Two-year-old children's production of multiword utterances: A usage-based analysis

ELENA LIEVEN,* DOROTHÉ SALOMO and MICHAEL TOMASELLO

Abstract

Children generate novel utterances from the outset of multiword speech. In this study, we apply a usage-based method called 'traceback' to the multiword utterances of four two-year-olds to see how closely related these utterances are to their previous utterances. Data was collected from the age of 2;0 until 6 weeks later on a relatively dense sampling schedule. We attempted to match each novel multiword utterance in a two-hour corpus to lexical strings and schemas that the child had said before. Matches were found for between 78–92 percent of all multiword utterances. Between 62–91 percent of the slots in schemas created by these tracebacks were for referring expressions and were filled with nouns or noun phrases. For one child, recording continued throughout his third year and we compared his data at MLUs matched with the other three children to investigate developmental changes. We found that, with increasing MLU, and developmentally, children were less repetitive within sessions, the tracebacks required a wider range of semantic slots and the material placed in these slots increased in complexity.

Keywords: *language acquisition; usage-based approaches; high-density developmental corpora; multiword utterances; individual differences.*

---

* Correspondence address: Elena Lieven, MPI-EVA, Deutscher Platz 6, 04103, Leipzig, Germany. E-mail:⟨lieven@eva.mpg.de⟩. Acknowledgements: We would like to thank Ewa Dąbrowska who was central to developing the 'Traceback' method; Katja Hummel, Kristin Wolter and Silke Harbodt who put many hours of careful work into the data analysis; Sascha Hoppe and Franklin Chang for developing the Autotracer; Jeannine Goh, Ellie O'Malley and Dimitra Doumpioti, the Dense Database team in Manchester; and, most importantly, the children and families who contributed so much of their time to the collection of the 'Dense databases'. Finally, two anonymous reviewers made extremely helpful comments on the original manuscript.

Cognitive Linguistics 20–3 (2009), 481–507 0936–5907/09/0020–0481
DOI 10.1515/COGL.2009.022

1. Introduction

By the time of their second birthday, almost all children are producing multiword utterances. This is reflected in the norms for the mean length of utterance (MLU) of English-speaking children at 2;0, which are between 1.47–2.37 (Miller and Chapman 1981). In this paper, we investigate how close these multiword utterances are to what children have said previously.

In the ongoing debate between nativist-linguistic and constructivist approaches to language development, it is the constructivist approach that needs to focus on the relationship between a child's utterance and what she or he has said and heard before since this is seen as the way in which grammar is built. This contrasts with nativist-linguistic approaches that emphasise the initial abstractness of children's grammars. In these latter approaches idioms, rote-learned phrases and semi-formulaic structures are by now acknowledged as parts of linguistic representation but they are seen as external to the essential abstractness and generativity of the grammar and of no interest to understanding the process by which the child maps his or her ambient language onto this abstract Universal Grammar. However, in the usage-based approach there is no sharp divide between lexically-specific, schematic and fully abstract constructions and, indeed, the same utterance could be generated from any one of these representations if they coexist in the grammar. In development, schematicity is an emergent property of breaking down concrete utterances and developing the slots created within them.

The child learns strings for achieving communicative functions from the input and gradually elaborates them. Children are seen as building up an inventory of constructions derived from an interaction between what they hear and what they want to say. Constructions are mappings between form, initially highly lexically and phonologically specific, and meaning. Semantics is emergent, with child-based meanings becoming increasingly conventionalised with development. The degree of idiosyncrasy in the meaning of the same lexical forms will depend on differences and similarities in how these are used to the child. Therefore a young child's grammar is hypothesized to be more lexically-specific and less schematic and abstract than that of an older child or adult. If the way in which the child generates utterances is based more closely on specific strings that have already been learned, the implication is that we should find close relationships between what is said and what the child has said before and that these should be closer earlier in development than later or for adults (Goldberg 2006; Lieven and Tomasello 2008; Tomasello 2003).

Both Lieven et al. (1997) and Tomasello (1992) related a corpus of utterances produced by the children to utterances that those same children had already produced. In Tomasello's (1992) diary study of his daughter's development of verbs, he showed that each new structure that his daughter produced with a particular verb was more closely related to what she had previously produced with that same verb, than to any general 'across the board' development of verb argument structure. Lieven et al. (1997) looked for antecedents to utterances produced by the 12 children in their study and showed that a mean of 60 percent of the utterances could be accounted for by limited scope schemas. The results from both studies are suggestive of a close relationship between the child's current and previous utterances but both had limitations in terms of sampling. In the case of Tomasello, almost all of the child's utterances with verbs were captured but at the expense of all her other utterances. In the Lieven et al. study, the sampling was only one hour of recording every 2–3 weeks probably representing only between 1–2 percent of what a child says or hears (Rowland et al. 2008; Tomasello and Stahl 2004). This makes identifying commonalities between a corpus of utterances and what has been said before very difficult. However the corpora used in the present study are very much denser, based on recordings of 5 hours a week and, we estimate, probably represent about 7–10 percent of the child's input and utterances (Lieven et al. 2003). Using these corpora and a method we call traceback (Dąbrowska and Lieven 2005; Lieven 2006; Lieven et al. 2003) we identify a set of multiword utterances for each corpus and investigate how closely related they are to what the child has said before.

2. The traceback method

The traceback method involves defining a test corpus of utterances and searching for precedents that contain all or part of the same strings in the preceding corpus. These precedents are called 'component units' and two types are defined: fully lexically-specific strings of one or more words and schemas with slots. 'Operations' are then defined that allow matching the child's utterances in the test corpus to component units in the main corpus. In Lieven et al (2003) this was done for one two-year-old child. The method showed that a large proportion of the child's test corpus utterances were either identical to previous utterances or could be derived by one single operation from component units in the main corpus. However, there were two problems with this method. First, five different operations were used (including 'insert' and 'rearrange'). This meant that every utterance could be traced back, though in a few cases only with large numbers of operations. This made the method completely unconstrained. The second problem was that the analysis was purely distributional and took no account of semantics: provided it fulfilled the distributional criteria, any component unit could be placed in any slot. Both these features made the method rather psychologically and linguistically unrealistic and Dąbrowska and Lieven (2005) changed it to take account of this, first, by reducing the number of operations to two and, second, by stipulating that slots could only be created by fitting strings with broadly similar semantics. This method was used to examine the novel syntactic questions of two children at 2;0 and 3;0. The authors found that, even at 3;0, most of these children's novel syntactic questions could be traced back to previous questions in the main corpus using only one operation. They suggested that the gradual development of lexically-based formulae accounted better for the syntax of the children's questions than did an account based on an highly abstract grammar involving, for instance, 'movement', a conclusion supported by Rowland (2007) and Rowland's and Pine's (2000) studies of wh-question development.

However it could be argued that looking at questions is bound to show up a relatively high degree of lexical specificity because of the limited number of wh-words and auxiliary combinations possible in English. In the present study, therefore, we explore the full range of multiword utterances produced by four 2-year-old children using a slightly adapted version of the Dąbrowska and Lieven (2005) method. We investigate how closely related a corpus of each child's utterances is to what has been said before; how previous utterances would have to be modified to produce the target utterances in the test corpus; and, in particular, the types of slots that would have to be hypothesised. For one of the children, Brian, we have the same dense sampling over the entire year from 2;0–3;0, so we also examine his development at MLUs matched to the other three children.

In attempting to establish how close a child's utterance is to what has been said before, there is the issue of whether to look only at the child's previous corpus or whether to include the input. In previous traceback studies the input was included on the grounds that it is from the input that the child is learning (see also Lieven 2006). However the argument for tracing back only to the child's own corpus is also strong: we then know that the string is (or rather was, when it was uttered) part of the child's linguistic representation and this obviously makes the comparison between present and prior utterances much more straightforward. In the present study, therefore, the children's utterances are traced back only to their own previous corpora. The differences between trackbacks based on only the child's previous utterances and those including the input will be briefly raised in the Discussion section.

3. Method

3.1. *Participants*

The data consist of high-density developmental corpora for four children, Annie, Brian, Fraser and Eleanor. Each was recorded for 5 hours per week for 6 weeks from the age of 2;0. In addition we recorded Brian on the same schedule for the entire year between 2;0 and 3;0. The children lived in a large metropolitan area in England and came from middle-class backgrounds.

3.2. *Data*

The recordings took place in home settings and were done by the mothers on four days of each week. The most typical activities during taping were playing with toys and having a snack. Once a week, a research assistant visited to make a video recording. For three of the children, this six-week period resulted in 30 hours of recording. Illness prevented two of Annie's sessions, so this corpus contains only 28 hours.

Research assistants transcribed all of the tapes in SONIC CHAT format (MacWhinney 2000). All speech was transcribed with the exception of the speech not directed to the child (i.e., speech between adults, telephone calls etc.).[1] The children's mean length of utterance (MLU) was measured in words across all their utterances in the corpus.[2] In increasing order, the children's MLUs were: Brian, 1.65, Fraser 1.8, Annie, 2.19 and Eleanor, 2.22. According to Miller and Chapman (1981) this places Brian and Fraser in late Stage I and at roughly the predicted chronological age for their MLU (Fraser one month ahead), with Annie and Eleanor in stage II, roughly 4 months ahead of the predicted age for their MLUs. At MLUs matched with the other three children Brian was 2;2 (MLU 1.8), 2;6 (MLU 2.19) and 2;7 (MLU 2.22).

3.3. *Traceback procedure*

Each corpus was divided into two parts: a test corpus, which consisted of recordings made on the last two days of the six-week period, and a main

1. The Annie 2;0 corpus is the same as that used in Lieven et al. (2003), the Brian and Annie 2;0 corpora are those that were used in Dąbrowska and Lieven (2005). For full details of the method of recording and of transcription, see Lieven et al. (2003).
2. We prefer to measure in words at this age rather than to make assumptions about the productivity of morphemes: measures of children's MLUs in words and morphemes are highly correlated (Parker and Brorson 2005).

corpus, which contained the remaining transcripts of all the preceding recordings. Information about the size of the subcorpora is given in Appendix A.

In the intelligible, multiword utterances of the test corpus, we first found all multiword utterance types by identifying those utterances that were immediate imitations or self-repetitions and those that were repeated by the child on different occasions during the two hours (see Table 5 below). The multiword types were the target utterances, that is to say, the utterances to be traced back. These were compared to utterances previously produced by the child in the main corpus in terms of whether they matched exactly or whether they could be derived by lexical or syntactic changes ("operations") from prior ones.

3.3.1. *Component units.* For each target utterance the closest prior strings in the main corpus were identified. These closest matches were putative component units of the target utterance. A component unit is an expression which shares lexical material with the target utterance (excluding imitations). Two types of units were identified: schemas with slots and fixed strings.

3.3.1.1. *Schemas with a slot.* If a string matched the novel utterance in the same way, with variation in the same position, this was identified as a potential schema with a slot. A slot was established if the string in the target utterance and the string in the same position in the potential schema in the main corpus belonged to the same broad semantic category. Six types of semantic slots were distinguished which are shown in Table 1

Table 1. *Types of slots*

Type of slot	Example utterances	Schema with slot
REFERENT (REF)	More choc choc **on there**. Bow's food **on there**.	REF **on there**.
PROCESS (PRO)	**I want to** get **it**. **I want to** roll **it**.	**I want to** PRO **it**.
ATTRIBUTE (ATT)	Pilchard there **he's** hungry a toast. **He's** upside down.	**He's** ATT.
LOCATION (LOC)	**I sit** on Mummy's bike. **I sit** there.	**I sit** LOC.
DIRECTION (DIR)	**Going** under bridge. **Going** down.	**Going** DIR.
UTTERANCE (UTT)	**Mummy**, here go. **Mummy**, what these.	**Mummy**, UTT.

and demonstrated by examples. Note that the slots can be filled by single words or by strings longer than one word but for all except UTTERANCE slots, the semantics of the slot and the semantics of the potential filler must match. UTTERANCE slots were almost always created by utterances that had been said in their entirety followed by or following a vocative, an initial conjunction (e.g., *and, but, so, then*) or a final locative (e.g., *here, there, outside*) and the strings creating them were therefore more semantically heterogeneous.

3.3.1.2. *Fixed strings.* A fixed string is any word, or continuous string of words, produced by the child, corresponding to a "chunk" of semantic structure (i.e., designating a REFERENT, PROCESS, ATTRIBUTE, LOCATION, DIRECTION or UTTERANCE)[3] which occurs in the main corpus. Again the string does not have to occur in isolation—so the following two utterances are regarded as evidence that the expression *make a cake* is available to the child as a unit (fixed string) which can be placed into a PROCESS slot:

(1) Fixed string as component unit (Eleanor, 2;0)
*CHI: oh let's <u>make a cake</u>.
*CHI: Mama you <u>make a cake</u>.

3.3.2. *Deriving target utterances from component units.* For each target utterance, all potential component units are identified in the main corpus and an attempt is then made to derive the target utterance using the operations of SUBSTITUTE and ADD[4]. SUBSTITUTE allows the placement of a component unit into the matching slot of a schema. ADD allows the placement of component units to one or other end of an utterance (vocatives and adverbials such as *now* and *then* being examples).

Table 2 shows the traceback for Fraser's novel utterance *A baby dragon*. The format of these tables is as follows: the target utterance is given in the title; in the left hand column are component units for the utterance; the right-hand column shows the number of times each of these has appeared in the main corpus for the child.

3. Abbreviated to REF, PRO, ATT, LOC, DIR and UTT.
4. In Dąbrowksa and Lieven (2005), the operation of SUPERIMPOSE allowed both for direct substitution into a slot and for superimposing a string over a slot as long as the semantics matched. In the present method only direct substitution is allowed. JUXTAPOSE in Dąbrowksa and Lieven is the same as ADD.

Table 2. *Target utterance*: A baby dragon *(Fraser 2;0)*

Component units	Number in main corpus
a baby REF	7
dragon	18

Only utterances where the relevant strings match the semantics of the target utterance are counted. So in the case of the target utterance *A baby dragon*, there is a schema in the main corpus for *A baby REF* (which occurred 7 times with two types in the REFERENT slot: *frog* and *car)* and the child has also said *dragon*. So this is a successful one operation traceback using SUBSTITUTE into a REFERENT slot.

In Table 3, we give an example of a SUBSTITUTE operation into a PROCESS slot

Table 3. *Target utterance*: I can't see it *(Annie 2;0)*

Component units	Number in main corpus
I can't PRO it	17
see	47

Annie's *I can't see it* was not said in the main corpus but matches 17 examples of *I can't X it* where the semantics of X are those of a 'process' (e.g., *do it, sing it, get it out, reach it, open it, pull it up*) and 47 of *see*. This allows *see* to be substituted into *I can't PRO it* to give the target utterance in a one-operation traceback.

Some tracebacks required more than one operation to arrive at a match for the target utterance. Table 4 gives an example from Brian's 2;7 traceback:

Table 4. *Target utterance*: No press the yellow one *(Brian 2;7)*

Component units	Number in main corpus
no PRO the REF	7
press	5
yellow one	17

Step	Component unit	Operation	Filler of slot	Result
1	no PRO the REF	SUBSTITUTION (PRO)	press	no press the REF
2	no press the REF	SUBSTITUTION (REF)	yellow one	no press the yellow one

No press the yellow one is traced back to *No* PRO *the* REF which occurs 7 times in the main corpus (*No climb the man, No open the lid, No get the Mummy, No take the binbag lorry* etc.). Brian also says *yellow one* (17 times) and *press* (5 times) so this is a two-operation traceback with substitutions into a PRO and a REF SLOT.

For some utterances, a number of alternative derivations from component units were possible. We wanted to reduce the tracebacks to the minimum number of operations and the following rules ensured this:

1. The longest possible schemas were used.
2. The slots were filled by the longest available units.
3. The minimum number of operations were taken.

A full discussion of these decisions and their justification can be found in Dąbrowska and Lieven (2005).

Component units were identified using a computer program[5] but all results established by the program were manually checked and revised where necessary, taking semantics into account. Semantic coding was done, after extensive training, by two research assistants.

3.3.3. *Failed derivations.* Utterances that cannot be derived from component units are called 'fails'. They are categorised into two types: lexical fails and syntactic fails, though an utterance could fail on both.

3.3.3.1. *Lexical fails.* A lexical fail occurred when the child used a word in the target utterance that was not found in the main corpus. However, if the word occurred in the immediately preceding discourse (defined as the previous five utterances), this frequency criterion was relaxed and we assumed that the word was available to the child even if it did not occur at all in the main corpus. This rule was restricted to single words.

3.3.3.2. *Syntactic fails.* If a child produced a novel utterance for which no relevant component units were found in the main corpus, this was coded as a syntactic fail.

3.4. *Analyses*

3.4.1. *Basic traceback analysis.* Utterance types were first identified in each child's test corpus and then traced back in the main corpus to give the following:

5. The program, Autotracer, was designed by Sascha Hoppe under the supervision of Franklin Chang.

1. The proportion of successful tracebacks, together with the proportion of lexical and syntactic fails
2. The number of operations required for the successful tracebacks
3. The range and proportion of different types of slots used in the traceback

For reliability, 38 percent of all tracebacks were coded twice. Agreement was very high (96.5% kappa = 0.89). The few disagreements were resolved by discussion.

3.4.2. *The nature of the* REFERENT *slot.* As we will see by far the largest number of slots required for the tracebacks were REFERENT slots. This allowed us to investigate the nature of the REFERENT strings in more detail. We conducted two analyses. In the first, we examined the different types of strings that went into the REF slots and in the second, the grounding of REFERENT slots. The term 'grounding' comes from Langacker's (1991: 318 et seq.) discussion of noun phrases and verb phrases, in which he points to the functional similarity of the 'X-bar' structure in 'grounding' bare nouns and verbs (with determiners in the case of nouns and finiteness marking in the case of verbs).

We coded the string placed in the REFERENT slot into the following categories:

– bare noun or pronoun
– *a/the* + noun
– other determiner + noun
– *a/the* + adjective + noun
– adjective + noun
– possessive (pronoun or *'s* clitic) + noun

In coding the target utterances for potential matches, we allowed for substitutions of both grounded (e.g., *my cat, the black cat*) and ungrounded (e.g., *cat, black cat*) nominals into the REFERENT slot, and both grounded (e.g., *sits, sat, is sitting*) and ungrounded (e.g., *sitting*) predicates into the PROCESS slot. This reflects the fact that the children often omitted determiners and used untensed verb phrases where tensed forms were required. There were too few PROCESS slots for anything but a very preliminary analysis but we explored the grounding of REFERENT slots further by analysing all the ungrounded slots in the successful tracebacks to see whether they were filled with grounded or ungrounded fillers. Frames were defined as grounded if they had correct determiners or were vocatives into which proper names could be substituted. Fillers were coded as grounding if they contained determiners, proper names, mass nouns or plural nouns that correctly grounded the REFERENT slot[6].

3.4.3. *Developmental analysis.* We used these comparisons between the four children at the same age but with different MLUs, and of Brian at four different ages and MLUs, to address the issue of development in terms of the traceback results.

4. Results

4.1. *Multiword utterances in the test corpus*

The total number of multiword utterances in the test corpora ranged between 201 for Brian at 2;0 to 476 for Annie at the same age (see Table 5). The number of types ranged between 23 percent (Brian at 2;2) and 63 percent (Eleanor at 2;0)

Table 5. *Types of tokens of multiword utterances in the test corpora*

(a) at age 2;0	Brian	Fraser	Annie	Eleanor
Number of multiword utterances	201	424	476	249
Number of types	66	181	210	157
% types	33%	43%	44%	63%
(b) Brian developmental		Brian (age 2;2)	Brian (age 2;6)	Brian (age 2;7)
Number of multiword utterances		205	419	366
Number of types		47	213	174
% types		23%	51%	48%

4.1.2. *Development.* At 2;0, Brian, with the lowest MLU, produces the lowest proportion of multiword utterance types while Eleanor, with the highest MLU, produces the highest proportion, with Fraser and Annie between. Note that this is not an obvious function of the relative sizes of the test corpora (i.e., more utterances leading to a greater chance of repetition), since from Appendix A, we can see that, at 2;0, Fraser and Eleanor have somewhat more utterances in their test corpora than Brian and Annie respectively. This suggests that the degree of repetitiveness

6. Not all ungrounded frames could be grounded (for instance ungrounded adjective phrases such as *Big REF* could not be grounded given the traceback methodology. These frames were not counted when analyzing the proportion of frames that became grounded once the slot was filled)

may decrease as children become more sophisticated language producers. Brian's developmental tracebacks support this conclusion to some extent: there are more multiword utterance types at the two later ages.

4.2. *Matching target utterances to the main corpus*

We applied the traceback method to all the multiword utterance types identified in the test corpus. In what follows we present the number of operations required to match these utterances (including exact matches), the proportions of lexical and syntactic fails and the relative proportions of the SUBSTITUTE and ADD operations.

4.2.1. *Number of operations required for tracebacks.* Figure 1a shows the number of operations required to build the target utterances from component units in the main corpus, including the proportion of fails, for the four children at 2;0. Figure 1b is the equivalent for Brian's developmental MLU matches. We can see that for all four children at 2;0 and all Brian's tracebacks, the great majority of the target utterances can be traced back either to an exact repeat of something said before (exact matches) or require only one operation. The combined figure ranges from 58 percent for Eleanor's traceback up to 92 percent for Brian's at 2;2. Almost all the remaining target utterances require only two operations, with only between 2 percent (Fraser's traceback) to 5 percent (Eleanor's) requiring more.

Development: Across the four children's 2;0 corpora, the number of target utterances which are exact matches goes down with increasing MLU while the number of multi-operation tracebacks goes up. This is also true for Brian's developmental MLU matches. Although utterance length will affect the number of operations possible in a traceback since shorter target utterances, schemas and fixed strings reduce the likelihood of multiple operations, inspection of Appendix A suggests that this is unlikely to be the whole explanation, since Brian's tracebacks have fewer multi-operations than Fraser's and Annie's at the same MLUs, a matter to which we return in the Discussion.

4.2.2. *Fails.* Across the four children's 2;0 corpora there are 136 utterances (15%) that cannot be derived from attested component units. Overall 4.6 percent are lexical fails. If the children use a word, they have presumably heard it before and we have simply failed to sample it. The proportion of syntactic fails ranges from 5.9 percent for Fraser to 16.3 percent for Eleanor. Brian's syntactic fails range between 4.3 percent (at 2;2) to 14.6 percent (at 2;6). Fails mainly result from (a) the unclear

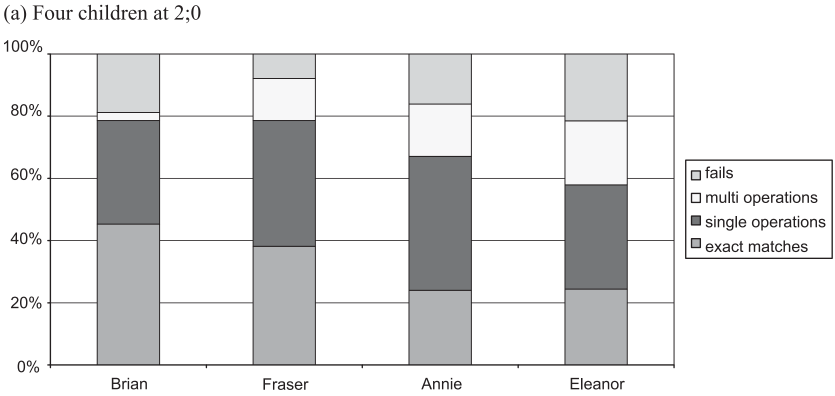


Figure 1. *Number of operations required to build target utterances from the main corpus*

meaning of what the child is trying to say, (b) the failure to find matching strings in the main corpus, often because the utterances seemed to be juxtaposed strings with parts missing, and (c) semantic non-compatibility between the strings in the putative slot. The first reason characterizes Brian's fails at 2;0 and is probably related to the shorter length of his utterances. Since we are only sampling about 7–10 percent of what the children say this level of failure to find component units in the main corpus is perhaps not surprising.

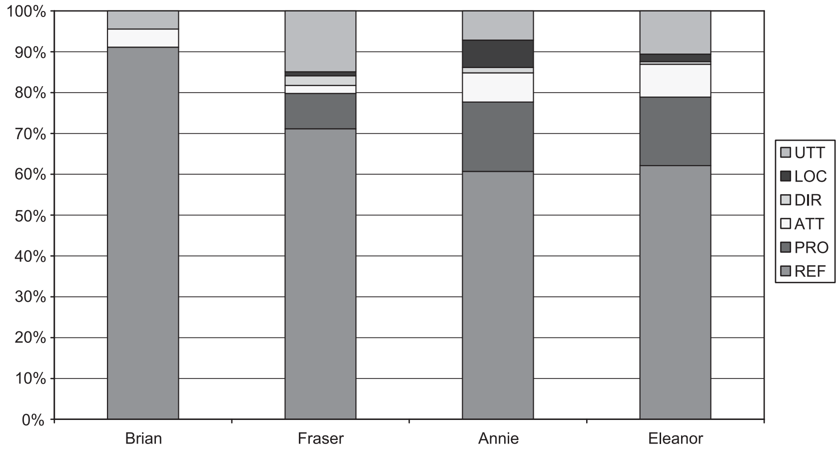
4.2.3. *Types of operations.* Over 95 percent of all operations were SUBSTITUTIONS with ADDS never more than 5 percent. Although there were some cases in which a traceback using ADD was an alternative, the requirement to use the longest possible schema usually resulted in a component unit with an UTTERANCE slot (e.g., *and UTT, UTT then*) rather than in an ADD operation.

4.3. *Types of slots*

Figure 2 shows the types of slots as a percentage of all slots, including those in multiword utterances. By far the largest proportion of slots in all tracebacks is of REFERENTS (61–93%). Brian's 2;0 traceback has no PROCESS slots, Fraser's has 9 percent, while they form 17 percent of all slots in Eleanor and Annie's tracebacks. This increase in higher proportions of PROCESS slots at higher MLUs also shows up in Brian's two later tracebacks where they form 15 percent and 21 percent respectively. Thus with increasing MLU the proportion of REFERENT slots in the tracebacks reduces and the proportion of other slots, most notably PROCESS slots, increases. Annie and Eleanor's tracebacks and Brian's at the two higher MLUs also have somewhat higher proportions of ATTRIBUTE slots. We will discuss REFERENT slots in more detail in Section 3.4 below. Here we will briefly consider the types of strings that were traced back to PROCESS, ATTRIBUTE and UTTERANCE slots.

4.3.1. *UTTERANCE slots.* These constituted between 4 percent (Brian at 2;0) to 15 percent (Fraser at 2;0) of all slots. The most frequent appearance of UTTERANCE slots was with proper names (e.g., *Mummy UTT, UTT Daddy*) or *no* or *yes* at one or other end. All tracebacks had strings of this nature (including almost all of the 15 percent of Fraser's). With increasing MLU there were also tracebacks to schemas with conjunctions (*And UTT, So UTT, Then UTT*) as well as adverbs (*UTT too*, *UTT again*), though there were somewhat fewer of these in Brian's 2;6 and 2;7 tracebacks than in those of Annie and Eleanor. Note that, in tracing back to an UTTERANCE slot, the string in the target utterance had to have occurred at least once in the main corpus. Thus the target utterance could have been produced by repeating a previous utterance and adding a vocative, conjunction or adverbial to one or other end. In the case of adding vocatives or interjections, this will increase the MLU without any necessary added syntactic complexity.

(a) Four children at 2;0



(b) Brian developmental

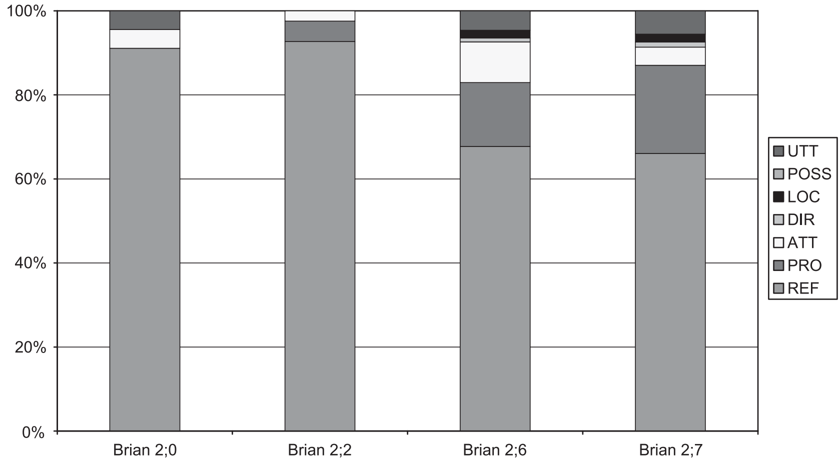


Figure 2. *Types of slots*

4.3.2. *ATTRIBUTE slots.* There were never more than 10 percent of these slots (Brian, 2;6). Tracebacks to slots mainly involved adjectives though there were a few adverbs (e.g., *very ATT = very loudly*, Brian 2;2). Adjectives in prenominal slots mainly occurred in strings traced back to

REFERENT slots (*Baby REF, Green REF, The big REF*) and most ATTRIBUTE slots were for predicative adjectives (*I'm ATT, It's ATT, That ATT*). Although there were also prenominal slots (*A ATT one, ATT light, Want a ATT one*). In the later tracebacks there are also strings where these predicative adjectives are modified (*Bit ATT*, *Too ATT*). Almost all the strings traced to ATTRIBUTE slots were of single words (usually colours, numbers or size adjectives) but there were a few longer strings (e.g., *Big blue, Not cold*).

4.3.3. *PROCESS slots.* The majority of PROCESS slots in Fraser's and all Brian's tracebacks are filled with single verbs (*No PRO it, PRO in on, I've PRO it, PRO them, PRO one there*). Annie and Eleanor's tracebacks contain a higher proportion of PROCESS slots which are filled with longer strings (*I can't PRO, PRO = put it there; No I PRO, PRO = sit over here; That's PRO, PRO = come out; I wanna PRO, PRO = make a REF Don't PRO, PRO = be sad*). All tracebacks except Brian's at 2;0 and 2;2, contained PROCESS slots following auxiliaries (*Can PRO, I can't PRO, Don't PRO it)* and 'semi-auxiliaries' *(Wanna PRO, I want to PRO it, Let's PRO it*) but this was much more frequent for Annie and Eleanor's tracebacks and therefore meant that the PROCESS slots in these tracebacks tended to be grounded by the frame, unlike Brian's whose tracebacks contained a lot of ungrounded negative frames (e.g., *No PRO it, Not PRO*).

4.3.4. *Multi-operation utterances.* As noted above the proportion of 2;0 tracebacks requiring more than one operation goes up with increasing MLU (Brian 3%, Fraser 14%, Annie 17%, and Eleanor 21%). Brian's tracebacks have only one multi-operation at 1;8, roughly the same proportion as Annie at 2;6 and considerably fewer than Eleanor at 2;7.

Most of these require only two operations (from 77% for Annie and Eleanor's tracebacks to 96% for Brian's at 2;7) and there are only 4 utterances in all that require 4 or more. Multi-operation tracebacks can be divided into three main types. First there are the units, discussed above, of the form *X, UTT* or *UTT, X* where X is usually a vocative or conjunction and the UTTERANCE itself contains a slot (often a REFERENT slot). These constitute 42 percent of Fraser's multi-operation tracebacks but about 25 percent of Annie and Eleanor's and 12–15 percent of Brian's at 2;6 and 2;7. A second type is where the component units contain more than one REFERENT slot. The proportion of these ranges from 20 percent for Fraser's traceback to roughly a third for all other tracebacks except Brian's at 2;6 where 61 percent of the multi-operation tracebacks contain only REFERENT slots. (see Section 3.4. below). Finally there are tracebacks containing units other than just UTTERANCE and/or REFERENT slots, occasionally

LOCATIONS or ATTRIBUTES but more often combinations of REFERENT and PROCESS slots. In the case of some of the more complex multi-operation tracebacks, we have the impression that, in fact, the utterance is being repeated as a whole and we have just not managed to pick it up: for instance, the only 6-operation traceback in the entire study, Annie's *Mummy's just trying to lie down on the bed here*, sounds exactly the sort of thing her mother would say as a whole utterance! With only 7–10 percent of the child's speech, we are obviously not going to be able to find straightforward tracebacks for every utterance, though we also have to recognise that a larger corpus of test utterances might also have contained more utterances requiring multi-operation tracebacks.

4.3.5. *Development.* The tracebacks show an overwhelming preponderance of REFERENT slots with PROCESS slots more evident at higher MLUs. This is unlikely to simply be due to an increase in MLU, however, since the semantics of the slots is not a function of string length. As we have seen, slots can be filled with single- or multi-word strings. The proportion of UTTERANCE slots varies between tracebacks, but tracebacks to UTTERANCE slots look somewhat more sophisticated at higher MLUs, with conjunctions and adverbials as opposed to simple vocatives and interjections.

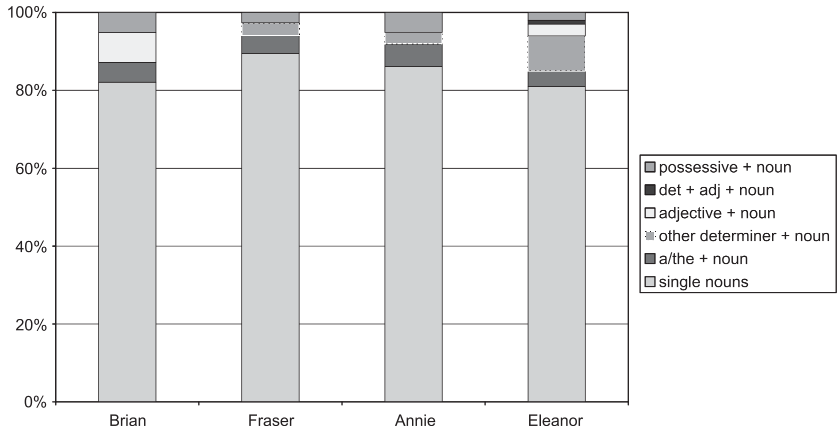
4.4. *Building the noun phrase*

We have seen referents are by far the most frequent category of slots. Here we ask two questions: first, do strings traced back to REFERENT slots change with increasing MLU and, second, does the syntax of REFERENT slots in these tracebacks change with increasing MLU?

4.4.1. *Types of REFERENT string.* As we can see from Figure 3a, between 70–85 percent of all REFERENT slots are filled with single nouns or pronouns at 2;0. All the children's tracebacks at 2;0 have a few 'possessive + noun' strings, and the proportion of strings with determiners other than *a* and *the* goes up with MLU. Brian's tracebacks always contain some 'adjective + noun' strings. Simple 'determiner + noun' strings also form the majority of Annie and Eleanor's multiword REFERENT strings but Eleanor's also show a wider range of other string types, for instance, of 'other determiners + nouns' (e.g., *those REF, lots of dinosaurs*).

4.4.2. *Grounding in REFERENT slots.* We have seen that the range of REFERENT strings in the tracebacks increases with increasing MLU but does this mean that their syntax is also more 'NP-like' at higher MLUs?

(a) Four children at 2;0



(b) Brian developmental

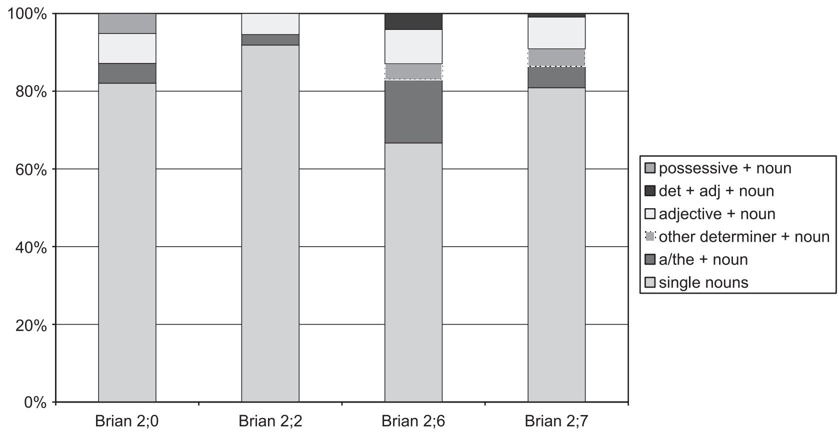


Figure 3. *Types of fillers in* REFERENT *slots*

When a REFERENT string is combined with a schema, is the outcome a 'grounded' noun phrase? Much of the time schemas are already grounded for instance because there is a determiner prior to the slot (e.g., *Where's the* REF or *There's a* REF). So we first identified all ungrounded slots in schemas and then analysed whether these were filled with grounded or ungrounded strings. Figure 4 presents these data.

We can see that in Annie and Eleanor's tracebacks and Brian's at his two higher MLUs, over 80 percent of slots in ungrounded frames become grounded when filled. The figures for the four tracebacks at lower MLUs are between 62–74 percent. This is largely due to the fact that bare nouns are more likely to fill slots at lower MLUs, though note that at the same MLUs repectively as Fraser and Eleanor, Brian's tracebacks ground fewer REFERENT slots.

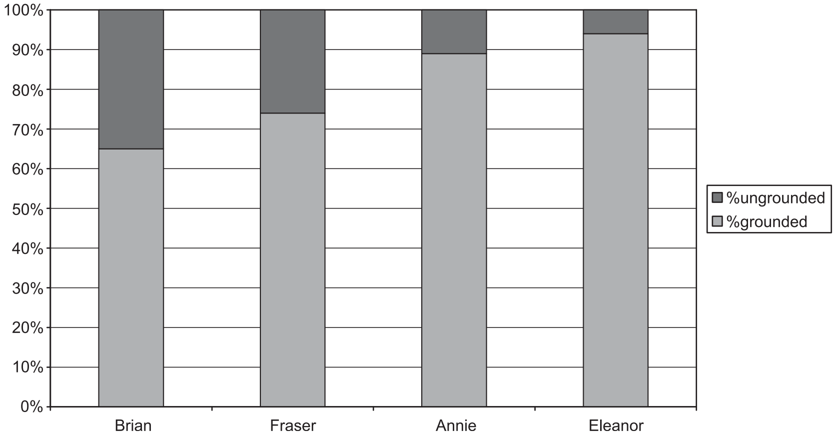
5. Discussion

In this study we identified the multiword utterance types said by four two-year-old children in the last two hours of a 6-week corpus collected from their second birthday and from three further corpora collected from one of the children over the next 6 months. We found that the children's corpora contained a greater proportion of multiword types with increasing MLU, suggesting that maintaining conversation by repeating whole utterances may reduce with development. Of these types, 58–92 percent could be traced back either to exact repeats of what the children had said at least once in their main corpora or to an utterance that required only one operation (almost always a substitution into a semantically similar slot) to arrive at a match. Exact repeats were less likely at higher MLUs and the proportion of trackbacks requiring more than one operation increased with MLU. The great majority of slots created in the trackbacks were for REFERENTS, very largely single nouns or pronouns, with the proportion of slots for PROCESSes, increasing with MLU. For multiword REFERENT strings, there was a wider range of determiners at higher MLUs. We first discuss issues relating to the trackback method before turning to the implications of our findings.

5.1. *Issues with the trackback method*

One major issue is that of a baseline with which to compare these results. If we conducted trackbacks on the mothers' corpora, would we find similar levels of lexical specificity? If so, since we know that the mothers have a more schematic and abstract grammar, we could not necessarily conclude that the children do not. We know that adult speech to two-year-olds is also very repetitive (Cameron-Faulkner, Lieven and Tomasello 2003; Stoll, Abbot-Smith and Lieven, in print). However, since the degree of repetitiveness in the children's utterances reduces with increasing MLU and with age, this is not the whole story. In Lieven et al. (2003) when the mother's multiword utterance types were traced back on her corpus,

(a) Four children at 2;0



(b) Brian developmental

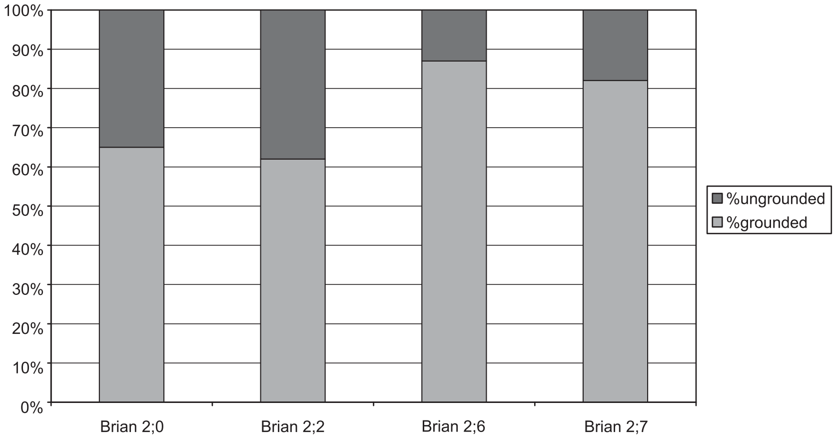


Figure 4. *Grounding of ungrounded* REFERENT *slots*

many fewer prior schemas and many more multi-operation tracebacks were found. In addition, the three operations that were hardly ever used for the child's traceback ('insert', 'drop' and 'rearrange') were needed. However the mother's corpus was not controlled, by comparison with the child's, for sample size, vocabulary or utterance length, and we cannot therefore be sure that, were we to make these controls, the caretakers' tracebacks might not look equally lexically specific, especially given the rather routinised contexts of play and interaction at this age. Since controlling for each of the above factors would have differing effects on our traceback measures, developing such baselines is a complex matter and beyond the scope of this paper. But we can point out that limiting the adult speech in this way is precisely what we did not do for the children's corpora and, therefore, our working assumption is that their corpora are better samples of everything that they can do with language than are those of the adults speaking to them.

A related issue with traceback as implemented here is that the sizes of the children's test and main corpora vary in the number of words and utterances. Without developing the types of controls mentioned above, we cannot be sure that the differences we have found do not result from corpus size differences. For instance, the size of the main corpus could impact the numbers of operations needed to trace an utterance back, with more utterances or words in the main corpus leading to more exact matches and fewer multi-operation tracebacks. However, we can note that each corpus of Brian's at an older age contains more words and utterances than the previous one despite the fact that, between his first two, and last two, tracebacks, the number of exact matches goes down and the number of multi-operation tracebacks goes up. Equally, Fraser has a much larger main corpus than Brian at the same MLU but somewhat fewer exact matches and more multi-operation tracebacks.

In previous tracebacks, we have usually included the caretakers' speech when tracing back the children's utterances and used a criterion of two precedents in the main corpus for both schemas and fixed strings. Clearly the more data one takes out of the main corpus, the more fails there will be. When Dąbrowska and Lieven (2005) compared tracebacks of syntactic questions for Brian and Annie at 2;0 and 3;0, they found a 15 percent reduction in successful tracebacks if the input was removed but the requirement for 2 precedents was also relaxed to 1, and a further 12 percent reduction if the requirement for 2 precedents was maintained. Even on the most restricted main corpus, therefore, they still managed to successfully trace back 62 percent of the children's novel utterances. Although the study was only of syntactic questions which may show more lexical specificity than the full range of utterances that a child can produce, this result was still quite striking because two of the four corpora were collected when the children were 3;0 and tracebacks were somewhat more complex. In the present study, without tracing back to the input, and using the criterion of one precedent for schemas and fixed strings in the main corpus, we successfully traced back 85 percent of the 2;0 utterances and 90 percent of Brian's utterances in the MLU-matched corpora[7]. For exact matches and one-operation tracebacks, the figures are 70 percent and 84 percent respectively. These results certainly suggest a very close relationship between children's utterance production at 2;0 and what they themselves have said before.

The final methodological issue is that the whole analysis is of course constrained by what the child produces in the test corpus. Since we only trace back from the multiword utterance types that we find in the test corpus, it is possible that what we find is not representative of the full range of the child's linguistic repertoire. One solution to this problem would be to increase the size of the test corpus, but the intrinsic problem would always remain. A more radical alternative is to 'trace forward', i.e., to extract a grammar from the main corpus and then see if it can predict the novel utterances that the child produces in the test corpus. We are currently developing this method.

5.2. *Children's grammars and the usage-based approach*

However, despite the methodological issues raised above, the broad pattern of our results fits well with other studies coming from within a usage-based approach which suggest that what children say is closely related to what they have said previously, not only because of extraneous factors such as the repetitiveness of interactions and of life for a two-year-old but precisely because this is how they build up their grammars. In particular, while, of course, each of these utterances could have been generated by highly abstract algorithms, there is no necessity for such a proposal. We are not suggesting that the children's utterances are actually constructed by the operations given here, nor that the schemas and fixed strings that are identified are necessarily present in the child's linguistic representations—this would require a very different approach, possibly experimental, and focussing on the phonetic characteristics of each utterance. Just because an utterance can be traced back to a fixed string in the main corpus does not mean that this string itself was not produced from a schematic construction. In order to arrive at one traceback for each utterance, we used a particular set of criteria outlined in the method section,

7. Note that we cannot make a strict comparison of these figures because Dąbrowska and Lieven (2005) also used the SUPERIMPOSE operation—an extended form of substitution—while the present study only allowed direct substitution into slots.

but changing these could have meant a somewhat different traceback for a number of utterances. For instance Annie's *Where's Mummy?* was an exact match to an utterance in the main corpus but could have been traced back to a *Where's* REF*?* schema which was also found in other of her tracebacks. In principle, these various possibilities might be distinguished by using a prosodic analysis to compare utterances of similar form and seeing whether fully lexically-specific strings can be separated from schematic strings, using a method similar to that of Bybee and Scheibman (1999) for *I dunno* (see also Verhagen 2002). But, from a cognitive-linguistic perspective, both forms could be present in the child's construction inventory, the more frequent and entrenched as fully lexically specified, and the schema as emergent. Indeed this is what our data suggests: with increasing MLU and age, the children's tracebacks involved more schematic constructions with a wider range of slots of increasing complexity as the analysis of REFERENT slots shows.

The development of utterances with more than one slot and the development of PROCESS slots, would greatly increase the child's ability to produce novel utterances and, at the same time, schematise constituents in ways that allow more flexibility. The schemas we found in the tracebacks with PROCESS slots after lexically-specific auxiliaries and semi-auxiliaries are very similar to the early schemas identified in Lieven's study of 6 children's auxiliary development between 2;0–3;0 (Lieven 2008). Over the year of that study these initially independent schemas developed into a more interconnected and abstract auxiliary syntax. An example of increasing flexibility with the internal structure of utterance production is Annie's use of a *Just* UTT schema at 2;0. In almost every case, *just* occurs at the beginnings of utterances in Annie's 2;0 corpus but by 3;0, she can use *just* flexibly before a variety of utterance-internal constituents (*She does just like it like that, My boots are just up to there, When we leave them just now they will cry just now*).

That children learn multiword strings and not just single words for subsequent assembly is suggested by a recent study by Bannard and Matthews (2008). Children repeated highly frequent strings from the input (e.g., *a drink of milk*) more fluently and with less error than matched strings in which the final word was changed (e.g., *a drink of tea)*. The degree of repetitiveness that we have found suggests that children are initially learning not only words but strings as 'big words' which subsequently start to be internally analysed. If they are sufficiently entrenched they may also remain in the lexicon/grammar as fully lexically specific strings. A number of studies suggest that this internal analysis is an ongoing process that can both give rise to errors and protect the child from error. Thus the studies of wh-questions by Rowland and Pine (2000) and Rowland (2007) show that high-frequency wh-auxiliary combinations in the input are correlated with the production of correctly inverted sentences while children are more likely to make non-inversion errors on low frequency combinations. Kirjavainen, Lieven and Theakston (in press) show that children are more likely to produce accusative for nominative errors in first person subjects before verbs which they have heard more frequently in the input in non-finite constructions (e.g., *Let **me have it***, *Did you see **me doing it***).

Finally we should note that REFERENT slots filled with single nouns or pronouns are by far the most frequent slots created in the tracebacks and this is particularly true of the lower MLU tracebacks. However it is not just a matter of utterance length since both PROCESS and ATTRIBUTE slots could be filled with single words and often were. This certainly suggests that, for English-learning children, it is easiest to schematise a slot for REFERENTS and that, once this is achieved, the child can start to further elaborate this slot linguistically. This initial primacy of referring expressions supports experimental findings showing that English-speaking children are able to place novel nouns that they are taught into utterances well before they are able to do this with novel verbs (Tomasello, Akhtar, Dodson and Rekau 1997).

5.3. *MLU and individual differences*

Although the differences in corpus sizes mean that comparisons between children can only be highly tentative, some of the differences between the four children at the same age, and between the MLU-matched tracebacks are interesting. One possible individual difference is the extent to which children might use utterance repetition to get things said. Brian has both a lower proportion of multiword utterance types in his test corpora, a greater number of exact matches in his tracebacks and fewer multi-operation tracebacks by comparison with Fraser and Eleanor at the same respective MLUs. In Fraser's tracebacks, the placement of vocatives and interjections before exactly repeated utterances is another way in which longer utterances could be produced by greater reliance on memory.

Another possible difference is in terms of the semantics of utterances and slots at similar MLUs. At the same MLU as Fraser, Brian's tracebacks have a narrower range of non-REFERENT slots. In his tracebacks at the same MLUs as Fraser and Eleanor, fewer of his REFERENT slots are grounded, while Eleanor's tracebacks show a wider range of determiners. Finally, more of Brian's PROCESS slots are ungrounded because of his method of producing negated utterances using *no* or *not* before verbs, while Eleanor and Annie's tracebacks have a higher proportion of negative schemas with auxiliaries which, therefore ground the slots (e.g., *Don't* PRO *Can't* PRO*)*. Since Cameron-Faulkner, Lieven and Theakston (2007), in an analysis of Brian's negation between 2;0–3;0, found that *no* and *not* were the only preverbal negators up to 2;6, this suggests that results from the tracebacks may well be reflecting some properties of the childrens' current grammatical systems.

As well as the variety of ways in which a particular utterance can be produced, there may also be differences between individuals in their grammars—their construction inventories. We see here that the children, at the same MLU, may be basing the construction of their utterances on somewhat different schemas. Most of this is undoubtedly due to different stages in development, despite similar MLUs, but adults too, may not arrive at identical grammars and, in particular, may differ in the schematicity of their constructions (Dąbrowska and Street 2006).

5.4. *Conclusion*

In this paper we traced back the children's utterances only to their own previous corpora, and found that these utterances could be very closely related to what the children had said previously. The data is suggestive of the development of schematised slots in constructions, initially for referring expressions, and that a wider and more abstract range of slots develops with increasing language experience. We interpret these results as supporting an account of language development based on learning pieces of language from the input mapped to child-based meanings, with the development of a more schematic and abstract inventory of conventionalised constructions.

Received 17 March 2008
Revision received 23 September 2008

Max Planck Institute for Evolutionary Anthropology, Leipzig

Appendix A: Corpus sizes

(a) at 2;0

	Brian		Fraser		Annie		Eleanor	
	main	test	main	test	main	test	main	test
Nr. of utterances	12,895	1,114	21,022	1,298	12,762	870	10,225	988
Nr. of words	20,010	1,855	35,951	2,475	23,633	1,999	23,100	2,361

(b) Brian developmental

	Brian (age 2,0)		Brian (age 2;2)		Brian (age 2,6)		Brian (age 2,7)	
	main	test	main	test	main	test	main	test
Nr. of utterances	12,895	1,114	13,110	885	14,493	1,228	15,932	1,180
Nr. of words	20,010	1,855	21,286	1,762	26,770	2,462	29,935	2,458

References

Bannard, Colin and Danielle Matthews
2008 Stored word sequences in language learning: The effect of familiarity of children's repetition of four-word sequences. *Psychological Science* 19, 241–248.
Bybee, Joan and Joanne Scheibman
1999 The effect of usage on degrees of constituency: The reduction of 'don't' in English. *Linguistics* 37, 575–596.
Cameron-Faulkner, Thea, Elena Lieven and Michael Tomasello
2003 A construction based analysis of child directed speech. *Cognitive Science* 27(6), 843–873.
Cameron-Faulkner, Thea, Elena Lieven and Anna Theakston
2007 What part of no do children not understand? A usage-based account of multiword negation. *Journal of Child Language* 34, 251–282.
Dąbrowska, Ewa and Elena Lieven
2005 Towards a lexically specific grammar of children's question constructions. *Cognitive Linguistics* 16, 437–474.
Dąbrowska, Ewa and James Street
2006 Individual differences in language attainment: Comprehension of passive sentences by native and non-native English speakers. *Language Sciences* 28, 604–615.
Goldberg, Adele
2006 *Constructions at Work*. Oxford: Oxford University Press.
Kirjavainen, Minna, Elena Lieven and Anna Theakston
in press Can input explain children's me-for-I errors? *Journal of Child Language.*
Langacker, Ronald
1991 *Concept, Image and Symbol*. Berlin: Mouton de Gruyter.
Lieven, Elena
2008 Learning the English auxiliary: A usage-based approach. In Behrens, Heike (ed.), *Trends in Corpus Research: Finding Structure in Data*. Amsterdam: Benjamins, 61–98.
Lieven, Elena
2006 Producing multiword utterances. In Kelly, Barbara and Eve Clark (eds.), *Constructions in Acquisition*. Stanford, CA: CSLI Publications, 83–110.
Lieven, Elena and Michael Tomasello
2008 Children's first language acquisition from a usage-based perspective. In Robinson, P. and Nick Ellis (eds.), *Handbook of Cognitive Linguistics and Second Language Acquisition*. New York: Routledge, 168–196.

Lieven, Elena, Heike Behrens, Jennifer Speares and Michael Tomasello
2003 Early syntactic creativity: A usage-based approach. *Journal of Child Language* 30, 333–370.
Lieven, Elena, Julian Pine and Gillian Baldwin
1997 Lexically-based learning and early grammatical development. *Journal of Child Language* 24, 187–219.
MacWhinney, Brian
2000 *The CHILDES project: Tools for analyzing talk*. Mahwah, NJ: Lawrence Erlbaum.
Miller, Jon and Robin Chapman
1981 The relationship between age and mean length of utterance in morphemes. *Journal of Speech and Hearing Research* 24, 154–161.
Parker, Matthew and Kent Brorson
2005 A comparative study of mean length of utterance in morphemes (MLUm) and mean length of utterance in words (MLUw). *First Language* 25(3), 365–376.
Rowland, Caroline
2007 Explaining errors in children's questions. *Cognition* 104(1), 106–134.
Rowland, Caroline and Julian Pine
2000 Subject-auxiliary inversion errors and wh-question acquisition: What children do know? *Journal of Child Language* 27, 157–181.
Rowland, Caroline, Sarah Fletcher and Daniel Freudenthal
2008 How big is big enough? Assessing the reliability of data from naturalistic samples. In Behrens, Heike (ed.), *Trends in Corpus Research: Finding Structure in Data*. Amsterdam: John Benjamins, 1–24.
Stoll, Sabine, Kirsten Abbot-Smith and Elena Lieven
in print Lexically restricted utterances in Russian, German and English child directed speech. *Cognitive Science* 33(1), 75–103.
Tomasello, Michael
1992 *First Verbs: A Case Study of Early Grammatical Development*. Cambridge: Cambridge University Press.
Tomasello, Michael, Nameera Akhtar, Kelly Dodson and Laura Rekau
1997 Differential productivity in young children's use of nouns and verbs. *Journal of Child Language* 24, 373–387.
Tomasello, Michael
2003 *Constructing a Language: A Usage-based Theory of Language Acquisition*. Harvard, MA: Harvard University Press.
Tomasello, Michael and Daniel Stahl
2004 Sampling children's spontaneous speech: How much is enough? *Journal of Child Language* 31, 101–121.
Verhagen, Arie
2002 From parts to wholes and back again. *Cognitive Linguistics* 13, 403–440.