the number of determiners in that individual's repertoire. We also predict that the same patterns will be seen even at the lowest MLUs included in our study (below MLU 2). In other words, there will be no qualitative change or development in children's determiner usage, contrary to predictions from a constructivist perspective. Finally, we predict that the results of Valian (1986) will be replicated: children will make few determiner errors.

OVERALL METHOD

Participants, audiotaping, transcription

Data from Valian (1991) were obtained from the CHILDES database (MacWhinney, 1991). The twenty-one children ranged in age from 1;10 to 2;8 and in MLU from 1·53 to 4·38. They were audiotaped in conversation with their mothers at the child's home, in a daycare center or in the laboratory. In general, there were two tapings, no more than two weeks apart; the first approximately 30 minutes in length and the second approximately 60 minutes in length. Tapes were transcribed by one listener and checked by a second. Disagreements were resolved either via discussion between the two listeners or by a third listener. One child's data (child 18) were taken from Valian (1986).

Corpora

Transcripts for each individual child were merged, yielding 21 transcripts for analysis, with a mean of 764 useable utterances per child. The corpus for each child consisted of all the child's utterances and an approximately equal number of the mothers' utterances (while some of the children's transcripts included other adults in addition to the mother, for consistency we used only the mother's data). In most transcripts there were substantially more

'mother' utterances than 'child' utterances; in these cases a subsample of the mother's utterances was uniformly drawn from the whole transcript. For example, if there were three times as many mother utterances as child ones, we selected every third one of the mother's utterances. In 4 transcripts the child produced slightly more utterances than the mother; in these cases all 'child' and 'mother' utterances were retained for analysis. Details on each child's and mother's corpus are provided in Appendix 1.

Procedure

On CHILDES, the children's and mothers' speech has been morphologically tagged via an automated part-of-speech tagger. We used the tagged XML files to construct a list of noun phrases (NPs) uttered by the child and her mother, using a specially developed program (created by JS). The extraction was driven by a search for items tagged as nouns (N) by the CHILDES part-of-speech tagger. To identify NPs, the program examined each utterance and each word W in the utterance. When the procedure encountered a W labeled as a common noun (N) by the tagger (i.e. a noun but not a proper noun), it then searched for the determiner. The search procedure was the following: (1) the search worked backwards from W, starting with the immediately preceding word and testing each word for its identity as a determiner. The search was terminated if a verb or comma was encountered, in order to establish a left boundary, but continued if an adjective was encountered until arriving at a determiner or the beginning of an utterance. That procedure allowed tabulation of all nouns with or without determiners. (2) The procedure set the noun of the NP to the (minimally one-membered) sequence of common nouns starting with W. The sequence, now going forward, ended when a comma, non-common noun or possessive morphology was encountered. It excluded items that appeared in a small stop-word list including 'mommy', 'daddy', 'lot', 'uh'. (3) The procedure listed all the NPs for each transcript, along with the utterance in which each occurred.

Since the tagger was not completely accurate, the full list of NPs produced for each child and mother was reviewed manually and any errors in the automated tagging were corrected. Specifically, words miscoded as nouns or determiners were removed, and determiners which were not accurately tagged as such were added. Since the program identified the full utterance from which the potential NP was extracted, error identification was aided. When necessary, the full original transcripts were consulted to aid in classifying ambiguous words.

The output also compiled other useful information, such as all utterances containing determiners. The extracted files were then categorized so that, for each child and for each mother, we had counts of (a) how many

determiner types there were, (b) how many tokens there were of each determiner type, (c) how many noun types there were, (d) how often each noun appeared with a determiner and (e) how often each noun appeared with each determiner type.

To code errors, we separately retrieved all possible examples of (a) two determiners in a row (either the same determiner twice or two different determiners), (b) a determiner as the sole content of an utterance, (c) a determiner as the last word of an utterance and (d) a determiner appearing before an adjective which was not followed by a noun. We analyzed each possible error by hand, returning to the original transcripts as necessary. Many of the putative errors were eliminated because they occurred in utterances that were incomplete or had unintelligible portions; others were eliminated because the punctuation indicated that they were false starts. The remainder were categorized along the same lines as Valian (1986).

Using these data, six analyses of children's and mothers' determiner usage were conducted. We present the data by child MLU, rather than age, although the two are highly correlated in this sample ($r = 0{\cdot}72$, $p < 0{\cdot}001$): our focus is on the nature of children's grammars when they first start to put words together and the subsequent development.

ANALYSIS 1

METHOD

In Analysis 1, we conducted a partial replication of the overlap analyses conducted by Pine & Lieven (1997) and Pine & Martindale (1996), examining overlap between *a* and *the* only. For each individual in the study (child or mother), the numerator in the overlap calculation was the number of nouns that occurred with BOTH *a* and *the* in that individual's corpus; the denominator was the number of nouns that occurred at least once with EITHER *a* or *the* in that individual's corpus.

RESULTS

Figure 1 displays the percentage of overlap of *a* and *the* for each of the twenty-one children and mothers in our study. The first analysis column of Table 1 presents the same data for children grouped into four broad ranges of MLU (1·53–1·99, 2·24–2·76, 3·07–3·72, 4·12–4·38), as well as mean child data and mean mother data (see Appendix 2 for raw data). Both children and mothers show little overlap – just under 11%. In other words, nearly nine in ten nouns used with either *a* or *the* appear with only one of those two determiners. There is no significant difference between children and mothers ($F(1, 20) = 0{\cdot}039$, n.s., partial eta squared $= 0{\cdot}002$), but for two children below MLU 2·5, overlap is zero.

TABLE 1. *Determiner overlap and determiner types* (*standard deviations are given in parentheses*)

Children by MLU	% *a/the* overlap	% all Dets overlap	Mean Det types	Mean N types	% matched overlap	
					All nouns	2+ nouns
1·53–1·99 ($n=5$)	11·9 (9·6)	19·6 (15·1)	6·8 (1·6)	36·4 (20·4)	20·6 (18·2)	54·6 (31·4)
2·24–2·76 ($n=5$)	8·1 (7·1)	18·4 (3·7)	10·6 (2·8)	54·4 (23·3)	15·2 (9·5)	45·4 (26·9)
3·07–3·72 ($n=8$)	10·3 (5·1)	22·4 (4·7)	14·0 (3·2)	82·4 (18·1)	17·5 (5·2)	55·8 (10·0)
4·12–4·38 ($n=3$)	12·7 (3·1)	29·3 (8·6)	15·7 (2·1)	82·0 (36·3)	29·4 (6·0)	69·4 (8·4)
Total ($n=21$)	10·5 (6·4)	21·8 (8·7)	11·7 (4·1)	64·7 (28·9)	19·4 (10·9)	54·9 (21·0)
Mothers ($n=21$)	10·9 (3·7)	23·6 (4·4)	22·6 (3·0)	136·4 (36·5)	20·4 (9·2)	59·4 (17·4)

Note: The final column represents nouns occurring two or more times with a determiner.

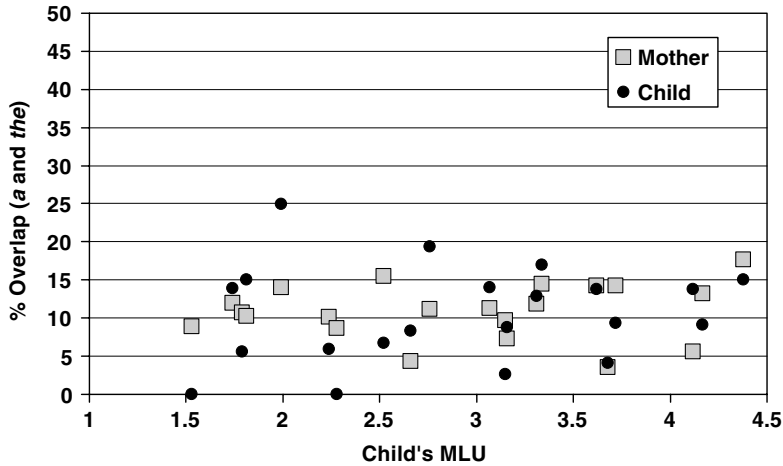

Fig. 1. Analysis 1: Overlap of *a* and *the* for children and mothers as a function of child MLU.

DISCUSSION

Our results differ from those of Pine & Lieven (1997) and Pine & Martindale (1996) in two respects. First, we found equivalent overlap of determiners before nouns for children and their mothers, while Pine & Martindale (1996) reported significantly less overlap for children compared to the adults who were in conversation with them. Lack of opportunity to detect overlap appears to have been the main culprit in the earlier studies.

In Pine & Martindale's (1996) study, children's overlap increased substantially – from 16% to 33% – between the first analysis period (when data from one recording period were analyzed) and the second analysis period (when data from two combined recording periods two months apart were analyzed). But adults' overlap also increased – from 30% to 44%. Since adults could hardly have increased their knowledge of the determiner category between the two recording periods, we infer that the same data limitation was at work for both children and adults: data from a single recording period did not offer a sufficient opportunity to detect overlap. In the analysis that combined the two recording periods, there was no difference between children and adults. Thus, even for Pine & Martindale (1996), children passed the overlap test when a sufficiently large database was at hand. The corpora that Pine & Lieven (1997) used were even more limited and thus even more subject to constraints related to opportunity to detect overlap.

The second difference between our findings and Pine & Martindale's (1996) is in the amount of overlap reported. We found roughly 11% while they reported 33% to 44%. A methodological difference between the studies provides an explanation. Pine & Martindale restricted their analysis to nouns appearing with both *a* and *the* – hence, only singular count nouns – in the combined corpora of all children in their study. In contrast, we examined all nouns in a given individual's corpus that occurred with a determiner. By allowing all nouns (whether singular count, plural count or mass) to be included, we effectively increased the denominator of the overlap calculation and thus reduced the level of overlap calculated.

Two children (children 1 and 7) had zero overlap of *a* and *the* before nouns. These two children might appear to have demonstrated no evidence of presence of the abstract category determiner; the same might be said for another two (children 12 and 17) who showed less than 5% overlap. But two mothers also showed less than 5% overlap; we would not want to suggest that they lack the determiner category. We suggested earlier that consideration only of *a* and *the* could systematically understate the degree of true determiner overlap in children's (or mothers') speech, and lead incorrectly to the conclusion that determiner usage derives primarily from lexically specific formula. That hypothesis is supported by a child-level analysis of our twenty-one children. In aggregate, *a* and *the* were the two most frequently used determiners, accounting for 64% of children's total determiner usage (similar to Valian's (1986) results of 72%). For seven of our twenty-one children, however, including the two with zero overlap, *a* and *the* were not the two most frequent determiners. For these children, measuring overlap only between *a* and *the* is particularly underrepresentative. In Analysis 2 we address this problem by examining overlap among all determiners.

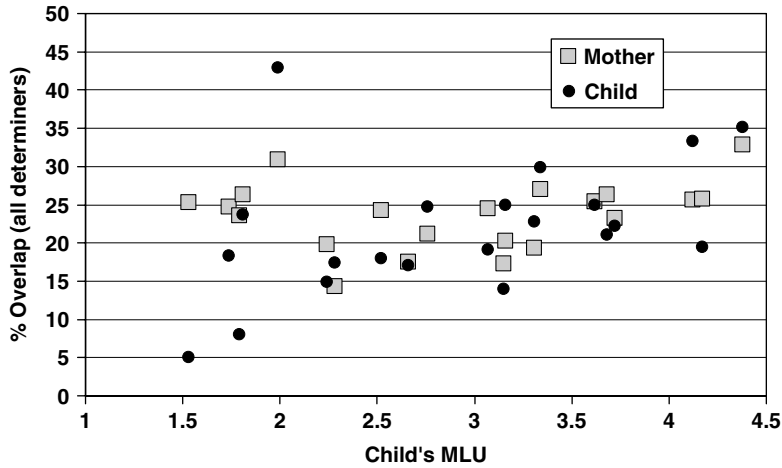

Fig. 2. Analysis 2: Overlap of all determiners for children and mothers as a function of child MLU.

ANALYSIS 2

METHOD

In Analysis 2, we expanded the overlap analysis to include all determiners (see Appendix 3 for all words that were classified as determiners). In this version of the overlap calculation, the numerator for each child or mother was the number of nouns that occurred with more than one determiner type in that individual's corpus; the denominator was the number of nouns that occurred at least once with any determiner.

RESULTS

Figure 2 and the second analysis column of Table 1 display overlap percentages for children and mothers based on all determiners (see Appendix 2 for raw numbers). Overlap calculated in this manner is more than double that calculated on the basis of *a* and *the* alone: children −21·5%; mothers −23·7%. That confirms our hypothesis that examining only *a* and *the* has the effect of understating the true degree to which an individual uses multiple determiners with the same noun. As with the *a* and *the* analysis, there is no significant difference between children and mothers ($F(1, 20) = 1{\cdot}75$, n.s., partial eta squared = 0·08).

DISCUSSION

The results of Analysis 2 provide further evidence that children and their mothers do not differ significantly in the degree to which they use multiple

determiners with the same noun. Rather, as we predicted, observed overlap is a function of opportunity to observe overlap. Overlap was greater in Analysis 2 than Analysis 1 because all determiners, rather than just *a* and *the* were included. Calculated overlap will understate true overlap if only *a* and *the* are included.

Two children below MLU2, however, exhibited less than 10% overlap in Analysis 2 (child 1: 5%; child 3: 8%). Overlap for these children was substantially lower than that for their mothers (mother 1: 25%; mother 3: 24%), and also lower than that for any other participant, child or mother, in the analysis. For these two children, it may seem that we cannot rule out a constructivist limited-scope account of their determiner usage.

Analysis 2 did not, however, take account of another aspect of opportunity to observe overlap, namely the frequency with which an individual uses a given noun with any determiner. As discussed earlier, the corpus under investigation must contain at least two occurrences of a noun with any determiner for overlap to be detectable. We suggested that the failure to exclude nouns occurring only once with a determiner was a potential source of underestimation in previous implementations of the overlap test (Pine & Lieven, 1997; Pine & Martindale, 1996). Furthermore, if an individual does in fact use a given noun with multiple determiners, the opportunity to detect that overlap will increase with the number of times that noun occurs with a determiner in the corpus studied. We investigate this in Analysis 3 by performing an analysis that is stratified by the number of times a given noun appears with a determiner.

ANALYSIS 3

In Analysis 3 we conducted a stratified analysis in order to test our hypothesis that observed overlap is a function of the opportunity to observe overlap. By stratifying each child's and mother's nouns according to how often that noun appeared with a determiner, we could test that hypothesis. On the syntactic account of children's determiner usage that we argue for, observed overlap should increase with the number of times an individual uses a noun with a determiner. If, instead, a lexically specific account of children's early determiner uses is correct, then a child should use the same determiner with a noun no matter how often she uses that noun.

We first calculated, for each child and for each mother, how often each noun occurred with a determiner. We then tabulated, for each child and mother, the percentage of nouns that occurred 1 time, 2 times, 3 times, 4 times, 5 times and 6 or more times with a determiner in that individual's corpus. Finally, we performed a fine-grained version of Analysis 2, calculating all-determiner overlap as a function of the number of times that a noun occurred with a determiner. For each number N of noun + determiner

TABLE 2. *Percentage (standard deviations) of nouns appearing with a determiner 1–6+ times*

	Number of times noun occurred with any determiner					
Children by MLU	1	2	3	4	5	6+
1·53–1·99 ($n=5$)	64 (14)	17 (5)	7 (6)	4 (4)	3 (2)	5 (5)
2·24–2·76 ($n=5$)	64 (10)	17 (10)	7 (5)	4 (2)	3 (1)	5 (3)
3·07–3·72 ($n=8$)	57 (2)	22 (3)	10 (3)	5 (2)	1 (1)	5 (5)
4·12–4·38 ($n=3$)	53 (13)	19 (5)	8 (5)	6 (3)	4 (1)	10 (5)
Total ($n=21$)	59 (10)	20 (6)	9 (4)	5 (3)	2 (2)	6 (5)
Mothers ($n=21$)	62 (6)	17 (3)	8 (2)	5 (2)	2 (1)	5 (3)

Note: Only nouns occurring at least once with a determiner are included. Percentages are rounded to the nearest whole number.

tokens, (where N ranged from 1 to 6+) the denominator of the overlap calculation was the number of nouns that occurred N times with a determiner and the numerator was the number of those nouns that occurred with more than one determiner type.

RESULTS

Table 2 shows the percentage of nouns occurring one, two, three, four, five and six or more times with determiner for children and mothers in our study. As seen here, for both children and mothers, over half the nouns used appeared only once with a determiner in that individual's corpus, allowing no possibility for overlap. For children below MLU 3, this percentage is even higher.

Figure 3 plots the percentage of overlap of all determiners separately for children and adults as a function of the number of times a noun appears with a determiner and Table 3 shows these same data for children grouped by MLU (see Appendix 2 for raw data). As Figure 3 shows, observed overlap is a function of opportunity: the more times a noun is used with a determiner, the more overlap children and mothers demonstrate. Note, for example, that for both children and mothers overlap occurs with over 40% of nouns used just twice with a determiner but with over 80% of nouns used 6 or more times with determiners. Furthermore, this pattern is seen even among children at MLUs below 2·0.

To further establish which factors best predict overlap among children and adults, we conducted a series of correlation and regression analyses.

Correlations

Children's MLU correlated significantly with overlap ($r(21)=0{\cdot}44$, $p<0{\cdot}05$), with the number of determiner types that a child produced

TABLE 3. *Determiner overlap (standard deviations) by Det + Noun frequency*

	% overlap					
	Number of times noun occurred with any determiner:					
Children by MLU	1	2	3	4	5	6+
1·53–1·99 ($n=5$)	0 (0)	38 (26)	81 (24)	39 (35)	67 (58)	89 (19)
2·24–2·76 ($n=5$)	0 (0)	45 (22)	28 (32)	40 (49)	100 (0)	77 (31)
3·07–3·72 ($n=8$)	0 (0)	42 (6)	47 (32)	82 (21)	57 (43)	88 (17)
4·12–4·38 ($n=3$)	0 (0)	40 (13)	65 (17)	76 (22)	67 (58)	79 (19)
Total ($n=21$)	0 (0)	42 (16)	52 (33)	64 (35)	74 (42)	84 (20)
Mothers ($n=21$)	0 (0)	45 (10)	64 (15)	79 (13)	78 (42)	87 (20)

Note: Percentages are rounded to the nearest whole number. For some cells, only a subset of children provided data, because some children occasionally produce no examples of a noun occurring a given number of times with a determiner.

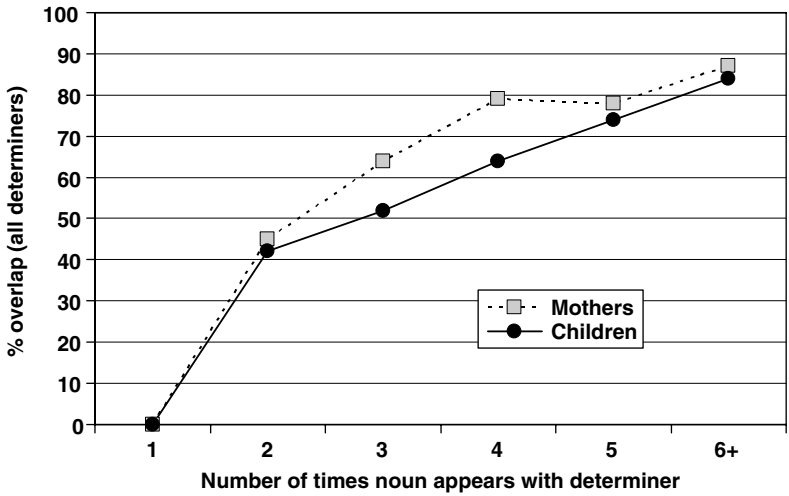

Fig. 3. Analysis 3: Overlap of all determiners for children and mothers as a function of opportunity for overlap.

($r(21)=0{\cdot}78$, $p<0{\cdot}001$), with the number of determiner tokens per type that a child produced ($r(21)=0{\cdot}46$, $p<0{\cdot}04$), with the total number of determiner tokens per 100 utterances ($r(21)=0{\cdot}79$, $p<0{\cdot}001$) and with the percentage of nouns that a child produced more than once with a determiner ($r(21)=0{\cdot}48$, $p<0{\cdot}03$). Low MLUs are correlated with low opportunities to observe overlap. Number of determiner types did not correlate significantly with nouns used more than once with a determiner ($r(21)=-0{\cdot}27$) or with

number of tokens per type. But, as would be expected, the number of determiner tokens per type correlated with number of nouns used more than once with a determiner ($r(21)=0{\cdot}83$, $p<0{\cdot}01$) and with overlap ($r(21)=0{\cdot}44$, $p<0{\cdot}05$). The percentage of nouns occurring more than once with a determiner and overlap were highly correlated for both child and mother (child $r(21)=0{\cdot}80$, $p<0{\cdot}001$; mother $r(21)=0{\cdot}83$, $p<0{\cdot}001$).

Predictors of overlap

We predicted that two variables – number of nouns used more than once with a determiner and number of determiner types – would account for a large and significant portion of the variance in overlap of determiners for both children and their mothers. To test those predictions, we conducted three regression analyses (simultaneous method) on children alone, mothers alone, and on children and mothers together. The predictor variables were percentage of nouns appearing more than once with a determiner and number of determiner types. The criterion variable was all-determiner overlap (i.e. the output of Analysis 2). For children alone, the adjusted R^2 was 0·62 ($F(2, 18)=20{\cdot}44$, $p<0{\cdot}001$). Nouns appearing more than once with a determiner was a significant predictor variable (standardized $\beta=0{\cdot}74$, $p<0{\cdot}001$); number of determiner types was not significant (standardized $\beta=0{\cdot}24$, $p=0{\cdot}095$). Once these factors were included in the analysis, children's MLU did not explain any further variance in overlap.

For mothers alone, the adjusted R^2 was 0·68 ($F(2, 18)=19{\cdot}29$, $p<0{\cdot}001$). Nouns appearing more than once with a determiner was the only significant predictor variable (standardized $\beta=0{\cdot}83$, $p<0{\cdot}001$); number of determiner types was not a significant predictor (standardized $\beta=-0{\cdot}04$, n.s.). For children and mothers combined, the adjusted R^2 was 0·66 ($F(2, 39)=41{\cdot}32$, $p<0{\cdot}001$). Nouns appearing more than once with a determiner was a significant predictor variable (standardized $\beta=0{\cdot}76$, $p<0{\cdot}001$), as was number of determiner types (standardized $\beta=0{\cdot}33$, $p=0{\cdot}001$).

Thus, as Figure 3 suggests, the best predictor of overlap for both children and mothers is the percentage of nouns appearing more than once with a determiner. What accounts for most of the variance in children's and adults' overlap of determiners is the opportunity for overlap, as measured by how often nouns are used with a determiner. Number of determiner types was not a significant predictor for children or mothers alone. But mothers on average used substantially more determiner types than did children (mothers: 19–28; children: 5–21). For the combined group, as the difference in the curves in Figure 3 suggests, the role of number of determiner types helps to account for the difference in overlap that remains between children and adults once the percentage of nouns used more than once with a determiner has been accounted for.

Control partial correlations

In a given conversation, children and mothers are talking about the same thing and thus are likely to influence each other. Children's and mothers' overlap correlated significantly ($r(21)=0{\cdot}51$, $p<0{\cdot}02$). If the relation between nouns occurring more than once and overlap were due to child–parent similarity, then partialling out child–mother overlap should eliminate it. But the partial correlation between the percentage of children's nouns occurring more than once with a determiner and children's overlap was the same as the zero-order correlation ($r(21)=0{\cdot}81$, $p<0{\cdot}001$). The main determinant of children's overlap is how often they use a noun with a determiner more than once.

DISCUSSION

The results of Analysis 3 confirm our prediction that, for both children and adults, the degree of overlap that is observed is a function of the opportunity for overlap. The more often nouns are used with any determiner, the more overlap there is. For children, a greater number of determiner types in the individual's repertoire also correlates with higher overlap. The overlap results support our hypothesis that the same syntactic structure, namely abstract categories, underlies both children's and adults' usage of determiners. There is no support for a limited scope account of children's early determiner usage.

Up to some small limit, a constructivist might also predict that observed overlap will increase with increased opportunity to observe overlap. Although it is not clear what that limit might be, our data implicitly address the point by showing that there appears to be no limit: overlap is over 80% for nouns that appear 6 or more times with a determiner. That is incompatible with a limited-scope formulae account. A constructivist might also predict a correlation between amount of overlap observed and number of opportunities to observe it, but that correlation could only hold for the very beginning part of the curve; once the limit was reached, it would plateau. Again, our results show that there is no plateau. Constructivism cannot account for our data.

Recall that in Analysis 2 two children (numbers 1 and 3) exhibited less than 10% overlap. These same two children also showed among the highest percentages of nouns used only once with a determiner (child 1: 80%; child 3: 72%), and each only used 1 noun more than twice with a determiner. Thus these children had very little opportunity to show overlap. Their low measured overlap appears to be an artifact of their low opportunity for overlap, not a consequence of deficient syntactic knowledge.

The results of Analysis 3 thus provide evidence of continuity. There is no evidence of development in the syntactic structure underlying children's

determiner usage. Once we take into account opportunity to detect overlap (namely, percentage of nouns used more than once with a determiner and number of determiner types), the child's MLU has no further value in predicting overlap. Although their low overall use of determiners makes their data somewhat less stable, the five children under MLU 2 exhibit the same general pattern of overlap as do their higher-MLU peers and their mothers: the more frequently a noun is used with a determiner, the higher the overlap that is calculated (up to 89% for nouns used 6+ times with a determiner).

As children's MLU increases, children use more determiner types and tokens. But there is no development in the NATURE of their determiner usage. A syntactic account predicts continuity: since children begin the acquisition process with an abstract determiner category, development will be limited to adding determiner types to the child's lexicon and to adding knowledge of the particulars of each determiner's behavior. The absence of any indication of development is evidence against a limited-scope or usage-based account, which predicts change in the child's representations, detectable by changes in the degree of overlap children show.

To summarize, Analysis 1 showed no differences between children and their mothers in overlap of *a* and *the* before nouns; Analysis 2 showed no differences between children and their mothers in overlap of all determiners before nouns; Analysis 3 showed that the same factors predict overlap among both children and mothers.

It might be suggested that child–mother differences on measures that could affect overlap, such as number of determiner and noun types, affected the results of the overall overlap calculations in Analysis 2, in effect masking a true difference in the level of overlap between children and their mothers. We address this possibility in Analysis 4.

ANALYSIS 4

In Analysis 4, we control for potential child–mother differences that could affect the results of the overlap calculation. The mothers in our sample used, on average, nearly twice as many different determiner types as did the children – 23 vs. 12 (see Table 1, analysis column 3). Recall that number of determiner types was a significant predictor of overlap in the combined sample of mothers and children. Other things being equal, the child–mother difference on this parameter would lead to higher measured overlap among mothers than children: the more different determiners types an individual makes use of, the higher the probability that two different ones will occur with the same noun. Children below MLU 2 have a particularly small repertoire of determiner types, 6·8 on average; that small repertoire could

be partially responsible for the small amount of overlap detected among low MLU children.

Mothers also used significantly more different noun types than children – 136 vs. 65 (see Table 1, analysis column 4). That difference would tend to result in lower overlap for mothers than children, since the more different nouns an individual uses, the less likely it is that he or she will use the same one multiple times with a determiner. Furthermore, if children and mothers simply use different nouns, overlap could be affected, since some nouns combine more freely with different determiners than others.

Finally, as demonstrated in the stratified analysis (Analysis 3), even if two individuals use the very same nouns and determiners, the number of times a given noun occurs with a determiner is likely to affect the level of observed overlap. For example, if one individual uses a noun ten times with a determiner, and another individual uses that same noun just twice with a determiner, we are much more likely to observe overlap for the first individual than the second.

Analysis 4 matches children and their mothers on three variables to avoid those potential confounds: (1) determiner types; (2) noun types; and (3) number of times each noun type occurred with a determiner. (1) For each child–mother pair, any determiner type that was used by only one member of the pair was excluded from the analysis (i.e. all determiner + noun tokens including that determiner were removed); (2) for each pair, any noun type that was used by only one member of the pair was likewise excluded; (3) for each noun type that remained, child and mother were matched on the number of determiner + noun tokens included in the analysis, by removing tokens at random from the corpus of whichever member of the pair used that noun more frequently with a determiner. Overlap was then calculated as in Analysis 2. In addition, a second version of the overlap calculation was performed in which the analysis was restricted to nouns occurring at least twice with a determiner. As we noted earlier, overlap will trivially be zero for any noun that occurs just once with a determiner.

RESULTS

The results of the matched analysis are shown in analysis columns 5 and 6 of Table 1 and in Figure 4 (see Appendix 2 for raw numbers). 'Matched' overlap was 19·4% among children and 20·4% among mothers, with again no significant difference between the two groups ($F(1, 20) = 0{\cdot}8$; n.s., partial eta squared = 0·04). Restricting the calculation to nouns occurring more than once with a determiner results in substantially higher overlap: child – 54·9%; mother – 59·4% ($F(1, 20) = 1{\cdot}43$; n.s., partial eta squared = 0·07). Thus, when possible sources of extraneous child–mother differences

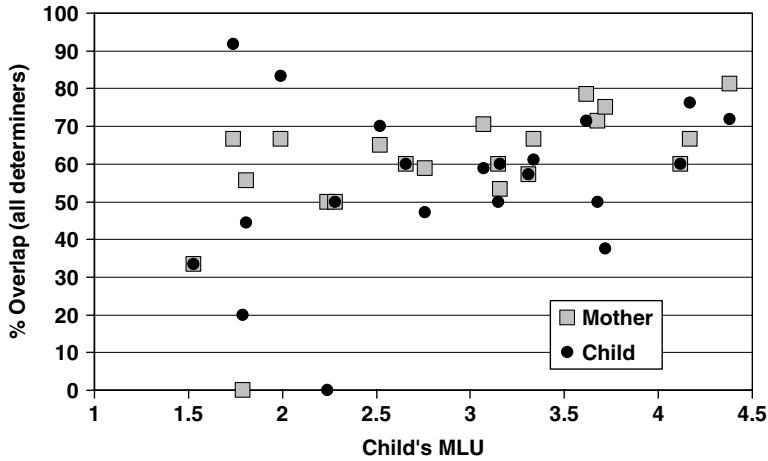

Fig. 4. Analysis 4: Matched overlap of all determiners (on nouns occurring at least twice with a determiner) for children and mothers as a function of child MLU.

are controlled for, children and their mothers use multiple determiners with a given noun to the same degree.

DISCUSSION

In this fourth and most rigorously controlled implementation of the overlap test, in which children and their mothers were matched on several factors that could influence overlap, the two groups demonstrated comparable levels of overlap. Young children and their mothers show no difference in the degree to which they use multiple determiners with the same noun. To the extent that the overlap test provides evidence for the presence of the syntactic category determiner in young children's grammars, the results of the matched analysis provide the most conclusive evidence yet for a syntactic rather than limited scope account of children's early determiner usage.

Convincing as these results are, there is one final possibility that must be excluded before concluding that the overlap analysis supports the syntactic account. Specifically, we need to rule out the possibility that overlap in children's determiner use is due to the occurrence of determiners in lexically specific formulae. Analysis 5 addresses that possibility.

ANALYSIS 5

Analysis 5 examines the extent to which children's determiner tokens occur in multiple lexically specific formulae (frames). If children's first uses of

determiners are locked into particular, frequently encountered locutions, that might be largely responsible for any overlap that might be measured: two formulae used together, each with a different determiner (say, 'where's the X' and 'that's a X'), could give a misleading appearance of productivity. On this view, one would predict formulae to be particularly evident at low MLUs. If, however, children have an abstract notion of determiners, even very early uses of determiners should be relatively independent of lexically specific slots. It could even be the case that use of formulae would increase with MLU, as children become more knowledgeable about the contexts in which determiners can occur.

To test the contribution of frames, we adapt from Eisenbeiss (2000) the definition of a frame as any word + determiner (e.g. *on the*, *in a*, *and a*) or phrase + determiner (e.g. *where's the*, *that's a*) sequence that occurs three or more times in a child's corpus. For each child, we identify all frames used and the percentage of determiner tokens that they account for. We then exclude from our analysis any determiners that appear within a frame, and recalculate overlap. We also correlate the percent of determiners occurring within frames with MLU to determine whether frames are more common at low MLUs, as a frame-based hypothesis would predict.

RESULTS

Table 4 lists the frames (as defined above) identified for each child, and indicates the percentage of that child's determiner tokens that they accounted for. The lowest MLU children – the five children below MLU 2 – showed no evidence of frames. Only 12% of their determiner tokens appeared in a context that occurred 3 or more times. The average across all children was 25%. Individual rates ranged from 0% (Child 1 at MLU 1·53) to 60% (Child 10 at MLU 3·07). A sample of four mothers across the children's MLU range (children 3, 9, 15 and 20) showed frames accounting for an average of 29% of determiner tokens, ranging from 25% to 33%.

When determiners that appear in frames are excluded, overlap declines from 21·8% (the value calculated in Analysis 2) to 19·8%. The decrease is primarily due to the fact that there are now fewer nouns that occur more than once with a determiner: in Analysis 2, 41% of children's nouns occurred more than once with a determiner, but in Analysis 5, fewer nouns – 35% – occurred more than once with a determiner. When the analysis is restricted to nouns occurring more than once with a determiner (as in Analysis 4), the exclusion of frames results in an increase in overlap (from 52·1% with frames to 55·5% without frames).

Use of frames correlates positively with MLU: as children's MLU increases, so does their use of frames ($r(21) = 0{\cdot}58$, $p = 0{\cdot}006$). Use of frames is not correlated with extent of overlap ($r(21) = 0{\cdot}26$, n.s.). When MLU is

TABLE 4. *All frames occurring three or more times per child*

Child	Frames (number of tokens)	% of Det tokens in frames
1		0
2	and a (4), that a (3), where the (3)	8
3	in the (5), on the (3)	23
4	on that (5), in the (4)	10
5	that a (6), that the (4), in the (4)	18
6	got a (4), got two (3), on my (4), want some (10)	23
7	close the (3)	7
8	find the (4), in the (3), made a (6), on my (3), see a (3)	12
9	in the (3),where's my other (4), this the (3), where's the (4)	20
10	at the (11), get the (3), in the (5), it's a (9), on the (4), see the (9), that's a (9), that's the (5), there's a (11), this a (4), want some (3), want the (4), where's another (5), where's the (20)	60
11	and a (5), has a (3), have some (4), in the (9), in my (3), is a (7), is the (6), it's a (9), of the (4), on the (8), that's a (5), there's a (3), this a (4)	38
12	have a (8), he's a (3), in my (3), in the (5), it's a (13), make a (5), make the (3), on my (4), on the (4), see the (3), that's a (12), that's his (3), there's a (6)	40
13	all the (3), are the (3), in the (5), is a (6), is the (3), like a (3), put the (3), see the (5), is the other (3), up the (3)	26
14	and a (3), and the (4), at the (4), in the (3), is a (15), is the (4), on a (3), on my (4), on the (15), there's a (3), this a (3), under the (3), want a (3), what's this (3), where the (3)	38
15	at the (3), found another (3), got the (4), in my (4), in the (9), is the (3), it's a (5), see any (3), that your (5), this a (3), this the (4), where's the (5)	23
16	and the (4), he's a (7), in my (3), in the (8), on the (5), taking a (6)	22
17	at the (4), get the (3), I'm a (3), in my (5), in the (3), it's a (4), on my (4), that's a (6), want my (5)	28
18	take a other (5), has other (3), have a (6), not a (5), that a (5)	22
19	and the (3), go our (3), got some (3), have some (8), in the (6), in this (4), want some (3), where's the (6)	30
20	all the (3), and a (7), and the (6), Dick the (8), has a (4), he a (3), he's a (5), in her (3), in my (6), in the (15), is a (7), on the (10), see her (3), this a (5)	38
21	and a (6), and this (4), build a (9), close the (3), get another (3), got a (3), here's a (5), here's the (7), in the (6), into the (3), is a (6), is the (4), make a (8), that's a (4), this a (7), up the (5), weeding the (6), what a (4), where's the (4)	35
	Total ($n=21$)	25

partialled out, use of frames remains uncorrelated with extent of overlap ($r(21)=0{\cdot}008$, n.s.).

DISCUSSION

Analysis 5 showed that children's overlap is not due to the use of lexically specific formulae. Children under MLU 2 used frames sparingly, despite

our broad definition of a frame – any Word + Det or Phrase + Det sequence that occurred 3 or more times. Removing the liberally defined formulae from the data resulted in only a small decrease in the percentage of overlap, and that decrease was due to a reduction in the number of nouns used more than once with a determiner. There was also no relation between the percentage of a child's determiners that occurred in formulae and the percentage of overlap.

To the extent that frames play a role in children's productions, they play a greater role LATER rather than EARLIER. MLU was positively correlated with use of frames. And, had we restricted frames to phrases occurring before a determiner, no frames would have been visible until MLU 2·66. The use of expressions like *where's* and *there's* and *that's* requires the ability routinely to produce multimorphemic utterances. With limited planning and memory capacities, the lowest MLU children produce no such examples. Our results differ from Eisenbeiss (2000), who found more frames early than late in acquisition of German determiners. The reasons for this discrepancy are not clear, but the fact that our sample of four mothers showed 29% use of frames suggests that frame usage is common in speakers with a full command of syntax.

In summary, children's overlap is not due to the use of multiple formulae. When taken together with Analyses 1–4, Analysis 5 supports our hypothesis that young children's use of determiners is based on a syntactic representation of the category determiner rather than limited scope formulae.

We argued earlier, however, that 'passing' an overlap test is not sufficient evidence of a child's possession of the determiner category. In addition, the child should pass distributional tests, using determiners only in contexts where they are allowed. That is the focus of our final analysis.

ANALYSIS 6

Analysis 6 follows the procedure of Valian (1986), examining the children's utterances for possible determiner errors in five categories: (a) two determiners in a row (either the same determiner twice or two different determiners); (b) a determiner as the sole content of an utterance; (c) a determiner as the last word of an utterance; (d) a determiner appearing before an adjective which is not followed by a noun; and (e) a determiner appearing after an adjective or noun.

RESULTS

Analysis 6 identified remarkably few errors overall, confirming Valian's (1986) findings. Across the twenty-one children studied, out of 2838

TABLE 5. *Summary of possible determiner errors*

Error type	Number of children producing error	Range of number of errors per child	Range of percent errors per child	Total errors	Example
Det as last or only word	1	0–3	0–3·2	3	'Old McKaw had a'
Det + Adj with no noun					
missing copula	3	0–4	0–8	7	'this heavy'
color terms	8	0–4	0–2·6	13	'that is a yellow only'
other types	8	0–2	0–5·7	11	'that is for the sunny'
Dets in sequence					
counting	8	0–3	0–12	13	'one two three four five'
missing copula	16	0–13	0–17·1	66	'that a shoe'
all + Det	12	0–14	0–9·3	33	'all my toys'
a + *other*	8	0–15	0–13·9	26	'do you have a other toy?'
uncategorizable	9	0–3	0–2·8	18	'where the a other car?'
Det after N or Adj	0				

Note: Error types in bold are classified as genuine determiner errors. See text for additional examples.

determiner uses, only 190 were flagged as POSSIBLE errors, a possible error rate of 7%, ranging across children from 1% (child 11) to 30% (child 18). Of those errors, only 21 were classified as GENUINE determiner errors, an actual error rate of less than 1% (whether calculated from the pooled errors or by child), ranging across children from 0 to 3·2%. Errors were scattered across the MLU range.

Table 5 summarizes all possible errors, including sequences that we conclude do not involve determiner errors but that we include for completeness. The error types which we have placed in bold type can be taken to be genuine errors. There are a total of 21 genuine errors: 3 where the determiner was the last word or the only word of an utterance (child 4, 'Old McKaw had a'; 'my') and 18 errors where children illegitimately sequenced determiners (e.g. child 2, 'here's a more one, mo'; child 8, 'this part a this'; child 18, 'want to ask him to find other one of these'; as is apparent, even some of these may not be true errors). Most children – thirteen of the twenty-one – made no errors; of the nine children who did make genuine errors, the largest number was 4 errors, accounting for 4% of that child's determiner tokens. Three of the five children below MLU 2 made no errors.

Of the possible errors that we excluded from the final count, the largest category was 73 cases where a copula seems to be missing – 7 involving an adjective and 66 involving a different term; seventeen of the twenty-one children made one or both errors. Three children made missing copula errors with Determiner–Adjective combinations; those errors averaged less than 4% of those children's determiner tokens; no child made more than 4 such errors. Examples of Determiner–Adjective errors include: child 1, 'this heavy'; child 14, 'that green'. Sixteen children made sequencing errors that almost always involved a demonstrative (*this* or *that*; once *the other*) plus a phrase. Since these phrases never occurred as either the subject or the object of an utterance with a verb, unlike other determiner phrases, we categorized them as errors in which a copula is missing rather than as determiner errors. For the sixteen children, missing copula errors averaged 3·3% of children's tokens; one child made 13 such errors, accounting for 17% of determiner tokens. Examples include: child 5, 'that a bear' and 'that the baby dolly'; child 8, 'this a scissor'; child 14, 'that my screwdriver'; child 16, 'this my soft bag'; child 19, 'and that a hole in there'; child 21, 'this a color'.

Thirteen of the twenty-one children made one or the other of the two Det–Adj error types; three made both types. Adult speech typically requires a noun or pronoun after a determiner, as in 'a purple one' or 'the hard one'. Although the Det–Adj sequences are clearly errors, it is not clear whether they are determiner errors or adjective errors. We excluded them from the rank of determiner errors, but including them would not materially change the findings: instead of an error rate of less than 1%, the error rate would be less than 2%. Eight children combined a determiner with a color adjective without a following noun; those errors averaged less than 4% of those children's determiner tokens; no child made more than 4 errors. Examples include: child 8's reply to the parental query 'What kind of animal is that?' with 'a purple'. Similarly, eight children combined a determiner and a non-color adjective but did not supply a following noun; those errors averaged 2% of those children's determiner tokens; no child made more than 3 errors. Examples include: child 3, 'the dark', where 'dark' was used as the opposite of 'sun'; child 12, 'I have some curly'; child 17, 'he do the hard'.

Two related errors involved *all* and *a other*. Twelve children combined *all* with *the*. Since such locutions are common in adult speech, we did not categorize them as errors. *All* errors averaged less than 2% of those children's determiner tokens. Child 16 made the largest number – 14 errors, accounting for 9% of determiner tokens. Examples include: child 12, 'see all the animals'; child 15, 'all the tea's in here'; child 16, 'excuse me, Mom, I wanna get in all the babies?'; child 20, 'gonna stand all the people up'.

Eight children combined *a* with *other*. We classified this as a morphophonological error, since it relates to an idiosyncratic fact of English: *a* + *other* is realized as *another*. Those errors averaged less than 3% of those children's determiner tokens. Child 18 made the largest number – 15 errors, accounting for 14% of that child's determiner tokens. Examples include: child 7, 'get a other one?'; child 18, 'this a other zebra?'; child 21, 'let me put a other red one in'. Finally, there was the category we call 'counting'. Eight children said numbers in a row; those errors averaged less than 3% of those children's determiner tokens. No child produced more than 3 such examples.

In sum, there were very few errors, even if one rejects the distributional reasoning on which we classified only 21 of the errors as genuine determiner errors.

DISCUSSION

Children in the present study had extremely few errors with determiners, in keeping with previous results (Abu-Akel, 2004; Ihns & Leonard, 1988; Pine & Lieven, 1997; Pine & Martindale, 1996; Valian, 1986). Using Valian's tests, we found that children put determiners in the right place vis-à-vis adjectives and nouns, did not sequence them except where allowed, did not use them as the sole content of an utterance, and did not use them as the last word of an utterance. By this test, as with the versions of the overlap test, children therefore show evidence of the abstract functional category determiner from the outset of combinatorial speech. The data on children's comprehension of determiners also dovetail with that conclusion.

Pine & Lieven (1997) and Pine & Martindale (1996) note other types of errors that their children made, but few if any of those errors are relevant to the claim that children have the abstract category of determiner. For example, 'agreement errors' (e.g. *a cars*, *a pants*) might suggest the child does not have full knowledge of the English system of determiner–noun agreement, or might reflect difficulty with nouns that exhibit a mismatch between syntax and semantics (e.g. *pants*, which is semantically singular but syntactically plural), but they have no bearing on knowledge of the distributional properties of determiners.

Another type of error Pine & Lieven observed is 'word class errors' such as 'a pull it', 'a doing', or 'this a mine'. Only 6 errors of this kind, however, were found in the combined corpora of the seven children Pine & Lieven (1997) analyzed. That number is many fewer than would be predicted if children's utterances were based on formulae such as 'a X' or 'this a X'. If the child really thought that any *X* could be substituted, such word class errors should be rampant; the fact that determiners overwhelmingly precede adjectives or nouns could not be explained. Finally, we note that

those errors, as well as all of the sequencing errors (e.g. 'a my other book', 'put a those on'), involved the determiner *a*, presumably realized phonetically as [ə]. Whether those productions were *a*, a filled pause (*uh*) or a filler syllable cannot be determined. The fact that *the* was never involved, however, makes *a* less likely.

GENERAL DISCUSSION

We used four types of test to determine if two-year-olds had the syntactic category determiner. One type (Analyses 1, 2 and 4) measured the extent to which children used more than one determiner in front of a given noun – refined versions of the overlap test suggested first by Pine & Martindale (1996). The second type (Analysis 3) examined overlap as a function of how many times a child or mother used a noun with a determiner. The third type (Analysis 5) was an analysis of the prevalence of lexically specific formulae. The fourth type (Analysis 6) was the set of distributional criteria Valian (1986) had used. For all six analyses across the four types of test, children's speech demonstrated the category determiner.

Children used more than one determiner before nouns to the same extent that their mothers did. Whether we confined our analysis to the determiners *a* and *the* (Analysis 1), as did Pine & Lieven (1997) and Pine & Martindale (1996), or looked at all determiners (Analysis 2), children and mothers were indistinguishable in the amount of their overlap. Similarly, when children and mothers were matched on determiners, nouns and determiner–noun combinations (Analysis 4), there were no differences between them.

When analyses took into account the opportunity for overlap (Analysis 3), we found that, like adults, children (even those under MLU 2) frequently exhibited overlap of determiners before nouns. For both children and their mothers, multiple determiners were used with over fifty percent of all nouns used twice or more with a determiner; this increased to eighty percent of nouns used six or more times with a determiner. In contexts where overlap in determiner usage can be expected if children abstractly represent determiners, overlap is the norm rather than the exception even for children at the outset of combinatorial speech.

Further, the same factor that principally accounts for how often adults use multiple determiners with a given noun – the percentage of nouns that occur more than once with any determiner at all – also principally accounts for how often a child uses multiple determiners. Thus, overlap is best predicted by a measure of opportunity for overlap. Since children used many fewer determiner types than did their mothers – about half as many – they were additionally limited in their opportunity for overlap. Together, those two factors (percentage of nouns used more than once with any determiner

and number of determiner types) account for two-thirds of the variance in overlap across the total combined sample of children and mothers.

What develops in children is not the category determiner but how often children use determiners and how many determiners they know. Buttressing that conclusion is the very small percentage of errors – less than 1% of all determiner uses – in children's placement of determiners. When children use determiners they use them correctly.

Children's correct use is not due to reliance on formulae, contra Pine & Lieven (1997) and Pine & Martindale (1996). Lexically specific formulae were rare in the speech of our lowest-MLU children and tended to be used MORE often, rather than less often, with advancing MLU (Analysis 5).

Although our motivation for increasing the determiner types examined was primarily to improve our statistical base, we note that identifying overlap among multiple determiners lessens the possibility that the child is using lexically specific formulae. Imagine a child who only had the formulae 'it's an X' and 'where's the X' and used both formulae with a relatively small number of nouns. Such a child would pass Pine & Martindale's (1996) overlap test, without having the determiner category. The overlap test with *a* and *the* is compatible with lexically specific formulae; it itself is too lax. By considering the wide range of determiners that we do, we reduce the risk of mistaking lexically specific formulae for an abstract category. Consider how many formulae or frames the child would need to produce overlap between (say) *a* and *the*, *a* and *my*, *the* and *some*, and so forth.

We conclude that children begin multiword speech with abstract categories that are already well-articulated. We do not claim that NO children's – or adults' – utterances are formulaic. Some clearly are. But just as adults' formulaic speech is compatible with their having abstract syntactic categories, so is children's formulaic speech. The question is whether there is independent evidence of syntactic categories, and our six analyses answer that question affirmatively.

In their successful passage of overlap tests and distributional tests, and in their lack of reliance on lexically specific formulae, very young children demonstrate that they operate with the abstract category determiner. What has appeared in some studies to be children's failure to show evidence of abstract categories is better attributed to the researcher's limited opportunity to detect the evidence. We hypothesize that such limited opportunities are also responsible for doubts about children's representation of other functional categories (e.g. Radford, 1990) or the category verb (Tomasello, 2000; Theakston, Lieven, Pine & Rowland, 2001).

Even at the outset of combinatorial speech, children have abstract functional categories. Taken together with earlier comprehension data, our production data suggest that once children can begin to put words together, there is no point at which they rely on formulae to the exclusion of abstract

syntactic categories, either lexical or functional. Our data provide the strongest support to date that children begin talking with syntactic categories.

REFERENCES

Abbot-Smith, K. & Tomasello, M. (2006). Exemplar-learning and schematization in a usage-based account of syntactic acquisition. *Linguistic Review* **23**, 275–90.

Abney, S. (1987). The English noun phrase in its sentential aspect. Unpublished PhD dissertation, Massachusetts Institute of Technology, Cambridge, MA.

Abu-Akel, A., Bailey, A. L. & Thum, Y-M. (2004). Describing the acquisition of determiners in English: A growth modeling approach. *Journal of Psycholinguistic Research* **33**, 407–424.

Bohnacker, U. (1997). Determiner phrases and the debate on functional categories in early child language. *Language Acquisition* **6**, 49–90.

Eisenbeiss, S. (2000). The acquisition of the DP in German child language. In M.-A. Friedeman & L. Rizzi (eds), *Acquisition of syntax. Issues in comparative developmental linguistics*, 27–62. Harlow: Longman.

Gelman, S. A. & Taylor, J. (1984). How two-year-old children interpret proper and common names for unfamiliar objects. *Child Development* **55**, 1535–40.

Gerken, L. A., Landau, B. & Remez, R. E. (1990). Function morphemes in young children's speech perception and production. *Developmental Psychology* **26**, 204–216.

Goldberg, A. E. (1995). *Constructions: A construction grammar approach to argument structure*. Chicago: University of Chicago Press.

Höhle, B., Weissenborn, J., Kiefer, D., Schulz, A. & Schmitz, M. (2004). Functional elements in infants' speech processing: The role of determiners in the syntactic categorization of lexical elements. *Infancy* **5**, 341–53.

Ihns, M. & Leonard, L. (1988). Syntactic categories in early child language: Some additional data. *Journal of Child Language* **15**, 673–78.

Katz, N., Baker, E. & Macnamara, J. (1974). What's in a name? A study of how children learn common and proper names. *Child Development* **45**, 469–73.

Kupisch, T. (2007). Testing the effects of frequency on the rate of learning: Determiner use in early French, German and Italian. In I. Gülzow & N. Gagarina (eds), *Proceedings of the workshop on input frequencies in acquisition: Defining the limits of frequency as an explanatory concept*, 83–113. Berlin: Mouton de Gruyter: Studies on Language Acquisition (SOLA) **32**.

MacWhinney, B. (1991). *The CHILDES project: Tools for analyzing talk*. Hillsdale, NJ: Erlbaum.

Morgan, J. & Demuth, K. (eds) (1996). *Signal to syntax: Bootstrapping from speech to grammar in early acquisition*. Mahwah, NJ: Lawrence Erlbaum Associates.

Murphy, G. L. (2004). *The big book of concepts*. Cambridge, MA: MIT Press.

Pine, J. M. & Lieven, E. V. M. (1997). Slot and frame patterns and the development of the determiner category. *Applied Psycholinguistics* **18**, 123–38.

Pine, J. M. & Martindale, H. (1996). Syntactic categories in the speech of young children: The case of the determiner. *Journal of Child Language* **23**, 369–95.

Radford, A. (1990). The syntax of nominal arguments in early child English. *Language Acquisition* **1**, 195–223.

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 alternate account. *Journal of Child Language* **28**, 127–52.

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

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.

Tomasello, M. & Stahl, D. (2004). Sampling children's spontaneous speech: How much is enough? *Journal of Child Language* **31**, 101–121.
Valian, V. (1986). Syntactic categories in the speech of young children. *Developmental Psychology* **22**, 562–79.
Valian, V. (1991). Syntactic subjects in the early speech of American and Italian children. *Cognition* **40**, 21–81.
Waxman, S. R. & Booth, A. E. (2001). Seeing pink elephants: Fourteen-month-olds' interpretations of novel nouns and adjectives. *Cognitive Psychology* **43**, 217–42.
Zamparelli, R. (2000). *Layers in the determiner phrase*. New York: Garland.

APPENDIX 1

CHILD AGES AND MLUS, AND FREQUENCIES OF UTTERANCES, DETERMINER TOKENS, AND DETERMINER TYPES FOR CHILD AND MOTHER

			Child			Mother		
Child	Child age	Child MLU	Utterances	Det tokens	Det types	Utterances	Det tokens	Det types
1	2;1	1·53	846	25	6	846	403	28
2	1;10	1·74	979	122	8	979	385	19
3	2;2	1·79	822	35	6	822	276	21
4	1;10	1·81	1115	93	5	1115	482	23
5	2;2	1·99	525	76	9	525	195	20
6	2;3	2·24	492	92	10	492	191	19
7	2;3	2·28	616	42	6	616	175	21
8	2;7	2·52	1005	153	13	1005	410	27
9	2;8	2·66	646	70	12	646	257	24
10	2;5	2·76	732	171	12	732	324	23
11	2;3	3·07	881	184	14	867	193	22
12	2;6	3·15	593	180	11	526	182	20
13	2;3	3·16	803	143	15	803	222	21
14	2;5	3·31	831	190	13	779	319	26
15	2;5	3·34	807	218	21	766	254	27
16	2;5	3·62	710	150	14	710	231	20
17	2;5	3·68	706	132	13	706	231	20
18	2;5	3·72	563	108	11	563	175	20
19	2;8	4·12	549	122	14	549	174	21
20	2;8	4·17	771	223	18	771	329	28
21	2;6	4·38	1043	262	15	1043	336	24

APPENDIX 2

FREQUENCIES CONTRIBUTING TO OVERLAP MEASURES (NOUNS WITH OVERLAP/TOTAL NOUNS)

				Stratification by number of times noun occurs with a determiner						Matched	
	Child MLU	*a/the*	All determiners	1	2	3	4	5	6+	All nouns	2+ nouns
Children											
1	1·53	0/9	1/20	0/16	0/3	1/1	0/0	0/0	0/0	1/13	1/3
2	1·74	9/65	13/71	0/50	3/9	3/4	1/2	3/3	3/3	11/43	11/12
3	1·79	1/18	2/25	0/18	2/6	0/0	0/0	0/1	0/0	1/18	1/5
4	1·81	3/20	9/38	0/20	4/7	3/6	0/2	0/0	2/3	4/28	4/9
5	1·99	5/20	12/28	0/13	4/6	2/2	2/3	1/1	3/3	10/20	10/12
6	2·24	2/34	7/47	0/28	2/7	3/7	0/1	1/1	1/3	0/25	0/6
7	2·28	0/12	4/23	0/12	2/8	0/1	0/0	1/1	1/1	3/13	3/6
8	2·52	4/60	14/78	0/55	5/11	0/2	2/2	4/4	3/4	14/60	14/20
9	2·66	3/36	8/47	0/35	4/5	3/4	0/2	1/1	0/0	3/21	3/5
10	2·76	13/67	19/77	0/43	7/16	1/5	3/5	3/3	5/5	8/52	8/17
11	3·07	9/64	18/94	0/55	8/19	3/12	2/2	1/1	4/5	10/47	10/17
12	3·15	2/78	14/100	0/58	7/22	2/12	2/4	1/2	2/2	5/51	5/10
13	3·16	5/57	20/80	0/46	9/19	5/8	3/4	1/1	2/2	9/46	9/15
14	3·31	10/78	23/101	0/55	11/24	6/13	4/6	0/1	2/2	12/52	12/21
15	3·34	9/53	26/87	0/46	7/15	4/7	5/8	0/0	10/11	11/58	11/18
16	3·62	7/51	19/76	0/45	8/18	5/5	3/3	1/3	2/2	10/45	10/14
17	3·68	2/49	16/76	0/43	7/21	5/7	3/3	0/0	1/2	7/50	7/14
18	3·72	3/32	10/45	0/27	4/9	0/2	1/1	0/0	5/6	3/28	3/8
19	4·12	4/29	14/42	0/17	5/11	3/6	1/1	0/2	5/5	9/32	9/15
20	4·17	8/88	22/113	0/73	5/18	5/6	5/7	3/3	4/6	16/66	16/21
21	4·38	11/73	32/91	0/41	9/17	5/8	4/7	5/5	9/13	23/64	23/32
Mothers											
1		12/135	45/178	0/102	13/30	11/19	6/7	4/8	11/12	1/13	1/3
2		15/125	41/166	0/92	13/38	7/14	6/6	2/3	13/13	8/43	8/12
3		9/84	29/123	0/72	7/18	6/12	9/11	2/2	5/8	0/18	0/5
4		14/136	52/198	0/110	11/33	15/21	9/13	4/5	13/16	5/28	5/9
5		7/50	25/81	0/49	6/10	6/7	3/4	2/3	8/8	8/20	8/12
6		7/69	23/116	0/82	10/18	3/5	4/5	2/2	4/4	3/25	3/6
7		6/69	15/105	0/76	5/14	2/6	2/3	2/2	4/4	3/13	3/6
8		25/161	48/198	0/116	16/38	12/22	4/4	5/6	11/12	13/60	13/20
9		5/116	29/166	0/120	14/25	4/10	4/4	2/2	5/5	3/21	3/5
10		13/117	35/165	0/107	13/25	12/18	3/5	0/1	7/9	10/52	10/17
11		7/62	26/106	0/66	12/22	4/7	5/6	1/1	4/4	12/47	12/17
12		7/72	18/104	0/72	4/14	4/7	4/5	3/3	3/3	6/51	6/10
13		7/96	27/133	0/88	9/20	9/13	5/6	4/5	0/1	8/46	8/15
14		14/118	33/171	0/106	13/32	9/16	5/10	2/2	4/5	12/52	12/21
15		13/90	38/141	0/85	15/30	10/11	6/8	4/4	3/3	12/58	12/18
16		10/70	30/118	0/74	10/17	7/9	7/10	0/1	6/7	11/45	11/14
17		3/86	34/129	0/77	13/25	10/13	5/7	4/5	2/2	10/50	10/14
18		8/56	20/86	0/56	5/10	5/6	6/9	0/0	4/5	6/28	6/8
19		3/54	21/82	0/45	4/16	5/7	6/7	2/2	4/5	9/32	9/15
20		16/121	43/167	0/106	14/27	6/9	9/10	7/8	7/7	14/66	14/21
21		18/102	43/131	0/65	12/28	6/9	7/8	4/6	14/15	26/64	26/32

Note: In each analysis column, the numerator is the number of nouns showing overlap and the denominator is the number of total nouns. The final column represents nouns occurring two or more times with a determiner.

APPENDIX 3

FREQUENCY OF TOKENS FOR WORDS CLASSIFIED AS DETERMINERS

	Number of tokens		
	Children	Mothers	Total
the	812	1957	2769
a	983	1512	2495
your	56	554	610
my	294	103	397
that	89	229	318
this	98	198	296
some	110	126	236
what	16	215	231
his	60	147	207
other, the other, my/your/ her other, that/some/ these/a few other, which/ what other, a other	58	85	143
her	34	74	108
those	25	83	108
Cardinal numbers	31	66	97
two	*26*	*31*	*57*
three	*3*	*15*	*18*
four	*1*	*3*	*4*
five		*3*	*3*
six		*4*	*4*
ten	*1*	*3*	*4*
twelve		*1*	*1*
fifteen		*1*	*1*
thirty two		*1*	*1*
fifty		*2*	*2*
five hundred		*1*	*1*
a thousand		*1*	*1*
these	19	75	94
another	38	47	85
any	11	36	47
one	4	40	44
no	11	26	37
their	10	27	37
our	6	26	32
more	15	16	31
whose	1	21	22
which	2	15	17
all		13	13
how many		13	13
every		8	8
many	2	5	7
a little		5	5
too many	1	4	5
a few		4	4
its		3	3

APPENDIX 3 (*Cont.*)

	Number of tokens		
	Children	Mothers	Total
most	2	1	3
a couple		2	2
enough	1	1	2
last	1	1	2
much		2	2
every other		1	1
fifth		1	1
several		1	1