Research Article

Stored Word Sequences in Language Learning

The Effect of Familiarity on Children's Repetition of Four-Word Combinations

Colin Bannard[1] and Danielle Matthews[2]

[1]*Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany, and* [2]*Max Planck Child Study Centre, School of Psychological Sciences, University of Manchester*

ABSTRACT—*Recent accounts of the development of grammar propose that children remember utterances they hear and draw generalizations over these stored exemplars. This study tested these accounts' assumption that children store utterances as wholes by testing memory for familiar sequences of words. Using a newly available, dense corpus of child-directed speech, we identified frequently occurring chunks in the input (e.g.,* sit in your chair*) and matched them to infrequent sequences (e.g.,* sit in your truck*). We tested young children's ability to produce these sequences in a sentence-repetition test. Three-year-olds (*n *= 21) and 2-year-olds (*n *= 17) were significantly more likely to repeat frequent sequences correctly than to repeat infrequent sequences correctly. Moreover, the 3-year-olds were significantly faster to repeat the first three words of an item if they formed part of a chunk (e.g., they were quicker to say* sit in your *when the following word was* chair *than when it was* truck*). We discuss the implications of these results for theories of language development and processing.*

Perhaps the most remarkable thing about language is its compositional nature—the fact that a limited number of sounds can be arranged in unfamiliar combinations to produce novel meanings. So impressive is such productive grammar that researchers have often argued it would be impossible to learn—it must somehow be innate. However, a number of recent accounts propose that children acquire the grammar of their native language (or languages) simply by observing patterns and generalizations in the input. One such view that has been gaining in popularity is the constructivist, or usage-based, account (e.g., Goldberg, 2006; Tomasello, 2003).

According to this account, children begin with a restricted set of utterances taken directly from experience and acquired via the domain-general skills of imitation and intention reading. Children then advance to productive syntax by generalizing over these utterances. This account relies on certain controversial assumptions. The most fundamental is that children are able to store whole sequences of words taken directly from the input. Although naturalistic observation supports this assumption (e.g., Clark, 1970; Lieven, Behrens, Speares, & Tomasello, 2003; Peters, 1983), no experimental work has tested it. The study reported in this article tested experimentally whether children store and reuse sequences of multiple words.

One reason one might expect children to store more than individual words in memory is that they do not hear demarcated words in the input; words and phrases run into one another and must be detected in the speech stream. Recently, it has been argued that children could segment speech by observing how regularly sounds co-occur in the language. Sounds that occur together frequently can be taken to be words or components of words. Conversely, word boundaries can be posited at points where low-frequency transitions between sounds are observed (Saffran, Aslin, & Newport, 1996; Thiessen & Saffran, 2003). A child performing segmentation in this fashion is likely to arrive at an inventory of segments containing not just conventional words, but also a number of multiword sequences.

Figure 1 illustrates this argument by showing the frequency of units of between one and five words observed in a 1.7-million-word corpus of one mother's speech to her child (see the Method section for details). The figure plots frequency against rank on

Address correspondence to Colin Bannard, Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, Leipzig D-04103, Germany, e-mail: bannard@eva.mpg.de.

 Copyright © 2008 Association for Psychological Science

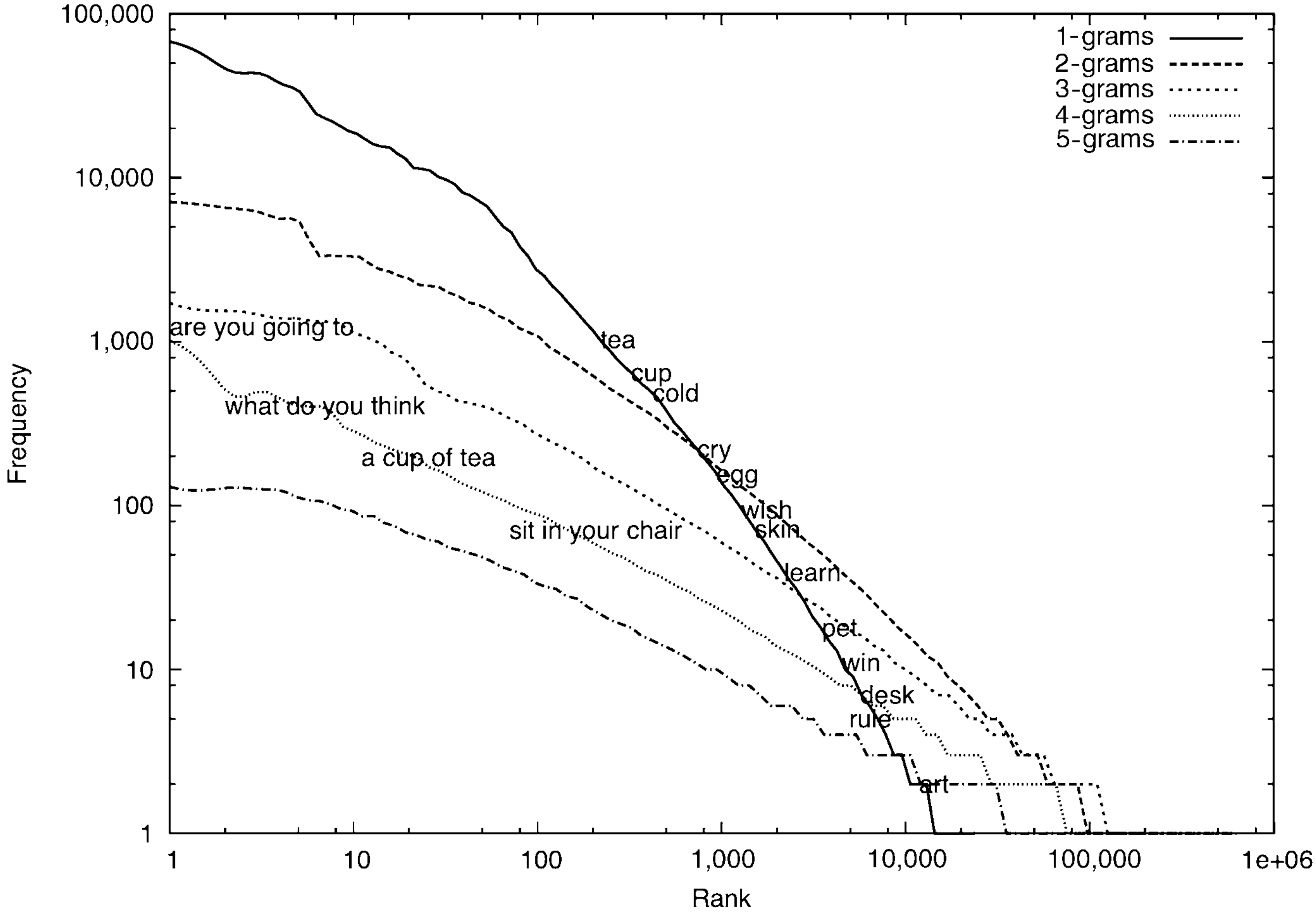

Fig. 1. Frequency of linguistic units plotted against their rank order, on logarithmic scales. Each line shows results for units of a different length (one through five words). Examples of four-word sequences and single words are also plotted. The data were derived from a corpus of 1.72 million words of child-directed speech.

logarithmic scales. The graph shows that there are many multiword sequences (e.g., *what do you think*) that occur more frequently than single words (e.g., *learn*, *pet*) that are part of the core vocabulary of English. Furthermore, there are multiword sequences (e.g., *a cup of tea*) that occur with a frequency almost equal to that of their component words (e.g., *cup*, *tea*). Given such a pattern of frequencies, an efficient language processor could be expected to extract form-meaning mappings for the whole phrase, as well as for individual words.

The prediction that children should extract multiword phrases from the speech stream is supported by models built to segment everything from artificial strings (e.g., Elman, 1990; Perruchet & Vinter, 1998) to child-directed speech (Brent, 1999; Brent & Cartwright, 1996), transcribed adult conversational speech (Cairns, Shillcock, Chater, & Levy, 1997), and written texts (de Marcken, 1996; Kit & Wilks, 1999). This work has consistently shown that if models are to avoid oversegmenting (erroneously identifying word parts as words), they must undersegment to some degree; that is, they must extract a large number of "common sequences that incorporate more than one word, but . . . co-occur frequently enough to be treated as a quasi-unit" (Elman, 1990, p. 193). From an information-theoretic point of view, there is also good reason to think that multiword storage continues to be the most efficient strategy available to the child even after full segmentation is possible. Shannon (1948) pointed out that when selecting a code to transmit information over some channel (a situation that has useful analogies to the child finding appropriate speech segments with which to produce and receive messages), an efficient strategy must consider the frequency with which different units occur. If one assumes that in language processing it is efficient to reduce the number of segment-retrieval operations required, then given the high frequency with which many multiword sequences appear, the additional storage cost of keeping such sequences in the lexicon in addition to individual words would be justified by the performance gains. That the best code, given the statistics of natural language, includes multiword elements has been repeatedly shown by both minimum-description-length modeling (e.g., Kit & Wilks, 1999) and related work in text compression (e.g., Ziv & Lempel, 1978).

Although these analyses are suggestive, there is little experimental work testing children's retention of frequent multiword sequences. Studies that have demonstrated effects of frequency on children's grammatical knowledge have done so only at the single-word level (e.g., Ambridge, Rowland, Pine, & Young, 2008;

Kidd, Lieven, & Tomasello, 2006; Marchman, 1997; Matthews, Lieven, Theakston, & Tomasello, 2005; Matthews & Theakston, 2006) or at most for combinations of two words (Rowland, 2007; for related findings with adults, see, e.g., McDonald & Shillcock, 2003; Pluymaekers, Ernestus, & Baayen, 2005; Reali & Christiansen, 2006). The current study tested whether frequency effects can be observed for longer sequences.

The logic of the current experiment is based on a classic study of an analogous problem in inflectional morphology. Taft (1979) sought to establish whether people always process words by decomposing them into their component morphemes (e.g., *dogs* > *dog* + *s*), or whether people sometimes process words directly as wholes. He reasoned that if the frequency of whole forms affects processing independently of the frequency of their components, then it must be the case that people store information about the whole forms. Identifying pairs of words in which one whole form had a very high frequency (e.g., *things*) and the other a low frequency (e.g., *worlds*), but in which the summed frequency of the base form and its inflectional versions was the same for the two words (e.g., frequency of *world* + frequency of *worlds* = frequency of *thing* + frequency of *things*), Taft reasoned that any effect of whole-form frequency would be evidence of whole-form storage. Participants performed a lexical decision task, and, indeed, a processing advantage was found for the high-frequency whole forms. These results provided support for the idea that people store some complex words as wholes.

Extending this logic, we chose as stimuli pairs of sequences that were identical except for the final word. In each pair, one sequence was highly frequent (e.g., *a drink of tea*), and the other was infrequent (e.g., *a drink of milk*), although the final words (*tea*, *milk*) and also the final bigrams (*of tea*, *of milk*) within each pair were matched for frequency. We hypothesized that children would show a processing advantage for the frequent over the infrequent sequence in each pair. As the first three words (the "stems") within each pair were identical and the final words and bigrams were matched for frequency, any such effect would necessarily result from the whole combinations of words and not the component words or pairs of words. This effect would therefore be evidence that children store information about frequent sequences of words.

We used a repetition task to probe children's knowledge. There is evidence that when asked to repeat sequences of words, children analyze what they hear and reproduce it as they would a regular utterance (Potter & Lombardi, 1990; see also Kidd et al., 2006; Valian & Aubry, 2005). In the current study, we hypothesized that children would more easily and accurately repeat high-frequency than low-frequency combinations.

METHOD

Participants

Thirty-eight normally developing, monolingual, English-speaking children were included in the study (12 boys, 26 girls). There were seventeen 2-year-olds (range = 2 years 4 months through 2 years 9 months, mean age = 2 years 6 months) and twenty-one 3-year-olds (range = 3 years 1 month through 3 years 6 months, mean age = 3 years 4 months). Five additional children were tested but not included because of fussiness. The children were tested in the Max Planck Child Study Centre, Manchester, United Kingdom. Parental consent was obtained.

Materials and Design

We created stimuli using the Max Planck Child Language Corpus, collected by the Max Planck Child Study Centre, Manchester. This corpus contains the speech addressed to and produced by a single child, Brian, when he was between the ages of 2 and 5. It was the largest corpus of child-directed speech available to us (1.72 million words of maternal input over 0.33 million utterances). For this experiment, we were interested in the language that children hear, and consequently took frequencies from Brian's mother's speech.

Using the method of Yamamoto and Church (2001), we extracted all repeated sequences of words from the corpus. This provided us with the distribution of events shown in Figure 1. We chose four-word sequences as stimuli, as four words was the greatest length yielding a wide enough frequency range, and four-word sequences would be sufficiently long to elicit variance in participants' performance in a repetition task (cf. Valian & Aubry, 2005). In selecting our test sequences, we applied a number of additional constraints. We required that (a) each sequence had been produced by Brian's mother at least once as a whole utterance, and not only as part of an utterance (so repeating the string in isolation would not be unnatural); (b) no sequence formed a question (children might be tempted to answer a question rather than repeat it); and (c) no sequence consisted of repetitions of the same word (e.g., *no, no, no, no*). Our most frequent item was *I don't know what*, which occurred 260 times (a natural-log frequency of 5.56). Our log frequency range was 0 through 5.56.

Within this set of candidates, we looked for high-frequency sequences that could be matched with low-frequency sequences and obtained 13 such pairs (see Table 1). All high-frequency sequences came from the top third of the frequency range (e.g., *a lot of noise*: log frequency = 4.66). All low-frequency sequences were from the bottom half of the frequency range, with all but one item being in the bottom third of the frequency range (e.g., *a lot of juice*: log frequency = 0.69). The final words of matched sequences were controlled for (a) the frequency of the final word (e.g., *juice* and *noise* are roughly equally frequent), (b) the frequency of the final bigram (e.g., *of juice* and *of noise* are roughly equally frequent), and (c) the length of the final word in syllables. Mann-Whitney tests confirmed that the high- and low-frequency sequences did not differ significantly in the final word's frequency ($U = 68$; $Z = -.847$; $p = .397$, $p_{\text{rep}} = .573$) or

TABLE 1
Stimulus Sequences Used in the Experiment, With Corresponding Logarithmic Frequencies

Four-word sequence	Frequency category	Sequence frequency	Final-word frequency	Final-bigram frequency	Final-trigram frequency
When we go out	High	3.69	8.44	5.92	4.13
When we go in	Low	1.10	9.80	6.05	2.20
A drink of milk	High	4.04	6.69	5.06	4.37
A drink of tea	Low	2.40	6.94	5.84	2.40
Sit in your chair	High	4.26	6.95	5.48	4.78
Sit in your truck	Low	0.00	6.78	3.53	2.08
We haven't got any	High	4.23	7.11	5.47	4.88
We haven't got enough	Low	0.69	6.30	3.64	1.61
Know what you mean	High	3.76	7.16	6.65	4.78
Know what you need	Low	0.00	7.68	6.26	2.89
Back in the box	High	4.14	7.31	6.57	5.46
Back in the car	Low	1.61	7.10	6.03	5.16
A piece of cheese	High	3.85	6.81	4.93	4.41
A piece of food	Low	0.00	6.60	4.22	0.00
A lot of noise	High	4.66	6.88	4.83	4.66
A lot of fruit	Intermediate	2.08	5.87	3.99	2.40
A lot of juice	Low	0.69	7.01	4.65	0.69
Up in the air	High	4.28	5.40	5.11	4.81
Up in the sky	Intermediate	3.04	5.54	5.41	5.16
Up in the bath	Low	0.00	6.04	5.38	5.12
We've got to go	High	4.36	8.70	7.51	5.53
We've got to look	Intermediate	2.40	9.08	5.93	2.94
We've got to eat	Low	1.39	7.52	6.68	3.95
You want to play	High	4.13	7.06	6.09	4.52
You want to eat	Intermediate	2.48	7.52	6.68	4.03
You want to work	Low	0.00	6.80	5.77	0.69
Go to the shop	High	3.87	6.78	5.97	4.94
Go to the door	Intermediate	1.95	6.98	6.51	4.43
Go to the top	Low	0.00	6.72	6.08	4.17
Out of the way	High	4.17	6.88	5.89	4.19
Out of the house	Intermediate	2.30	7.21	5.67	3.97
Out of the side	Low	0.00	6.46	5.11	0.00

Note. Intermediate-frequency items were included in regression analyses only.

the final bigram's frequency (U = 68.5; Z = −.821; p = .412, p_{rep} = .562).

For 6 of the 13 pairs of sequences, we identified a third, intermediate-frequency sequence that was matched on the same three criteria (e.g., *a lot of fruit*: log frequency = 2.08). Because we were not able to identify an intermediate-frequency sequence in every case, we did not include these items in our factorial design, but we presented these 6 additional sequences as test stimuli so as to include the time it took children to repeat these items in an additional regression analysis, as detailed in the Results section. The order of presentation of the sequences was fully counterbalanced across the participants. All sequences were read by a female British English speaker with normal declarative intonation. They were recorded in a soundproof booth onto a computer disk using SoundStudio Version 3.5 (Freeverse, New York). The sampling frequency was 44,100 Hz. To ensure that the first three words of each pair were identical, we took one sequence as a base and created the matched sequence by splicing in the final word using the open-source software Audacity Version 1.2.4 (available on the Web at http://audacity.sourceforge.net/). We used randomly selected high-frequency sequences as bases half the time and low-frequency sequences as bases the other half of the time.

Procedure

The experimenter sat with the child at a table in front of a computer (the child sat alone or on his or her parent's knee). The experimenter produced a picture of a tree with stars in the branches and explained that she and the child would cover each star with a sticker. She explained that to get the stickers, they needed to listen to what the computer said and then say the same thing. The experimenter offered to go first. She then clicked on a mouse to play the first of six example sequences, repeated the sequence, and awarded herself a sticker. She repeated this procedure for the next two example sequences and then offered the child a turn for the remaining three examples. The experimenter helped the child or replayed the practice sound files once each if necessary. Each time the child attempted to repeat a sequence, he or she received a sticker.

The experimenter then played the test sequences in exactly the same manner except that no help was given and no sound files were replayed. If the child did not spontaneously repeat a sequence after a reasonable delay, the experimenter prompted the child once (saying, "Can you say that?"). If the child did not then respond, or if anything other than this prompt came between the stimulus sequence and the repetition, the response was excluded. A response was also excluded if the child did not hear the stimulus sequence (e.g., if the child spoke unexpectedly as the sound file played). If a response was excluded for one sequence in a pair, results for the other sequence in the pair were also excluded. In the case of triads, the exclusion of either the high- or the low-frequency item (but not of the intermediate-frequency item) resulted in the exclusion of the rest of the triad. The procedure continued until all 32 sentences had been presented. Responses were recorded onto a computer disk using Audacity Version 1.2.4.

Transcription and Error Coding

Two research assistants blind to the hypothesis of the experiment transcribed and coded the children's responses from audio files. Each word in each sequence was coded for the presence or absence of the errors listed in Table 2. If a child did not make a single error in an entire sequence, this sequence was coded as *correctly repeated*. If a child correctly repeated a pair of sequences, or made only errors of pronunciation, then this pair was included in the duration analysis detailed in the next section.

TABLE 2
Error Codes Used for Responses

Code	Description of error
Repetition	Whole word or one syllable of the word is repeated
Deletion	Whole word is missing
Insertion	A word or isolated phonetic material is inserted between words
Substitution	A word in a sequence is replaced by another word
Mispronunciation	A word is missing a phoneme, has an extra phoneme inserted (e.g., "a loft of noise"), or is a morphological variant of the target word (e.g., "going" instead of "go")

Note. A missing phoneme that yielded a pronunciation compatible with adult speech and regional dialect (e.g., dropping -*'ve* in *we've*, producing a glottal stop instead of word-final *t*) was not scored as an error. The pronunciation of *the* as /də/ was also accepted.

Agreement between the coders was moderately good (Cohen's $\kappa = .586$). In all cases in which the first two coders did not code a word identically, a third research assistant, also blind to the hypotheses of the experiment, listened to the relevant response and resolved the discrepancy.

Duration Coding

We also coded how long, in milliseconds, it took each child to say the first three words of each sequence. We coded all pairs of matched items that a child repeated without error or with errors of pronunciation only (items with other errors were excluded, as the durations of such items would have differed dramatically depending on the nature of the error). The duration of intermediate-frequency items was coded when the corresponding high- and low-frequency sequences had been repeated successfully. Applying these criteria did not leave enough data for the 2-year-olds (68% excluded), and thus only the 3-year-olds' responses were coded for duration (only 34% excluded).

A research assistant blind to the hypothesis of the experiment measured the duration of each sequence from the onset of the first word to the onset of the fourth word using Audacity software. The paired sequences *a piece of cheese* and *a piece of food* were excluded because it was impossible to find the offset of *of* and the onset of *food* in the low-frequency sequence. We excluded the intermediate-frequency item *a lot of fruit* for the same reason. Two participants (approximately 10% of the data) were randomly selected, and their responses were coded by a second blind coder. Reliability was assessed by computing the Pearson's correlation coefficient between the two coders ($r = .99$, indicating high reliability).

RESULTS

All children attempted repetition of most of the items (90% of items repeated). We report analyses based on the mean number of items children repeated correctly and then discuss analyses of the duration of the repeated sequences.

Correct Repetition of Sequences

For 2-year-olds, the mean proportion of correctly repeated sequences was .42 ($SD = .28$) for high-frequency sequences and .32 ($SD = .27$) for low-frequency sequences. For 3-year-olds, the corresponding proportions were .69 ($SD = .19$) and .65 ($SD = .21$). To analyze these data, we arcsine-transformed the proportions and conducted 2 (age) × 2 (sequence frequency) analyses of variance by participants (F_1) and by items (F_2). There was no significant interaction between age and frequency, $F_1(1, 36) = 0.401$, $p = .530$, $p_{rep} = .479$, $\eta^2 = .001$; $F_2(1, 24) = 0.121$, $p = .731$, $p_{rep} = .332$, $\eta^2 = .005$. Older children repeated items better than younger children did, $F_1(1, 36) = 14.344$, $p < .001$, $p_{rep} = .989$, $\eta^2 = .285$; $F_2(1, 24) = 21.878$, $p < .001$, $p_{rep} = .996$, $\eta^2 = .477$. Yet children in both age groups were more likely to repeat high-frequency sequences correctly than to repeat low-frequency sequences correctly, $F_1(1, 36) = 6.358$, $p = .016$, $p_{\mathrm{rep}} = .935$, $\eta^2 = .15$; $F_2(1, 24) = 3.561$, $p = .071$, $p_{\mathrm{rep}} = .85$, $\eta^2 = .129$. The same pattern of results emerged when we conducted analyses without arcsine-transforming the proportions and when we counted items that had errors on the fourth word only as correctly repeated.

Duration of Three-Word Stems

The mean duration of the first three words of the repeated sequences was 6,895 ms ($SD = 96.51$) in the high-frequency condition and 7,167 ms ($SD = 100.65$) in the low-frequency condition. One-tailed paired-samples t tests by participants and by items revealed that children repeated the first three words significantly faster for high-frequency than for low-frequency sequences, $t_1(19) = 1.923$, $p = .035$, $p_{\mathrm{rep}} = .900$, $\eta^2 = .156$; $t_2(11) = 1.998$, $p = .034$, $p_{\mathrm{rep}} = .902$, $\eta^2 = .266$.

To further investigate the relation between sequence frequency and duration, we fitted a simultaneous multiple regression model to the duration data, including final-word frequency, final-bigram frequency, and four-word-sequence frequency as predictor variables. Following the standard procedure for regressions over repeated measures data, we entered children into the model using dummy variables (Lorch & Myers, 1990). Because high-, low-, and intermediate-frequency items were matched for the first three words, we also entered item group into the model using dummy variables (e.g., items beginning with *a lot of* were coded as one item group). The outcome variable was the duration (in milliseconds) of the first three words of each sequence. These durations were log-transformed to correct for heteroscedasticity. The results are reported in Table 3. Four-word sequence frequency was a significant predictor of duration. The negative beta value indicates that the more frequent a sequence, the less time children took to produce its first three words. The positive beta value for final-bigram frequency suggests that the more frequent the final bigram of a sequence, the longer the repetition duration. We did not predict this result, but it might suggest that the existence of a frequent bigram inhibits the production of the whole sequence because of competition for activation (see Sosa & MacFarlane, 2002, for a similar finding in adults). However, this hypothesis would also predict a significant effect for individual word frequency, which was not found.

TABLE 3

Summary of the Simultaneous Regression Analysis for Variables Predicting Repetition Duration

Variable	b	$SE\ b$	β
Constant	−0.766	0.192	
Final-word frequency	0.014	0.027	.052
Final-bigram frequency	0.060	0.025	.183*
Sequence frequency	−0.019	0.007	−.123**

Note. $R^2 = .588$.
*$p < .05$. **$p < .01$.

One aspect of our stimuli that we have not yet discussed is the frequency of the final three words of our sequences. This variable was found to have little impact when entered as an additional predictor into our simultaneous model; the frequency of the whole sequence remained a significant predictor ($b = -0.021$, $SE = 0.009$, $\beta = -.130$, $p = .021$, $p_{\mathrm{rep}} = .925$), and trigram frequency had a small positive beta ($b = 0.003$, $SE = 0.014$, $\beta = .015$, $p = .848$, $p_{\mathrm{rep}} = .234$). We should, however, be cautious in interpreting this result because the frequency of the final trigram was positively correlated with overall sequence frequency in our stimuli, giving rise to collinearity in the model. Although these analyses suggest that frequency of the four-word sequence is the stronger predictor of duration in our data, we should certainly not conclude that the frequencies of smaller components do not affect language processing. The results are more likely a consequence of the selection and control of our stimuli. We maximized the variance in frequency of our four-word sequences while minimizing variance in final-word and final-bigram frequency and ignoring final-trigram frequency. We expect that stimuli with a greater range of frequency for the component n-grams would produce effects of frequency at those alternative levels of granularity.

DISCUSSION

The analyses we have reported reveal that the frequency with which word sequences occur in the linguistic environment determines the speed and accuracy with which children are able to produce them in a repetition task. This effect is independent of any effect of syntax, of the frequency of the component words, or of transitional probabilities between pairs of words. This finding is consistent with the general tendency for linguistic events that are encountered more often to be processed more quickly (e.g., Howes & Solomon, 1951; see Ericsson & Kintsch, 1996, for a discussion of expertise effects in language and other domains). Following Taft's (1979) reasoning about morphology, we take our results as evidence of whole-form storage. Our speakers seem to have had some experience-derived knowledge of specific four-word sequences. It seems probable to us that this knowledge was in addition to their knowledge of the individual component words (all words in our stimuli were likely to be familiar to the children independently of the particular sequences), which suggests that the children had complementary representations at different levels of granularity.

What are the implications of this finding for models of language processing and learning? A popular perspective in linguistics assumes a clear distinction between lexicon and grammar (or words and rules). In the words of Ullman (2001), "[People's] use of language depends upon two capacities: a mental lexicon of memorized words and a mental grammar of rules that underlie the sequential and hierarchical composition of lexical forms into predictably structured larger words, phrases and sentences" (p. 37). Because this view incorporates a distinction between memory-based processing at the word level and algorithm-based processing at the multiword level, it is clearly incompatible with our finding. A number of models do give memorized sequences a role, varying in the extent to which they break with the words-and-rules perspective. Some allow for storage of a very large number of sequences, but still posit a special mode of grammatical rule- or constraint-based processing (e.g., Culicover & Jackendoff, 2005; Jackendoff, 2002). More radical is the usage-based approach (e.g., Langacker, 1987), which disregards the distinction between lexicon and grammar altogether, seeing all of language production and comprehension as based on previously experienced exemplars and proposing that grammar is emergent from "the cognitive organization of language experience" (Bybee, 2006, p. 730). The present results do not allow us to determine which of these two kinds of models is correct. However, it is worth noting that our finding that children use multiword sequences in processing even when one would expect individual words to be available to them supports Langacker's (2000) prediction that processing language using concrete, exemplar-based knowledge will be the preferred strategy when such knowledge is available.

The claim that children simultaneously store information about units of language at differing levels of granularity (words, bigrams, and so on) raises some important questions for future research. It was convenient in our introduction to refer to items such as those we studied as "multiword" sequences, but it is not clear whether the children's representations of the sequences overlapped with or were completely disjoint from their representations of the words that a standard linguistic analysis indicates the sequences consist of. It will be crucial to explore what relationships exist between representations at different levels of granularity. For example, are they mutually reinforcing, in competition, or entirely unrelated? To answer this question, one would need to use stimuli in which the frequencies of components of different lengths are systematically varied and pitted against each other.

Our results also raise fundamental developmental questions. Although the older children we tested were significantly better than the younger children at repeating sequences, we observed no interaction between age and frequency. This suggests continuity in frequency effects across development. Would the effect of frequency diminish in older children or adults? Last, researchers need to explore whether the storage of multiword units affects not only the processing of the exemplars, but also the processing of similar items. It is possible that the multiword statistics of the input determine the kinds and extent of children's generalization, thereby shaping the development of productive syntax. The findings we have presented here are an essential first step in developing an input-driven account of the ontogenesis of multiword speech.

Acknowledgments—The authors would like to thank Jess Butcher, Anna Roby, Manuel Schrepfer, and Elizabeth Wills for help in data collection; Roger Mundry for statistical guidance; and Morten Christiansen, Elena Lieven, and Mike Tomasello for helpful comments.

REFERENCES

Ambridge, B., Rowland, C.F., Pine, J.M., & Young, C.R. (2008). The effect of verb semantic class and verb frequency (entrenchment) on children's and adults' graded judgements of argument-structure overgeneralization errors. *Cognition*, *106*, 87–129.

Brent, M.R. (1999). Speech segmentation and word discovery: A computational perspective. *Trends in Cognitive Sciences*, *3*, 294–301.

Brent, M.R., & Cartwright, T.A. (1996). Distributional regularity and phonotactic constraints are useful for segmentation. *Cognition*, *61*, 93–125.

Bybee, J.L. (2006). From usage to grammar: The mind's response to repetition. *Language*, *82*, 711–733.

Cairns, P., Shillcock, R.C., Chater, N., & Levy, J. (1997). Bootstrapping word boundaries: A bottom-up corpus-based approach to speech segmentation. *Cognitive Psychology*, *33*, 111–153.

Clark, R. (1970). Performing without competence. *Journal of Child Language*, *1*, 1–10.

Culicover, P.W., & Jackendoff, R. (2005). *Simpler syntax*. Oxford, England: Oxford University Press.

de Marcken, C. (1996). *Unsupervised language acquisition*. Unpublished doctoral dissertation, Massachusetts Institute of Technology, Cambridge, MA.

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

Ericsson, K.A., & Kintsch, W. (1996). Long-term working memory. *Psychological Review*, *102*, 211–245.

Goldberg, A.E. (2006). *Constructions at work: The nature of generalization in language*. Oxford, England: Oxford University Press.

Howes, D., & Solomon, R. (1951). Visual duration threshold as a function of word-probability. *Journal of Experimental Psychology*, *41*, 401–410.

Jackendoff, R. (2002). *Foundations of language: Brain, meaning, grammar, evolution*. Oxford, England: Oxford University Press.

Kidd, E., Lieven, E.V.M., & Tomasello, M. (2006). Examining the role of lexical frequency in the acquisition and processing of sentential complements. *Cognitive Development*, *21*(2), 93–107.

Kit, C., & Wilks, Y. (1999). Unsupervised learning of word boundary with description length gain. In M. Osborne & E.T.K. Sang (Eds.), *Computational Natural Language Learning (CoNLL-99)* (pp. 1–6). Retrieved October 2005 from http://www.cnts.ua.ac.be/conll99/programme.html

Langacker, R.W. (1987). *Foundations of cognitive grammar: Vol. 1. Theoretical prerequisites*. Stanford, CA: Stanford University Press.

Langacker, R.W. (2000). A dynamic usage-based model. In S. Kemmer & M. Barlow (Eds.), *Usage based models of language* (pp. 1–63). Stanford, CA: CSLI Publications.

Lieven, E.V.M., Behrens, H., Speares, J., & Tomasello, M. (2003). Early syntactic creativity: A usage-based approach. *Journal of Child Language*, *30*, 333–370.

Lorch, R.F., & Myers, J.L. (1990). Regression analyses of repeated measures data in cognitive research. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*, 149–157.

Marchman, V.A. (1997). Children's productivity in the English past tense: The role of frequency, phonology, and neighborhood structure. *Cognitive Science*, *21*, 283–304.

Matthews, D.E., Lieven, E.V.M., Theakston, A., & Tomasello, M. (2005). The role of frequency in the acquisition of English word order. *Cognitive Development*, *20*, 121–136.

Matthews, D.E., & Theakston, A. (2006). Errors of omission in English-speaking children's production of plurals and the past tense: The effects of frequency, phonology and competition. *Cognitive Science*, *30*, 1027–1052.

McDonald, S.A., & Shillcock, R.C. (2003). Low-level predictive inference in reading: The influence of transitional probabilities on eye movements. *Vision Research*, *43*, 1735–1751.

Perruchet, P., & Vinter, A. (1998). PARSER: A model for word segmentation. *Journal of Memory and Language*, *39*, 246–263.

Peters, A. (1983). *The units of language acquisition*. Cambridge, England: Cambridge University Press.

Pluymaekers, M., Ernestus, M., & Baayen, R. (2005). Articulatory planning is continuous and sensitive to informational redundancy. *Phonetica*, *62*, 146–159.

Potter, M.C., & Lombardi, L. (1990). Regeneration in the short-term recall of sentences. *Journal of Memory and Language*, *29*, 633–654.

Reali, F., & Christiansen, M.H. (2006). Word-chunk frequency affects the processing of object and subject relative clauses. *Quarterly Journal of Experimental Psychology*, *60*, 161–170.

Rowland, C. (2007). Explaining errors in children's questions. *Cognition*, *104*, 106–134.

Saffran, J.R., Aslin, R.N., & Newport, E.L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*, 1926–1928.

Shannon, C.E. (1948). A mathematical theory of communication. *The Bell Systems Technical Journal*, *27*, 379–423.

Sosa, A.V., & MacFarlane, J. (2002). Evidence for frequency-based constituents in the mental lexicon: Collocations involving the word *of*. *Brain and Language*, *83*, 227–236.

Taft, M. (1979). Recognition of affixed words and the word frequency effect. *Memory & Cognition*, *7*, 263–272.

Thiessen, E.D., & Saffran, J.R. (2003). When cues collide: Use of stress and statistical cues to word boundaries by 7- to 9-month-old infants. *Developmental Psychology*, *39*, 706–716.

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

Ullman, M.T. (2001). The declarative/procedural model of lexicon and grammar. *Journal of Psycholinguistic Research*, *30*, 37–69.

Valian, V., & Aubry, S. (2005). When opportunity knocks twice: Two-year-olds' repetition of sentence subjects. *Journal of Child Language*, *32*, 617–641.

Yamamoto, M., & Church, K. (2001). Using suffix arrays to compute term frequency and document frequency for all substrings in a corpus. *Computational Linguistics*, *27*, 1–30.

Ziv, J., & Lempel, A. (1978). Compression of individual sequences via variable length coding. *IEEE Transactions on Information Theory*, *24*, 530–536.

(RECEIVED 5/18/07; REVISION ACCEPTED 8/6/07)