Lexically-based learning and early grammatical development

ELENA V. M. LIEVEN, JULIAN M. PINE and GILLIAN BALDWIN

Journal of Child Language / Volume 24 / Issue 01 / February 1997, pp 187 - 219
DOI: null, Published online: 08 September 2000

J. Child Lang. **24** (1997), 187–219.

Lexically-based learning and early grammatical development*

ELENA V. M. LIEVEN
University of Manchester

JULIAN M. PINE
University of Nottingham

AND

GILLIAN BALDWIN
University of Manchester

(*Received* 10 *January* 1995. *Revised* 28 *June* 1996)

ABSTRACT

Pine & Lieven (1993) suggest that a lexically-based positional analysis can account for the structure of a considerable proportion of children's early multiword corpora. The present study tests this claim on a second, larger sample of eleven children aged between 1;0 and 3;0 from a different social background, and extends the analysis to later in development. Results indicate that the positional analysis can account for a mean of 60% of all the children's multiword utterances and that the great majority of all other utterances are defined as frozen by the analysis. Alternative explanations of the data based on hypothesizing underlying syntactic or semantic relations are investigated through analyses of pronoun case marking and of verbs with prototypical agent–patient roles. Neither supports the view that the children's utterances are being produced on the basis of general underlying rules and categories. The implications of widespread distributional learning in early language development are discussed.

[*] An earlier version of parts of this paper was given at the Stanford Child Language Research Forum, 1995, and is published in the Proceedings. We would like to thank all the families who took part in this study and to acknowledge gratefully the assistance of Tameside Regional Health Authority in collecting the sample. Thanks are also due to Helen Dresner Barnes who was centrally involved in the original data collection. The research reported here was partly funded by The University of Manchester Research Support Fund and partly by the ESRC, grant number: R000221285. Address for correspondence: Elena V. M. Lieven, Department of Psychology, University of Manchester, M13 9PL, UK.

INTRODUCTION

The problems encountered with the semantic relations approach to early structure have led to increasing emphasis on distributional learning as exemplified by Maratsos (1988) on the one hand and innate grammatical knowledge on the other (Valian, 1986, 1991; Radford, 1990). Recent work demonstrating the possibilities of computational approaches to categorization (Elman, 1990; Finch & Chater, 1992) and morphology (Plunkett & Marchman, 1991) have also suggested that distributional regularities between words and morphemes may play a more significant role in governing how the child starts to generate productive utterances than has previously been thought. However, arguing in principle that children can use distributional information in the input to work out language structure is not the same as a developmental account showing how they do so. The present study attempts to contribute to such an account.

Early productivity

There are three contrasting approaches to the question of how children structure their early multiword utterances: the semantic relations, the syntactic and the distributional. The semantic relations position sees the underlying categories of children's speech as arising from universal cognitive underpinnings. Thus the child starts by developing lexical categories such as person, animate agent or concrete object. In interaction with the language heard s/he starts to use these categories in rule-governed ways for generating utterances. Evidence for this should come from patterns of under- and over-marking both of lexical categories (e.g. noun and verb) and of case relations (subject, direct object) and from regularities of word order (i.e. English-speaking children should always place agents before actions and actions before objects acted upon). However, an account based centrally on early semantics has faced increasing problems as it has become apparent (1) that while children do seem to use narrower classes than those of the adult grammar, these often do not appear to be based on some simple, unitary and universal cognitive concept; (2) that children's morphological marking seems to observe distinctions which have no obvious semantic bases in their early word classes (e.g. noun gender in German (Mills, 1986), French (Karmiloff-Smith, 1979) and Spanish (Pérez-Pereira, 1991) and noun/verb distinctions in Hebrew (Levy, 1988)) both of which appear to have a formal rather than a semantic basis; and (3) that children do not appear to start with prototypical semantic categories (e.g. of agent) and expand these into more syntactic categories (e.g. SUBJECT, Bowerman, 1990). These problems have led the field in two directions, one a re-emphasis on very early, formal, syntactic abilities and the second an emphasis on lexically-based distributional learning.

Although there is much disagreement among those suggesting an early and

innate basis for formal aspects of syntax, there seems to be consensus amongst linguists in the Universal Grammar tradition that knowledge of X-bar structure is innate. In Radford's (1990) theory this takes the form of arguing that as soon as the child has developed lexical categories, s/he knows about X-bar projection, i.e. the ordering of adjuncts, complements and specifiers in the language; movement rules have to await the development of functional categories and the I- and D-systems. Valian (1991) too suggests that there is innate knowledge of clause structure and that this gives rise to early and correct use of subjects except where constrained by performance factors. Both theories also depend on the rapid, early and errorless learning of lexical categories such as verb and noun. While the details of these theories lie beyond the scope of the present study, neither would predict that children's early utterances were organized either on the basis of any underlying semantic categories or on the basis of distributional learning of frames around particular words.

An approach which emphasizes distributional learning suggests that the passive registration of distributional regularities between words in utterances may play a more important part in language learning than has been recently thought to be the case. There is considerable evidence to support this in relation to morphological learning where there have been strong cases put forward for a purely distributional approach to learning, at least initially, though this is, of course, hotly contested from the structuralist position. In other words, the child may not initially understand either the meaning of the morphemes s/he is starting to insert nor their grammatical relationship with other related morphemes, either syntagmatic or paradigmatic. In extending these ideas to the ordering of words in utterances we have suggested that the establishment of frames with slots in them may account for more of early multiword structure than is usually accepted. Which frames are learned and which slots develop in them is accounted for by what is salient to the child i.e. (*a*) what it is that s/he wishes to talk about and (*b*) his/her knowledge of the regularities between utterances and the contexts in which they are spoken. This approach sees lexical categories and other grammatical phenomena as emergent from these regularities rather than as necessarily arising either from pregiven knowledge of grammatical structure or from semantically transparent cognitive categories. In part this position is methodological: how much of early multiword structure can be accounted for on the basis of distributional predictability around particular lexical items without having to infer any posited underlying structure? But answering this methodological question focuses on the possibility that the methods by which children generate their early utterances may be on the basis of low-scope, idiosyncratic formulae the relation of which to adult grammar is given by regularities in the child's environment and input, which of course reflect the structural characteristics of the language being learned.

Structure in multiword utterances

The fundamental distinction between multiword utterances in children's corpora is on the basis of whether they are wholly rote-learned or are novel combinations. Beyond this, there is the question of how the novel combinations are put together:

(1) Rote learned utterances: the methodological issue with these is to distinguish them from novel combinations. Until relatively recently this was not done with any systematicity, coders using their intuitions and largely confining the coding of rote-learned utterances to 'stock phrases' e.g. *here you are; ready*, *steady*, *go*. This has resulted in a potentially large underestimate of children's rote-learned phrases which in turn can make the task of identifying any structure in the novel combinations more difficult. More recently, Plunkett (1993) has used a prosodic analysis to distinguish rote-learned from segmented utterances and Lieven, Pine & Dresner Barnes (1992) developed a coding scheme which defined utterances as rote-learned on the basis of the existence or otherwise of the constituent words in the child's single-word lexicon and any positional regularities between the same word in different utterances.

(2) Novel combinations where there is no intrinsic relation between the words: Veneziano, Sinclair & Berthoud (1990), Lieven & Pine (1992) and Tomasello (1992) all suggest that children juxtapose words before there is evidence of any relation between them other than that explicitly generated by the discourse or that the child is thinking of them in that order. Some examples of discourse-generated utterances are where the first word is a vocative, an instruction or an answer to a previous question (e.g. *mummy book* [Mummy, I want a book]; *yes bird* [in answer to: *Is he singing*?]).

(3) Novel combinations where there is some relation between the words:

(i) Lexically-based patterns: These are productive positional patterns in which a constant item (e.g. *more*, *theres a*, *I want*) occupies a constant position in relation to variable items with which it is combined. Braine (1976) suggested that positional associative patterns (a group of similar rote-learned phrases) might give rise to positionally productive patterns through the child 'noticing' the regularities that united the group. The question of whether this will result in a new slot inside an old frame or the creation of a slot on one or other side of a word will presumably depend on the size of the initially segmented unit and its relationship to other regularities in the input (Peters, 1997). Most authors recognize the existence of these patterns but differ greatly in how much significance they give to them. For some, their significance lies only in the possibility that they may suggest that a child is more advanced syntactically than is actually the case and they have no developmental importance (Radford, 1990). On the other hand Braine (1976) sees them as crucial evidence that many children's early combinations are

based on patterns of much lower semantic scope than is suggested by a semantic relations approach and he suggests that this provides a more realistic basis for deciding how the child moves on to further stages of grammatical learning.

In fact Braine's approach is partly semantic and partly distributional. While he suggests that much generativity arises from the positional regularity of lexical items and he certainly argues forcibly against any need to posit general syntactic categories to explain the regularities in children's corpora, he organizes the children's utterances into semantic groups and leaves open the categorial status or otherwise of the words that fill the open slot in juxtaposition to the fixed lexical item. He does, however, point out the idiosyncrasy of these slots in different children's speech and the difficulty, therefore, of producing a general theory of semantic relations from them. Clearly there is a certain semantic coherence to the words that fill the slots, and often, too, they belong to one syntactic (lexical) category, but it is an open question whether this seeming systematicity is produced simply by the nature of what the child wishes to talk about or, alternatively, by an underlying organization of words into categories. Indeed it is possible that at any one point in a child's development both these descriptions could be true of different parts of her/his corpus and certainly that over time the surface juxtaposition of words by the child could develop into the emergence of a putative category.

This then raises the question of the place played by these patterns in the development of structure. Once a positionally associative pattern becomes productive, it is possible to investigate the relationship between different items placed in the same slot. These might have more or less semantic or syntactic coherence and this, in turn, would give some indication of the child's development of grammatical categories. Thus the analysis of a slot can act as a diagnostic of grammatical development.

(ii) Utterances which are structured on some rule-governed basis: as stated above, the contrast between this and the lexically based approach is that here the child is said to be generating utterances on the basis of underlying categories and rules for combining them. Both syntactic and semantic approaches start from this position though they differ markedly in what they suggest the underlying categories and rules to be. The critical difference from the lexically-based approach is that in these approaches two categories are combined rather than the combination occurring between a fixed lexical frame and other lexical items. This generativity is central to both the semantic and syntactic approaches, though in one, the categories derive from cognitive-lexical development and in the other, they derive from innate grammatical knowledge. On either approach, however, one would expect to see little systematicity accounted for by lexical items in themselves. Of course, rote-learned and semi-rote-learned formulae are acknowledged to

exist in both approaches and many child language researchers have discussed the existence of pivot-type formulae (Brown, 1973; Bloom, Lightbown & Hood, 1975) but it is essential to both approaches that these types of utterances reflect the earliest and most primitive stages of language development and, moreover, that they are not involved in the child's developmental path to language. Demonstrating that lexically-based positional patterns could account for a major proportion of children's multiword utterances up to the point where they started to produce large numbers of three- and four-word utterances, would, we suggest, present serious problems for approaches based centrally on the idea either of general semantic or of general syntactic categories and rules.

In the first part of this study we analyse the data from a purely distributional point of view, asking how many of each child's first 400 multiword utterances fit the first 25 positional formulae which are productive (by our definition), that is how much of the data can be accounted for by the first 25 productive patterns in which one constant word has a fixed position. Since all these utterances will conform to the word order of some fragment of the adult language, it will of course, always be possible to argue that any positional consistency can be accounted for by the interaction between underlying rules and the limitations of the child's lexicon. In the final analysis, this may not be an empirical matter since it may turn on different philosophical interpretations of what is a parsimonious explanation, but we explore it further by contrasting the goodness of fit of a positionally based analysis with an analysis based on the idea that, in principle, more general underlying categories and rules should predict less positional ordering of particular lexical items and should produce the appearance of more freedom of ordering of particular words. In other words if the child has underlying rules which either order agents, actions and patients or specifiers, heads and complements, there should be less positional consistency at the lexical level and particular words should appear in a more varied set of positions. However, since the consistent positioning of particular lexical items cannot, of course, rule out that this is caused by underlying structure of more general scope, we subsequently undertake two further analyses which attempt to examine each potential level of structural organization (semantic and syntactic) in more detail:

(*a*) The syntactic approach: here predictions differ substantially depending on the theorist. Valian in her 1986 paper argues that children develop very early the lexical categories of determiner, noun, noun phrase and adjective and that they use these independently of any semantic underpinnings. In her 1991 paper she suggests that children have an early knowledge of X-bar structure which is manifested in the correct use of SUBJECT. Where there is incorrect usage, either errors or omissions, this can be accounted for by performance limitations. Radford's position can be summarized as claiming

that (i) children innately know the correct ordering for specifiers, adjuncts and complements in the X-bar phrase and (ii) that the child's earliest system is a lexical grammar without the syntax provided by the emergence of the I-, D- and C-systems which only come in later. The presence of inflections, determiners and complementizers in the early stages is accounted for by formulaic or partially-formulaic learning. However, there is a clear general prediction from both approaches that where a syntactic category is hypothesized to be already present, the child should be able to use this category or rule in a flexible and highly generative way, limited only by what words are in the lexicon. Thus if the determiner system is present, the child should be able to use any determiner present in its vocabulary with any appropriate noun also present. If the child already has the relation of Subject and the category of pronoun, then s/he should be able to use all pronouns correctly inflected for Subject and Direct Object positions. Evidence for restricted usage and for lack of full grammatical flexibility would count against these positions.

(*b*) The semantic approach: if children start to generate their multiword utterances on the basis of underlying semantic categories e.g. agent (or actor) + dynamic action; action + patient; agent (or actor) + action + patient; only subsequently developing these categories into the more formal categories of adult grammar by a process of semantic assimilation, we would expect: (i) that children's use of nouns in initial position should be restricted initially to those fulfilling a prototypical agent or actor role with respect to the semantics of the verb and (ii) that utterances containing a verb and two arguments should first only occur with prototypical action verbs and only later appear with non-prototypical verbs.

The present study seeks to investigate these issues using:

(1) A quantitative comparison between an analysis based solely on distributional regularities in the positioning of particular lexical items and an analysis which allows more freedom in the placing of specific words.
(2) A quantitative analysis of subject and direct object marking on pronouns.
(3) An analysis of all the children's utterances which have a Subject–Verb–Direct Object structure with a view to comparing syntactic, semantic and lexical accounts of these structures.

METHOD AND PROCEDURE

Subjects

Subjects for the present study were 12 children (six male and six female; six first-born and six later-born) whose families were asked to participate with

their children in a study of their 'language development'. These are the same subjects whose earlier language development is reported in Lieven *et al.* (1992) where fuller details of their recruitment and the method of conducting the recordings may be obtained. With the exception of one child, the children all came from lower middle-class or working-class backgrounds. Data collection started when each child was approximately twelve months old and continued until each child reached 3;0, except for one child who was lost from the sample at 2;3 due to a family bereavement. Most of the analyses in the present study were, therefore, conducted on 11 children.

Data collection

Parents were instructed in how to keep a continuous record of their child's speech, recording each new instance of a spontaneous word or phrase. Visits to the family took place every three weeks and were supplemented in the early stages by weekly telephone calls to see how record keeping was progressing. Audio recording started when the child had a lexicon of 20 different utterance types (single words or phrases) and continued at six-weekly intervals until a total lexicon of approximately 500 utterance types was reached, when recordings were increased to three weekly intervals. When the children's lexicons reached about 100 utterance types it became impossible for the parents to write down every new utterance and they were then asked to use a 'sampling diary'. This involved writing down everything that the parents thought the child had not said before in one of three time slots which covered the whole of the child's day. The time slot remained the same for one week and then changed for the next week. This ensured that the same time slot in the child's day was sampled once every three weeks. By this time in a child's development, one cannot hope to have complete coverage of all new utterance types produced but it was hoped that this method together with the data collected during research visits and tape recordings in which, by now, all the children were speaking a considerable amount, would provide a reasonably accurate reflection of the child's range of multiword utterances. MLU calculations (for words) were made using the CLAN programs of the CHAT transcription system in CHILDES (MacWhinney & Snow 1985, 1990). This, of course, does not distinguish between frozen and non-frozen utterances.

All the multiword utterance types collected using the procedures described above were entered into a computerized database in the chronological sequence in which they had occurred. This included utterances recorded on parental diary forms, utterances written down by the researchers during visits and utterances taken from the one-hour audio recordings. The present study is based on an analysis of the first 400 of these multiword utterance types for each of the 11 children.

The coding scheme

The children's multiword utterances were coded as either frozen, intermediate or constructed by using the coding scheme outlined in full in Pine & Lieven (1993). The coding scheme relies on (*a*) the positional regularity (P)[1] of particular lexical items (the putative 'frame') and (*b*) evidence that a particular lexical item is present independently (I) in the child's lexicon rather than embedded in a rote-learned phrase (i.e., is occupying a 'slot'). A multiword utterance is defined as frozen if there is no evidence of such independence for its constituent words (I−). Intermediate utterances arise from one of two sources: (i) they are the second instance of positional regularity for one of the words (P+[2]) and there is evidence of independent existence for the other word(s) in the utterance (I+) or (ii) all the words in the utterance have already occurred independently (I+:I+) i.e. they are part of the child's lexicon. Utterances are defined as constructed when they are the third instance of positional regularity of one of the words and there is evidence for the independent existence of the other word(s) in the child's vocabulary (I+:P+[3]). Subsequent instances of constructed utterances do not have to have items in the slot for which there is independent evidence in the vocabulary since, by definition, these words are now appearing in a generated utterance (P+[4]). Pine & Lieven (1993) give examples of how the coding works and analyse its validity.

Full definitions are as follows:

(1) MULTIWORD UTTERANCES: utterances which consist of more than one word (as defined in the adult language) including forms which are frozen in adult speech or in the child's, but excluding word compounds and reduplications. These are subdivided as follows:

(A) FROZEN PHRASES: utterances which contain two or more words which have not previously occurred alone in the child's vocabulary or which contain one such word, provided it has not occurred in the same position in a previous multiword utterance (I−:I−:P−) or (I+:I−:P−).

(B) INTERMEDIATE UTTERANCES: utterances in which (i) all the words or phrases have previously occurred independently in the child's vocabulary, provided none of them has occurred in the same position in two previous multiword utterances, (I+:I+:P−/P+[1]); and/or (ii) one or more of the words or phrases has occurred in the same position as in one but only one other previous multiword utterance, provided the word or phrase which makes up the remainder of

[1] In what follows, it is assumed for ease of exposition that the utterances consists of only two items. We are grateful to an anonymous reviewer for suggesting this method of symbolizing the coding scheme.

the utterance has already occurred independently in the child's vocabulary (I+:P+[2]).

(C) CONSTRUCTED UTTERANCES: utterances which (i) contain one or more words or phrases which have occurred independently in the child's vocabulary together with a word or phrase which has occurred in the same position in at least two other previous multiword utterances (I+:P+[3]) the second of which must be classed as intermediate (B) (i) or (ii) above; or (ii) conform to a positional pattern as in (C) (i) above which is already established in the child's vocabulary, regardless of whether the variable word has occurred previously (P+[4]). Once a pattern is defined as constructed it forms a 'template or frame' consisting of a particular lexical item in a fixed position with a 'slot' on one or other side of it. New instances of the pattern consist of lexical items being placed in the slot which have not been placed there before. The difference between constructed patterns is therefore in the nature of the frame (position and identity of regularly recurring word or words).

Reliability

For two of the children, all the utterances collected by the method described above were coded independently by two coders, up to the point at which 25 patterns defined as constructed was reached. Reliability expressed as percentage agreement was 99% for the first child (kappa = 0·98) and 95% for the second (kappa = 0·93).

RESULTS

The children reached the 25 pattern point at ages ranging from 1;8 to 2;8 and at a MLU (words) with a range of 1·41 to 3·75 (see Appendix A). The time it took them to get from vocabularies of 100 words to 25 patterns ranged from 3 to 9 months. The fact that MLU at 25 patterns can be as high as 3·75 with a mean of 2·41 provides a good test for the idea that these children are constructing their utterances more from lexically-based, low-scope frames than on the basis of underlying semantic or syntactic categories, since most previous research assumes that MLUs this high will not include a large number of formulaic or semi-formulaic phrases but do not systematically exclude any but the most obvious of these (e.g. *here you are*, *thank you*).

Table 1 shows the numbers of frozen phrases, intermediate and constructed utterances up to the point at which each child had 25 constructed patterns. At this point the children had a mean of 300 multiword utterances (range 207–402) including those coded as frozen. What proportion of these multiword utterances fit one of the first 25 patterns? Table 2 shows that the mean is 60·3% with a range of between 51·3% and 71·9%. Given the way the coding scheme operates each of the 25 patterns defined as constructed have

TABLE 1. *The number of frozen, intermediate and constructed utterances for each child at 25 patterns*

Child	Frozen	Intermediate	Constructed	Total
1	72	44	91	207
2	166	57	108	331
3	81	49	81	211
4	198	45	159	402
5	87	43	100	230
6	185	65	89	339
7	140	50	115	305
8	179	43	92	314
9	206	77	124	407
10	174	54	116	344
11	84	37	93	214
Mean	143	51	106	300
Range	[72–206]	[37–77]	[81–159]	[207–407]

TABLE 2. *The number of utterances accounted for by the first 25 constructed patterns*

Child	Utterances fitting a pattern (inc. frozens) (%)	Utterances not fitting a pattern but classed as frozen (%)	Columns 1+2 combined	Remainder (%)	Utterances not fitting and 'juxtaposed' (%)
1	68·1	22·2	90·3	9·7	6·8
2	57·1	33·8	90·9	9·1	6·0
3	64·9	24·6	89·5	10·5	7·1
4	54·7	39·8	94·5	5·5	1·5
5	70·9	21·7	92·6	7·4	3·5
6	51·3	38·1	89·4	10·6	5·3
7	61·3	30·8	92·1	7·9	3·3
8	55·2	38·5	93·7	6·3	3·5
9	53·3	35·4	88·7	11·3	5·7
10	61·3	31·7	93·0	7·0	3·8
11	65·4	27·1	92·5	7·5	1·9
Mean	60·3	31·2	91·6	8·4	4·4

by definition to be preceded by at least two others of the same lexical-positional consistency and this would mean that at least 75 multiword utterances for each child must fit one of their 25 patterns. However, this is only 25% of the mean (300) and thus there are a substantial additional number of utterances which are consistent with one or other of the 25 patterns. Given that there must be at least some of these utterances which are definitely frozen (as well as being defined as such by the coding scheme) and

TABLE 3. *The number and percentage of productive positional patterns for which additional instances were found*

	+1		+2	
Child	Number	Percent	Number	Percent
1	22	88	18	72
2	22	88	16	64
3	20	80	17	68
4	21	84	16	64
5	24	96	19	76
6	17	68	15	60
7	20	80	18	72
8	22	88	17	68
9	18	72	14	56
10	18	72	15	60
11	19	76	15	60
Mean	20·3	81·1	16·4	65·5

that there will be other patterns which may well be on the way to productivity, this is suggestive evidence for substantial organization around lexical frames.

Further evidence for the proposal that many of the 25 patterns identified by the coding scheme are playing a functional role in generating the children's utterances is provided by looking forward to see how many further instances of them occur. This is done by examining the recording session after that in which 25 patterns is reached. Table 3 shows that an average of 81 % of these patterns have one further instance and 65 % have two.

The overwhelming proportion of utterances which do not fit one of the 25 patterns are classified as frozen by the coding scheme. This means that there is no evidence for the words that make up these utterances occurring as independent vocabulary items. If such evidence existed, there would be a much higher proportion of non-positionally-based utterances in the intermediate category. If the coding scheme is correct in defining these utterances as frozen, the proportion of utterances accounted for rises to a mean of 91·6 %, leaving only 8·4 % on average which neither fit a pattern nor are defined as frozen. On average about half of these consist of words in the child's vocabulary which are combined but show no evidence of positional consistency and about half show positional consistency (i.e. may be on the way to forming a lexical frame).

These findings support and extend to a later point in development, the study by Pine and Lieven (1993) which showed that an analysis in terms of each child's first 10 low-scope patterns could account for a large proportion of their multiword utterances. Of course, the coding scheme provides a

purely mechanical method of defining frozen utterances and productive slot-and-frame patterns and we would not want to claim psychological plausibility for every definition of an utterance as frozen or productive. Instead we suggest that it is instructive to see how much of the data can be accounted for by such a mechanical procedure and that the result of this should lead us to consider seriously an analysis in terms of low-scope organization and raises some doubt over whether any more abstract characterization of children's speech at this stage of development is required.

However, objections to these conclusions might be twofold. First, it could be argued that many of the utterances defined as frozen in our analysis are actually the product of more powerful underlying rules and that we are, therefore, seriously underestimating the child's grammatical knowledge and abilities. At the limit this position becomes irrefutable: provided the rules are general enough, a corpus of two- and three-word utterances will always provide positive evidence of conforming to the structure of the adult language, however defined. But the empirically important issue is how to decide which utterances are frozen or partially frozen and which not. Without an explicit method for systematically excluding rote-learned utterances from a consideration of the basis for underlying generativity, it is difficult to have confidence in any hypothesised underlying structure.

It is, of course, true that the difficulties with our coding scheme do become more serious for those children who reach 25 patterns after the average length of their utterances has increased well beyond MLU = 2·0. The problem is that it becomes increasingly difficult to sample fully enough to check evidence for whether words appearing in the multiword utterances have an independent existence in the child's vocabulary. And there is also the problem that, as a child's language develops, s/he may be less likely to use words which are independently available in the lexicon as single word utterances. Thus there is, indeed, the danger that utterances may be defined as frozen when they are, in fact, the results of productivity of one sort or another.

The second objection is that, despite the fact that lexical positioning accounts well for the data, it is an epiphenomenon of more abstract structural representations. Thus the 25 patterns could be seen simply as representing what the child can do with a core grammar and a limited lexicon, and, it could be argued that when the utterances accounted for by the 25 patterns are combined with the frozen utterances, evidence for the core grammar or some other deeper level of structural organization can be demonstrated. The challenge is therefore to see whether an analysis based on lexical positioning gives evidence not just of accounting for the data in a numerical sense but of methodological and psychological plausibility. This challenge and the issues we raise above concerning the status of frozen utterances suggest that we should look in detail for greater evidence of underlying organization and

TABLE 4. *Summary of pronoun use in subject and object position for all children*

	Subject position		Object position	
Pronoun	Correct	Incorrect	Correct	Incorrect
I	257	0	0	0
me	0	62	24	0
we	6	0	0	0
us	0	0	4	0
you	82	0	23	0
he	28	0	0	0
she	15	0	0	0
him	0	0	11	0
her	0	0	1	0
it	77	0	193	0
them	0	0	26	0
they	6	0	0	0

TABLE 5. *First person pronoun use*

	Subject position			Object position
Child	I	me/my for I	Error rate (%)	me
1	8	0	0·0	1
2	17	2	5·6	3
3	19	2	9·5	1
4	12	28	70·0	2
5	14	11	44·0	3
6	27	1	3·6	6
7	24	0	0·0	2
8	6	4	40·0	1
9	70	10	12·5	1
10	51	4	7·3	4
11	9	1	10·0	0

more general rules than is indicated by the results based on a low-scope analysis.

We do this by investigating the organization of two different aspects of the children's multiword utterances, irrespective of whether they are coded as frozen or as potentially productive by our scheme: (i) pronoun case-marking for subject and object and (ii) the verb argument structures of the children's first utterances which contain a verb and explicit subject and direct object arguments.

These analyses are conducted on the first 400 multiword utterances for each child collected either from maternal diaries, from visits to the family or on tape. While this still results in a wide MLU range for the sample

(1·89–3·75), it does ensure that even those children who reach 25 patterns at a relatively early stage (e.g. child 1, MLU = 1·41 at 25 patterns) are producing a sufficient number of multiword utterances to make an analysis of underlying structure feasible.

Pronouns and case assignment

Table 4 summarizes the children's pronoun use when each child had 400 recorded multiword utterances, together with any errors where they are identifiable. It can be seen that first person singular is the most frequently used subject pronoun while *it* is the most frequently used pronoun in object position. These findings are broadly in accordance with Valian (1991). The only identifiable errors which occur are for first person subjects. These are on 25 % of first person subject uses and consist of substituting *me* or *my* for *I*.[2]

TABLE 6. *Child 5: utterances with* I *and* me/my *in subject position*

Age	I	Age	me/my
1;8	I draw	1;10	me go
1;8	I got it	2;4	me have some
1;8	I do it	2;4	me taste
1:11	I go	2;4	me help, Mummy
1;11	I come	2;4	me want that one Mum
1;11	I carry it	2;4	me put it back
1;11	I blow	2;4	me put in, watch, done it
2;0	I ride bike	2;4	my shut gate
2;1	I done it	2;4	my shut door
2;1	I want drink	2;2	my do it
2;1	I first	2;4	my mend it
2;1	I want do it		
2;1	I'm a baby		
2;4	I'm not		

This contrasts with Valian's findings but is very similar to other reports in the literature (Budwig, 1989; Rispoli, 1994). In Table 5 we present the use of first person singular pronouns in subject position for each child. We can see that only two children make no errors and three have extremely high error rates of 44 %, 40 % and 70 % (children 5, 8 and 4). These errors suggest that the children may not control the basic case marking for first person subject pronouns.

We now look in more detail at the children with high error rates and also

[2] The following utterances have been excluded from the analyses: *bless you*, *there you/we go*, *see you soon/later*, *here/there you/we are*, *thank you*, *here/there s/he/it is*, *pardon me*, *oh dear me*.

at the other child who produces a substantial proportion of errors (12% – child 9). The MLU (word) of these children at 400 utterances is 2·05 (child 5), 2·56 (child 8), 2·38 (child 9) and 3·75 (child 4) respectively. Starting with the child with the lowest MLU of these four, we can see that across the same developmental period he appears to use *I* or *me* at random to express first person subject. Note that on three occasions the child uses the same verb (*go*, *do*, *want*) with both *I* and *me*/*my* and that it is hard to see any semantic difference between the group of utterances used with *me*/*my* and those used with *I*.

The second child (child 8) produces only five utterances with *me* and six with *I* but they fall across the same time period and shows some overlap (*I have a go*, *me a go*, *me have some*). The third child (child 9) provides a contrast since there is some evidence of a developmental shift from the use of *me* to *I*, most uses of *me* occurring before *I* emerges in large numbers. However, this child still says *me go down*, *me have shower*, *me go to see lady* and *me pay lady* at least seven weeks after producing *I go home*, *I'll have some*, *I go to bed* and *I'll fetch Thomas*. The fourth child (child 4) is more like the first two in that her uses of *I* and *me* cover the same developmental period of 13 weeks. She is interesting in that she has more uses of *me* than of *I* but, again, there is considerable overlap in which predicates they are used with as can be seen in Table 7. It is not the case that utterances with *I* are more advanced than

TABLE 7. *Child 4: utterances which overlap* I *and* me

Age	me	Age	I
2;5	me do it	1;10	I'll do it
2;5	me like new car Daddy	2;5	I like ice cream, Mummy
2;7	me go in Carlton	2;8	I gonna go on jigsaw
2;7	me have it potatoes	2;7	I have one

those with *me*, the child using both pronouns with the past tense (*I finished*, *me fell*) and with 'complex' verb constructions (*I gonna go*, *me want to do*).

None of the children have more than six utterance types with *me* in direct object position, the figures for the four children detailed above being three, one, one and two utterances respectively (see Table 5 above). This is important because by focusing on the children's use of the subject position, there is a danger of forgetting that they are not only using the wrong pronoun in this position but are in some cases using this pronoun (*me*) more in subject position, incorrectly, than in object position, correctly. This makes it difficult to conclude that the children are operating with a clear idea of contrastive case assignment for first person singular and highlights the danger of inferring an understanding of case by only looking at one category such as

subject. Rather, we would maintain that the data support the claim that much of these children's speech is still based around particular lexical items rather than on underlying categories of more general scope and we suggest that, despite their very different MLU, the children's underlying grammatical grasp may not be dissimilar.

What about the other pronouns in which case marking is detectible (*s/he*, *we*, *they; her/him us*, *them*)? The numbers are very small, but nevertheless, it is possible to ask whether those children who show evidence of a pronoun in one case also show evidence of it in the other. If we take the six children who produced five or more utterances containing identifiably case-marked pronouns, we can see from Table 8 that only two of them showed any overlap

TABLE 8. *Nominative and accusative use of case-marked pronouns* (*excluding first person singular*) *for children with more than five instances*

		Number of instances			
Child	Pronoun usage*	he	she	we	they
1	Nominative	1	0	2	2
	Accusative	0	0	0	0
6	Nominative	4	1	0	1
	Accusative	0	0	1	5
7	Nominative	4	1	2	1
	Accusative	0	1	0	4
8	Nominative	4	1	0	0
	Accusative	2	0	0	0
9	Nominative	4	0	0	0
	Accusative	8	0	0	6
10	Nominative	2	11	1	0
	Accusative	0	0	0	6

* Usage of these pronouns was errorless. Occurrences of the same pronoun and verb were only counted once.

between nominative and accusative uses. Child 8, who has an MLU of 2·56 and a 40% error rate on his 10 first person pronouns, produces four utterances with *he* as subject (*what he say*, *Daddy he shout*, *there he is*, *he's fell on grass*) and two with *him* (*Mummy tell him*, *find him now*). Child 9, the other child with overlap, uses *them* six times without using *they* and has six utterances with *he* (three of them containing *he want*) and 10 with accusative *him* (three containing *get him*). There is no overlap in the period of use of these pronouns, the last instance of *he* occurring one month before the first utterance containing *him*. All the utterances with *him* refer to this subject's baby brother who was born two weeks before the first recorded use of *him*. While these data cannot rule out the possibility that child 9 has contrastive control of the masculine third person singular pronoun, neither can it rule

out the possibility that he has learned where to place *him* on the basis of a new set of utterances heard in a new set of circumstances but without reference to his past use of *he*.

TABLE 9. *Child 9: use of third person singular pronoun*

Age	he	Age	him
1;7	he want to get out	1;10	[I'll] see him.
1;7	he want go home	1;11	get him
1;7	he want go out	1;11	Percy get him
1;8	he at home	1;11	I'll get him
1;8	he down	1;11	found him
1;9	he stuck	1;11	pick him up
		2;0	kiss him
		2;0	no, give him back, Adam
		2;0	I'll splash him
		2;0	dog barking wake him up Mummy
		2;0	I'll play with him
		2;0	don't shout at him George

To summarize the argument so far, despite the fact that all the children produce some first person singular subject pronouns and most produce a considerable number, they also produce errors by using the direct object pronoun in subject position. While they do not show similar case errors for the other overtly marked pronouns in subject position, they do show a patchy and non-systematic pattern of production. Both these results suggest that we cannot reliably conclude that these children have any overall grasp of the case assignment system for subject and object pronouns. This is despite the fact that some of these children have quite sophisticated-looking speech with auxiliaries on the *I* (*I've*, *I'll*), double verb constructions (*I gonna go*, *I want to* +*x*) and various verbal inflections (*fell*, *finished*, *barking*, *going*). Many of these children also use their own name in subject position to refer to themselves and they also frequently leave out the subject. Other researchers have suggested explanations for these phenomena (Budwig, 1989; Rispoli, 1994) but, for the purposes of assessing the children's underlying grammatical knowledge, the conclusion would appear to be that while they know that they can refer to themselves in any of these ways they have no grasp of the special grammatical status of the first person pronoun nor of the grammatical relationship between subject and direct object argument positions on the verb. In distributional terms, these children know that *I* refers to themselves and know where to place it in the utterance as they have never heard it in direct object position and do not place it there. Their use of *me* is much rarer and they have not yet worked out its position – after all they have evidence from the input that it can come both before and after the verb (*give it to me*; *let me see it*). The other subject and direct object pronouns

TABLE 10. *First 20 utterances with subject, verb and direct object structure*

	Prototypical			Non-prototypical		
		Age at:			Age at:	
Child	*N*	first utt.	last utt.	*N*	first utt.	last utt.
1	12	1;5	1;9	8	1;6	1;11
2	12	2;1	2;3	8	1;1	2;3
3	4	1;1	2;0	16	1;9	2;1
4	12	1;1	2;7	8	2;5	2;7
5	11	1;8	2;5	9	1;8	2;4
6	5	1;9	2;5	15	1;10	2;3
7	7	2;5	2;11	13	2;5	2;11
8	12	2;2	2;7	8	1;10	2;5
9	7	1;8	1;1	13	1;8	1;10
10	4	2;2	2;4	16	1;8	2;7
11	13	1;3	1;1	7	1;8	1;10

suggest the same conclusion: the children show positive evidence of having picked up their correct position from the input but not of understanding the relationship between them.

It may be, however, that children are organizing their utterances on the basis of the combination of underlying semantic categories rather than either in terms of syntactic knowledge or the use of lexically based templates of limited scope. To investigate this we consider the issue of semantic prototypicality for those of the children's utterances in which both the verb and two arguments are expressed.

Prototypical verb argument structures

Here we investigate the proposal that verbs for which agent and patient roles are canonically mapped into subject and direct object position might appear earlier with both argument slots filled than would other types of two-argument verbs (i.e. verbs that have themes, locations or goals as their arguments). Bowerman (1990) discusses this suggestion in relation to Pinker's hypothesis of innate linking rules (Pinker, 1984), but it has also been made in a similar form by researchers who suggest that children might arrive at syntactic relations by bootstrapping from semantics and that the basic agent–action–patient scenario would provide a good basis for this (Schlesinger, 1982; Slobin, 1982). In the following analysis we use Bowerman's definitions of prototypical agent–patient verbs on the one hand and other two-argument verbs on the other, to see if there is any evidence for the earlier emergence of the former in utterances with two arguments:

Prototypical agent-patient verbs:

(*a*) Strings expressing the causation by an agent of a change of state or location, with verbs such as *open*, *close*, *break*, *fix*, *put away*, *throw away* and *pick up*;
(*b*) Utterances with kinetic verbs expressing events in which the agent acts on the patient in a physically obvious way; e.g. *push*, *wash*, *bite*, *eat*, *tickle*, *spank*, and *get* (in the sense of *grab* or *take*).

Other two-argument verbs:

(*a*) Utterances with Theme-subject verbs and Location/Source/Goal subject verbs (e.g. *have*, *got*, *get*, *lost*);
(*b*) Stative transitive verbs (e.g. *see*, *hear*, *like*, *need*, *scare*);
(*c*) Non-prototypical agent–patient strings expressing events in which the agent is relatively static (*hold*, *read*, *ride*) or does not control the event (*drop*, *spill*) or does not act physically on the object (*talk* (*to*), *say*, *look at*) or does not act on the object in any specific way (*find*, *play* (*with*)) or in which there is no pre-existing object (*draw*, *write*, *make*) (Bowerman, 1990: 1273).

We start by examining the first 20 utterances for each child that contain a verb together with subject and direct object arguments (see Table 10). Each utterance was coded either as prototypical or non-prototypical using Bowerman's definitions. Utterances with particles/prepositions (i.e. *put X in*, *put in X*, *take X out*, *take out X*, *go to X*) or in which particles/prepositions had been omitted were excluded (*go school*, *put biscuit*) as were utterances in which it was unclear whether the postverbal noun should be considered as the indirect or direct object. There is no evidence for the earlier emergence of prototypical agent–patient structures in these children's utterances on a Wilcoxon test ($W = 22$, $N = 11$ n.s.). Indeed if we exclude utterances with *do it* from the prototypical group on the grounds that *do* might stand for both prototypical and non-prototypical verbs, this becomes even clearer. A further issue is that in many of these utterances the direct object is *it* and this is true for both prototypical and non-prototypical verbs. Again, though it might be argued that this pervasive use of *it* suggests that the children do not have real control of the direct object position, removal of utterances with *it* in direct object position would not yield an advantage to utterances with prototypical agent-patient relations (see Appendix B for a full list of these utterances). The reasons for this become clear if we look back at earlier utterances containing a verb and a single argument in direct object positions that the children were producing. Again there are a number of *Verb + it* utterances but they appear in both columns. However, we can also see where

many of the other non-prototypical two-argument utterances in Table 10 are coming from, since many of them are obviously related to the Verb–Direct object constructed patterns in Table 11 (see Appendix C for a full list of these

TABLE 11. *First 10 utterances with verb and direct object structure*

	Utterances with verb and direct object structure			
Child	Number of prototypical	Number of non-prototypical	Age at last utterance	Constructed patterns
1	6	4	1;9	x+it drink+x
2	5	5	2;3	x+it
3	7	3	1;10	want+x x+it x+that
4	4	6	2;6	—
5	3	7	1;11	want+x x+it x+Daddy
6	5	5	1;10	x+it want+x
7	5	5	2;1	x+it
8	4	6	2;2	want+x
9	3	7	1;10	x+it
10	1	9	1;10	see+x want+x
11	5	5	1;8	—

utterances). Thus many of the children have a constructed pattern of *want+x*, which subsequently becomes *I/me/my want+x*; and there are also *draw+x*, *see+x* and *get+x* patterns which reach the criterion of 'constructed' according to the definitions outlined in our coding scheme. Again this suggests that it may be more parsimonious at this stage of these children's development to think of them as building up patterns based on specific lexical items rather than as generating utterances on the basis of any wider scope underlying rules and categories of either a semantic or syntactic nature. Clearly this raises the question of when such underlying structure does start to develop, by what means it develops and how it should be characterized. We turn to these issues in the next section.

DISCUSSION

In this paper we have applied a distributional analysis to the utterances of a group of children whose language as measured by MLU and vocabulary range is considerably more developed than those studied in Pine & Lieven

(1993) and who would be expected by most child language researchers to be well beyond the phase in which pivot-type formulae could form a central feature of their language. The application of the coding scheme suggests that, in fact, the building up of distributional patterns around specific lexical items may be the best way of characterizing many of these children's utterances, even at a mean MLU of 2·41. However, by the time children reach 25 constructed patterns and have a total of 400 different multiword utterances, it becomes increasingly difficult to be sure that utterances defined as frozen really are being learned as unanalysed wholes as opposed to having words in them which, though the child has full control over them, have not been written down in the diaries or heard by the researchers as independent items. For this reason, it was considered important to see if there was other evidence suggesting that the children were operating with more sophisticated language than was apparent through the use of the coding scheme.

The first of these analyses involved pronoun case-marking. Here we saw that though all children use *I* correctly, nine of the 11 also use *me* or *my* for *I*, while only six use *me* in object position more than once and even for these children the numbers are relatively small. The other case-marked pronouns are used correctly but sporadically, often without subject–object contrast. We concluded that there was no evidence for systematic control of case-marking across the pronoun system for any of the children, despite the extensive use of at least some pronouns by most of them. In this case it is not possible to conclude that underlying competence is being disguised by performance considerations since the errors in pronominal first person singular are so in evidence.

Secondly, we investigated the idea that children's utterances which expressed a verb and two arguments might, early on, be underpinned by an underlying semantic rule based on the prototypical or highly salient event structure of an agent acting on a directly affected patient (Slobin, 1982). Again we found no evidence for the primacy of these verbs or patterns. The children's early utterances with two arguments and a verb were as often non-prototypical as prototypical and seemed to have developed from previous verb and direct object patterns, based around specific verbs like *do*, *get*, *want*, *see* and *make* which fulfilled the criteria of constructed positional patterns in our coding scheme. Again the basis of organization and development seems to be more related to specific lexical items than to any more generally underlying systematicities.

We are not arguing that children never move beyond producing language organized around specific lexical items, merely that this is where they start from in that movement. Our metaphor would be of language developing initially as a number of different islands of organization which gradually link up, as Tomasello has argued for early verbs (1992). These islands are initially segments (either words or phrases) which the child has identified to the

extent that s/he can start analysing other systematic relations between what comes before, after or within them. Such an approach fits well with the evidence of amalgam learning in highly agglutinative languages such as Hungarian (MacWhinney, 1985) or West Greenlandic (Fortescue & Olsen, 1992), where children initially learn a set of inflections attached as a whole to highly frequent words and subsequently start to break them down, at first not generalizing these newly isolated segments to other words but learning to substitute other appropriate contrastive inflections in the new slot.

However, we would not wish to claim that the output of our coding scheme can be taken as a literal account of the child's generative patterns or rote-learned phrases nor that organization around specific lexical items is the only method of generating utterances available to children, even in the early stages. It is quite possible that various discourse and contextual factors might lead to systematically generated utterances (Veneziano *et al.* 1990; Lieven & Pine, 1992) and, indeed that such organization might underpin some of the slot-and-frame structures identified by the scheme (e.g. vocatives). And certainly, as development proceeds, methods of generating utterances must become more general and abstract. However, the coding scheme has two main virtues. First, it gives a methodological handle with which to estimate the degree of surface positional consistency in a child's corpus – how one then accounts for this consistency requires further analysis. Second, it brings into clear focus the problem of how to distinguish rote-learned utterances from those which are productive. This is critically important. While the coding scheme may well overestimate the amount of rote-learning, it defines a set of utterances the status of which we can check against other tests. Without this there is a real danger that the definition of utterances as rote-learned or productive is either arbitrary or, even more worryingly, the result of systematic theoretical biases.

Many of the children in the present study had groups of utterances united by wh-frames such as *wheres x* and *whats x doing*, and structures such as *its x*, *its a x*, *theres a x* and *theres the x* but our approach would suggest that we are not likely to find that individual children have a fully flexible grasp of wh-questions or of copula construction. Instead, they are building up utterances around one particular frame while ignoring others which, on the grounds of syntactic theory, ought to be related. However, these frames do develop and in the end they must join up to give a more powerful generative machine. An example both of lexical specificity and development comes from the use of *x-gone* structures by one of the subjects in our sample (child 2). This child uses a much wider range of subjects before *gone* than he does for any other verb – these include people, pronouns, non-humans and inanimate objects. He also starts to add a complement to *gone* well before there is evidence for any general addition of complements to other appropriate verbs. Our suggestion is that the flexibility with which this child initially treats the subject position

in utterances with *gone* may then be generalized to other verbs, though the precise nature of this generalization remains to be explored. Of course one could argue either that the child has an abstract rule for the expression of subjects but is prevented from its full generalizability on other verbs by performance factors or, alternatively, that s/he has not developed the method of expressing subjects and complements and that the behaviour with *gone* represents semi-formulaic usage but neither account seems to be satisfactorily developmental or grounded in the actual evidence. We are arguing, then, that semi-formulaic usage should be seen as part of the developmental process and not simply as a diagnostic of the level of development.

A final example of a possible locus for development is the frequent use of *it* post-verbally by many of the children in our sample. This has, of course, been noted many times by child language researchers working on the acquisition of English (Brown & Bellugi, 1964; Tomasello, 1992). Indeed Brown and Bellugi suggested that, for Adam, *it* represented some form of post-verbal inflection and Ninio (1994) also draws attention to the importance of post-verbal *it* in the development of the predicate position. In our data too, the fact that *it* appears so widely on a range of verbs suggests that we may be seeing the development of an emergent category of verb. Some of the children also start postposing a direct object NP after the *it* towards the end of the period under study. Further research will address the ways in which the direct object slot loses its 'obligatory' *it* and starts to be filled by direct object NPs but clearly here we do have some possible evidence of system-wide development around the category of verb.

We are suggesting, in an approach which starts from the data, that specific lexical items can provide a focus of development for a period well beyond the early stages of multiword utterances. In our view, the evidence from these children's language at this stage does not require the postulation of underlying syntactic or semantic structure of a very abstract or a very general nature. We think, rather, that the data can support a view of structure as emergent. Finally, we want to suggest that we are much more likely to capture a psychologically realistic account of the development of any more powerful underlying generative rules and structures if we develop careful and accurate methodologies to test for their existence.

APPENDIX A

The mean length of utterance (word) for subjects at 25 patterns and 400 utterances

	25 Patterns		400 Words	
Subject	Age	MLU	Age	MLU
1	1;7	1·41	1;10	1·89
2	2;4	2·83	1;10	2·83
3	2;0	1·97	2;1	2·92
4	2;8	3·75	2;8	3·75
5	2;1	1·65	2;5	2·06
6	2;3	2·30	2;5	2·23
7	2;9	3·24	2;10	2·94
8	2;7	2·56	2;7	2·56
9	2;1	2·38	2;1	2·38
10	2;8	2·94	2;8	2·94
11	1;8	1·84	1;10	2·24

APPENDIX B

Child 1

Prototypical	*Non-prototypical*
Daddy burnt it	Mummy read it
Teddy bump head	I saw a see-saw
Mummy drink it	We found it
Mummy do it	I want dummy
Daddy do it	No, Amy wants some biscuits
Helen do it	Mummy got headache
Sweep do it	Dolly wants some drink
Amy dry it	Amy wants some now
Amy do it	
I do it	
I drink lemonade	
No, Daddy do that	

Child 2

Prototypical	*Non-prototypical*
Mummy get it please	I love you
John wash windows	I want a drink
I drive it	I want a banana please
Mummy burst ball	Alan needs new pennies
I do it	Jean need pennies
Wind shut door	Daddy need pennies
Hoover, Mummy push it	Mummy need pennies

I've dropped apple
I'll get it
Tuff chase birds
Emma do it
I draw ball
Daddy stop it now

Child 3

Prototypical
Tom made that
Daddy fix car
Charlie rip that
Rupert got dressing gown

Non-prototypical
I want that
I want cornflakes
I want Noddy
Pat found Jess
Pat got key
Charlie broke that
I want my potty
Rupert had a sweet
I want my Mummy
I want it Mummy please
Me got that drill
Thomas got a tractor
I got a duck
I want that one
I want juice apple
I want drink

Child 4

Prototypical
I'll do it
No, Mummy rub it
Daddy pinched my toast
Me do it
I did it
Mummy do them
Daddy do a gee-gee
No Wes tickle my tummy
Me fasten my coat
I drink it
Me, I drink lager
Mummy carry that

Non-prototypical
I like ice cream, Mummy
Me like new car, Daddy
Wes have cucumber
Graeme see tractor
Me like tomato and ham
Me like ham and tomato
Me have it potatoes
Me had hula hoops

Child 5

Prototypical	*Non-prototypical*
I do it	I got it
Mummy do it	I ride bike
I carry it	I done it
No, Daddy do it	I want drink
My do it	Me have some
My shut gate	Me want that one, Mum
My shut door	My have it
My mend it	Me have chip, Mum
My drink it	I'll share them
Girl eat it	
My done it	

Child 6

Prototypical	*Non-prototypical*
Mummy pinched it, bow-wow cake	Karen want one
I do it	I answer it
Lady pinch it	I don't like black ones
I'll take him	Daddy stop it
Mummy take book	I like that
	I'm warning you
	I'm watching this first
	I want that one
	I want bingo, Mum
	I love it
	I want something
	I've got special ones today
	I missed them
	I answer it
	Poo, bow-wow got a smelly bum

Child 7

Prototypical	*Non-prototypical*
Slug get you	I want crisp
Mummy, I getting the trolley	You got two
Kids get toys	I got one
I do it	I want blue one
I doing it	I'm finding blue one
Gordon's take them	I find blue one

I ripped that

I want bike
I want that cup
I want that box
I want push door
I want book
Chrissie's got some
Chrissie got book

Child 8

Prototypical
Joss did it
I get it
Nancy take it
I do it
Mummy get it
Let me fix it
Nancy pinched it, sweetie
Daddy do that one
You get it
No, Leroy do it
Leroy do that one
Daddy do it

Non-prototypical
I have a go
Joss had a wee-wee
Daddy wants a cup of char
Mummy play that one
Nancy done that one now
I have it
Thomas Tank see you
Me have some

Child 9

Prototypical
Me do it
Me do that
I'll do that
I've broken it
Grandma's eaten it
I throwed it
Mummy get me

Non-prototypical
I want it
I want a drink
I got drink
You get it, Mum
I'll have some
Mummy got one
I want some
I want a bath
I want more
I want some more
I read it
Me have shower
I ride bike, Mummy

Child 10

Prototypical	*Non-prototypical*
Mummy get drink, please	You say it
I go get my bag	Me see baby doll
I go get my bike	Aunt Cora get a baby cousin
I go take my bike	You want ice cream and cigarettes
	I go see Helen
	I want dummy
	I want get dummy
	Me see book
	I don't want cornflakes
	I want my lunch
	I love you
	I go phone Nana
	I want toast and jam
	I want my dummy
	I want my cup of tea
	I'd like a biscuit

Child 11

Prototypical	*Non-prototypical*
I'll do it	Yes, Mummy got bag
I get it	Dad, you want a button?
I do it	I love you
Mummy smack Dad	Julie doing jam butty
Mummy smacking my knee	Helen do a wee-wee
Helen smack me	I want my dummy
Helen smack my knee	Hey, Julie stop it now
	Julie got the toffees
	Julie got a book
	Andrew got a book
	Julie got new shoes
	I want doggie
	Mummy want one

APPENDIX C

Child 1

Prototypical	*Non-prototypical*
Dropped it	Caught it
Do it	Bump head
Change it	Want that

Do drawing	See Mummy now
Drink lemonade	
Doing a jump	

Child 2

Prototypical	*Non-prototypical*
Burst balloon	Phone Tracy
Drink tea	Finished sugar puffs
Spilled it	Draw ball
Fix it	Want Jenny
Do it	Miss it

Child 3

Prototypical	*Non-prototypical*
Bite nail	Want wee-wee
Poke eye	Want dinner
Throw that	Want book
	Want ball
	Want drill
	Want shoes
	Bump head

Child 4

Prototypical	*Non-prototypical*
Yeah do it	Draw pussycat
Carrying penguin	See doctor now
Smack me	Have my medicine, Mummy
Bite a boy	Ask Wes
	Want a more drink
	Cut my knee

Child 5

Prototypical	*Non-prototypical*
Open door	Want my Daddy
Broke it	Want dummy
Open it	Want a yoghurt
	Want that
	Want dinner
	Want bite
	Want breakfast

Child 6

Prototypical	*Non-prototypical*
Dropped it	Don't want it
Stroke it	Don't like it
Do it	See you
Carry me	Want dummy
Wash hands	Phone Michael

Child 7

Prototypical	*Non-prototypical*
Get it	See you
Stop it	See it
Warm it	Done it
Take it	Got it
Fix it	Hold it

Child 8

Prototypical	*Non-prototypical*
Catch it	See it
Stop it	Want Daddy
Wash teeth	Don't touch it
Open doors	Want drink
	Have a go
	Leave it

Child 9

Prototypical	*Non-prototypical*
Shut door	Want that
Dropped it	Hold it
Change that	Find it
	Want more biscuit
	Want more drink
	Read it
	See him

Child 10

Prototypical	*Non-prototypical*
Catch them	See you
	See bag
	See Pat
	See ball
	See baby
	Want my dummy
	See dummy
	See bike
	See daddy

Child 11

Prototypical	*Non-prototypical*
Shut door	Want wee-wee
Open gate	See you
Dropped it	Read that
Touch it	Want pen
Open door	Want drink

REFERENCES

Bloom, L., Lightbown, P. & Hood, L. (1975). Structure and variation in child language. *Monographs of the Society for Research in Child Development* **41**, 164.

Bowerman, M. (1990). Mapping thematic roles onto syntactic functions: are children helped by innate linking rules? *Linguistics* **28**, 1253–89.

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

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

Brown, R. & Bellugi, U. (1964). Three processes in the child's acquisition of syntax. *Harvard Educational Review* **34**, 133–51.

Budwig, N. (1989). The linguistic marking of agentivity and control in child language. *Journal of Child Language* **16**, 263–84.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science* **14**, 179–211.

Finch, S. & Chater, N. (1992). Bootstrapping syntactic categories. In *Proceedings of the 14th Annual Conference of the Cognitive Science Society*.

Fortescue, M. & Olsen, L. L. (1992). The acquisition of West Greenlandic. In D. I. Slobin (ed.), *The crosslinguistic study of language acquisition*, Vol. 3, pp. 111–220. Hillsdale, NJ: Erlbaum.

Karmiloff-Smith, A. (1979). *A functional approach to child language: a study of determiners and reference*. Cambridge: Cambridge University Press.

Levy, Y. (1988). The nature of early language: evidence from the development of Hebrew morphology. In Y. Levy, I. Schlesinger & M. D. S. Braine (eds), *Categories and processes in language acquisition*, pp. 73–98. Hillsdale, NJ: Erlbaum.

Lieven, E. V. & Pine, J. M. (1992). Productivity in early multiword utterances: analytic and synthetic solutions. Paper presented at British Child Language Seminar, Annual Conference, Glasgow.

Lieven, E. V., Pine, J. M. & Dresner Barnes, H. (1992). Individual differences in early vocabulary development: redefining the referential–expressive distinction. *Journal of Child Language* **19**, 287–310.
MacWhinney, B. (1985). Hungarian langauge acquisition as an exemplification of a general model of grammatical development. In D. I. Slobin (ed.), *The crosslinguistic study of language acquisition*. Vol. 2, pp. 1069–156. Hillsdale, NJ: Erlbaum.
MacWhinney, B. & Snow, C. (1985). The child language data exchange system. *Journal of Child Language* **12**, 271–96.
—— & —— (1990). The child language data exchange system: an update. *Journal of Child Language* **17**, 457–72.
Maratsos, M. (1988). The acquisition of formal word classes. In Y. Levy, I. M. Schlesinger & M. Braine (eds), *Categories and processes in language acquisition*, pp. 31–44. Hillsdale, NJ: Erlbaum.
Mills, A. E. (1986). *The acquisition of gender: a study of English and German*. Berlin: Springer-Verlag.
Ninio, A. (1994). Words with holes: the acquisition of the predicates. Paper presented at the Emory Conference on Cognitive and Functional Approaches to Grammatical Development, Emory, Atlanta.
Pérez-Pereira, M. (1991). The acquisition of gender: what Spanish children tell us. *Journal of Child Language* **18**, 571–90.
Peters, A. (1997). Language typology, individual differences and the acquisition of grammatical morphemes. In D. I. Slobin (ed.), *The crosslinguistic study of language acquisition*, Vol. 4. Hillsdale, NJ: Erlbaum.
Pine, J. M. & Lieven, E. V. (1993). Reanalysing rote-learned phrases: individual differences in the transition to multiword speech. *Journal of Child Language* **20**, 551–71.
Pinker, S. (1984). *Language learnability and language development*. Cambridge, MA: Harvard University Press.
Plunkett, K. (1993). Lexical segmentation and vocabulary growth in early language acquisition. *Journal of Child Language* **20**, 43–60.
Plunkett, K. & Marchman, V. (1991). U-shaped learning and frequency effects in a multi-layered perception: implications for child language acquisition. *Cognition* **38**, 43–102.
Radford, A. (1990). *Syntactic theory and the acquisition of English syntax*. Oxford: Basil Blackwell.
Rispoli, M. (1994). Pronoun case overextensions and paradigm building. *Journal of Child Language* **21**, 157–72.
Schlesinger, I. (1982). *Steps to language: toward a theory of native language acquisition*. Hillsdale, NJ: Erlbaum.
Slobin, D. I. (1982). The origin of grammatical encoding of events. In W. Deutsch (ed.), *The child's construction of language*. London: Academic Press.
Tomasello, M. (1992). *First verbs: a case study of early grammatical development*. Cambridge: Cambridge University Press.
Valian, V. (1986). Syntactic categories in the speech of young children. *Developmental Psychology* **22**, 562–79.
—— (1991). Syntactic subjects in the early speech of American and Italian children. *Cognition* **40**, 21–81.
Veneziano, E., Sinclair, H. & Berthoud, I. (1990). From one word to two words: repetition patterns on the way to structured speech. *Journal of Child Language* **17**, 633–50.