J. Child Lang. **30** (2003), 333–370. © 2003 Cambridge University Press
DOI: 10.1017/S0305000903005592 Printed in the United Kingdom

Early syntactic creativity: a usage-based approach*

ELENA LIEVEN, HEIKE BEHRENS,
JENNIFER SPEARES AND MICHAEL TOMASELLO
Max Planck Institute for Evolutionary Anthropology, Leipzig

(*Received* 1 *February* 2000. *Revised* 25 *June* 2002)

ABSTRACT

The aim of the current study was to determine the degree to which a sample of one child's creative utterances related to utterances that the child previously produced. The utterances to be accounted for were all of the intelligible, multi-word utterances produced by the child in a single hour of interaction with her mother early in her third year of life (at age 2;1.11). We used a high-density database consisting of 5 hours of recordings per week together with a maternal diary for the previous 6 weeks. Of the 295 multi-word utterances on tape, 37% were 'novel' in the sense that they had not been said in their entirety before. Using a morpheme-matching method, we identified the way(s) in which each novel utterance differed from its closest match in the preceding corpus. In 74% of the cases we required only one operation to match the previous utterance and the great majority of these consisted of the substitution of a word (usually a noun) into a previous utterance or schema. Almost all the other single-operation utterances involved adding a word onto the beginning or end of a previous utterance. 26% of the novel, multi-word utterances required more than one operation to match the closest previous utterance, although many of these only involved a combination of the two operations seen for the single-operation utterances. Some others were, however, more complex to match. The results suggest that the relatively high degree of creativity in early English child language could be at least partially based upon entrenched schemas and a small number of simple operations to modify them. We discuss the implications of

[*] Our thanks to the mother and child who made this study possible. We also thank Dimitra Doumpioti who served as research assistant on the project, Heike Forwergk who did much of the initial analysis, Anna Theakston for back-up support, and Elaine Cooper and Lorna Burke for their transcribing. For helpful comments on the manuscript we thank Nameera Akhtar, Shanley Allen, Ewa Dabrowska, Michael Israel, Kelly Jaakkola, Ann Peters, Julian Pine and two anonymous referees. Address for correspondence: Elena Lieven, Max Planck Institute for Evolutionary Anthropology, Inselstrasse 22, D-04103, Leipzig, Germany. e-mail: lieven@eva.mpg.de

https://doi.org/10.1017/S0305000903005592 Published online by Cambridge University Press

these results for the interplay in language production between strings registered in memory and categorial knowledge.

INTRODUCTION

When adults produce speech there is, presumably, a complex interplay between generating each utterance, or part of an utterance, afresh and retrieving a multi-word unit which has become automatized. The question of how this interplay between assembling units from more general and abstract underlying representations, and producing the units as 'prefabricated chunks' works is not well understood either for issues of real time speech production or for its implications for underlying language representation. Clearly, adults have a powerful generative system for producing utterances which depends in part on representations which are abstract: i.e. not tied to specific words or narrow meanings. However, a theory of language representation that is intended to form the basis for language production has also to deal with the fact that speech contains many more or less 'frozen' strings, the meaning of which cannot be straightforwardly recovered from the words or constructions into which they could be analysed using an abstract grammar. There is a growing body of evidence that, for adults, both forms of storage and access coexist, such that speakers access highly frequent units by 'shortcut' without having to construct them from scratch. Eventually, over time in the linguistic community, the link to the underlying parts may become lost. Such processes become visible in 'misparsings' and diacronic change (see, for example, Bybee & Scheibmann, 1999).

This is dealt with differently as a function of the particular theory of grammar. For 'usage-based' theorists (e.g. Langacker, 1987; Bybee, 1995; Croft, 2001) a particular utterance will be multiply represented at levels in the speaker's system ranging from the lexical specificity of the actual utterance through to its place in an abstract network of related constructions and paradigms. Exactly how it is represented and at which levels will depend on issues of type and token frequency. More recently, theorists working from perspectives that start from the assumption of an abstract and generative universal grammar quite separate from the lexicon, have also been centrally concerned with the relationship of idioms, partial idioms and low-scope constructions to this grammar (e.g. Lebeaux, 1988; Jackendoff, 1996; Culicover, 1999).

The connection between these issues and children's early syntactic development has not often been made but may, in fact, be very relevant. Researchers have shown that much of children's early language can be analysed in terms of specific linguistic items and phrases (e.g. Clark, 1974; Braine, 1976; Ewing, 1982; Peters, 1983; Lieven, Pine & Dresner Barnes, 1992; Tomasello, 1992; Pine & Lieven, 1993). These researchers agree that children's utterances are constructed in a variety of ways of which rote-learning forms

https://doi.org/10.1017/S0305000903005592 Published online by Cambridge University Press

one end of the continuum and the generation of utterances from abstract categories and rules forms the other, with a variety of semi-formulaic patterns lying somewhere in between. All also agree that the frequency of particular sequences in the language the child hears will partially determine whether a string is learned as an unanalysed whole or as a 'slot-and-frame' formula. Some examples are Brown's (1973) low-scope 'operations of reference', Kuczaj's (1979) observation that for some children the cliticized copula in *It's a X* might be unanalysed, and Pinker's (1989) decision to exclude utterances containing concatenated negative auxiliaries (*can't*, *don't*) from his study on the grounds that they are probably unanalysed. However, the significance of these unanalysed and semi-formulaic utterances varies for researchers in ways that are similar to those outlined in the paragraph above. For those who see grammar as an emergent property of the child's development, abstract representations arise from an interaction between the learning of these specific frames, their increasing connectedness to other frames and socio-cognitive development (Tomasello, 2000). For others, the relevance is viewed more as a methodological problem: if some strings are learned as fully or partially unanalysed wholes, how can we distinguish these from strings that are generated from the already existing underlying grammar (Radford, 1990).

Corpus data cannot resolve the question of whether a string is learned as a whole or assembled – this requires experimentation under conditions that control what the child has heard. What researchers can do with corpus data and have traditionally done, is to relate a particular string to what the child has said before in order to estimate the likelihood that the string is rote-learned or partially lexically-specific. Sometimes this is done using a fairly *ad hoc* procedure, other times it is more systematic. For instance, Pine & Lieven (1993) and Lieven, Pine & Baldwin (1997) used explicit criteria for 'frozen', 'semi-constructed' and 'constructed' (outlined in Lieven *et al.*, 1992) to analyse the relation of each multi-word utterance to every previously occurring utterance containing the same words in the children's corpora. These studies showed that a very large proportion of these multi-word utterances could be related to small numbers of lexically based patterns. While these types of studies cannot, of course, prove that this is actually how the child produces her/his utterances, they can yield important information about the 'natural history' of particular utterances and lexemes in terms of their frequency and range of use. In principle this can then be used to inform decisions about the definition of productivity and the relative likelihood that the string is based more on utterance level material retrieved from memory or built up from underlying categories in the child's current linguistic representations.

However, the usual sampling frame for these and almost all other studies makes the quality of the information gained about the history of an utterance

https://doi.org/10.1017/S0305000903005592 Published online by Cambridge University Press

and the decisions about how to define its status extremely problematic. In most existing databases, the sampling intervals are too far apart to determine what is creative in any given child utterance. Assuming that a child is awake and talking 8–10 hours per day, biweekly or weekly data collection intervals means that we are able to examine something like 1–2 per cent of her/his language during the period of sampling.

With these considerations in mind, our approach in the current study was to record the language of one two-year-old child using dense taping intervals (supplemented with maternal diary notes). We taped this child's linguistic interactions with her mother for one hour per day, 5 days per week, for 6 weeks – making the taped data roughly 5 to 10 times denser than most existing databases of child language. Using reasonable assumptions about how many hours per day this child was awake and talking (and given that the taping sessions were 'quality' conversational time but only for 5 days per week), we estimate that we have recorded on tape close to 10% of all the utterances that the child produced during this 6 week period.

We took as our target all of the child utterances produced in the last one-hour taping session at the end of the 6 week period. We then searched the database going back through the entire corpus and attempted to determine the relationship of each of these utterances to what the child had said before. Was the utterance an immediate repeat of what either the mother or child had just said? Had the child said exactly the same utterance before in previous taping sessions? Or was the utterance novel in terms of the corpus we had available? If the utterance was novel, what was its relationship to previous utterances?

There is a large number of ways in which one might set about answering this last question. For the present study we used a procedure in which we searched for the closest match to the novel utterance in the preceding corpus. Having found this closest match, we analysed the ways in which the novel utterance differed from it. If a number of prior utterances had occurred that matched the novel utterance in the same way, with variation in the same position, we identified this as a schema with a slot. Each novel utterance is therefore related, first, to the utterance in the previous corpus which shares the greatest number of morphemes sequenced in the same way and, second, to all other utterances with a similar sequence which show variation from the closest match at the same point as the novel utterance. In principle this gives some indication of the flexibility with which the child can both manipulate prefabricated sequences and substitute words or phrases into slots. In a subsequent analysis, we took this a step further by identifying two potential 'constituents' in the novel utterances that, in the adult grammar, would be NPs and PPs. We felt that allowing this limited degree of constituency would make for an interesting comparison with the mechanical matching scheme.

https://doi.org/10.1017/S0305000903005592 Published online by Cambridge University Press

In the current study, we make no claims about the nature of the child's 'stored linguistic experience' in terms of either its level of abstractness or its functional-semantic basis. We perform our analysis with the specific linguistic forms and schemas used by the child as the point of departure and look at the extent to which this child's utterances can or cannot be related to what she has said before.

METHOD

Participants

The child was an English-speaking girl, Annie,[1] who lived in a large metropolitan area in England. Annie's parents were both university-educated professionals, and the only sibling was a teenage brother. Annie's mother spent virtually all day with her daughter. Annie was 2;0 at the beginning of the study. She was a relatively precocious language learner, with a MacArthur CDI vocabulary at 2;0 of 391, just below the 75th percentile, and an MLU from 1·95 (first tape at 2;0) to 2·72 (last tape at 2;1.11). The mother was employed by the investigators as a research assistant during the course of the study.

Observational procedure

Beginning on Annie's second birthday, her mother made one-hour tapes of herself and Annie in relatively typical interactions in their home five days per week over a period of 6 weeks.[2] Audiotapes were recorded four days per week, and a videotape was made once a week (with a research assistant also present) – mostly on weekdays but occasionally on a weekend. The tapes were most typically recorded during the morning hours, but sometimes at other times of the day, and on some days two half-hour tapes were made during different parts of the day. The tape recorder was always stationary in one room in the house, with two cordless directional microphones on opposite sides of the room. The most typical activities during taping were playing with toys and/or having a snack.

During this same 6-week period the mother also kept a diary of all of the child's syntactically 'new' and 'most complex' utterances ('emergent structure' as defined by Braunwald & Brislin, 1979, and Tomasello, 1992).

For purposes of the current study we focused on the last day of the study, when Annie was 2;01.11. The taping session on that day was a typical session, with recording taking place continuously for one hour during the morning in a dining room/playroom. During the rest of the day, the mother kept the diary as usual.

[1] All proper names have been changed.
[2] The analysis is based on 28 tapes: illness prevented 2 sessions of recording.

https://doi.org/10.1017/S0305000903005592 Published online by Cambridge University Press

Analytic procedures

Transcription. Research assistants transcribed all of the tapes using standard CHAT procedures (MacWhinney & Snow, 1990). Transcribers were specifically trained to deal with the following:

- Intonation contours were used to judge the beginning and end of utterances.
- We kept the phonetic transcription of schwa for reduced central vowels and refrained from attributing adult word meanings.[3]

Transcription was 'morphemized' to allow the subsequent application of the MOR automatic parser of the CHILDES system.[4] The following morphemes were transcribed: *-s* (plural), *-'s* (possessive), *-'am*, *-'is*, *-'are* (BE copula, auxiliary), *'ll* (*shall*, *will*), *-es* (3rd person singular), *-ed* (past tense), *-ing* (progressive), *-'has*, *-'have*, *-'had* (HAVE auxiliary), *-'nt* (contracted negation), *-'us* (in *Let's*). For homonyms such as *Daddys*, transcribers were instructed to use context for morphemizing into, for example, *Daddy's*, *Daddy-s* (plural), *Daddy-'is* and *Daddy-'has*. For utterances that are ambiguous between a perfect and copula reading (e.g. *It's gone*, *It's finished*), transcribers were instructed to morphemize for the perfect unless context made it clear that the other meaning was intended.

All transcripts were checked for consistency by the third author. Because of its special importance, the tape for the last day was checked especially carefully by the authors together, and a few corrections were made.

Identifying novel utterances. The total number of child utterances on tape and in the diary on the last day was 537. Interrupted and untranscribable utterances, and utterances which the transcriber had marked as containing uncertain transcriptions of words were excluded. There were 231 one-word utterances on the tape and one in the diary. This left for analysis 295 multi-word child utterances on tape and 10 multi-word utterances in the mother's diary for that day (see Table 1).

In the 295 intelligible, multi-word utterances on tape, we first identified those utterances that were immediate imitations or self-repetitions following standard CHAT procedures: exact repetitions or reduced repetitions (exact repetitions of a subpart of the utterance) of one of the last 5 utterances of the mother or child. The remainder of the utterances on this tape were potentially novel. For each of these we searched through the entire corpus leading up

[3] If the transcriber thinks that the schwa approximates a morpheme, this is transcribed as such i.e. in CHAT format: *a@sc* or *the@sc*. Otherwise it is transcribed as @sc. When utterances are cited, this method of transcribing the original schwas has been retained. However, for reasons of readability, the transcribed morphemizations in utterances have been removed.

[4] The transcription conventions for these morphemes can be found on pp. 54–56 of the CHILDES manual (MacWhinney, 2000).

https://doi.org/10.1017/S0305000903005592 Published online by Cambridge University Press

TABLE 1. *Numbers of different types of utterances for Annie at* 2;01.11

Utterance type	Number
N utterances	537
N single-word diary utterances	1
N multi-word diary utterances	10
N single-word utterances on tape	231
N multi-word utterances on tape	295

to the last day for previous uses of each of the morphemes in each utterance. Some of the utterances, including one diary utterance, turned out to have been said in full on previous occasions in the past, but others were truly novel in the sense that Annie had never before produced that exact multi-word utterance during the previous 6 weeks of recording. Utterances spoken similarly but transcribed with different morphemizations are treated as different. Restarted utterances do not count as separate words: thus *more* [/] *more here* is analysed as *more here*.

Finding the closest match to the novel utterance. We were concerned to define: (1) the closest prior utterance to the target novel utterance; (2) the changes that would be required to turn the closest prior utterance into the target novel utterance; (3) the frequency with which parts of the target novel utterance had been said before in similar combinations.

First, we looked for a positive match between every morpheme in the novel utterance and one or more preceding utterances. For this purpose utterances starting with vocatives such as *Mommy*, and negatives such as *No*, were ignored if they had been transcribed with the vocative comma. The transcribers were instructed to do this if there was a clear intonation break between the vocative and the main utterance.

This procedure could result in one or more closest prior utterances. We started by finding the prior utterance(s) that contained the greatest number of morphemes in common with the novel utterance. If there was a tie, we took the order of morphemes and the frequency of previous utterances of similar form into account as summarized in the procedure outlined below.

- Step 1: In the case of an exact positive morpheme match and only one preceding utterance, this utterance became the closest match.

Example 1:
Target utterance 7: I got the butter
Match: I got the door

Morphemes in common are underlined. Note that this method of matching morphemes ignores the potentially rote nature of some morpheme combinations. Closest matches for 72% of the novel utterances were found using this step. If there were a number of utterances with the same positive

https://doi.org/10.1017/S0305000903005592 Published online by Cambridge University Press

morpheme match and the same overall number of morphemes, this was defined as a schema.

Defining a schema. If there were a number of utterances with the same positive morpheme match, the same overall number of morphemes and type variation in the same position as the target novel utterance, this was defined as a schema and the position of the morphemes that varied was noted with W for each slot (see Example 5 below). Note that the identification of a schema depends on exact morpheme matching. Thus in Example 2 below, the matching schema is *Where's Annie's W?* (where W = one morpheme/word) and not, for instance, *Where's X's Y?* (where Y = any number of morphemes/words). Similarly in Example 3, the schema is *Want a W* and not *Want Y*. The schema (i.e. exact positive morpheme match and exact match with overall number of morphemes) was defined as the closest match to the target utterance. See Example 5 below.

In the case that Step 1 produced a tie between two preceding utterances or schemas, which shared the same number of morphemes, but not the same morphemes with the target utterance, Step 2 was implemented.

- Step 2: The utterance with the most CONSECUTIVE morphemes that matched the target utterance is defined as the closest match.

Example 2:
Target utterance 10: Where's Annie's plate?
Matches: (1) Where's Annie's W? = CLOSEST MATCH
(2) Where's Mummy's plate?

W indicates a single morpheme slot in a schema (see below). Each of the matches in Example 2 shares 4 morphemes with the target but in the first, these are consecutive and so this is defined as the closest match.

A further 14% of closest matches were found using this step. In the cases where there was still more than one prior match, frequency was used to define the closest match.

- Step 3: The utterance with the most frequent exemplars of the matching morphemes is the closest match.

Example 3:
Target utterance 6: Want a knife
Matches: (1) Want a W (24) = CLOSEST MATCH
(2) Not a knife (1)

This resolved a further 13% of the novel utterances.

If there was a tie both between the positively matching consecutive morphemes and in the frequency of the prior utterances, Step 4 was implemented.

- Step 4: Prior utterances with an exact match in the OVERALL number of morphemes (matching and non-matching) are defined as the closest matches.

https://doi.org/10.1017/S0305000903005592 Published online by Cambridge University Press

Example 4:
Target utterance 5: I have some toast
Matches on number and order of morphemes and frequency:
(1) I have some coke (1) CLOSEST MATCH
(2) Nan and Dad have some toast (1)

This was the only utterance that needed to be resolved using this criterion.

Defining the relationship between the target novel utterance and its closest match. Having identified the closest utterance or schema for each of the novel utterances, the question was then: how would this have to be changed to arrive at the target novel utterance? In some cases the substitution of a new word into the slot of an established utterance schema or the addition of a word or previously existing phrase to an utterance or schema was required. We called these single-operation utterances. For example, we analysed Utterance 92 (U92), *I want a paper*, as coming from the well-established utterance schema *I want a W* (occurring 16 times in the past) with the word *paper* (occurring 15 times previously) substituted. Similarly, we analysed U23, *It's burning here* as coming from the utterance *It's burning* (said 8 times previously) to which the high frequency locative *here* was added. In other cases, however, more was needed to get from the closest previous utterance to the novel utterance; we designated these as multiple-operation utterances. For example, for U67, *I want a@sc crying tissue*, *crying* and *tissue* have each to be substituted in separate operations – since Annie had never said this phrase before (although she had used both components on previous occasions before, *crying* 14 times and *tissue*, 6):

Example 5: Utterance 67: *I want* a@sc *crying tissue*

I want a@sc	crying	tissue
↓	↓	↓
I want a@sc this one (1)	crying (14)	tissue (6)
I want a@sc Annie drink (1)		
I want a@sc W (17)		

Had she said *crying* and *tissue* together before, this would have counted as only one substitution into the schema *I want a@sc W* (which had been said 17 times previously). However, *crying* and *tissue* had only occurred separately, thus the closest matching schema was, in fact, *I want a@sc W1 W2* (said on two previous occasions: *I want a@sc this one*, *I want a Annie drink*).

As well as substituting or adding on a previously produced morpheme, word or phrase, the utterances could also differ through insertions between words which had always occurred together before in the closest matching utterance and/or by one or more morphemes or words being dropped in the target utterance by contrast with its closest match. Finally, the target and its

https://doi.org/10.1017/S0305000903005592 Published online by Cambridge University Press

closest match could differ through a re-arrangement of items that were in both utterances but in a different order.

The following is the set of possible ways in which the target utterance could differ from its closest match:

SUBSTITUTE – a morpheme, word or group of words (if the words have occurred together previously as a group), replaces a word or fills a slot.

Example 6: Target utterance: I want a paper
Closest match: I want a W **SUBSTITUTE *paper***

ADD ON – a morpheme, word or group of words (if the words have occurred together previously as a group) is added on to the beginning or end of a previous utterance or schema (including affixation and stem changes).

Example 7: Target utterance: Let's move it around
Closest match: Let's move it **ADD ON *around***

DROP – a morpheme, word or group of words (if the words have occurred together previously as a group) is dropped from a previous utterance or schema. Any consecutive words or phrases dropped from an utterance is counted as only one DROP operation.

Example 8: Target utterance: And horse
Closest match: And a horse **DROP *a***

INSERT – a morpheme, word or group of words (if the words have occurred together previously as a group) is inserted into a previous utterance or schema between two words that have previously always occurred together.

Example 9: Target utterance: Have you finished with your book?
Closest match: Have you finished your book? **INSERT *with***

REARRANGE – two single- or multi-morpheme strings (if the morphemes have occurred together before as a string) in a previous utterance or schema, exchange position.

Example 10: Target utterance: Away it goes
Closest match: It goes away **REARRANGE *away***

If there was an exact match in terms of length, then this utterance was always taken, rather than defining the utterance as an instance of an ADD ON. Thus for the novel utterance *Right here* there were many instances of the schema *W here*, which was therefore identified as the closest match. The operation leading to the novel utterance then became a SUBSTITUTE rather than an ADD ON of *Right* to *here*.

In summary, two sets of mechanical procedures were used: (1) a morpheme matching procedure to identify a prior utterance or schema as closest to the

https://doi.org/10.1017/S0305000903005592 Published online by Cambridge University Press

novel utterance, (2) an 'operations' procedure to identify the nature of the differences between the novel utterance and its closest match. The utterances were first coded by the third author and each coding was then checked by the other authors.

Using this procedure we first made a quantitative assessment of the relationship between Annie's 295 intelligible, multi-word utterances spoken on the last day of taping (2;01.11) together with the 10 multi-word diary entries from that same day and the utterances she had said before in the previous 6 weeks (Analysis 1).

Defining a level of 'constituency'. In a second analysis, we altered this matching procedure to one in which the morpheme matching requirement was relaxed to allow for the possibility that the child's productions might show some degree of 'constituency' with 'noun phrases' and 'locative prepositional phrases'. While the precise representations that children of this age have of NPs and locative phrases need further research, there is not much doubt that they know how to refer to objects and locations in a fairly flexible manner and it therefore seemed reasonable to assume some semantically or conceptually limited constituency without making too many assumptions about the generality of the constituents present at this point in development. Details are provided below (Analysis 2).

RESULTS

The overall level of novel utterances

Of the 232 single-word utterances spoken on the last day of taping at 2;01.11, only 4 were novel. Of the 295 multi-word utterances on tape, 72 had been said before at some point during the previous 6 weeks of taping, 28 were immediate imitations of the mother and 86 were exact or reduced self-repetitions. These utterances therefore account for 63% of all the multi-word utterances that Annie produced during the hour-long target taping session. The remaining 109 (37%) of Annie's taped multi-word utterances were novel. Of the diary utterances, one had been said before and 9 were novel. All 118 novel utterances and their codings can be found in the following tables.

We checked to see that there was nothing particularly unusual about this session by making the same initial analysis for the 4 preceding sessions. The number of novel utterances in this last session is broadly in line with the numbers in the previous four sessions with proportions varying between 35%–47%. Thus between about one third and one half of Annie's multi word utterances at this age are novel in that they have never been said in this exact way in our relatively dense sample of her speech over the previous 6 weeks (although, of course, our sample is only about 10%).

https://doi.org/10.1017/S0305000903005592 Published online by Cambridge University Press

TABLE 2. *Operations required to change the target novel utterance to its closest match*

Number of operations	On TAPE			On DIARY		
	Operations	Number	Per cent	Operations	Number	Per cent
1	SUBSTITUTE*	68	62			
	ADD-ON	12	11	ADD-ON	1	
	DROP	1	1	INSERT	1	
	Total: single-operations	81	74		2	22
2		22	20		3	
3 or more		6	6		4	
	Total: multiple-operations	28	26		7	78
	Total	109	100		9	100

* U36, *Daddy's big*, remained a tie after all four steps outlined in the Method section had been applied: *Daddy's big* (tie: *Daddy's W* (6) *W's big* (6)) Both of the tied utterances required a single substitution and this is how it has been counted.

Analysis 1: matching on a morpheme-by-morpheme basis

Table 2 shows the operations required to find the closest matching utterance for each of the 109 taped novel utterances on the last day of recording. 74% of the novel utterances differed from something Annie had previously produced in our 6-week taped sample by a single-operation, with an additional 26% utterances requiring multiple-operations. The single-operations consisted of either SUBSTITUTIONS (the great majority) or ADD-ONS with the exception of one utterance, which required a DROP (U90), and one diary utterance, D1, which required an INSERT. Thus, in the 81 single-operation utterances on tape, it was never necessary to insert a word between two words which had always occurred adjacently in previous utterances or to rearrange the order of words in a previous utterance that otherwise matched the novel utterance. One word could either be substituted for another, often into a slot in a pre-existing schema, or a word could be added on to the beginning or end of a previous utterance in a way that did not interrupt any previously invariant sequential strings. We will consider these single-operation changes in more detail before turning to the utterances requiring multiple-operations to match the closest target utterance.

Single-operation utterances: substituting. Sixty-eight of the utterances requiring one operation to arrive at their closest match were substitutions of one word for another. These are shown in Tables 3 and 4. Forty-seven of these substitutions consisted of placing an object or person label ('noun') where only other 'nouns' had previously occurred before (Table 3). Where utterances appear to be closely related they are listed together (e.g. the utterances with *Where's* and *want*). Otherwise utterances are listed in the order in

https://doi.org/10.1017/S0305000903005592 Published online by Cambridge University Press

TABLE 3. *Single-operation substitutions into noun slots*

Utterance and number	Closest match	Frequency of closest matching schema or utterance	Frequency of substituted item
1 Where's the bus-s?	Where's the W-s	9	0
9 Where's Mummy's plate?	Where's Mummy's cup+of+tea?	1	14
18 Where's Mummy's knife?	Where's Mummy's W?	2	11
73 Where's Mummy's car?	Where's Mummy's W?	3	27
10 Where's Annie's plate?	Where's Annie's W?	2	15
19 Where's Annie's knife?	Where's Annie's W?	3	12
11 Where's the butter?	Where's the W?	53	5
49 Where's the tape?	Where's the W?	54	1
64 Where's the box?	Where's the W?	57	9
80 Where's the hippo?	Where's the W?	58	1
81 Where's the elephant?	Where's the W?	59	2
87 Where's the dog?	Where's the W?	60	2
108 Where's the people?	Where's the W?	61	0
112 Where's the drawing?	Where's the W?	65	13
89 Where's sheep?	Where's W?	114	2
109 Where's my seat?	Where's my W?	52	3
2 Not a hat	Not a W	6	12
4 Hello toast	Hello W	25	28
5 I have some toast	I have some coke	1	29
6 Want a knife	Want a W	24	9
8 I want a toastie	I want a W	14	0
92 I want a paper	I want a W	16	15
22 Mummy wants butter	Mummy wants it	3	9
48 Want a@sc football	Want a@sc W	7	8
60 I want one+two … +six	I want W	44	58
63 I want one+two … + four	I want W	45	59
68 I want tissue	I want W	47	7
84 Want some water	Want some W	4	26
93 I want some water	I want some W	28	28
7 I got the butter	I got the door	1	3
20 I got no Puddleduck	I got no W	2	0
21 Got home	Got W	5	15
16 I don't like jam	I don't like W	2	5
17 Just butter	Just W	18	5
26 The toaster	The W	14	0
88 The sheep	The W	15	1
37 Dave's big	W's big	7	56
42 Cally's big	W's big	8	24
39 And baby smaller	And Annie smaller	1	97
43 Cally smaller	Baby smaller	1	11
45 Two penny-s	Two W-s	20	0
66 I need a tissue	I need a jacket	1	4
82 That's a big lion	That's a big step	1	1
91 That's a pig	That's a W	20	2
115 That's my bag	That's my W	30	23
85 And a chocolate	And a W	12	72
101 There's the rolling+pin	There's the W	7	2
Mean frequency and range		23·0 (1–114)	15·4 (0–97)

https://doi.org/10.1017/S0305000903005592 Published online by Cambridge University Press

TABLE 4. *Substitutions into non-noun slots*

A. Non-noun substitutions with more than one exemplar

Utterance and number	Closest match	Frequency of closest matching schema	Frequency of substituted item
32. It's upside+down	It's W	(88)	1
34. It's lovely		(110)	1
35. Mummy's big	Mummy's W	(14)	66
36. Daddy's big [TIE]	Daddy's W	(6)	67
	W's big	(6)	49
40. Baby's smaller	Baby's W	(5)	16
105. It's not rolling	It's not W-ing	(3)	10
25. Right here	W here	(244)	6
51. From downstairs	W downstairs	(7)	10
61. Six paint-s	W paint-s	(3)	23
3. Mummy, have toast	W toast	(4)	9
111. Going to get the people	Going to get W	(2)	4
Mean frequency and range		44·7 (2–244)	23·8 (1–67)

B. Non-noun substitutions with only one closest matching utterance

Utterance and number	Closest match	Frequency of closest match	Frequency of substituted item
74. I sit on here	I sit on Rocking+Doggy	(1)	980
58. I want to paint	I want to get it	(1)	21
99. I want to roll it	I want to get it	(1)	2
95. I nearly finished	I not finished	(1)	2
96. I want more paper-s	I want all paper-s	(1)	338
54. All on floor	Down here on floor	(1)	46
78. You sorry tiger	I'm sorry tiger	(1)	160
79. This ones go here	This ones Piglet here	(1)	307
44. To Penny	To wind	(3)	1
13. Can have it	I have it	(2)	6
Mean frequency and range		1·3 (1–3)	186 (1–980)

Words underlined indicate that they have been said together before and count as 1 for the purpose of slot filling.

which they were produced during the recording. Table 4 lists novel utterances requiring substitutions of words other than 'nouns'. In both tables the novel utterance and its number is given in the first column. The closest matching previous utterance or schema is shown in the second column and the frequency with which this has previously occurred in the third. The final column shows the token frequencies with which the substituted word has occurred previously in the corpus. If there is more than one example in the session of a novel utterance in that schema, then the frequency of previous occurrences of the schema will increase through the recording session and

https://doi.org/10.1017/S0305000903005592 Published online by Cambridge University Press

this can be seen in the third column (see for example U18 and U73 in the first cell).[5] Where there is only one previous exactly matching utterance (i.e. no schema) the frequency shown in the third column will be 1 (e.g. U9), unless the closest matching utterance has been said in its entirety more than once before (e.g. U22). It can be seen that the closest matching utterance for 50 of these 68 substitutions contains a 'slot' (indicated by W), which tells us that Annie has produced more than one similar utterance previously with variation in the same position as in the novel utterance. These substitutions show the prototypical 'pivot look' characteristic of the early multi-word speech of many children speaking the world's languages (Brown, 1973; Braine, 1976), with copula frames (*There's the W, That's —, Where's —?*) phrases which are probably whole utterances in the adult corpus as well (*Hello W, And a W*), and frames around *got* and *want*, with or without *I* as Subject. The widespread substitution of 'nouns' into these patterns supports our decision to define a category of 'object/person' in Analysis 2 below.

Adding across Tables 3 and 4, the mean frequency of the closest matching utterances is 23·3 (with a median of 7) and of the substituted items is 48·9 (median = 9). Thus most of the material that Annie is using in the novel utterances, identified by our analysis as substitutions, has been said on a number of previous occasions (range 1–244 for the closest matching utterances and 0–980 for the substituted words).

Single-operation utterances: adding-on. The remaining 15 of Annie's single-operation utterances largely involved adding a word onto either the beginning or end of the closest matching previous utterance together with one example each of DROP and INSERT. They are listed in Table 5. Eight of the utterances involve Annie adding extra material to a just preceding utterance (noted as SR in Table 5), which suggests that the characterization of this operation as 'adding-on' may have some psychological reality.[6] Note that the words added on are almost all extremely frequent in the corpus (*put, here, now, I want, no, it's, and*). While the closest utterances are much less frequent, most of them can be related to highly frequent groups of utterances already identified in Tables 3 and 4 – a matter to which we shall return below (see Table 10).

Multiple-operation utterances. There are 28 utterances on tape requiring more than one operation to arrive at the closest matching utterance and 8 diary utterances. These are listed in Table 6. Part A of the table shows the taped utterances in the order in which they were spoken by Annie and the

[5] This will not always increase by 1, since utterances which are exemplars of the schema but have been said before, may intervene between two novel utterances which are also exemplars of the schema.

[6] It would also be interesting to look at the utterances immediately preceding substitutions to see if there is discourse support for these. However, this would have added unacceptably to the length of the present study and waits further analysis.

https://doi.org/10.1017/S0305000903005592 Published online by Cambridge University Press

TABLE 5. *Other single-operation utterances*

Utt. no.	Novel utterance	Closest match	Frequency of closest match	Frequency of ADD-ON	
ADD-ONS					
15	Put a bit more here	A bit more here	(1)	put = 121	
23	It's burning here	It's burning	(8)	here = 954	
27	Where's Daddy now?	Where's Daddy?	(9)	now = 69	SR#
56	The girls not playing tennis here	Girls not playing tennis here	(1)	the = 357	SR
59	I got one here	I got one	(4)	here = 972	
62	I want six paints	Six paints	(1)	I want = 181	SR
69	I want tissue lounge	I want tissue	(1)	lounge = 3	SR
70	Where's cars?	Where's car?	(1)	cars = 10	
75	No I sit on here	I sit on here	(1)	no = 808	SR
76	It's going up	Going up	(3)	it's = 223	SR
83	And that's a big lion	That's a big lion	(1)	and = 309	SR
106	It's not rolling it	It's not rolling	(1)	it = 737	SR
D5.	I want my drink home	I want my drink	(1)	home = 16	
OTHER					
90 DROP	And horse	And a horse	(1)	a = 477	
D1 INSERT	Man outside working	Man working	(2)	outside = 14	

SR = partial self-repetition of just preceding utterance.

diary utterances are in Part B. Twenty-six utterances require two operations and 10 require three or more operations to match the closest previous utterance. It is interesting that the majority of the diary utterances require two or more operations and thus have relatively distant closest matches. This suggests that the mother may indeed have been capturing Annie's most complex, novel utterances in the diary.

For some of the novel utterances in Table 6, the analysis looks similar to what we saw in the single-operation utterances. For instance, there are some SUBSTITUTIONS into previous, and frequently occurring, schemas (*Where's*: U28, U29, U57; *I want*: U67, U114, D2, D5, D7) and also the ADD-ON of *no* (U46)[7] or *here* (U55, U72). In addition, DROP operations involve omitting words that were ADD-ONS in the single-operation utterances. This suggests that Annie does indeed have creative flexibility with these words (e.g *here* in U50, *and* in D8). However, DROP, INSERT and REARRANGE play a larger role in these multi-operation utterances than they did in the single-operation utterances. This is because the closest matching utterances are more distant from the novel utterance and, particularly in the case of INSERT and REARRANGE, matching the closest previous utterance

[7] This is the one utterance for which there was a tie between operations depending on the order in which they were applied: thus *No* can be an ADD-ON, SUBSTITUTE or INSERT depending on the point at which *You —and* is 'dropped'.

https://doi.org/10.1017/S0305000903005592 Published online by Cambridge University Press

involves words that come from a variety of grammatical categories in the adult system.

However, it is also because the requirement of our mechanical matching is that EACH morpheme in the novel utterance is matched to find the closest previous utterance. This means that in this first analysis we ignore constituency of any kind, be it semantic or syntactic. Examples are the analysis of U28 as DROP *my*, U31 as DROP *a* and U29 as DROP '*s* (the possessive marker) when, in the adult system, we would simply talk of the placing of novel NPs. In our second analysis of Annie's novel utterances, we therefore relaxed the mechanical requirement by allowing a degree of 'constituency' based on an examination of items which would be defined as nouns or NPs and as prepositions, adverbs or PPs in the adult language. What difference would this assumption make to the frequency of prior exemplars and the number and type of operations required to match the novel utterance to the closest previous utterance?

Analysis 2: abstracting across OBJECTS/PERSONS and LOCATIONS

All of the noun phrases and pronouns in this child's novel utterances referred to objects or persons and almost all her prepositions, prepositional phrases and adverbs referred to locations. We therefore replaced nouns and noun phrases with the category OBJECT/PERSON and all locative adverbials, prepositions and prepositional phrases with the category LOCATION in the novel utterances and the rest of the 6-week corpus. For the novel utterances, Tables 7 and 8 show which words or phrases were replaced. From Table 7, we can see that all Annie's 'nouns' and 'noun phrases' in the novel utterances refer to people and objects. These are largely restricted in Subject position to pronouns and proper names but references to objects and people in other positions and in *wh*-questions are more varied and suggest that Annie has the ability to use some determiners and the possessive morpheme, though we cannot decide from naturalistic data the level of abstractness at which this ability is represented.

Table 8 shows those novel utterances in which words or phrases are replaced by locative adverbials, prepositions or PPs (utterances that contained verbal particles were not included e.g. *lying down* in U53). Annie's locative references mainly involve the single words *here* or *down* or phrases with *down* or *on* as the 'head' and she shows signs of being able to concatenate her frequent locative expressions: *down*, *here*, *on X*, into longer phrases, for instance *down on the bed here* in U72. Note that there are only two non-spatial adverbs and no non-spatial PPs in her novel utterances.

In what follows we refer to these potential 'constituents' as OBJECT/PERSON and LOCATION so as not to prejudge the issue of the level of abstractness at

https://doi.org/10.1017/S0305000903005592 Published online by Cambridge University Press

https://doi.org/10.1017/S0305000903005592 Published online by Cambridge University Press

TABLE 6. *Multiple-operation utterances*

Novel utterance and number	Closest match and number of operations	Frequency of closest match	Operations
(A) *Utterances on tape*			
12. Toast is ready	Toast is smile-ing [2]	(1)	SUBSTITUTE DROP- *mor -ing*
14. Mummy's toast not ready	Mummy's blue white [2]	(1)	SUBSTITUTE SUBSTITUTE
24. A little bit burnt here	Tiny little bit here [2]	(1)	SUBSTITUTE INSERT- *burnt*
28. Where's Daddy's work?	Where's my Daddy's cup+of+tea? [2]	(1)	SUBSTITUTE DROP- *my*
29. Where's my Daddy now?	Where's my Daddy's cup+of+tea? [2]	(1)	SUBSTITUTE DROP- *mor 's*
31. Little bit burning here	A little bit burnt here [2]	(1)	SUBSTITUTE DROP- *a*
38. And Ann smaller	And Annie's tired [2]	(1)	SUBSTITUTE DROP- *mor -'is*
46. No put it on my plate	You take your nappy off and put it on my shelf [3]	(1)	SUBSTITUTE ADD-ON DROP- *You … and*
47. In this house	Still in the house [2]	(1)	SUBSTITUTE DROP- *still*
50. From my room	From my bottom here [2]	(1)	SUBSTITUTE DROP- *here*
52. Not on lounge	W1 on W2 [2]	(29)	SUBSTITUTE SUBSTITUTE
53. What's that lie-ing down?	Mummy's lie-ing down [2]	(1)	SUBSTITUTE INSERT- *that*
55. Girls not play-ing tennis here	Girls play-ing falling down [3]	(1)	INSERT- *not* SUBSTITUTE ADD-ON

https://doi.org/10.1017/S0305000903005592 Published online by Cambridge University Press

57. Where's the boy's tennis + racket?	Where's the W1 W2? [2]	(19)	SUBSTITUTE SUBSTITUTE
65. Which ones go right on here?	Which ones go here? [2]	(2)	INSERT- *right* INSERT- *on*
67. I want a@sc crying tissue	I want a@sc W1 W2 [2]	(2)	SUBSTITUTE SUBSTITUTE
71. What about this big chair?	What about this one? [2]	(4)	SUBSTITUTE ADD ON
72. Mummy's trying to lie down on the bed here	Mummy's lie-ing down [6]	(1)	SUBSTITUTE INSERT REARRANGE ADD ON ADD ON ADD ON
77. Mummy's car down	Like Mummy's car [2]	(1)	ADD ON DROP- *like*
86. And your tummy	And W1 W2 [2]	(24)	SUBSTITUTE SUBSTITUTE
94. Red and blue	W1 and W2 [2]	(35)	SUBSTITUTE SUBSTITUTE
100. Lets roll it together	Lets W it [2]	(3)	SUBSTITUTE ADD ON
102. Got it back on	Mummy's got it on the floor [3]	(1)	INSERT- *back* DROP- *Mummy's* DROP- *the floor*
103. I can't put it back on	I don't put it on there [3]	(1)	SUBSTITUTE INSERT- *back* DROP- *there*
107. I can't go and just get it	I can't get it [2]	(2)	INSERT- *go and* INSERT- *just*
110. Get off down the stairs	W1 down the W2 [2]	(2)	SUBSTITUTE SUBSTITUTE
113. What's that funny drawing down there?	What's that lie-ing down? [3]	(1)	SUBSTITUTE INSERT- *drawing* ADD-ON

https://doi.org/10.1017/S0305000903005592 Published online by Cambridge University Press

TABLE 6. *(Cont.)*

Novel utterance and number	Closest match and number of operations	Frequency of closest match	Operations
114. I want a chair stand up	I want a W1 W2 [2]	(6)	SUBSTITUTE SUBSTITUTE
(B) *Diary entries*			
D2. I want Moggie come	I want my Mummy come back [2]	(1)	SUBSTITUTE DROP- *back*
D3. What's that moving ladder	What's that *lie-ing* down [2]	(1)	SUBSTITUTE SUBSTITUTE
D4. I got a poo+poo change the nap toilet	Change the nap he's done a poo [4]	(1)	ADD ON REARRANGE- *a poo+poo* SUBSTITUTE DROP- *done*
D6. I don't put it in Dave	I don't put it on there [2]	(1)	SUBSTITUTE SUBSTITUTE
D7. I want some water pour it a@sc flower	I want some water [3]	(4)	ADD ON ADD ON ADD ON
D8. That's Cally's boobie	And that's Cally outside [3]	(1)	DROP- *and* INSERT- *'s* SUBSTITUTE
D9. Look what I made	I made that [3]	(1)	ADD ON ADD ON DROP- *that*

Words underlined indicate that they have been said together before and count as 1 for the purpose of slot filling.
mor=morpheme.

TABLE 7. *Replaced nouns and NPs*

	Nouns and NPs replaced
As SUBJECT	● I (24), it (7), that (4), you, all
	● Mummy (3), Daddy, Dave, Cally(2), Ann, baby (2), toast, girls
	● a (bit more/little bit), this one, little bit
	● the (girls/rolling pin), Mummy's [toast/car]
OTHER	● it (8), toast (2), home, pennies, tennis (2), tissue (2), horse, jam, floor, one, chair, tiger, penny, butter (2), toilet, Moggie
	● a (hat/knife/football (2)/chocolate/toastie/pig/paper/chair/poo-poo, crying tissue/big lion (2))
	● the (toaster/sheep/people/stairs/butter)
	● my (room/bag/plate/drink), your tummy, Callie's boobie
	● this house, (2), some (water (3)/toast), more papers, six paints
	● tissue lounge, one, two … (2)
In Wh-questions	
Where's	● Daddy
	● the (buses/butter/tape/box,/hippo/elephant/dog/people/drawing)
	● cars, sheep
	● my (seat, Daddy) (2)
	● Mummy's [plate/knife/car], Annie's [plate/knife], Daddy's work, the boy's tennis + racket
What about	● this big chair
What's	● that, that funny drawing, that moving ladder
Which	● ones

TABLE 8. *Replaced adverbs, prepositions and PPs*

● here (8), down (2), up
● on (floor/lounge/here/my plate/the bed here)
● down (there/the stairs)
● from (downstairs/my room)
● in (this house/Dave)

which they are represented. Thus U46, *No put it on my plate* becomes *No put* OBJECT LOCATION.

Having made these replacements we then searched as before for the closest previous utterance and compared the results to those of Analysis 1. A simple example of the effects of Analysis 2 can be seen for U90, *and horse*. It will be recalled that this was the only single-operation utterance that required a DROP operation to match the closest previous utterance: *and a horse*. However, if we substitute OBJECT for *horse*, the schema, *and object*, becomes the closest match and there are 120 previous exemplars of this schema.

There were three main outcomes of Analysis 2:

(1) the number of single-operation utterances increased from 71% under Analysis 1 to 78% under Analysis 2;

https://doi.org/10.1017/S0305000903005592 Published online by Cambridge University Press

TABLE 9. *Summary of changes in Analysis 2*

Total tape + diary utterances	*Analysis 1* Exact morpheme matching	*Analysis 2* Object/person and location substitutions	% utterances with changes in number of prior exemplars Anal. 1→Anal. 2	Changes in number of operations between Analysis 1 and Analysis 2	
Single-operation	84 [71%]	92 [78%]	Increase (76) 64% No change (33) 28%	Multiple-operations in 1 ↓ Fewer operations in 2	18
Multiple-operations	34 [29%]	26 [22%]	Decrease (9) 8%	Single-operations in 1 ↓ Multiple-operations in 2	7
Overall total	118	118	118		

(2) for 64% of the novel utterances, the number of prior exemplars forming the closest match increased;
(3) the number of operations to arrive at the closest matching utterance decreased for 18 utterances and increased for 7. For some of the novel utterances this resulted from a change in the closest matching utterance (10 utterances) while for others it arose directly from the change from morpheme matching to 'phrase' matching.

These changes are summarized in Table 9. We first discuss the changes in numbers of exemplars and then, in more detail, the utterances for which the numbers of operations changed.

Changes in the number of exemplars. The most obvious consequence of this procedure is to increase dramatically the number of prior exemplars. Many of the novel utterances which, under Analysis 1, had only one prior closest match, now have a schema as the prior match. The number of prior exemplars increased for 76 of the novel utterances, stayed the same for 33 and decreased for 9. Clearly, if we generalize across putative Noun Phrases and Locative phrases, the number of prior exemplars for schemas such as *Where's* OBJECT/PERSON (345) and *Put* OBJECT LOCATION (42) will increase. Thus almost all the utterances in Table 3 increase the number of prior exemplars. With the exception of U39 and U43, all utterances that, under Analysis 1, had only one prior exemplar now have 5 or more. In Table 4B, 5 out of the 9 utterances which had only one closest matching utterance under Analysis 1 now have more than one. The situation is more complex for the ADD-ON utterances in Table 5 and for some of the multiple-operation utterances in Table 6, but for many of these latter utterances the number of prior exemplars also greatly increased, as we shall see below.

Changes in the closest match and/or the number of operations. 13 multi-word operations, which required two operations to get from the closest match, now require only one. This is because in Analysis 1, the 'noun phrases' or

https://doi.org/10.1017/S0305000903005592 Published online by Cambridge University Press

'locative phrases' had to be assembled and then substituted into the schema, whereas, in Analysis 2, these count as one-operation substitutions. In three cases, three-operation utterances reduce to two for similar reasons. D8: *That's Callie's boobie* is an example of an utterance for which the closest matching utterance changes. Under Analysis 1, this was *And that's Callie's outside* as a result of the matching morphemes criterion. Under Analysis 2, the closest match becomes *That's OBJ*, which means that D8 becomes a one-operation substitution.

The seven utterances for which the number of operations to reach the closest match increases have changed their status because the closest match has changed in some way. For example, in U58, *I want to paint*, *get it* had been said as a whole before so the substitution of *paint* was one operation. However, in the new analysis, while the closest match is the same, it now becomes *I want to get OBJ* and this results in two operations (OBJECT dropped from *I want to get it*, and *paint* substituted for *get*). Two of the utterances cease to be classified as simple ADD-ONS (of *Its* to *Going up* (U76) and *And* to *That's a big lion* (U83)). In terms of their actual production, each of these utterances was an immediate repeat of what Annie had just said with the extra word at the beginning of the utterance (see Table 5). However, it is interesting to note that by defining possible constituents, Analysis 2 may give us additional clues to the underlying segmentations in some of these utterances. Developing this further would require the use of methods other than distributional analysis, as discussed below. Full details of those utterances for which the numbers of operations changed between Analysis 1 and Analysis 2 can be found in Appendices A and B.

Finally, Tables 10 and 11 summarize all of Annie's novel utterances using these, more abstract, object/person and locative categories. In both tables, the utterances are entered in columns corresponding to their status under Analysis 1: thus each column represents data from Tables 3–6 above. In Table 10 we list all those utterances the status of which changed under Analysis 2 and in Table 11, those utterances for which it did not. Since Annie's use of NPs in Subject position was rather restricted, we have summarized these in the tables rather than replacing them by the more general term, OBJECT. The numbers in parentheses after entries in Table 10 represent the number of novel utterances in the last session that are identical. Novel utterances in the same row share lexical material, usually the verb.

In Table 10, ten of the multiple-operation utterances now show the same schemas as single-operation utterances. We can also see that a number of utterances are related to each other by the presence or absence of *And*, *No* or LOC. There is a variety of methods of negation, including negating the utterance by placing *No* at the front (sometimes correctly and sometimes incorrectly from the point of view of the target language) and one instance each of *can't* and *don't*.

https://doi.org/10.1017/S0305000903005592 Published online by Cambridge University Press

https://doi.org/10.1017/S0305000903005592 Published online by Cambridge University Press

TABLE 10. *Utterances which changed their status under Analysis 2*

Table 3 Single Ops SUBSTITUTE (obj/pers)	Table 4 Single Ops SUBSTITUTE (other)	Table 5 Single Ops ADD ON	Table 6A Multiple Ops TAPE	Table 6B Multiple Ops DIARY
OBJ (3)	1			
Just OBJ (1)				
Not OBJ (1)				
Hello OBJ (1)				
And OBJ (1)			1	
That's OBJ (3)		And that's OBJ		1
There's OBJ (1)				
OBJ's big (2)	2			
And/- OBJ/('s) smaller (2)	1		1	
Where's OBJ? (16)		1	3	
		Where's OBJ now?	1	
I/- have OBJ (1)	1			
	Can have OBJ			
I don't like OBJ (1)				
I need OBJ (1)				
I/Mummy want OBJ (10)	1	1	1	
		I want OBJ LOC (2)		
			I want OBJ stand up	I want OBJ come
				I want OBJ pour it a@sc flower

https://doi.org/10.1017/S0305000903005592 Published online by Cambridge University Press

I/- got OBJ (3) I got no OBJ		I got OBJ LOC	Got OBJ back on	I got OBJ change the nap toilet
	Going to get OBJ LOC OBJ LOC		I can't go and just get OBJ 2 1	
	I sit LOC	No I sit LOC		
	This ones go LOC		Which ones go right LOC	
		It's burning LOC	1 OBJ burnt LOC	
		Put OBJ LOC	No put OBJ LOC I can't put OBJ back on OBJ not ready OBJ is ready Not LOC Get off LOC	I don't put OBJ LOC
			What's that OBJ LOC What's that lying down	Whats that OBJ
		The girls not playing OBJ LOC	Girls not playing OBJ LOC	

Subject NPs have been left as lexemes or summarized: /=alternative Subjects; -=Subject sometimes absent.

TABLE 11. *Utterances which did not change their status under Analysis 2*

Table 3 Single Ops SUBSTITUTE (obj/pers)	Table 4 Single Ops SUBSTITUTE (other)	Table 5 Single Ops ADD ON	Table 6A Multiple Ops TAPE	Table 6B Multiple Ops DIARY
You sorry OBJ (tiger)				
	It's upside+down			
	It's lovely			
	I want to paint			
	I want to roll OBJ (it)		Let's roll OBJ (it) together	
	It's not rolling	It's not rolling OBJ (it)		
		It's going up		
	I nearly finished			
	Right here			
	To Penny			
			Red and blue	
				Mummy's trying to lie down LOC (on the bed here)

Table 11 consists mainly of copulas and progressives for which we have seen there are a large number of prior exemplars, together with a few other utterances for which either the interpretation is unclear (*To Penny*) or the OBJ/LOC abstraction was irrelevant (*Red and blue*) and two utterances with infinitival complements after *I want* (*I want to paint ; I want to roll it*).

Although we could, of course, go much further in defining underlying constituents for the utterances in Tables 10 and 11, there does not seem to be a great deal in the data to justify this. The most likely candidates would be a category of attributives (in the copulas) and perhaps of actions (in the verb complement utterances with *want*). These latter may be related to the utterances with *want* in Table 10.

If we examine all Annie's novel utterances with *want* on this day (see Example 11), we can see that she clearly has a well-established way of expressing her desire for an object, and also shows some evidence of being able to use *want* as a 'semi-auxiliary'. However, she may only be beginning to express more complex ideas with *want* syntactically.

Example 11: Utterances with *want*
I want OBJ (13)
I want to paint
I want to roll it
I want tissue lounge (from/in the lounge?)

https://doi.org/10.1017/S0305000903005592 Published online by Cambridge University Press

I want my drink home (at home?)
I want chair stand up (to stand (something) up on?)
I want Moggie come (to come?)
I want some water pour it@sc flower (to pour on the flower?)

In particular, she may add-on material that is connected to the object of *want* rather than knowing how to organize the syntax. An alternative explanation is that the syntax is available to her but material is omitted as a consequence of performance limitations. As always with corpus data we cannot definitively decide between these interpretations but we will take them up in the Discussion below.

DISCUSSION

Our goal in this study was to make a preliminary attempt to assess how much novelty there was in a child's speech using a new dense corpus, which gave us relatively rich information about what the child had said before. We defined novelty in terms, first, of utterances that had not been said before and second, in terms of how these utterances related to utterances that HAD been said before. The analysis took all the utterances recorded on the last day of data collection and assessed their relationship to utterances in the 28 hours of sampling in the previous six weeks. Our results show that, for this child, between one-third to one-half of her multi-word utterances are novel when compared to a roughly 10% sample of everything she has said before. By this we mean that the utterance was neither an immediate imitation nor a repetition of her own or another's utterance, nor had it been said in identical form before.

Using a mechanical procedure of matching each novel utterance to prior utterances on the basis of number and order of shared morphemes and the frequency of exemplars, we identified a closest matching previous utterance or schema for each novel utterance. We then used a set of defined operations to work out what would have been required to change the prior utterance to the novel one. The results for Annie's 109 novel utterances on tape were distributed as follows. About three-quarters of the novel utterances (74%) required only a single operation to get from a previously produced utterance to the target utterance. The majority of these (84% of the three-quarters) involved substituting one word for another – mostly a word referring to an object, and mostly in an utterance position where type variation had previously occurred. A minority (15% of the three-quarters) were cases in which a very high frequency word had to be added on (such things as *here*, *now*, and *and*) to a previous utterance. The remaining 26% of Annie's novel utterances on tape represent cases in which we needed to modify a previously said utterance in more than one way in order to produce the target utterance. Twenty-two of the utterances on tape and three diary utterances required

https://doi.org/10.1017/S0305000903005592 Published online by Cambridge University Press

two operations, mostly substituting, adding on, and dropping elements, but with five cases of insertions. There were ten utterances in all that required three or, in two cases, more operations in order to arrive at the target utterance. When, in a second analysis, we treated references to objects and locations as putative 'constituents' substitutable within the same reference domain, not surprisingly, the number of prior exemplars for each utterance schema increased and the number of utterances requiring multiple operations for the closest match decreased. We discuss these results under three headings: the estimation of 'novelty' in this child's utterances; the role of memory in language production; and, finally, the issue of linguistic representation.

Estimating the degree of novelty

In this study we defined novelty simply in terms of whether an utterance had been said before and left aside the issue of how the previous utterances had been produced, to which we return below. Would the number of utterances identified as novel have increased or decreased as a function of more data? Given the same sample of child utterances (i.e. all those recorded on the last day of the study), a larger database of previous utterances could only reduce the degree of novelty. We also have only 6 weeks of data and, although Annie was developing very fast, there will still have been multi-word utterances that she was producing before recording started that might contribute to reducing the number of utterances identified as novel. However, we also have to consider that there were equivalent numbers of 'novel' utterances on each of the previous four recordings from the last week of data collection. In addition, while Annie's speech was certainly sampled more broadly than that of almost all children so far studied, it remains the case that, with the exception of the mother's diary record, most recordings took place in approximately the same contexts every day. In terms of our criterion of novelty, a wider range of contexts could have increased the number of novel utterances but it could also, of course, have provided more previous exemplars for utterances we had trouble in matching (e.g. U72: *Mummy's trying to lie down on the bed here*). There are a number of possibilities for empirically and mathematically estimating the relative effects on measures of novelty of adding or reducing the amount of data which we intend to investigate in future research.

Accepting our criterion of novelty, HOW novel were the novel utterances as compared to those that Annie had already produced? Under our analysis, which depends on the matching of specific lexical and morphological forms, there was a very close relationship. Most of the novel utterances required the substitution or addition of only one word – usually into a previous schema which itself had a number of previous exemplars. Even in the case of the 26% of utterances requiring multiple operations, most can be related fairly straightforwardly to previous utterances and schemas.

https://doi.org/10.1017/S0305000903005592 Published online by Cambridge University Press

The question of whether the degree of novelty identified by our criterion should be considered as indicating a high level of productivity or not, rests to some extent in the eye of the beholder. On our criterion, approximately one in five of all Annie's utterances on tape and one in three of her multi-word utterances are novel. However, many are very closely related to utterances that have been said before. Talk with children of this age can be very repetitive and in addition, these children have limited lexicons. Both these factors will contribute to restricting the amount of novelty identified using a lexical matching procedure such as ours. Indeed when we applied the same criterion of novelty to the mother's utterances on the last tape, we found that the overall number of utterances identified as 'novel' was comparable. Conversation with small children about the same child-friendly topics does, indeed, encourage repetition. However, once we moved to attempting to match the material in the mother's novel utterances to her utterances in the previous 6 weeks, we found many fewer prior schemas and exemplars. In addition, more of the matches required not only multiple operations, but more than two operations and more use of INSERT, DROP and REARRANGE. Not surprisingly the mother's utterances were more complex in terms of overall length, the length of constituents and their semantic complexity, and the range of syntactic constructions used.

It is interesting to reflect on the question of how much novelty we would see in speech between adults using the criteria adopted for this study. This raises the issue of which genre of conversation one would choose to compare. Is there anything like mother–child conversation in adulthood? Would a conversation at a bus stop or one between two academics who had been discussing the same issues for many years make a better comparison? Intuitively it seems that both types of conversation could show a combination of high degrees of repetitiveness, narrow-scope productivity and complex syntactic construction. This raises the issue of how these factors might be interwoven in language production.

The role of memory in language production

The issue of the relationship between linguistic representations and their role in language production is much debated for both child and adult language production. All would agree that adults have access to relatively abstract representations. However, almost all would probably also agree that, in on-line speech production, people can pull out strings which are more or less 'pre-packaged' either because they cannot be (fully) built up from more underlying representations (Jackendoff, 1996; Culicover, 1999) or because they have become automatized as a result of frequent use (Bybee & Scheibmann, 1999; Verhagen, 2001). When we come to children the same issues arise. All researchers accept that young children's speech contains rote-learned

https://doi.org/10.1017/S0305000903005592 Published online by Cambridge University Press

phrases and lexically based (pivot-type) formulae (e.g. Bloom, 1970, 1973; Brown, 1973; Peters, 1983) and that frequency is important in determining what these are. The implication is that whole utterances, and sequences within utterances, are being registered in memory and that this gives rise both to utterances learned as a whole ('frozen' or 'rote-learned' utterances) and utterances which 'overlap' in lexical content and semantics (pivot-type patterns). Leaving aside for a moment the question of linguistic representation, if these utterances and patterns are, indeed, stored in memory, what processes might they undergo as they are retrieved to form the basis for new utterances and are the operations that we have used realistic in this respect?

The operations we used – SUBSTITUTE, ADD-ON, DROP, INSERT, and REARRANGE – are basically all of the possibilities. Establishing their psychological reality will require some other kind of research, presumably experimental investigations in which a variety of factors are carefully controlled. However, we might tentatively propose that the operations of SUBSTITUTE and ADD-ON may be realistic in terms of how the child builds up some novel utterances at this point in development. Despite the fact that some of the substitutions do not accord with our grammatical intuitions, the great majority do and this is, of course, even more the case in Analysis 2. Certainly the suggestion that children substitute items into semi-formulaic patterns has been accepted for many years. The plausibility of ADD-ON as an operation is supported by the fact that the majority of single-operation add-ons occurred in the context of adding-on a word to the whole of a just previous utterance, i.e. an utterance which we know is in immediate active memory (Clark, 1977; Snow, 1981). Whether DROP is equally plausible is more questionable though we should note that, for some matches, the same morphemes that were added on to arrive at the target utterance were dropped for other matches. Some of the matches requiring the DROP operation in Analysis 1 did not do so under Analysis 2, once matches were made to the 'constituents' that we had defined (Appendix A). Other matches, while perhaps more satisfying in linguistic terms, required new DROP operations (Appendix B). This seems to us to reflect the idea that the underlying representations upon which utterances are constructed can be multiple and at different levels of lexical specificity and abstractness. Which representations contribute to the production of a particular utterance will depend on factors such as context, what is stored in memory and its recency. Determining which representations are involved will require methods other than distributional analysis, for instance psycholinguistic measures of speech production in natural contexts together with more controlled experimental designs.

It is also telling that our five possibilities did not occur with equal frequency. In particular SUBSTITUTE was highly frequent, REARRANGE was virtually absent and INSERT occurred only in 11 utterances (all but one,

https://doi.org/10.1017/S0305000903005592 Published online by Cambridge University Press

multiple-operation utterances). It is possible that INSERT reflects a more sophisticated skill, perhaps reflecting the child's beginning grasp of some aspects of constituency, for instance inserting *not*, *just* or *and* before whole constituents in the utterance. Most of the INSERTs are indeed of this form e.g. the insertion of *not* before the 'predicate' in U55 and of *right* before *on here* in U65. We should also note that when, in Analysis 2, we allow for a degree of constituency in object and locative phrases, which reduces the requirement for morpheme by morpheme matching, most of the INSERTs drop out e.g. in U65 *on here* becomes a locative phrase rather than requiring the insertion of *on* before *here*. The lack of REARRANGE may reflect the very linear, word-order dependency of English. In this case, it will be important to see, when we conduct a similar analysis for children learning other languages, whether we have to use more single-operation REARRANGEs to match the novel utterance to a prior utterance. This is research that we are currently undertaking. Alternatively REARRANGE may also be much more relevant to operating with constituent or word order changes of a pragmatic (e.g. topic focus) or syntactic (e.g. verb final infinitives in German) nature and the child may not yet have reached this stage.

This discussion of the possible psychological reality of these operations clearly raises the issue of how, in the production of utterances, strings stored in memory might interact with more abstract categories that have traditionally been the central focus of linguistic study. On the one hand, we are currently developing psycholinguistic tests to ascertain whether our distinctions between 'novel' utterances and 'utterances said before' and between the different operations have independent validity. On the other, our second analysis indicates that introducing a degree of 'constituency', at whatever level of abstractness this is conceptualized, confirms many of the analyses arrived at under the first analysis and gives a more intuitively plausible account for those, relatively few utterances that required the operations to apply across linguistic 'equivalence classes'. This is hardly surprising given that the first analysis was purely mechanical, devoid of semantic and pragmatic considerations, based on only about 10% of what the child had said before and did not take into account the speech of the adults around her. However, from our point of view, it is important, when introducing a level of categorial organization to do this without assuming pre-established categories for which we have no independent evidence. A further alteration to the procedure would be to allow the target to be matched to more than one preceding utterance. This would then allow a 'blending' process which might be considerably more psychologically realistic than the current procedure that only allows matching to one previous utterance. Clearly there is much further research to be done before we can assess the possible psychological reality of these and related suggestions. But the methodological approach that we have taken in this paper is a beginning in terms of the kind of methodologies that it

https://doi.org/10.1017/S0305000903005592 Published online by Cambridge University Press

might be possible to develop when dealing with the much larger corpora that are likely to become available in the future.

Linguistic representations

Finally, there is the question of how this analysis might bear on the issue of linguistic representation. In this paper we have made no assumptions about this. It is possible that each of the novel utterances and, indeed, each of the prior exemplars is either assembled afresh each time from more abstract representations or, at least, that the child represents the structure of the utterance in terms of the more abstract and general categories of the adult, or some intermediate, grammar, even if this is not always how they are constructed. When dealing with naturalistic data, it is a matter of theoretical choice at which level to postulate the linguistic representations which underpin children's production of utterances, and while it is important that researchers utilize an objective definition of productivity, this will largely be a methodological decision rather than one that can be independently arrived at from the data.

Our own position is that we think it unlikely that children, or indeed adults, project each of their utterances from scratch out of the most elementary parts possible. It is, of course, the case that each of the prior matching utterances was itself a novel production for Annie at some time in the past. Some novel utterances may have been put together by combining individual words that have already been learned. Others will have been learned as wholes and combined as wholes with other words or constructions – a procedure that can result in an ungrammatical utterance (as also noted by Clark, 1974, Ewing, 1982 and Peters, 1983). Others will have been learned as wholes and then develop slots into which the child places words, sometimes belonging to the same semantic-pragmatic category and sometimes not. Some may have been assembled by the juxtaposition of underlying semantic or pragmatic categories (for instance utterance-initial negation). Clearly, the meaning of linguistic items (morphemes, words, utterance-level constructions) for the child, the distributional facts of the language being learned and the relative frequencies of words and constructions in the ambient speech will affect precisely how this proceeds and the nature of the underlying representations that the child is combining.

We suggest that different levels of abstractness can coexist within and across utterances in terms of contributing to productivity (which Tracy also suggested for German, 1991: 402–407) and that this is a developmental phenomenon. Our second analysis, in which we introduced a level of 'constituency', provides some indication of how this might work. The data suggested that, for this child at this stage, these constituents may be more narrowly based than are NPs and PPs in the adult language but that a level

https://doi.org/10.1017/S0305000903005592 Published online by Cambridge University Press

of construction based on more general categories was developing. Thus for both functional reasons (objects and locations are highly salient for children) and formal reasons (in English they may be relatively easily identifiable by small number of frequently repeated morphemes) children may be able to generate some types of novel NPs and Locative phrases at a relatively early point in their development of syntax. They can then place these into schemas which are much more lexically-specific. If we are right, the complexity of these slots should develop over time. For example, for the lexically-specific frame *Where's —?*, we noticed that the material going into the slot became more complex over the 6 week period of the study. Initially Annie used only the names of family members and object nouns preceded by a schwa. By the end of the period a number of 'determiners' were clearly present, together with a semantically limited set of 'adjectives'. Thus our hypothesis is that the more abstract categories of the adult language are EMERGENT in this process rather than being pre-given (Behrens, 2000; Tomasello, 2000). We have attempted to provide an initial snapshot of this creativity 'on the move' in terms of what the child has previously produced and this is the sense in which we describe our approach as 'usage-based'.

Our hypothesis is that while Annie is able to construct relatively large numbers of novel utterances with her language, the most revealing account of her syntactic competence is one that focuses on the specific linguistic content – the actual words, phrases, and utterances – that she hears and produces. Indeed the general approach to the representation of the language in terms of usage-based 'pieces' dates from Bolinger (1968) and has formed a central feature of a number of accounts, most notably those of Clark (1974), Ewing (1982), Peters (1983), Lebeaux (1988), Tomasello (1992), Pine & Lieven (1993) and Lieven *et al.* (1997). In addition, more recent theories of syntactic categories have suggested that these can emerge from the co-occurrence relationships in lexical material (see, for instance, Hopper, 1987; Behrens, 2000; Croft, 2001; and MacWhinney's edited volume, 1999). If these approaches do begin to capture the processes involved in language development, they have the potential for eliminating the 'linking problem' (Pinker, 1989) i.e. the problem of how children link the specific language that they hear to their innately endowed Universal Grammar. Of course, these approaches stand in contrast to other accounts of early child grammar that assume the presence of more abstract knowledge with which children generate their utterances, although some of these accounts (e.g. Lebeaux, 1988; Radford, 1990) take the lexical specificity and early phrasal structure of children's grammar more seriously than do others.

However, it is certainly not possible to prove this one way or the other with naturalistic data. Since Annie is learning English her language will be describable in terms of the target grammar – particularly if elements are allowed to be missing on the surface. Thus the fact that Annie shows evidence

https://doi.org/10.1017/S0305000903005592 Published online by Cambridge University Press

of a highly frequent and productive *Where's NP?* pattern, that she produces almost no other *wh*-questions and that she never uses the non-cliticized copula with *Where* except in *Where are you?* (said 4 times) and *Where is it* (all imitations) could still mean that she has the underlying representations which allow *wh*-movement or that *wh*-operators initially reside in SpecIP (Radford, 1988). We can propose, alternatively, that the pattern of the data suggests that this is a lexically-specific pattern based around *Where's* and that the auxiliary is not analysed. However, while many researchers have adopted one or other of these positions, children's production of a limited range of correct *wh*-questions cannot not provide evidence one way or the other as to the level of morphosyntactic abstraction. Only experiments which control the conditions under which children hear particular words and constructions can answer the question of the level of abstraction at which the child's linguistic system is represented at a particular developmental point.

All we can say on the basis of the present study is that many of Annie's novel utterances are closely related to utterances that have been said previously, that these previous utterances are highly frequent, and have the appearance of containing variable slots. And that, in so far as some of children's utterances are accepted as arising from the registration of strings in memory and the creation of slots in those strings, our results suggest that this may account for a considerable proportion of the data for an English-speaking child at this stage of development.

Finally, it is almost certainly the case that studies of children learning other languages or somewhat more advanced in development will reveal different proportional frequencies of these types of operations than those presented here, and this will give us additional information relevant to the question of psychological reality. It is also possible that there will be individual differences between children in the degree to which they depend on using prefabricated strings and generation from underlying categories in the production of utterances (Lieven, 1997; Peters, 1997). Future research should use both cross-linguistic and experimental methods to investigate the question of whether these operations represent psycholinguistic operations children actually use in constructing utterances 'on-line'.

REFERENCES

Behrens, H. (2000). Lexical representation and acquisition theory. *Psychology of Language and Communication* **4**(2), 33–44.

Bloom, L. (1970). *Language development : form and function in emerging grammars*. Cambridge, MA: MIT Press.

Bloom, L. (1973). *One word at a time*. The Hague: Mouton.

Bolinger, D. (1968). *Aspects of language*. New York: Harcourt, Brace & World.

Braine, M. D. S. (1976). Children's first word combinations. *Monographs of the Society for Research in Child Development* (*1, Serial No. 164*), **41**.

https://doi.org/10.1017/S0305000903005592 Published online by Cambridge University Press

Brown, R. (1973). *A first language: the early stages*. Cambridge, MA: Harvard University Press.
Braunwald, S. & Brislin, R. (1979). The diary method updated. In E. Ochs & B. Schieffelin (eds), *Developmental pragmatics*. New York: Academic Press.
Bybee, J. (1995). Regular morphology and the lexicon. *Language and Cognitive Processes* **10**, 425–55.
Bybee, J. & Scheibmann, J. (1999). The effects of usage of degrees of constituency: the reduction of 'don't' in English. *Linguistics* **37**, 575–96.
Clark, R. (1974). Performing without competence. *Journal of Child Language* **1**, 1–10.
Clark, R. (1977). What's the use of imitation? *Journal of Child Language* **4**, 341–58.
Croft, W. (2001). *Radical construction grammar*. Oxford: Oxford University Press.
Culicover, P. (1999). *Syntactic nuts*. Oxford: Oxford University Press.
Ewing, G. (1982). Word order invariance and variability in 5 children's three-word utterances. In Carolyn E. Johnson & C. L. Thew (eds), Papers from the Second International Congress for the Study of Child Language. University Press of America.
Hopper, P. (1987). Emergent grammar. *BLS, 13*. Berkeley: Berkeley Linguistics Society.
Jackendoff, R. (1996). *The architecture of the language faculty*. Bradford: MIT Press.
Kuczaj, S. (1979). Influence of contractibility on the acquisition of *Be*: substantial, meagre or unknown? *Journal of Psycholinguistic Research* **8**, 1–11.
Langacker, R. W. (1987). *Foundations of cognitive grammar, vol. 1: Theoretical prerequisites*. Stanford: Stanford University Press.
Lebeaux, D. (1988). Language acquisition and the form of grammar. Unpublished dissertation, University of Massachussetts.
Lieven, E. V. M. (1997). Variation in a crosslinguistic context. In D. I. Slobin (ed.), *The crosslinguistic study of language acquisition, volume 5: Expanding the contexts*. Mahwah, NJ: Erlbaum.
Lieven, E. V. M., Pine, J. M. & Baldwin, G. (1997). Lexically-based learning and early grammatical development. *Journal of Child Language* **24**, 187–219.
Lieven, E., Pine, J. & Dresner Barnes, H. (1992). Individual differences in early vocabulary development. *Journal of Child Language* **19**, 287–310.
MacWhinney, B. (ed.) (1999). *The Emergence of Language*. Mahwah, NJ: Erlbaum.
MacWhinney, B. (2000). *The CHILDES project: tools for analyzing talk, 3rd edition, volume 1: Transcription format and programs*. Mahwah, NJ: Erlbaum.
MacWhinney, B. & Snow, C. (1990). The child language data exchange system: an update.. *Journal of Child Language* **17**, 457–72.
Peters, A. (1983). *The units of language acquisition*. Cambridge: CUP.
Peters, A. (1997). Language typology, prosody and the acquisition of grammatical morphemes. In D. I. Slobin (ed.), *The crosslinguistic study of language acquisition, volume 5: Expanding the contexts*. Mahwah, NJ: Erlbaum.
Pine, J. & Lieven, E. (1993). Reanalysing rote-learned phrases: individual differences in the transition to multi-word speech. *Journal of Child Language* **20**, 551–71.
Pinker, S. (1989). *Learnability and cognition: the acquisition of argument structure*. Cambridge, MA: MIT Press.
Radford, A. (1988). *Transformational grammar*. Cambridge: CUP.
Radford, A. (1990). *Syntactic theory and the acquisition of English syntax*. Oxford: Blackwell.
Snow, C. (1981). The uses of imitation. *Journal of Child Language* **8**, 205–12.
Tomasello, M. (1992). *First verbs: a case study of early grammatical development*. Cambridge: CUP.
Tomasello, M. (2000). Do young children have adult syntactic competence? *Cognition* **74**, 209–53.
Tracy, R. (1991). *Sprachliche Strukturentwicklung*. Tübingen: Narr.
Verhagen, A. (2001). Subordination and discourse segmentation revisited, or: Why matrix clauses may be more dependent than complements. In E. Sanders, J. Schilperoord & W. Spooren (eds), *Text representation, linguistic and psychological aspects*. Amsterdam: John Benjamins.

https://doi.org/10.1017/S0305000903005592 Published online by Cambridge University Press

https://doi.org/10.1017/S0305000903005592 Published online by Cambridge University Press

APPENDIX A

UTTERANCES WITH REDUCED NUMBERS OF OPERATIONS UNDER ANALYSIS 2

Utt. no.	Utterance	Closest match 1	Operations 1	Closest match 2	Operations 2
14	Mummy's toast not ready	Mummy's blue white (1)	SUBSTITUTE SUBSTITUTE	OBJ not ready is@sc not ready (2)	SUBSTITUTE
28	Where's Daddy's work?	Where's my Daddy's cup + of + tea? (1)	SUBSTITUTE DROP- *my*	Where's OBJ? (353)	SUBSTITUTE
29	Where's my Daddy now?	Where's my Daddy's cup + of + tea? (1)	SUBSTITUTE DROP- *mor 's*	Where's OBJ now? (3)	SUBSTITUTE
46	No put it on my plate	You take your nappy off and put it on my shelf (1)	SUBSTITUTE SUBSTITUTE DROP- *You … and*	No put it LOC W put it LOC (49)	SUBSTITUTE SUBSTITUTE
47	In this house	Still in the house (1)	SUBSTITUTE DROP- *still*	LOC (in OBJ (8))	SUBSTITUTE
50	From my room	From my bottom here (1)	SUBSTITUTE DROP- *here*	LOC (from OBJ (4))	SUBSTITUTE
52	Not on lounge	W1 on W2 (29)	SUBSTITUTE SUBSTITUTE	Not LOC (18)	SUBSTITUTE
57	Where's the boy's tennis + racket?	Where's the W1 W2? (19)	SUBSTITUTE SUBSTITUTE	Where's OBJ? (357)	SUBSTITUTE
67	I want a@sc crying tissue	I want a@sc W1 W2 (2)	SUBSTITUTE SUBSTITUTE	I want OBJ (159)	SUBSTITUTE

https://doi.org/10.1017/S0305000903005592 Published online by Cambridge University Press

71	What about this big chair?	What about this one? (4)	SUBSTITUTE ADD ON	What about OBJ? (4)	SUBSTITUTE
72	Mummy's trying to lie down on the bed here	Mummy's lie-ing down (1)	SUBSTITUTE INSERT REARRANGE ADD ON ADD ON ADD ON	Mummy's lie-ing down LOC (1)	SUBSTITUTE INSERT REARRANGE ADD ON
77	Mummy's car down	Like Mummy's car (1)	ADD ON DROP- *like*	OBJ LOC (OBJ down (4))	SUBSTITUTE
86	And your tummy	And W1 W2 (24)	SUBSTITUTE SUBSTITUTE	And OBJ (119)	SUBSTITUTE
102	Got it back on	Mummy's got it on the floor (1)	INSERT- *back* DROP- *Mummy's* DROP- *the floor*	Got OBJ on (2)	SUBSTITUTE INSERT- *back*
113	What's that funny drawing down there?	What's that lie-ing down? (1)	SUBSTITUTE INSERT- *drawing* ADD-ON	What's that OBJ LOC? (1)	SUBSTITUTE SUBSTITUTE
D3	What's that moving ladder?	What's that lie-ing down (1)	SUBSTITUTE SUBSTITUTE	What's that OBJ (117)	SUBSTITUTE
D6	I don't put it in Dave	I don't put it on there (1)	SUBSTITUTE SUBSTITUTE	I don't put it LOC (3)	SUBSTITUTE
D8	Thats Cally's boobie	And that's Cally outside (1)	DROP- *and* INSERT- *'s* SUBSTITUTE	That's OBJ (37)	SUBSTITUTE

Words underlined indicate that they have been said together before and count as 1 for the purpose of slot filling. *mor = morpheme*.

https://doi.org/10.1017/S0305000903005592 Published online by Cambridge University Press

APPENDIX B

UTTERANCES WITH INCREASED NUMBERS OF OPERATIONS UNDER ANALYSIS 2

Utt no.	Utterance	Closest match 1	Operations 1	Closest match 2	Operations 2
13	Can have it	I have it (2)	SUBSTITUTE	Can have OBJ Can have have OBJ pleasey (1)	DROP- OBJ DROP- *please* SUBSTITUTE
54	All on floor	Down here on floor (1)	SUBSTITUTE	OBJ LOC (check)	SUBSTITUTE SUBSTITUTE
58	I want to paint	I want to get it (1)	SUBSTITUTE	I want to W OBJ (check)	SUBSTITUTE DROP- *it*
76	It's going up	Going up (3)	ADD ON - *It's*	It's going to fall LOC (1)	DROP- *to fall* SUBSTITUTE
78	You sorry tiger	I'm sorry tiger (1)	SUBSTITUTE	Sorry OBJ (8)	SUBSTITUTE ADD ON
79	This ones go here	This ones Piglet here (1)	SUBSTITUTE	Which ones go here? Which OBJs go LOC?	DROP- *Which* SUBSTITUTE
83	And that's a big lion	That's a big lion (1)	ADD ON- *and*	And OBJ's OBJ And this one's Winnie+thePooh (1)	SUBSTITUTE SUBSTITUTE