The Interaction Between Vocabulary Size and Phonotactic Probability Effects on Children's Production Accuracy and Fluency in Nonword Repetition

Jan Edwards
Mary E. Beckman
Ohio State University,
Columbus

Benjamin Munson
University of Minnesota,
Minneapolis

Adults' performance on a variety of tasks suggests that phonological processing of nonwords is grounded in generalizations about sublexical patterns over all known words. A small body of research suggests that children's phonological acquisition is similarly based on generalizations over the lexicon. To test this account, production accuracy and fluency were examined in nonword repetitions by 104 children and 22 adults. Stimuli were 22 pairs of nonwords, in which one nonword contained a low-frequency or unattested two-phoneme sequence and the other contained a high-frequency sequence. For a subset of these nonword pairs, segment durations were measured. The same sound was produced with a longer duration (less fluently) when it appeared in a low-frequency sequence, as compared to a high-frequency sequence. Low-frequency sequences were also repeated with lower accuracy than high-frequency sequences. Moreover, children with smaller vocabularies showed a larger influence of frequency on accuracy than children with larger vocabularies. Taken together, these results provide support for a model of phonological acquisition in which knowledge of sublexical units emerges from generalizations made over lexical items.

KEY WORDS: phonological development, lexical development, phonotactic probability

Traditional models of grammar posit that phonological knowledge is instantiated in the form of rules or constraints operating on abstract mental representations of words. A fundamental assumption of these models is that the rules and constraints of phonology exist in a module of the grammar that is quite separate from the words whose structure they govern. This assumption is difficult to reconcile with a growing body of research that suggests that phonological processing in adult speakers of English is tightly coupled to the phonological structures of the words that they know. In particular, it is sensitive to the relative frequencies with which different sublexical sequences occur in the lexicon. These relative frequencies are often called phonotactic probabilities or transitional probabilities, reflecting the fact that they are usually expressed as the probability that a sequence of sounds will occur in a lexical item.

Sensitivity to phonotactic probability has been demonstrated using a variety of measures of implicit or procedural knowledge. For example, adults are faster to repeat nonwords that contain high-frequency consonant–vowel and vowel–consonant sequences (Vitevitch & Luce, 1999; Vitevitch, Luce, Charles-Luce, & Kemmerer, 1997). Their speeded repetitions of nonwords containing high-frequency sequences also are more accurate, although this effect is not as robust or as consistently replicated across experiments as is the effect on response time. Phonotactic probability also influences speech perception in adults. For example, listeners are biased to hearing acoustically ambiguous consonant segments as members of high-probability sequences (Pitt & McQueen, 1998). Furthermore, when adults are asked to transcribe nasal-obstruent sequences embedded in nonwords, their transcription errors are more likely to "correct" a low-frequency sequence by writing a phonetically similar but more frequent sequence (Hay, Pierrehumbert, & Beckman, in press). Adults also have a better recognition memory for nonwords containing high-probability sequences of phonemes than for those containing low-probability sequences (Frisch, Large, & Pisoni, 2000).

Sensitivity to phonotactic probability is also reflected in explicit judgments of how well a nonword conforms to the phonological patterns attested in real words. When asked to judge how "wordlike" nonsense words are, adults give higher wordlikeness ratings to forms that contain phoneme sequences that are attested in many words. This result is extremely robust and has been seen in a large number of experiments (e.g., Coleman & Pierrehumbert, 1997; Frisch et al., 2000; Munson, 2001; Pierrehumbert, 1994; Vitevitch et al., 1997). Moreover, it interacts with vocabulary size (Frisch, 2001). Whereas adults with large vocabularies differentiate sequences with varying low frequencies by assigning them different (low) wordlikeness ratings, adults with small vocabularies assign the same (lowest) wordlikeness rating to many low-frequency sequences, as if they were all equally unattested in the lexicon.

Together, the results of these studies on implicit and explicit phonological knowledge support a rather different view of phonology than the encapsulated module of the traditional accounts of adult grammars. The kind of "probabilistic phonology" that they support was laid out most completely in Pierrehumbert (2003). In a probabilistic phonology, a familiar word is represented in the mental lexicon by a rich hierarchy of different types of phonological representation, each of which encodes some level of abstraction over representations at one or more other levels. Even parametric representations, such as the vowel formant trajectories or lingual contact patterns in different utterances of the word *daddy*, are abstractions away from the signal that is presented to the listener's ears or to the speaker's kinesthetic receptors. The phonetic encoding of these parametric representations into categories, such as the voiceless unaspirated alveolar stop that occurs at the beginning of many utterances of the words *daddy*, *dancing*, *dandelion*, and so on, is a further abstraction, based on generalizations over multiple experiences of utterances of these words and others like them. As the research on phonotactic probability effects shows, adults have also abstracted away a "phonological grammar" of generalizations about where different phonetic categories are likely to occur. The phonotactic probabilities discussed above are an abstraction of this type. These generalizations shape adults' perceptual parsing of new words (including the nonword stimuli in Pitt and McQueen, 1998, or in Hay et al., in press). They also affect how easy it is to accept a new word as a possible word of the language (in the wordlikeness rating experiments) and to establish memory traces for new words (as in Frisch et al., 2000). In a probabilistic phonology, then, phonetic categories, such as the category /d/ at the beginning of *daddy*, and phonological constraints, such as the fact that /dʊ/ and /di/ are much less likely sequences to begin a word than /dæ/, are simply two different kinds of abstraction that develop during the course of extended experience with learning and using words. Phonetic categories begin to emerge as the infant notices recurring patterns in the parametric representations of sets of utterances, whereas phonological constraints emerge as the language user notices commonalities among the sound shapes of words in the lexicon.

The idea that the lexicon plays a key role in phonological development is not new. Almost 30 years ago, Ferguson and Farwell (1975) proposed that "a phonic core of remembered lexical items and the articulations that produced them is the foundation of an individual's phonology…even though it may be heavily overlaid or even replaced by phonologically organized acquisition processes in later stages" (p. 437). However, while there is a large body of research on adults' sensitivity to generalizations over the lexicon in both perception and production, there is relatively little comparable research on young children. The few studies that have been done suggest that children, as well as adults, are sensitive to phonotactic probability. For example, Storkel (2001) found that 3–6-year-old children learned new words more rapidly when the words contained high probability sequences as compared to low probability sequences. Gathercole, Frankish, Pickering, and Peaker (1999) found that 7- and 8-year-old children repeated lists of nonwords more accurately in a serial recall task when the nonwords contained only high-frequency consonant–vowel and vowel–consonant sequences. Using a less demanding immediate repetition task, Beckman and Edwards (2000a) found that children 3–5 years of age repeated high-frequency two-phoneme sequences in nonwords more accurately than they repeated low-frequency two-phoneme sequences. Munson (2001) found an influence of phonotactic probability on production fluency as well as on accuracy. He used segment duration as a measure of fluency and found that children from 3 to 8 years of age produced shorter durations for the same segment when it was in a high-frequency consonant–consonant sequence, as compared to a low-frequency sequence.

In this article, we continue to explore the influence of sublexical sequence frequency on production accuracy and fluency in children. A second focus of the article is on the relationship between the effect of sublexical sequence frequency and estimates of the child's vocabulary size. Specifically, we wanted to determine whether this effect of frequency, if observed, was mediated by vocabulary size. Gathercole et al. (1999) found an effect of vocabulary size on accuracy overall, but no interaction of high versus low vocabulary scores with high versus low transitional probabilities. However, the claim that children acquire a phonological system based on generalizations over the lexicon predicts that children with larger lexicons should have more robustly generalized phonological systems. Their representations of familiar sublexical patterns can be more quickly accessed and more flexibly reapplied to less familiar but analogous patterns. Children with smaller vocabularies, conversely, will know fewer words that exemplify any particular sequence in a variety of larger contexts, as well as fewer words that exemplify the component segments in a variety of more or less similar sequences. Smaller vocabularies thus provide less support for abstracting knowledge about the acoustics and articulation of consonants and vowels away from the specific contexts in which they have been encountered. Representations of familiar sublexical patterns are more fragile and cannot be reapplied as flexibly to form production routines for less familiar but analogous patterns. This effect might be particularly evident in younger children, in whom the same absolute difference in vocabulary size means a proportionally larger difference in experience—that is, a proportionally larger difference in the support for a robust representation of the individual phonological components independent of specific contexts. This predicts that the effect of low transitional probability on a simpler repetition task might be especially pronounced in children with small vocabularies.

We tested these hypotheses using a nonword repetition task to measure production accuracy and fluency and two standard clinical tests to estimate vocabulary size. Our approach differs from most previous research on children's nonword repetition accuracy in two respects. First, we systematically controlled the phonotactic probability of the sublexical sequences within the nonword stimuli by matching each high-frequency sequence with a minimally different low-frequency sequence. Second, we measured both accuracy and fluency of production. This research also differs from our own previous work in that we tested a much larger group of children with a substantially larger set of stimuli. We found systematic effects of transitional probability on repetition accuracy and fluency, and a relationship between the accuracy effect and the size of the children's vocabularies.

Method

Stimuli

An important concern with the three stimulus sets used in our earlier studies was that they were small—only six item pairs in each of the stimulus sets in Beckman and Edwards (2000a) and only eight pairs in Munson (2001). Therefore, we devised a new stimulus set that was designed to test a much larger range of sublexical sequences in several different syllable and word positions, as well as a good range of transitional probabilities. In order not to make the stimulus set too large for the attention spans of our youngest participants, we kept the design of our earlier studies in which stimulus items were paired. One member of each nonword pair contained a low-frequency target that occurred in few or no words that would likely be familiar to children; the other member of the nonword pair contained a high-frequency target that occurred in many words familiar to children. The two sequences were placed in identical positions within similar nonwords. The final expanded set contained 22 nonword pairs, half of them disyllabic and half trisyllabic, with 7 nonword pairs containing target CV sequences contrasting in low versus high transitional probability, 7 pairs containing target VC sequences, and 8 pairs containing CC sequences, with the last including word-initial onset clusters and word-final coda clusters, as well as word-medial heterosyllabic clusters. The stimuli are listed in Table 1, along with wordlikeness ratings and two measures of the phonotactic likelihood of the target sequences.

In constructing the stimuli, we avoided including very late-acquired sounds, such as /ɹ/ and /ʧ/, in the target sequences (Smit, Freilinger, Bernthal, Hand, & Bird, 1990). Vowel errors are infrequent beyond the age of 3;0 (years; months) (Templin, 1957), and all of the consonants included in the target sequences are produced accurately in the relevant word position by 75% of children by age 3;0, except for initial /j/ (3;6) and initial /v/ (4;0) according to Smit et al. These two consonants were used in both the high-frequency and the low-frequency members of the relevant pairs (e.g., /jau/ vs. /ju/). More

Table 1. Nonword pairs, with the low- versus high-frequency (freq.) target sequences underlined. The third column lists segments (Seg.) from pairs for which we measured the duration of one or both target phonemes, and subsequent column pairs show mean wordlikeness rating (on a scale from 1 to 5) and log transitional probabilities for the embedded target sequences calculated from the MHR[a] database and from the HML[b] database.

Phonetic form			Wordlikeness		MHR[a]		HML[b]	
Low freq.	High freq.	Seg.	Low	High	Low	High	Low	High
/jugoin/	/bogib/		3.06	3.30	−12.42	−9.71	−12.92	−10.84
/moipəd/	/mæbɛp/	[m]	2.96	2.76	−13.11	−8.09	−12.00	−7.81
/vugim/	/vɪdæg/	[v]	3.19	2.91	−13.11	−8.73	−12.92	−8.53
/bodəjau/	/medəju/		2.35	2.96	−13.11	−8.37	−14.30	−7.56
/vukɑtɛm/	/vɪtəgɑp/	[v]	2.96	2.65	−13.11	−8.73	−12.92	−8.53
/gaunəpek/	/gitəmok/		2.78	2.64	−12.42	−9.71	−11.82	−10.84
/nʊbəmən/	/nɪdəbɪp/	[n]	1.68	1.88	−13.11	−8.26	−10.84	−7.79
/motauk/	/petik/		3.38	3.50	−13.31	−9.48	−14.59	−9.77
/donug/	/bedæg/		3.08	3.50	−13.31	−9.79	−14.59	−9.62
/tedaum/	/podaud/		2.90	3.11	−13.31	−10.67	−14.59	−11.81
/auptəd/	/iptən/	[pt]	3.79	3.60	−13.31	−9.68	−14.59	−10.67
/dugnəted/	/tʌgnədit/	[g]	2.68	3.03	−13.31	−9.98	−14.59	−10.53
/aukpəde/	/ikbəni/		2.41	2.06	−13.31	−9.48	−14.59	−9.77
/auftəgɑ/	/auntəko/	[au]	2.43	3.11	−13.31	−8.56	−14.59	−8.96
/nəfæmb/	/mɪnæmp/		2.49	3.03	−13.57	−9.32	−15.73	−11.08
/pwɑgəb/	/twɛkɛt/		1.69	2.28	−13.88	−9.93	−13.55	−10.78
/bufkit/	/kiften/	[f]	2.61	3.68	−14.00	−11.11	−15.57	−11.79
/dogdet/	/tæktut/		2.76	3.38	−14.00	−9.75	−15.57	−9.45
/kɛdəwəmb/	/fɪkətæmp/		2.14	3.13	−13.57	−9.32	−15.73	−11.08
/pwɛnətɛp/	/twɛdəmɪn/		1.90	2.13	−13.88	−9.93	−13.55	−10.78
/næfkətu/	/gʌftədaɪ/	[f]	2.73	2.44	−14.00	−11.11	−15.57	−11.79
/dɛgdəne/	/tiktəpo/		2.43	2.54	−14.00	−9.75	−15.57	−9.45

[a] The MHR database is from the study by Moe, Hopkins, and Rush (1982). [b] The HML database is the Hoosier Mental Lexicon (Pisoni et al., 1985).

generally, the consonants used in the two members of a nonword pair were either identical or highly similar in both phonetic identity (e.g., /gd/ vs. /kt/) and age of acquisition. This design ensured that the task assessed children's ability to produce the target sequences, rather than simply their ability to correctly produce a particular phoneme.

The sequences were developed using the MHR database (Moe, Hopkins, & Rush, 1982). This is an online list of pronunciations of the 6,366 most frequently occurring words in the spontaneous continuous speech of first grade children. It was created by making an electronic version of the word list resulting from the Moe et al. study and then extracting phonetic transcriptions for the words from the Carnegie Mellon University Pronouncing Dictionary (http://www.speech.cs.cmu.edu/cgi-bin/cmudict), which gives pronunciations from the same general dialect region as the central Ohio varieties spoken by the children in our study. Each low-probability sequence occurred in either none or very few words in this database, while each high-probability sequence occurred in many words in this database. For the two nonwords of each sequence pair, the sequence was placed in the same prosodic position in the two nonwords and the transitional probability of all other phoneme sequences within the two nonwords was matched as closely as possible.

We calculated the transitional probabilities of the target sequences based on the frequency of the segmental sequence in the target syllable position, adjusted by a factor representing the frequency of the sequence type in that target syllable position (e.g., "syllable-initial CV" for the /jau/ of /bodəjau/, "syllable-initial CC" for the /pw/ of /pwɑgəb/, "hetero-syllabic CC" for the /fk/ of /bufkit/, "syllable-final VC" for the /auf/ of /auftəgɑ/, and "syllable-final CC" for the /mp/ of /mɪnæmp/). The adjustment factor was intended to capture the effect of prosodic context. That is, because phonological acquisition involves developing representations for prosodic structure as well as for the segments that can fill different prosodic positions, frequency of the sequence type should contribute to accuracy of a two-phoneme sequence independently of the frequency of the sequence itself. For instance, both the syllable-initial CV pair of /ju/ and /jau/ target sequences and the syllable-final CC pair of /mp/ and /mb/ target sequences contrast in occurring in many versus no words. The high-frequency sequence /ju/ occurs in words such as *you*, *use*, and *uniform*, and the high-frequency sequence /mp/ occurs in words such as *jump*, *limp*, and *camping*; the low-frequency sequences /jau/ and /mb/ occur in no words at all. However, most English words contain at least one syllable-initial CV sequence, whereas syllable-final CC sequences are relatively more rare. Thus, although /jau/ is no more frequent as a sequence than /mb/, it might be "easier" simply because CV sequences are more frequent than CC sequences. Therefore, the transitional probability of each sequence included two terms—the first being a measure of sequence frequency and the second a measure of the likelihood of the position in which the sequence occurred, as shown in the following equation:

$$p = \ln\left(\frac{S}{D} * \frac{P}{D}\right),$$

where D is the total number of two-phoneme sequences in the database, S is the count of the number of instances in the database in which the particular target sequence occurred in the relevant syllable position (e.g., the number of /fk/ sequences separated by a syllable boundary when tabulating the transitional probability for /fk/, the number of syllables ending in /auf/ when tabulating it for /auf/), and P is the number of instances of the sequence type in that position (e.g., the number of instances of any heterosyllabic CC sequence when tabulating the transitional probability for /fk/, the number of instances of any syllable-final VC sequence when tabulating it for /auf/). As in other studies of the effects of frequency, we took the natural logarithm of this adjusted transitional probability. For sequences with a frequency of zero, we substituted a count of 0.5 for the numerator in the first term (the raw transitional probability of the sequence), because the natural log of 0 is undefined.

We calculated transitional probabilities first by counting occurrences in the MHR database for children, which was our source for the development of the low- and high-frequency sequences. We also calculated the transitional probabilities a second time, based on the Hoosier Mental Lexicon (HML; Pisoni, Nusbaum, Luce, & Slowiacek, 1985), an online 19,000-item database of words from *Webster's Pocket Dictionary* that many researchers have used to compute transitional probability (e.g., Vitevitch et al., 1997). We decided to include transitional probability counts based on the HML because we were concerned that the MHR database might underestimate children's productive vocabulary. Recall that the MHR database is a list of the 6,366 most frequently occurring words in the speech of first graders. The frequencies are based on number of occurrences in a corpus of 285,623 word tokens taken from spontaneous speech that includes both free-topic conversations between peers and more structured narratives elicited using prompts, such as "Tell me about your favorite TV show." This database probably underestimates the expressive vocabulary of many 6-year-old children and necessarily underestimates that of older children and adults. The frequency relationships in the HML were in accord with those in the MHR. Although many sequences that did not occur in any words in the MHR did occur in one or more words in the HML, paired comparison t tests revealed that transitional probabilities were significantly different between the two sequences of each nonword pair in the HML, just as they were in the MHR, $t(21) = 24.45$, $p < .001$ for MHR; $t(21) = 14.04$, $p < .001$ for HML.

These sequences were embedded in larger "frames" (i.e., nonwords) that were matched in relevant aspects for each pair of low- and high-frequency targets. In particular, the frames for any pair were identical in prosodic structure and very similar in segmental content. We did not use segmentally identical frames because our previous studies showed that this induced a practice effect. Instead, we controlled for any effect of the segments in the frame by matching for wordlikeness. We did this by creating a larger list of candidate nonwords for each pair and then choosing the final frames on the basis of a wordlikeness rating study.

Sixteen adults were presented with the larger list of nonwords over headphones in a sound-treated booth and were instructed to rate the nonwords on a 5-point scale (1 = *very unlike a real word* and 5 = *very like a real word*). Five randomized blocks of the nonwords were presented to each adult. The wordlikeness ratings in Table 1 are the mean ratings averaged over all five trials for all participants. Insofar as possible, the final 44 nonwords were selected to minimize differences in wordlikeness ratings across the 2 members of each nonword pair. The difference between each participant's ratings for matched pairs on any trial clustered around 0, with no difference in 34% of the blocks and a difference of only one point in either direction in 39% of the blocks. Thus, we were fairly successful in controlling for wordlikeness across the 2 members of each pair. Nonetheless, the nonwords containing high-frequency sequences were judged on average to be slightly more wordlike than the paired nonwords containing low-frequency sequences (M = 2.98 and 2.65 for nonwords with high-frequency targets and for those with low-frequency targets, respectively, $t[21] = 2.07$, $p = .02$). Given that our purpose was to contrast transitional probabilities at the target sequence itself, we would expect some difference in wordlikeness.

The one remaining question then is whether this difference in mean wordlikeness rating is due to the contrasting transitional probabilities at the target sequence or to the uncontrolled difference in total transitional probability of the frame. Regression analyses showed the mean wordlikeness rating to be significantly correlated with the total transitional probability of the frame ($R^2 = .274$), $F(1, 42) = 15.82$, $p < .001$, but not with the target sequence transitional probability, calculated from either the HML or from the MHR. These results replicate the analyses of Frisch et al. (2000), who showed that whole-form measures of goodness, such as the total log probability of all sequences in the nonword, were better predictors of wordlikeness than local measures of constraint violation, such as the transitional probability of the least likely sequence (i.e., the target sequence in the case of the low-frequency sequences in our stimulus set). At the same time, these results suggest that we need to be careful to correlate our accuracy results with wordlikeness, because Gathercole, Willis, Emslie, and Baddeley (1991), in a test much like the one we report here, found that young children more accurately repeat nonwords that adults have judged to be more wordlike.

Participants

The participants were 104 typically developing children ranging in age from 3;2 to 8;10 and 22 young adults ranging in age from 21 to 34 years. All participants were part of a larger study on phonological knowledge deficits in phonological disorder and were monolingual speakers of English. Each of the 104 children met the following four criteria for typical development: (a) normal articulatory development, as evidenced by a score no more than 1 *SD* below the mean on the Goldman-Fristoe Test of Articulation (GFTA; Goldman & Fristoe, 1986); (b) normal hearing, as evidenced by passing a hearing screening at 20 dB at 500, 1000, 2000, and 4000 Hz; (c) normal structure and function of the peripheral speech mechanism, as evidenced by a standard score no more than 1 *SD* below the mean on the oral movement subtest of the Kaufman Speech Praxis Test for Children (Kaufman, 1995); (d) normal nonverbal IQ, as evidenced by a standard score no more than 1 *SD* below the mean on the Columbia Mental Maturity Scale (Burgemeister, Blum, & Lorge, 1972). Each of the adult participants also passed a hearing screening and had no reported history of speech, language, or hearing problems. Table 2 provides descriptive information for the different participant groups. The last two rows of the table report standard scores on measures of expressive and receptive vocabulary that were administered to all participants.

Procedure

Three pseudorandomized lists of the stimuli were created. For each list, all two-syllable words were presented before the three-syllable words, the two members of a nonword pair were always separated by at least two words, and an equal number of words containing high-frequency sequences were presented before their paired words containing low-frequency sequences and vice versa. The nonwords were played to the participants over two external speakers. The participants were instructed to repeat the nonwords as accurately as possible. Training prior to the experiment consisted of the presentation of two practice nonwords by live voice and then the presentation of two additional digitized practice nonwords over the speakers. Training continued until the participant understood the task and repeated the practice nonwords accurately. No more than the four practice trials were needed for any of the participants. The participants' repetitions were recorded with a head-mounted microphone connected to a digital audiotape (DAT) recorder.

Table 2. Sample size and percentage of males in each age group, and mean age and test scores (with standard deviations in parentheses).

Group characteristics, including test scores	Age group			
	3–4-year-olds	5–6-year-olds	7–8-year-olds	Adults
Sample size	43	38	23	22
Age in months	50 (6)	66 (5)	97 (6)	303 (42)
Gender (% male)	63	61	57	45
GFTA percentile ranking[a]	65 (24)	70 (22)	79 (19)	
CMMS standard score[a,b]	109 (10)	111 (12)	108 (10)	
EVT standard score[b]	111 (9)	110 (13)	102 (7)	120 (11)
PPVT-III standard score[b]	114 (11)	114 (13)	112 (16)	119 (12)

Note. EVT = Expressive Vocabulary Test; PPVT-III = Peabody Picture Vocabulary Test–Third Edition.

[a] The Goldman-Fristoe Test of Articulation (GFTA) and the Columbia Mental Maturity Scale (CMMS) are not normed for adults. [b] Standard scores have a mean of 100 and a standard deviation of 15.

Analysis

Transcription

As a first step in coding the responses for accuracy, the recording for each participant was transferred from the DAT to a digital file on a computer, and the participant's responses were transcribed in the International Phonetic Alphabet at the level of a careful, broad phonemic transcription. That is, transcription was not done directly from the DAT, but using a waveform editor so that each nonword could be played as often as necessary without rewinding the tape. All of the responses were transcribed by a single transcriber. A second transcriber independently transcribed 10% of the data, comprising all responses by 4 participants from the 3–4-year-old group, 4 participants from the 5–6-year-old group, 3 participants from the 7–8-year-old group, and 2 adults. Phoneme-by-phoneme interrater reliability ranged from 86% to 99% for data from individual participants, with a mean of 94% across these 13 participants.

Coding. In coding responses on repetition tasks, researchers often use rather coarse-grained measures of accuracy, such as the number of tokens repeated without error in a string of seven repetitions of the target nonword (e.g., Gathercole et al., 1991) or the proportion of phonemes repeated accurately in the target sequence or syllable within the nonword (e.g., Beckman & Edwards, 2000a; Dollaghan, Biber, & Campbell, 1995; Fisher, Hunt, & Chambers, 2001; Munson, 2001). The use of such coarse-grained measures when coding responses from young children has several disadvantages. First, they do not distinguish between errors related to experimental conditions and "ordinary" mispronunciations that a very young child might make, such as the substitution of [θ] for /s/ or [d] for /g/. Second, they do not distinguish between small subtle errors, such as the place feature substitution that perceptually "corrects" /mt/ to /nt/, and more drastic errors, such as the deletion of the /t/ in the /mt/ cluster so that the /m/ is resyllabified as the onset of the following syllable. Our study covered an extremely large age range, and the larger study included children with phonological disorder who have habitual age-inappropriate mispronunciations. Because the severity and type of error might be more informative of the nature of phonological generalization than the gross error rate, we decided to code the transcriptions using a finer grained segmental accuracy score.

For this segmental accuracy score, each of the two phonemes in a target sequence was scored for accuracy on each of three features. For consonants, one point was awarded for correct place (labial, alveolar, or velar), one point was awarded for correct manner (stop, fricative, or glide), and one point was awarded for correct voicing (voiced or voiceless). For example, if the /k/ in the /kt/ sequence was produced as /s/, it would receive one point for correct voicing, but would lose two points, one for incorrect place and one for incorrect manner. For vowels, one point was awarded for correct production on the dimension front–back (front, central, or back), one point was awarded for correct vowel height (high, mid, or low), and one point was awarded for correct "length" (i.e., tense or lax for a monophthong target and monophthong or diphthong for a diphthong target). For example, an /u/ for /i/ substitution would receive two points, one for correct tenseness and one for correct height, but would lose one point for being a back rather than a front vowel. Thus, the maximum segmental accuracy score for any target sequence was six points, and the minimum score was 0.

Substitution Outcome Probabilities. Each sequence also received a "prosody score" of 1 or 0. The prosody score was coded either as 0, if the response changed the prosodic position of one or both of the target phonemes (e.g., deleting the consonant in a VC sequence to make the vowel the nucleus of an open syllable or inserting a vowel in a medial CC sequence to make the first consonant an onset rather than a coda), or as 1, if the response retained the target prosody (i.e., a mere feature-changing substitution if either phoneme was not produced accurately). Productions that were coded as not completely correct on the segmental accuracy score, but correct on the prosody score could then be identified as "substitution errors."

We were interested to know whether the participants might be more likely to produce substitution errors on low-frequency sequences, where the outcome could be a "correction" to a more likely sequence, suggesting a robust phonological generalization not just of phonemes but of the morpheme structure constraints stating which sequences are likely to occur within a monomorphemic word. Therefore, we analyzed the substitution errors to determine whether the produced sequence had a higher transitional probability than the target sequence. We calculated the transitional probability of the outcome sequence based on the HML database and using the same two-factor formula described in the previous section for the target transitional probabilities listed in Table 1.

Segment Duration. We were also interested in whether fluency of production is related to sublexical sequence frequency. Following Munson (2001), we used segment duration as our measure of production fluency, because duration is an acoustic measure of the speed with which a speech movement is executed (e.g., MacKay, 1982). All other factors being equal, shorter segment durations should indicate greater fluency than longer durations. Duration measurements could be made for 9 of the 22 nonword pairs. These were pairs where the same sound occurred in the target sequence of both members of a nonword pair, and this sound (or this sound and an identical neighboring nontarget phoneme) could be isolated on the waveform. The nonword pairs for which duration measurements could be made are indicated by listing the measured phoneme(s) in Table 1. Measurements were made from the waveform using conventional criteria for determining the onset and offset of each sound. Duration measurements were made only for productions that had completely correct segmental accuracy scores. Because of this restriction, the number of tokens per utterance type was not constant across types. Therefore, an utterance token was included in the statistical analysis only when the matched utterance token produced by the same participant also could be included.

Vocabulary Size Measures. Finally, we wanted to know whether differences in accuracy effects between younger and older participants reflect differences in typical vocabulary size across ages, as suggested above, or are due to some process of typical phonological development that is independent of vocabulary growth. To explore these two possibilities, we used two standardized tests to estimate vocabulary size. For receptive vocabulary size, the Peabody Picture Vocabulary Test–Third Edition (PPVT-III; Dunn & Dunn, 1997) was administered. This widely used measure was most recently revised and renormed in 1997 and has been shown to be much less culturally biased than previous versions (Washington & Craig, 1999). We used the Expressive Vocabulary Test (EVT; Williams, 1997) to measure expressive vocabulary size. These two tests were co-normed for participants aged 2 through 90 years. It can be observed in Table 2 that, overall, the participants have larger than average vocabularies for their ages. Also, the four age groups are well matched for standard scores on the test of receptive vocabulary, but less so for the test of expressive vocabulary. A one-way analysis of variance (ANOVA) showed a significant effect of age on the EVT standard scores, $F(3, 122) = 10.69$, $p < .001$, $\eta^2 = .21$, with adults having significantly higher scores than any of the groups of children, and the 7–8-year-olds having significantly lower scores than the 3–4-year-olds.

In our analyses of the nonword repetitions, we used these scores in two ways. First, we used the raw vocabulary scores as an independent variable in various regression analyses of each participant's mean segmental accuracy for high- versus low-frequency target sequences. Because the relationship between vocabulary size and age is exponential (i.e., vocabulary growth levels off as age continues to increase), we used the natural log of the raw vocabulary scores in all analyses. Second, we used the raw vocabulary scores to group the children by vocabulary size, and compared accuracy for different subsets of the target sequences between the large-vocabulary and small-vocabulary groups. For this second set of analyses, we divided the 104 children into two groups of 52, on the basis of EVT raw scores. The larger-vocabulary and the smaller-vocabulary groups were well separated by this estimate of expressive vocabulary size (for the larger-vocabulary group, $M = 73$, $SD = 11$; for the smaller-vocabulary group, $M = 48$, $SD = 6$).

Results

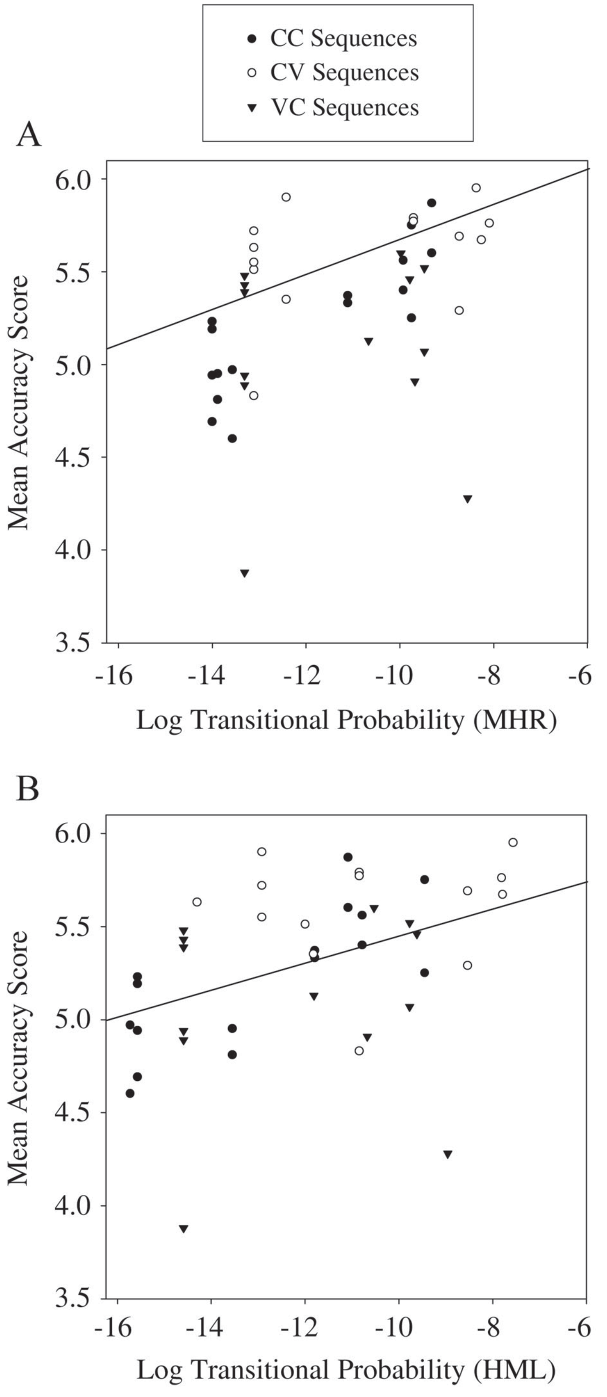
Segmental Accuracy Scores by Item

Accuracy scores were averaged across the 126 participants for each of the target sequences. A paired-comparisons t test on these scores for the 22 nonword pairs revealed a significant effect of frequency on accuracy, $t(21) = 2.89$, $p = .009$. That is, accuracy scores were significantly higher for the target sequences with high transitional probabilities than for the sequences with low transitional probabilities ($M = 5.46$, $SD = 0.38$ for high-frequency sequences; $M = 5.16$, $SD = 0.45$ for low-frequency sequences). The difference between the two sequence types was somewhat more pronounced when the accuracy scores for the adults were not included in the analysis, $t(21) = 3.10$, $p = .005$ ($M = 5.34$, $SD = 0.38$ for high-frequency sequences; $M = 5.03$, $SD = 0.50$ for low-frequency sequences).

Figure 1 shows mean accuracy scores plotted against transitional probability based on each of the two databases, with the three sequence types (CV, VC, and CC) represented by different symbols. The overall trend was for accuracy to be greater for sequences with higher transitional probabilities. Note also that the CV sequences were generally more accurate than would be predicted by transitional probability alone. This was so even though the transitional probabilities were adjusted to reflect the greater probability of the CV sequence type. There are also two outliers in these graphs, the low-frequency sequence /auk/ and the high-frequency sequence /aun/, both of which have lower accuracy scores than would be predicted by their transitional probabilities.

In order to determine whether this effect of transitional probability can be explained by the differences in wordlikeness of the overall nonwords rather than by the transitional probabilities of the target sequences themselves, we correlated the mean accuracy scores for the target sequences with each of the three stimulus properties listed in Table 1. That is, we correlated mean accuracy of the sequences with their transitional probabilities as measured in the child-sized MHR database and in the adult-sized HML database, and we correlated the mean accuracy of the sequences with the mean wordlikeness scores of the nonwords in which they were embedded. Accuracy was significantly correlated with both measures of the target sequence probability ($r^2 = .18$, $p = .004$ for MHR; $r^2 = .19$, $p = .003$ for HML), but not with

Figure 1. Mean accuracy for target sequence plotted against its transitional probability calculated from (a) the Moe, Hopkins, and Rush (MHR; 1982) database and (b) the Hoosier Mental Lexicon (HML; Pisoni et al., 1985) for all 44 nonwords.



wordlikeness scores ($r^2 = .07$, $p = .09$). Moreover, the correlations between mean accuracy and transitional probabilities remained significant even when the effect of wordlikeness was partialled out (partial $r^2 = .15$, $p = .009$ for MHR; partial $r^2 = .18$, $p = .005$ for HML).

Segmental Accuracy Scores by Participant

Figure 2 shows mean accuracy scores for the high-frequency and low-frequency sequences for the four age groups. A two-way (Frequency × Age Group) mixed-model ANOVA showed a significant main effect of frequency, $F(1, 122) = 128.30$, $p < .001$, $\eta^2 = .51$; a significant main effect of age group, $F(3, 122) = 23.30$, $p < .001$, $\eta^2 = .36$; and a significant Frequency × Age Group interaction, $F(3, 122) = 6.56$, $p < .001$, $\eta^2 = .14$. The interaction was due to the larger difference between low- and high-frequency sequences for the three groups of children, as compared to the adults. That is, post hoc tests of simple main effects found a significant main effect of sequence frequency for all four age groups. Measures of effect size, however, showed that target sequence frequency affected the segmental accuracy scores for adult repetitions less than it affected segmental accuracy for any of the three groups of children.

Outcome Sequence Frequency

Figure 3 shows the transitional probability for each of the 44 target sequences plotted against the mean outcome transitional probability (averaged across the 126 participants) for substitution errors. An $x = y$ reference

Figure 2. Mean accuracy scores (with standard errors) for the low- and high-frequency sequences for the four age groups.

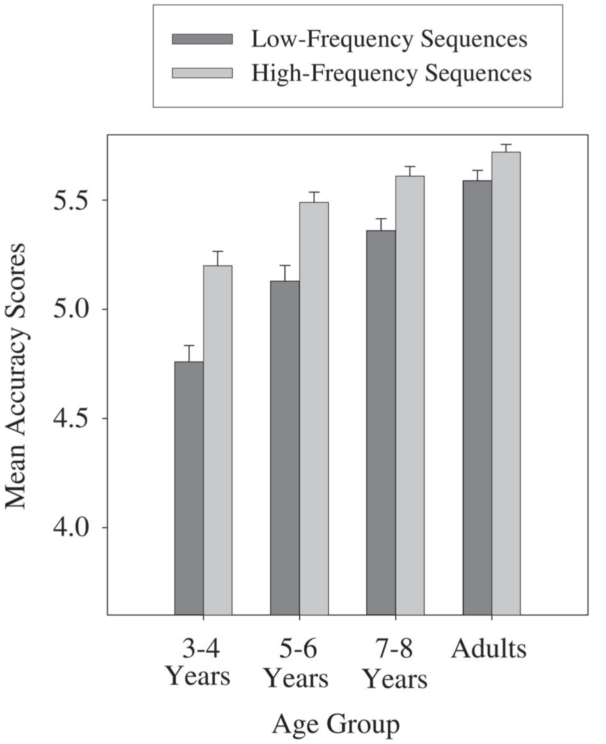
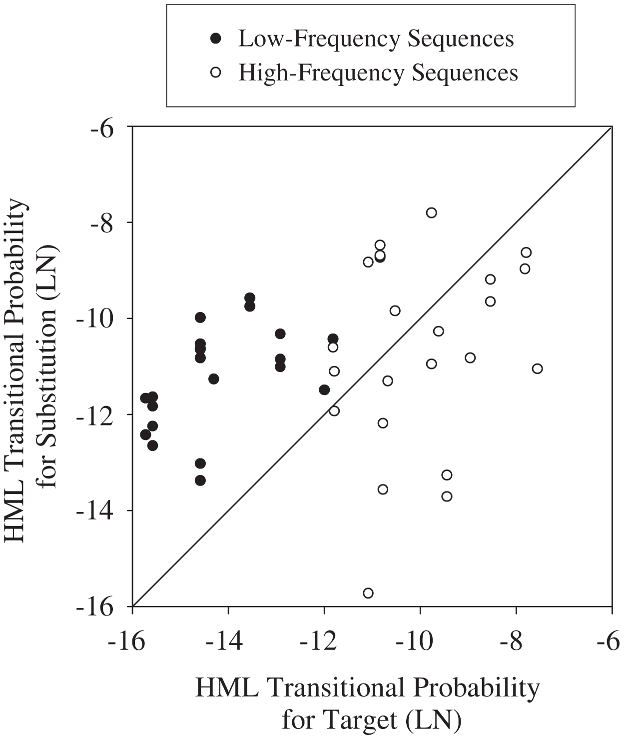


Figure 3. Mean outcome transitional probability (averaged across participants) plotted against transitional probability for target sequence for all 44 nonwords. Reference line shows *x* = *y*.



line is also drawn on the figure. For the low-frequency sequences, all of the substitutions resulted in higher frequency sequences as a rule (note that all data points lie above the reference line). By contrast, for the high-frequency sequences, about half of the substitutions resulted in higher frequency sequences and about half of the substitutions resulted in lower frequency sequences (i.e., some data points are above the reference line and some are below). A Wilcoxon matched-pairs signed ranks test found the difference in transitional probability between the target sequences and the errored productions to be significant for the low-frequency sequences ($z = 4.11, p < .001$) but not for the high-frequency sequences ($z = 1.38$, $p = .168$).

Duration Analysis

Because different segments have different inherent durations, the duration value for a particular segment produced by a particular participant was included in the analysis only if both the low-frequency and matched high-frequency target sequence containing the measured segment were produced correctly. For the younger age groups, therefore, this analysis necessarily overrepresents productions by those participants who behaved more like older participants in terms of error rates. Table 3 shows the number of token pairs and the mean durations of each segment type in the low- versus high-frequency sequence for each age group. A segment in a low-probability sequence is generally longer than in a high-probability sequence. This tendency is more consistent for the younger groups and not evident in the means for the adults.

The literature on segment durations in English suggests that nasals and voiced obstruents are inherently shorter than voiceless obstruents, which in turn should be shorter than sequences of two voiceless consonants or the two vowel targets of a diphthong. We therefore grouped [m], [n], [v], and [g] together as "short" segments and [pt] and [au] together as "double" segments in a three-way ANOVA with factors of segment type, age group, and sequence probability. As expected, there was a significant main effect of segment type, $F(1, 711) = 127.608$, $p < .001$. There were also significant main effects of age group, $F(3, 711) = 4.701$, $p = .03$, and sequence frequency, $F(1, 711) = 10.229$, $p = .001$, as well as a significant Age × Frequency interaction, $F(3, 711) = 5.807$, $p = .016$. The Age × Frequency interaction was due to the fact that duration of the low-frequency sequences

Table 3. Number of tokens and mean durations in milliseconds (with standard deviations in parentheses) for each measured segment type in low- (LF) versus high-frequency (HF) target sequences for each age group.

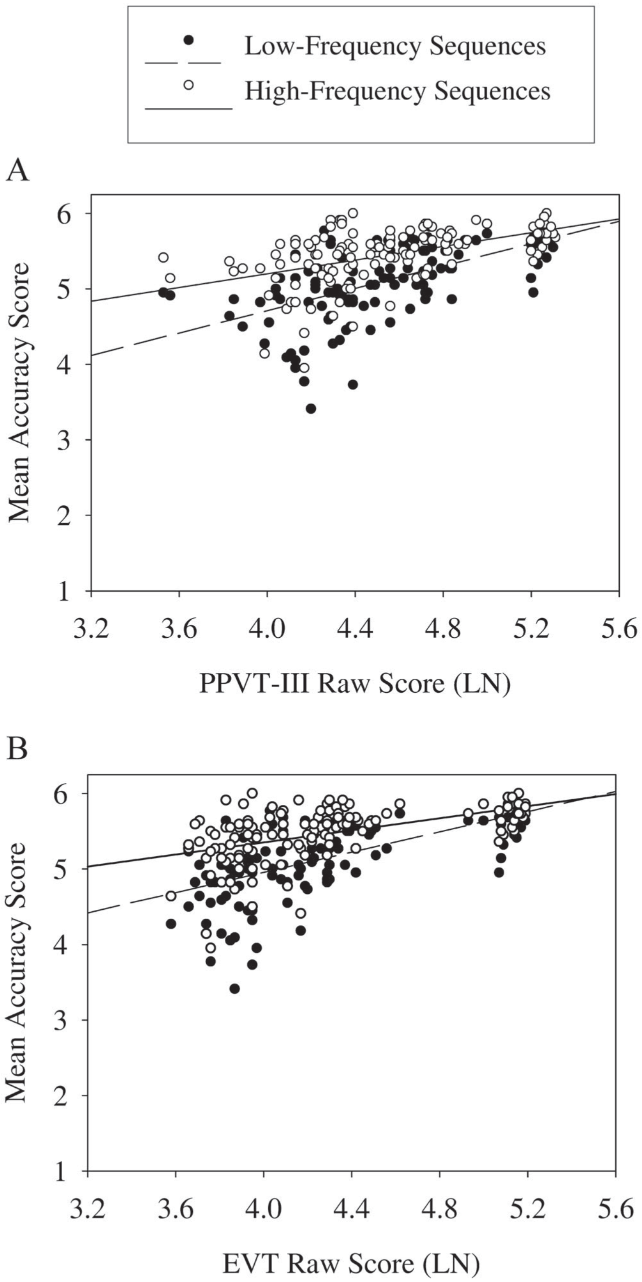
	Age group											
	3–4 years			5–6 years			7–8 years			Adults		
Sequence	*N*	LF	HF	*N*	LF	HF	*N*	LF	HF	*N*	LF	HF
au	39	189 (63)	186 (54)	31	195 (71)	184 (38)	21	173 (39)	180 (38)	15	165 (30)	173 (33)
f	23	117 (51)	114 (57)	19	131 (57)	108 (47)	18	126 (38)	124 (40)	30	92 (31)	105 (25)
g	18	108 (59)	107 (48)	27	107 (89)	85 (43)	17	60 (37)	76 (36)	17	77 (58)	59 (32)
m	38	82 (72)	53 (33)	33	79 (58)	68 (37)	23	71 (44)	64 (27)	16	75 (37)	70 (26)
n	36	91 (72)	77 (57)	31	132 (12)	88 (48)	22	102 (38)	85 (51)	16	99 (40)	124 (12)
pt	23	197 (47)	206 (45)	26	108 (72)	211 (74)	11	212 (122)	187 (43)	9	168 (17)	204 (28)
v	35	78 (48)	70 (47)	46	84 (106)	85 (79)	38	87 (51)	64 (49)	41	82 (34)	73 (46)

generally decreased with age across the four age groups. In contrast, the durations of the high-frequency sequences were comparable for the oldest group of children and the adults (see Table 3).

Segmental Accuracy Scores and Vocabulary Size

Figure 4 shows mean accuracy scores for high- and low-frequency sequences plotted against the two measures of vocabulary size. For both plots, the regression line for the high-frequency sequences lies above the line for the low-frequency sequences, and the distance between the two lines is the effect of target sequence frequency. This distance decreases as vocabulary size increases. The participants with the largest vocabularies were the adults, for whom there was the smallest effect of frequency on accuracy, as indicated by the significant Age × Frequency interaction observed in the repeated-measures ANOVA. To determine the quantitative relationship between vocabulary size and repetition accuracy more precisely, we correlated mean segmental accuracy scores for the low- and high-frequency sequences with our two measures of vocabulary size. These correlations were significant and were greater for low-frequency sequences, as compared to high-frequency sequences (for low-frequency sequences, $r^2 = .38$, $p < .001$ for PPVT-III, and $r^2 = .38$, $p < .001$ for EVT; for high-frequency sequences, $r^2 = .26$, $p < .001$ for PPVT-III, and $r^2 = .25$, $p < .001$ for EVT). When adults were excluded from the analysis, the correlations were somewhat smaller but were still significant (for low-frequency sequences, $r^2 = .25$, $p < .001$ for PPVT-III, and $r^2 = .30$, $p < .001$ for EVT; for high-frequency sequences, $r^2 = .18$, $p < .001$ for PPVT-III, and $r^2 = .21$, $p < .001$ for EVT).

Figure 4. Mean accuracy scores for low- and high-frequency sequences plotted against (a) receptive vocabulary size (Peabody Picture Vocabulary Test–Third Edition [PPVT-III]) and (b) expressive vocabulary size (Expressive Vocabulary Test [EVT]) for all participants.



Accuracy was related to both vocabulary size and age. Furthermore, vocabulary size and age were highly correlated with each other. Age was also correlated with measures of articulatory ability. To tease apart the interactions among these factors, we performed two stepwise multiple regression analyses on the data of the child participants. In both analyses, the independent variables were age, two measures of vocabulary size (the natural log of the EVT raw score and the natural log of the PPVT-III raw score), and one measure of articulatory ability (GFTA raw score). In the first analysis, the dependent variable was mean segmental accuracy averaged across all items for each participant, and in the second analysis it was the mean difference between the segmental accuracy scores for the high- versus low-frequency targets averaged across all item pairs. When the dependent variable was overall accuracy, EVT raw score accounted for 29% of the variance and GFTA raw score accounted for an additional 4.8%. When the dependent variable was the difference in accuracy between the high- and low-probability sequences, the only significant predictor was EVT raw score, accounting for 9.9% of the variance. Two additional hierarchical regression analyses were also performed. These analyses examined the relative predictive power of vocabulary size when age and GFTA raw score, our measure of articulatory ability, were controlled. Both analyses forced age as the first variable and GFTA raw score as the second variable. The results of these analyses were similar to those of the stepwise multiple regression analyses. In the first analysis, when the dependent variable was mean accuracy, EVT raw scores accounted for a significant proportion of the variance (10.1%) beyond that accounted for by age and GFTA. In the second analysis, when the dependent variable was the difference in accuracy between the high- and low-probability sequences, the only significant predictor was EVT, which accounted for 9.2% of the variance. The results of these regression analyses suggest that it is expressive vocabulary size, rather than age per se, that accounts for the higher accuracy and the smaller effect of transitional probability on accuracy for the older children and adults. Expressive vocabulary size is a better predictor of accuracy than age, and it is also a better predictor of accuracy than GFTA raw score, a direct measure of articulatory ability.

Mean Accuracy for Low-Frequency Versus Unattested Sequences

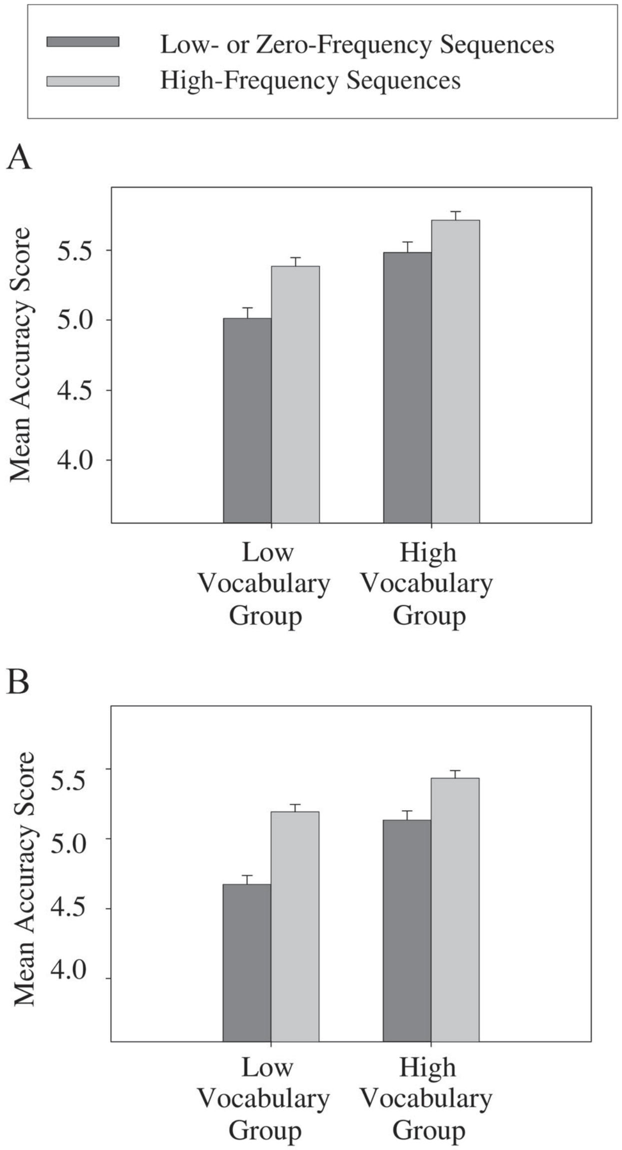
There are two possible reasons why increasing vocabulary size reduces the effect of transitional probability on nonword repetition accuracy. First, it is possible that children with larger vocabularies show a smaller effect because they are more likely to have encountered specific low-frequency sequences by learning real words containing them. That is, they are more likely to have practiced the auditory and motor representations necessary for perceiving and for fluently producing each of the two sounds in the context of the other sound in that particular sequence within a word. The second explanation is that the children with larger vocabularies show a smaller effect because they have robustly generalized a representation for each component phoneme that is relatively more independent of context and, hence, more extensible to new contexts. That is, their perceptual and/or motor representations are more robustly segmented into sublexical units or properties that are smaller than the sequence (cf. Walley, 1993), hence making the representation more flexible (i.e., more easily incorporated into a completely novel pattern). Of course, these two explanations are not mutually exclusive. Children with larger vocabularies may have both more practiced phonetic representations of the particular sequences from having encountered them previously and more robustly abstracted representations of individual phonemes.

In order to tease apart these two explanations, we compared performance of children with larger vocabularies to performance of children with smaller vocabularies on each sequence, differentiating two types of low-frequency sequences: the low-frequency but attested sequences versus the completely unattested, zero-frequency sequences. The first explanation (i.e., greater likelihood of familiarity with the specific sequence) predicts that we should find that children with larger vocabularies are more accurate than children with smaller vocabularies on the attested low-frequency sequences, but not on the zero-frequency sequences. By contrast, the second explanation (i.e., more robust phonological generalizations) predicts that children with larger vocabularies will be more accurate than children with smaller vocabularies on both the low-frequency attested sequences and the zero-frequency sequences.

As described earlier, we divided the children into a smaller- and a larger-vocabulary group on the basis of EVT raw scores, as this estimate of expressive vocabulary size was a significant predictor of overall accuracy and of the difference in accuracy between high-frequency and low-frequency sequences. The stimuli had not been designed a priori with this analysis in mind and most of the low-frequency sequences in our corpus were completely unattested in the MHR, but there were nine low-frequency sequences that were attested in one to six words in the HML. Four of these sequences were used in both two- and three-syllable nonwords, so this resulted in 13 words with low-frequency but attested sequences. We examined the effect of vocabulary group and frequency for the zero-frequency versus their matched high-frequency sequence pairs and for the low-frequency versus their matched high-frequency sequence pairs.

Figure 5 shows mean accuracy scores for the two sets of word pairs for both the larger-vocabulary and smaller-vocabulary groups. Two two-factor mixed-model ANOVAs were used to examine these differences. In both of these ANOVAs, the between-subjects factor was vocabulary group and the dependent variable was mean accuracy. In the first ANOVA, the within-subjects factor was frequency (zero-frequency sequences vs. their matched high-frequency sequences). In this ANOVA, the effect of group was significant, $F(1, 102) = 20.8$, $p < .01$, partial $\eta^2 = .17$—that is, the larger-vocabulary group was more accurate than the smaller-vocabulary group. The effect of sequence frequency was also significant, $F(1, 102) = 141.8$, $p < .01$, partial $\eta^2 = .58$. Moreover, these factors interacted, $F(1, 102) = 10.2$, $p < .01$, partial $\eta^2 = .09$. The difference between zero-frequency sequences and their matched high-frequency sequences was larger for children in the smaller-vocabulary group than for children in the larger-vocabulary group (see Figure 5a). In the second ANOVA, the within-subject factor was again frequency—in this case, low-frequency sequences versus their matched high-frequency sequences. In this ANOVA, the effect of group was significant, $F(1, 102) = 21.7$, $p < .01$, partial $\eta^2 = .18$. The effect of frequency was also significant, $F(1, 102) = 38.8$, $p < .01$, partial $\eta^2 = .29$. In this ANOVA, these factors did not interact. The

Figure 5. Mean accuracy scores (with error bars) for (a) the matched low-frequency versus high-frequency sequence pairs and (b) the matched zero-frequency versus high-frequency sequence pairs for the larger-vocabulary and smaller-vocabulary groups of child participants.



difference between low-frequency and matched high-frequency sequences was similar for both the larger- and the smaller-vocabulary groups of children.

These results support the second of the two explanations proposed above, that children with larger vocabularies have made more robust symbolic generalizations. Children with larger vocabularies were more accurate than children with smaller vocabularies for both the low-frequency and the zero-frequency sequences. Furthermore, the difference between the two groups was greatest for the zero-frequency sequences. The fact that children with larger vocabularies were more accurate than children with smaller vocabularies in repeating phonemes, even in sequences that they had never before encountered within a word, supports the idea that children with larger vocabularies have more flexible, context-independent representations.

Discussion

We found that participants repeated consonants and vowels more accurately in the context of target sequences occurring in many real words. Furthermore, substitution errors on low-frequency sequences were likely to result in higher frequency sequences. In pairs of productions of sequences containing an identical measurable phoneme segment where both the low- and the high-frequency sequence were produced accurately, participants also produced shorter durations in the high-frequency sequences. These effects of target sequence frequency on segmental accuracy and fluency were largest in productions by 3–4-year-old children and smallest in productions by adults. Given how much closer the young child is to the onset of lexical acquisition, it is not surprising that the child's representations of speech sounds are even more highly tied to the contexts in which these sounds occur in words in the lexicon. When the young child encounters a new word with a low-frequency sublexical pattern, there are fewer words in the lexicon that can be used by analogy to aid in the creation of acoustic and articulatory representations for the new word. This increased difficulty makes production of a new word less accurate and less fluent when it contains an infrequent phoneme, or a relatively frequent phoneme in an unfamiliar context.

An analysis of the relationships among target sequence frequency, age, articulatory ability, and vocabulary size showed that the effect of frequency on segmental accuracy is related to the massive vocabulary growth that normally occurs during early childhood rather than to some other aspect of normal maturation that is independent of vocabulary size. Interestingly, it was the estimate of expressive vocabulary size that was the best predictor of both overall accuracy and the effect of frequency on accuracy. Having a word in the expressive vocabulary requires that the child have a detailed articulatory–motor representation of the word, in addition to the acoustic–auditory representation that is required for a word that is only in the receptive vocabulary. Adding a newly encountered word to the expressive vocabulary also requires that the child do a "fast mapping" between the very disparate parametric representations in these two domains. In the context of performing this fast mapping for very many words containing a particular acoustic and articulatory configuration, well-practiced phonetic categories such as “voiceless unaspirated stop” might come to be robustly abstracted away from the specific words and sequences containing them to become almost quasi-symbolic categories, or “phonemes,” as discussed in Beckman and Pierrehumbert (in press). In much previous research, receptive vocabulary, rather than expressive vocabulary, has been found to be one of the best predictors of nonword repetition accuracy (e.g., Gathercole & Baddeley, 1989; Gathercole, Hitch, Service, & Martin, 1997; Gathercole, Willis, Emslie, & Baddeley, 1992; Metsala, 1999). However, all of these studies measured only receptive vocabulary, with the exception of Gathercole et al. (1997), and that study found similar correlations between nonword repetition accuracy and both expressive and receptive vocabulary size. This difference in methodology may also account for the differing results of Gathercole et al. (1999) relative to this study. Gathercole et al. (1999) found no interaction between vocabulary size and the effect of frequency, but again they measured only receptive vocabulary size.

These results support an account of acquisition in which the typically developing child gradually acquires more and more robust phonological knowledge as a consequence of learning to produce many words. That is, an increase in vocabulary size does not simply mean that the child knows more words, but also that the child is able to make more and more robust phonological generalizations. This claim was further supported by our finding that children with larger vocabularies were more accurate than children with smaller vocabularies, not just on low-frequency sequences that they might have encountered in precocious acquisition of words such as *pueblo* and *fugue*, but also on the subset of low-frequency sequences that were completely unattested even in the HML.

More generally, these results support the view that symbolic knowledge at all levels of phonology emerges from each individual speaker’s experience in acquiring and using words of the ambient language. In the mature language user, phonotactic constraints are patterns generalized over known words, which help the adult speaker pick out familiar words in connected speech and to recognize and remember new words. Similarly, at a younger age, phonemes, syllables, and the other symbolic structures specific to phonology emerge through interaction between the input forms that the child hears and the increasingly more complex hierarchy of representations that the child builds in order to recognize and produce words in connected speech.

Our results support a particular view of the relationship between grammatical knowledge and processing skills in general. Knowledge of more word forms is associated with more robustly generalized knowledge of how to learn to hear and say new word forms. This is consistent with a view of grammar as an emergent property of the history of interactions between the language user and the language events in the world (see Allen & Seidenberg, 1999; Bates & Goodman, 1999; Beckman & Edwards, 2000b; Pierrehumbert, 2001; Werker, Corcoran, Fennell, & Stager, 2002). In this view, the relationship between knowledge of the phonological grammar and processing of phonological patterns is a symbiotic one. Knowledge feeds on processing, and processing feeds on knowledge. The more often a child has heard and said a word, the better the child knows the word. The child can fluently incorporate the word into unfamiliar prosodic structures in productions of novel sentences. In the same way, the more words the child has heard and said that contain a particular phonological pattern, the more basis the child has for abstracting away a generalized knowledge of the possible patterns, to quickly access the same or similar patterns in other words. As the child gains more experience with more words, and as more specific instances of a pattern accumulate, fine-grained phonological knowledge becomes richer. At the same time, aspects of speech production and perception that are shared across sets of similar subparts of words and that contrast in analogous ways to subparts of other sets of words can become practiced as a relational pattern at another higher level of representation. If we recast Ferguson and Farwell’s (1975) idea of a “lexical core” in this view, it is not so much that a “pregrammatical” foundation of knowledge of how to produce a small core of words is overlaid by phonological knowledge. Rather, phonological knowledge incrementally emerges from the initial layer of first-learned words to build an increasingly structured scaffolding, an increasingly rich set of alternative paths to hearing and reproducing a novel word-form.

Acknowledgments

This work was supported by Grant 02932 from the National Institute on Deafness and Other Communication Disorders to Jan Edwards and by Training Grant T32 DC0051 from the National Institutes of Health to Robert A. Fox, which provided a traineeship to the third author. We thank the children who participated in the study, the parents who gave their consent, and the schools at which the data were collected. For assistance in stimuli preparation, data collection, and analysis, we thank Erin Casey, Lynn Carahaly, Lisa Draper, Melissa Epstein, Heidi Hochstetler, Maryann Holtschulte, Bridgett Isermann, Satako Katagiri, Laurie Vasicek, Amy Vitale, Pauline Welby, and David White.

References

Allen, E., & Seidenberg, M. S. (1999). On the emergence of grammar from the lexicon. In B. MacWhinney (Ed.), *The emergence of language* (pp. 115–151). Mahwah, NJ: Erlbaum.

Bates, J., & Goodman, J. C. (1999). The emergence of grammaticality in connectionist networks. In B. MacWhinney (Ed.), *The emergence of language* (pp. 29–79). Mahwah, NJ: Erlbaum.

Beckman, M. E., & Edwards, J. (2000a). Lexical frequency effects on young children's imitative productions. In M. Broe & J. Pierrehumbert (Eds.), *Papers in laboratory phonology V* (pp. 207–217). Cambridge, U.K.: Cambridge University Press.

Beckman, M. E., & Edwards, J. (2000b). The ontongeny of phonological categories and the primacy of lexical learning in linguistic development. *Child Development, 71*, 240–249.

Beckman, M. E., & Pierrehumbert, J. B. (in press). Interpreting 'phonetic interpretation' over the lexicon. In J. Local, R. Ogden, & R. Temple (Eds.), *Papers in laboratory phonology VI*. Cambridge, U.K.: Cambridge University Press.

Burgemeister, B. B., Blum, L. H., & Lorge, I. (1972). *Columbia Mental Maturity Scale*. New York: Harcourt Brace Jovanovich.

Coleman, J., & Pierrehumbert, J. (1997). Stochastic phonological grammars and acceptability. *Proceedings of the 3rd Meeting of the ACL Special Interest Group in Computational Phonology* (pp. 49–56). Somerset, NJ: Association for Computational Linguistics.

Dollaghan, C. A, Biber, M. E., & Campbell, T. F. (1995). Lexical influences on nonword repetition. *Applied Psycholinguistics*, *16*, 211–222.

Dunn, L., & Dunn, L. (1997). *Peabody Picture Vocabulary Test–Third Edition*. Circle Pines, MN: American Guidance Services.

Ferguson, C. A., & Farwell, C. B. (1975). Words and sound in early language acquisition: English initial consonants in the first fifty words. *Language*, *51*, 419–439.

Fisher, C., Hunt, C., & Chambers, K. (2001). Abstraction and specificity in preschoolers' representations of novel spoken words. *Journal of Memory and Language, 45*, 667–687.

Frisch, S. (2001). Emergent phonotactic generalizations in English and Arabic. In J. Bybee & P. Hopper (Eds.), *Frequency and the emergence of linguistic structure* (pp. 159–179). Amsterdam: John Benjamins.

Frisch, S. A., Large, N. R., & Pisoni, D. B. (2000). Perception of wordlikeness: Effects of segment probability and length on processing of nonword sound patterns. *Journal of Memory and Language, 42*, 481–496.

Gathercole, S. E., & Baddeley, A. D. (1989). Evaluation of the role of phonological STM in the development of vocabulary in children: A longitudinal study. *Journal of Memory and Language, 28*, 200–213.

Gathercole, S. E., Frankish, C. R., Pickering, S., & Peaker, S. (1999). Phonotactic influences on short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 84–95.

Gathercole, S. E., Hitch, G. J., Service, E., & Martin, A. J. (1997). Phonological short-term memory and new word learning in children. *Developmental Psychology, 33*, 966–979.

Gathercole, S. E., Willis, C., Emslie, H., & Baddeley, A. D. (1991). The influences of number of syllables and wordlikeness on children's repetition of nonwords. *Applied Psycholinguistics*, *12*, 349–367.

Gathercole, S. E., Willis, C., Emslie, H., & Baddeley, A. D. (1992). Phonological memory and vocabulary development during the early school years: A longitudinal study. *Developmental Psychology, 28*, 887–898.

Goldman, R., & Fristoe, M. (1986). *The Goldman-Fristoe Test of Articulation*. Circle Pines, MN: American Guidance Services.

Hay, J., Pierrehumbert, J., & Beckman, M. (in press). Speech perception, well-formedness, and the statistics of the lexicon. In J. Local, R. Ogden, & R. Temple (Eds.), *Papers in laboratory phonology VI*. Cambridge, U.K.: Cambridge University Press.

Kaufman, N. (1995). *Kaufman Speech Praxis Test for Children*. Detroit, MI: Wayne State University Press.

MacKay, D. (1982). The problems of flexibility, fluency and speed-accuracy trade-off in skilled behavior. *Psychological Review, 89*, 483–506.

Metsala, J. (1999). Young children's phonological awareness and nonword repetition as a function of vocabulary development. *Journal of Educational Psychology, 91*, 3–19.

Moe, S., Hopkins, M., & Rush, L. (1982). *A vocabulary of first-grade children*. Springfield, IL: Charles C Thomas.

Munson, B. (2001). Phonological pattern frequency and speech production in adults and children. *Journal of Speech, Language, and Hearing Research, 44*, 778–792.

Pierrehumbert, J. (1994). Syllable structure and word structure: A study of triconsonantal clusters in English. In P. A. Keating (Ed.), *Papers in laboratory phonology III* (pp. 168–190). Cambridge, U.K.: Cambridge University Press.

Pierrehumbert, J. (2001). Why phonological constraints are so coarse-grained. *Language and Cognitive Processes*, *16*, 691–698.

Pierrehumbert, J. B. (2003). Probabilistic phonology: Discrimination and robustness. In R. Bod, J. Hay, & S. Jannedy (Eds.), *Probabilistic linguistics* (pp. 177–228). Cambridge, MA: MIT Press.

Pisoni, D., Nusbaum, H., Luce, P., & Slowiacek, L. (1985). Speech perception, word recognition, and the structure of the lexicon. *Speech Communication*, *4,* 75–95.

Pitt, M., & McQueen, J. M. (1998). Is compensation for coarticulation mediated by the lexicon? *Journal of Memory and Language, 39*, 347–370.

Smit, A. B., Freilinger, J. J., Bernthal, J. E., Hand, L., & Bird, A. (1990). The Iowa Articulation Norms Project and its Nebraska replication. *Journal of Speech and Hearing Disorders, 55*, 779–798.

Storkel, H. (2001). Learning new words: Phonotactic probability in language development. *Journal of Speech, Language, and Hearing Research, 44*, 1321–1338.

Templin, M. C. (1957). *Certain language skills in children, their development and interrelationships* [Institute of Child Development and Welfare monograph series, no. 26]. Minneapolis: University of Minnesota.

Vitevitch, M. S., & Luce, P. A. (1999). Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language*, *40*, 374–408.

Vitevitch, M. S., Luce, P. A., Charles-Luce, J., & Kemmerer, D. (1997). Phonotactics and syllable stress: Implications for the processing of spoken nonsense words. *Language and Speech*, *40*, 47–62.

Walley, A. C. (1993). The role of vocabulary development in children's spoken word recognition and segmentation ability. *Developmental Review*, *13*, 286–350.

Washington, J. A., & Craig, H. K. (1999). Performances of at-risk, African-American preschoolers on the Peabody Picture Vocabulary Test–III. *Language, Speech, and Hearing Services in Schools*, *30*, 75–82.

Werker, J., Corcoran, K. M., Fennell, C. T., & Stager, C. L. (2002). Infants' ability to learn phonetically similar words: Effects of age and vocabulary size. *Infancy*, *3*, 1–30.

Williams, K. (1997). *Expressive Vocabulary Test*. Circle Pines, MN: American Guidance Services.

Received March 10, 2003

Accepted July 14, 2003

DOI: 10.1044/1092-4388(2004/034)

Contact author: Jan Edwards, PhD, Department of Speech and Hearing Science, Ohio State University, 110 Pressey Hall, 1070 Carmack Road., Columbus, OH 43210. E-mail: edwards.212@osu.edu