The interaction between vocabulary size and phonotactic probability effects on children’s production accuracy and fluency in novel word repetition

Jan Edwards
Dept. of Speech and Hearing Science, Ohio State University

Mary E. Beckman
Dept. of Linguistics, Ohio State University

Benjamin Munson
Dept. of Communication Disorders, University of Minnesota

Please send inquiries to:

Jan Edwards
Dept. of Speech and Hearing Science
Ohio State University
Columbus, OH 43210
FAX: 614 292-7504
Phone: 614 292-1742
Email: edwards.212@osu.edu

Running title: Vocabulary size and phonotactic probability
Abstract

Adult performance on a variety of tasks suggests that phonological processing of nonwords is grounded in generalizations about sublexical patterns over all known words. To test this account of phonological processing, production accuracy and fluency were examined in nonword repetitions by children and adults. Stimuli were 22 pairs of nonwords, in which one contained a low-frequency or unattested two-phoneme sequence while 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 larger vocabularies showed a smaller influence of frequency on accuracy than children with smaller vocabularies. These results support the claim that speakers develop a phonological system based on incremental generalizations over the lexicon.

Key words:

Phonotactic probability, phonological acquisition, nonword repetition, lexical development, language acquisition
In acquiring a spoken language, a child must learn to pick out words in the continuous speech signal, and recognize them in terms of the stored memory of the many words that she already knows. The child must also learn to reproduce the sound pattern of a novel word in some way that makes it recognizable to his conversational partners. With each newly encountered word, the child must determine how to map from a complex acoustic pattern produced by another speaker’s vocal tract to a complex motor control program that reproduces the form using her own vocal tract.

This mapping is non-unique and highly non-linear. It cannot be learned by simple pattern matching between the perceived acoustics of a new word and the child’s reproduction of it (see, e.g., Atal, Chang, Matthews, & Tukey, 1978; Jordan, 1990; Rose, Schroeter, & Sondhi, 1996; Nearey, 1996). Moreover, the young child must determine this mapping frequently. To go from the approximately 50 words that Hart and Risley (1995) reported at about 18 months in their longitudinal study (see also Bloom, 1993) to the 500 to 1500 words that they reported for the same children at 36 months, the young child must acquire at least one or two new words per day.

It would be impossible to account for this rapid vocabulary growth without positing that the child acquires an implicit phonological “grammar” to index sublexical patterns on both sides of the mapping and to access words that are already known to the child. That is, the child abstracts away a symbolic representation that allows her to decompose the input (acoustic or auditory) form of a word into smaller elements – low-level sublexical patterns such as characteristic burst spectra or formant patterns for a particular stop-vowel sequence, and higher-order patterns such as the abstract category for the particular stop and the particular vowel that follows it.

This representation is symbolic in at least three senses. First, it is symbolic in the sense that it indexes some synoptic memory of each word that the child already knows. This synoptic
memory must include some category structure abstract enough to unify the potentially very
different episodic memories of a word into a single lexeme, so that the child can access the lexicon
quickly to recognize an utterance as a new instance of a familiar word – for example, understanding
the word *cake* as the same word when produced by his grandfather rather than by his mother, or by
his mother when she is speaking to the baby rather than to him. Assuming this level of
representation means that a novel word can then be differentiated from a known word because it
activates only sublexical patterns.

The representation is also symbolic in the sense that it indexes other similar sublexical
patterns in the lexicon. For example, if the child hears an utterance of a new word, *cape*, produced
by a clerk selling her mother a Dracula costume for Halloween, the burst spectrum and formant
transitions for the word-initial /ke/ sequence should invoke the categories formed by the child’s
experiences of the same or similar sequences in all of the utterances that she has heard before of
words such as *cake*, *came*, and *Katie*, so that she can parse the acoustics of the initial CV sequence
in terms of the sound of the same or similar CV sequences in familiar words.

The representation is also symbolic in the sense that these categories invoked by the
acoustic input serve to index comparably tractable sublexical patterns in the articulatory-motor
representations of words that the child already knows how to produce. If the initial CV sequence of
*cape* is correctly identified with the initial /ke/ of *cake* and *Katie*, for example, the articulatory-
motor representation of the /k/ can then be recombined with other familiar elements from other
words, such as the /e/ and /p/ of *tape* and *staple*, into a novel motor plan for a more or less fluent
reproduction of the novel word. Thus, the process of recognizing and repeating a known word or a
new word invokes a grammar of “legal” combinations of symbolic elements and contexts on both
sides of the mapping between articulation and acoustics.
Where does this grammar of symbolic phonological categories and combinations come from, and what is it like? The answers to these two questions are closely intertwined, and it is impossible to address the first without addressing the second. Four closely related attributes of the phonological grammar are suggested by research on adult patterns of perception and production and by some infant studies.

First, the grammar refers to multiple levels of representation. This is obvious when we contrast the parameter spaces appropriate for capturing acoustic similarities relevant for models of perception with the parameter spaces appropriate for capturing articulatory similarities relevant for models of production. However, the multiplicity goes beyond this simple dichotomy between representations for perception and representations for production. For example, speech perception must involve more than simple acoustic pattern matching over the whole word. There must be a “fast phonological preprocessor” (Pierrehumbert, in press) that can decompose the speech stream into phonological patterns involving such gross sublexical categories as stressed versus unstressed syllables or the phonemes /t/ versus /k/. Otherwise, we would have no way to explain error patterns in adult lexical access in connected speech (see, e.g., Cutler & Butterfield, 1992), or the apparent degradation of attention to phonetic detail in very young children when they first begin to produce words (Werker, Corcoran, Fennell, & Stager, 2000). Decomposition into sublexical patterns also is consistent with infant studies suggesting some unified categorization in memory of disyllabic stimuli sharing a common initial syllable (Jusczyk, Jusczyk, Kennedy, Schomberg, & Koenig, 1995), and with adult studies showing differences in facilitative effects of partial versus complete identity in phonemic priming of words in isolation (see, e.g., Slowiaczek, Nusbaum, & Pisoni, 1987).
Second, phonological categories (such as the English phonemes /t/ versus /k/) clearly are not innate, but are acquired in learning a spoken language. Otherwise, we would have no way to explain the systematic differences in distribution of acoustic and articulatory values for roughly comparable sublexical categories across speakers of different languages and dialects (see, e.g., Bradlow, 1995, Caramazza & Yeni-Komshian, 1974, Ladefoged & Bhaskararao, 1983, among many others). These language-specific patterns emerge quite early in acquisition, even before a child’s productions are completely adult-like (see, e.g., Stoel-Gammon, Williams, & Buder, 1994).

Third, the grammar is something that becomes established as a more or less automatic and ingrained set of processing skills in the course of acquiring a first language. Otherwise, adult speakers would have no trouble generalizing the mapping ability to reproduce the different acoustic patterns or different combinations of articulatory elements that are appropriate for a second language or dialect (cf. reviews in Flege & Hillenbrand, 1987; Flege, 1995). Infant studies showing loss of attention to non-native contrasts over the course of the first year also suggest this attribute (e.g., Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992; Tsushima, Takizawa, Sasaki, Shiraki, Nishi, Kohno, Menyuk, & Best, 1994; Werker & Tees, 1984).

Finally, the abstraction of phonological categories and permissible combinations is based on generalizations over patterns that can be observed in the lexicon of actual words in the specific language. There are several sources of evidence for this last attribute. One is the research on cross-language differences in babbling. This work has shown that the sounds in babbling are influenced by the distribution of sounds in the lexicon of the ambient adult language. For example, measurements of vowel formants in early variegated babbling reflect the frequencies of different vowels in the lexicon of the ambient adult language (de Boysson-Bardies, Hallé, Sagart, & Durand, 1989). Transcribed consonants in later variegated babbling also reflect cross-language differences
in the relative frequency of different consonants in the ambient adult language (de Boysson-Bardies & Vihman, 1991). Moreover, although all infants produce multisyllabic babbling with simple consonant-vowel alternations in reduplicated babbling, by the time children are producing their first words, English-acquiring children produce relatively more monosyllabic babbles and babbling ending in consonants, relative to French-acquiring infants (Vihman, 1993). These differences reflect the frequencies of different word shapes in the two languages.

Another source of evidence for the claim that children develop a phonological system based on generalizations over the lexicon comes from an increasing body of research showing that adult speakers of English are sensitive to the relative frequencies with which different sublexical patterns occur in the lexicon. These relative frequencies are often called phonotactic 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. For example, they are faster to repeat new “words” (i.e. nonsense forms) that contain high-frequency consonant-vowel and vowel-consonant sequences (Vitevich, Luce, Charles-Luce & Kemmer, 1997, Vitevitch & Luce, 1998; Vitevitch & Luce, 1999). Their speeded repetitions of nonwords containing high-frequency sequences also are more accurate, although this effect is not as robust or consistently replicated across experiments. When asked to spell nasal-obstruent sequences embedded in nonsense words, 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). When asked to choose between forms that conjoin two nonsense words in a blending task, they are less likely to choose the preferred blending point (which takes the onset of the first word and the whole rime of the second) just in case the less preferred blending results in a more frequent consonant vowel sequence (Treiman, Kessler, Knewasser, Tincoff, & Bowman, 2000). When asked to judge how “wordlike”
nonsense words are, they give higher wordlikeness ratings to forms that contain phoneme sequences which are attested in many words. This last result is extremely robust and has been seen in a large number of experiments (e.g., Pierrehumbert, 1994; Coleman & Pierrehumbert, 1997; Vitevitch, et al., 1997; Frisch, Large, & Pisoni, 1999; Munson, in press).

This body of research with adults complements work showing that children are sensitive to “wordlikeness” in repeating nonsense words. For example, Gathercole, Willis, Emslie, and Baddeley et al. (1991) had 20 adults give ratings on a scale from 1 (for “not like a word at all”) to 5 (for “very like a word”) for each of thirty nonwords varying in length from two to five syllables. They also used these nonwords in a repetition task and found that 4-, 5-, and 6-year-old children are more accurate at repeating nonwords that adults have judged to be more wordlike. These two findings – the finding that adults judge nonsense words with high frequency sequences as more wordlike and the finding that children repeat nonwords with higher wordlikeness ratings more accurately – together suggest that children’s repetitions will be more accurate for nonwords containing high-frequency sequences. That is, the phonological and phonetic representations that allow a young child to acquire new words should be the same frequency-based generalizations that adults are using to judge wordlikeness. Moreover, given how much closer the young child is to the onset of lexical acquisition, we might expect her representations of speech sounds to be even more highly tied to the contexts in which these sounds occur in words in her lexicon. That is, we might expect repetition accuracy to be more vulnerable to transitional probability. When the young child encounters a new word with a low-frequency sublexical pattern, he might be less able to access the representations of parallel but slightly different familiar patterns, to apply in making an analogous acoustic or articulatory representation for new word. Production of a new word should be less accurate and less fluent if it contains an infrequent phoneme, or a relatively frequent phoneme in an
unfamiliar context such as a two-phoneme sequence in which the target phoneme does not occur in any words that the child already knows.

Some recent research suggests that this may be the case. Gathercole, Frankish, Pickering, and Peaker (1999) found that seven- and eight-year-old children repeat lists of nonwords more accurately on a serial recall task if the nonwords contain only high frequency consonant-vowel and vowel-consonant sequences. Using a less demanding immediate repetition task, Beckman and Edwards (2000a) found that children three to five years of age repeated low-frequency two-phoneme sequences in nonwords less accurately than they repeated high-frequency two-phoneme sequences. Schaadt (1997) found the same effect, albeit only for the consonant-consonant sequences, in children as old as seven years. Munson (in press) found an influence of sublexical sequence frequency on production fluency as well as on accuracy. He used segment duration as a measure of fluent production and found that children from three to eight years of age produced longer durations for the same segment when it was in a low-frequency consonant-consonant sequence, as compared to a high-frequency sequence.

In this paper, we continue to explore the influence of sublexical sequence frequency on production accuracy and fluency in children. A second focus of the paper is the relationship between the effect of sublexical sequence frequency and the actual vocabulary that the child commands. 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
might be more quickly accessed and more flexibly reapplied to less familiar but analogous patterns. This effect might be particularly evident in younger children, where the same absolute difference in vocabulary size means a proportionally larger difference — i.e., a proportionally higher probability of having established a robust representation of the individual phonological components independent of specific acoustic and articulatory 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. This work differs from most previous work on children’s nonword repetition accuracy in that the phonotactic probability of the sublexical sequences within the nonword stimuli was systematically controlled. 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

Participants

The participants were 104 typically developing children (73 males and 41 females, ranging in age from 3;2 to 8;10 years;months) and 22 young adults (10 males and 12 females, 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. All of the children met the following four criteria: (1) normal articulatory development, as evidenced by a score at or above one standard deviation below the mean on the Goldman-Fristoe Test of Articulation (GFTA, Goldman & Fristoe, 1986); (2) normal hearing, as evidenced by passing a hearing screening at 20 dB at 500, 1000, 2000, and 4000 Hz; (3) normal structure and function of the peripheral speech mechanism, as evidenced by a standard score at or above one standard deviation below the mean on
the oral movement subtest of the Kaufman Speech Praxis Test for Children (KSPT, Kaufman, 1995); (4) normal non-verbal IQ, as evidenced by a standard score at or above one standard deviation below the mean on the Columbia Mental Maturity Scale (CMMS, 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.

Stimuli

The stimuli were 11 disyllabic and 11 trisyllabic nonword pairs, designed to contrast a low-probability two-phoneme sequence to a high-probability two-phoneme sequence at the same relative position within the word. That is, one member of each nonword pair contained a sequence that occurred in few or no words that children would likely be familiar with and the other member of the nonword pair contained a sequence that occurred in many words familiar to children. Seven nonword pairs contained CV sequences contrasting in low versus high transitional probability, seven nonword pairs contained VC sequences contrasting in transitional probability, and eight nonword pairs contained CC sequences. The sequences were developed using the MHR database, an on-line list of pronunciations of the 6366 most frequently occurring words in the spontaneous continuous speech of first grade children. This database was created by making an electronic version of the word list resulting from Moe, Hopkins, and Rush’s (1982) 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 dialect region as the central Ohio varieties spoken by the children. 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 example, one CC sequence pair was /ft/ and /fk/. The medial cluster /ft/ occurs in many words, such as after, fifteen, and safety, while /fk/ does not occur
in any words at all. Sequences were then embedded in nonwords. For the two nonwords for 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.

Because wordlikeness is known to have an effect on nonword repetition accuracy in children (Gathercole et al., 1991), the nonwords were rated for wordlikeness by adult listeners. Sixteen adults were presented with a larger list of nonwords over headphones in a sound-treated booth and were instructed to rate the nonwords on a 5-point scale, with 1 corresponding to “very unlike a real word” and 5 corresponding to “very like a real word.” Five randomized blocks of the nonwords were presented to each adult. Insofar as possible, the final 44 nonwords were selected to minimize differences in wordlikeness ratings across the two members of each nonword pair. However, a paired t-test still showed a significant difference between the wordlikeness ratings for the high-versus low-probability words ($t[21] = 2.64, p < .05$). This is likely due to the fact that the perceived wordlikeness of a string of phonemes that does not itself constitute a word is strongly affected by how frequently the substrings it contains match parts of real words (see, e.g., Pierrehumbert, 1994; Frisch, et al., 2000). That is, while we tried to minimize differences due to the parts of the words outside the target sequences, it would be impossible to match nonwords for wordlikeness if there is even one sequence of phonemes that differs substantially in transitional probability.

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. The adjustment factor was intended to capture the effect of prosodic context. That is, since 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, just as heterosyllabic /ft/ and /fk/ contrast in occurring in many versus no words, syllable-initial /ju/ and /jau/ contrast absolutely. The familiar sequence /ju/ occurs in many words such as you, use and uniform, whereas the novel sequence /jau/ occurs in no words at all. However, most English words contain at least one syllable-initial CV sequence, whereas heterosyllabic CC sequences are relatively more rare. For one thing, they cannot occur in monosyllabic forms. Thus, although /jau/ is no more frequent as a sequence than /fk/, it should be “easier” simply because CV is more frequent than CC. Therefore, the transitional probability of each sequence included two terms. For the first term, we counted the number of instances in which a target sequence occurred in the relevant syllable position (i.e., syllable-initial for CV; syllable-final for VC; and onset, medial heterosyllabic, or coda position for the different types of CC sequences), and divided this frequency count by the total number of two-phoneme sequences in all of the words in the MHR to get the raw transitional probability. For the second term, we counted the number of instances of the sequence type (e.g., the number of heterosyllabic CCs for /ft/ and /fk/), and divided that by the same denominator. The adjusted transitional probability was then the raw transitional probability of the two-phoneme sequence multiplied by the probability of the sequence type. 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), since 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, & Slowiaczek, 1985), an on-line 20,000 word database 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 6000 most frequently occurring words in the speech of first grade children. 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 7- to 8-year old children. Table 1 gives a list of the stimuli, along with wordlikeness judgments and transitional probabilities for each word, calculated using each of the two methods. As expected, paired comparison t-tests revealed that transitional probabilities were significantly different between the two sequences of each nonword pair (t[21] = 24.45, p < .001 for MHR; t[21] = 14.04, p < .001 for HML).

***Insert Table 1 about here***

**Procedure**

The words were recorded by an adult male speaker of a MidAtlantic variety of Standard American English and were digitized at 20 KHz with 16-bit precision. Three pseudo-randomized 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 as vice versa. The words 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 two practice
words presented by live voice and then two additional digitized practice words presented over the
speakers. Training with digitized practice word pairs then continued until the participant
understood the task and repeated the two digitized practice words accurately. (No more than four
practice trials with digitized practice word pairs was needed with any of the participants.) The
participants’ repetitions were recorded with a head-mounted microphone connected to a digital
audio tape recorder. The participants’ repetitions were then digitized at 20 kHz with 16-bit
precision.

Transcription

The transcriptions were done from the digitized nonwords, using a waveform editor so that each
nonword could easily be played as often as necessary. All of the words were transcribed by a
single transcriber. A second transcriber transcribed 10 percent of the data (four participants from
the three-to-four-year-old group, four participants from the five-to-six-year-old group, three
participants from seven-to-eight-year-old group, and two adults). Phoneme-by-phoneme inter-rater
reliability ranged from 86 to 99 percent for data from individual participants, with a mean of 94
percent across the 13 participants.

Accuracy measures

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 score for any target sequence was six points, and the minimum score was 0. The transcriptions were entered into a database and a computer program was developed and used for automatic scoring.

**Duration measures**

We were also interested in whether fluency of production is related to sublexical sequence frequency. We used segment duration as our measure of production fluency since duration is an acoustic measure of the speed with which a speech movement is executed. 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 non-target 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 for correct productions only. 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

Standardized tests were used to measure vocabulary size. For receptive vocabulary size, the Peabody Picture Vocabulary Test-III (PPVT-III, Dunn & Dunn, 1997) was administered. This is a widely-used measure of receptive vocabulary which requires the participant to point to one of four pictures, given the prompt “show me __________.” It was most recently revised and renormed in 1997 and this most recent version 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. This is a relatively new test in which participants are asked to label pictures for the first 38 items (starting points for children aged 2.6 to 4.11) and then to provide synonyms for the items beyond that point, given a picture and word as a prompt (e.g., the child is shown a picture of a stone and is given the word “stone” and is expected to provide “rock” as a synonym). Both tests provide tables to convert raw scores (the number of items answered correctly) to standard scores, which have a mean of 100 and a standard deviation of 15. The two tests were normed together for participants aged 2 through 90.

Results

Accuracy scores were averaged over the 126 participants for each of the target sequences. A paired-comparison t-test on these scores for the 22 nonword pairs revealed a significant effect of familiarity on accuracy ($t[21] = 3.05, p < .01$). That is, accuracy scores were significantly higher for the target sequences with high transitional probabilities, as compared to the sequences with low transitional probabilities ($M = 5.44, SD = .39$ for high-frequency sequences, $M = 5.10, SD = .46$ 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 ($t[21] = 3.26,$
Vocabulary size and phonotactic probability 18

\[ p < .005, \text{ with } M = 5.37, SD = .39, \text{ for high-frequency sequences, } M = 4.99, SD = .50 \text{ for low-frequency sequences}. \]

There was also a small difference in mean duration for the high- versus low-frequency sequences for the sounds in each of the nine nonword pairs where it was possible to make measurements (\( M = 108 \text{ ms, } SD = 72 \text{ ms, for high-frequency sequences, and } M = 116 \text{ ms, } SD = 79 \text{ ms, for low-frequency sequences} \)). A paired comparison t-test showed this 8 ms difference to be significant (\( t[718] = 2.74, p < .01 \)). (For this analysis, individual duration measures were compared rather than mean data, as the number of tokens per sound was not consistent across sequence types or across participants.) That is, low-frequency sequences, in addition to being produced less accurately, were also significantly longer. Again, when the adults were excluded from the count, the difference was somewhat larger (\( t[575] = 3.41, p < .005 \), with \( M = 109 \text{ ms, } SD = 73 \text{ ms for high-frequency sequences, and } M = 121 \text{ ms, } SD = 84 \text{ ms for low-frequency sequences} \)).

In order to determine whether accuracy was systematically related to transitional probability and to wordlikeness judgments, we correlated the mean accuracy score for the target sequences with their transitional probabilities and with their wordlikeness scores. We did this analysis both for the measure of transitional probability based on the HML database and for the one based on the MHR database. Accuracy was significantly correlated with both measures of transitional probability (\( r^2 = .18, p < .005 \) for MHR and \( r^2 = .19, p < .005 \) for HML), but was not significantly correlated with wordlikeness (\( r^2 = .07, p = .09 \)). (The lack of a significant correlation between accuracy and wordlikeness ratings was probably due to the fact that we had tried, insofar as possible, to select words that had similar wordlikeness ratings across the two members of each nonword pair.) Figure 1 shows accuracy plotted against transitional probability based on each of the two databases, with the three sequence types (CV, VC, CC) differentiated by different plotting
symbols. The overall trend is for accuracy to be greater for sequences with higher transitional probabilities. Note also that the CV sequences are 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 /aunk/ and the high frequency sequence /aunn/, both of which have lower accuracy scores than would be predicted by their transitional probabilities.

Accuracy was also correlated with age. We computed an accuracy score averaged across items for each participant and correlated this measure with age in months. This correlation was significant ($r^2 = .18, p < .001$). Figure 2a shows accuracy for novel and familiar sequences plotted against age. Generally, accuracy increases as age increases. This relationship was stronger for low-frequency sequences ($r^2 = .21, p < .001$) than for high-frequency sequences ($r^2 = .12, p < .001$), so that the regression lines converge for the oldest participants (see Figure 3a). When adults were excluded from the analysis (see Figure 2b), correlations between age and accuracy were still observed, but were smaller in size ($r^2 = .11, p < .005$ for overall accuracy, with $r^2 = .11, p < .005$ for low-frequency sequences, and $r^2 = .09, p < .005$ for high-frequency sequences). We also calculated a measure for each subject of the size of the influence of sequence frequency on repetition accuracy, by summing over all of the differences between the accuracy scores for the high-frequency minus the low-frequency member of each word pair. We called this measure the “familiarity effect.” Figure 3c shows the familiarity effect plotted as a function of age. The familiarity effect was correlated with age when the adults were included in the analysis ($r^2 = .11, p < .001$), but not when they were excluded ($r^2 = .02, p > .1$).

We also examined the relationship between vocabulary size and age. As expected, vocabulary size, as measured by EVT and PPVT-III raw score, was strongly correlated with age ($r^2 = .84, p <
Vocabulary size and phonotactic probability 20

.001 for PPVT-III and \( r^2 = .91, \ p < .001 \) for EVT). The older the participant, the higher the raw score on our measures of expressive and receptive vocabulary. The relationship was still quite strong when the adults were excluded from the analysis (\( r^2 = .66, \ p < .001 \) for PPVT-III and \( r^2 = .69, \ p < .001 \) for EVT). These relationships are plotted in Figure 4. Because the relationship between vocabulary size and age is exponential (that is, vocabulary growth levels off as age continues to increase), we used the natural log of the raw vocabulary scores in all subsequent analyses.

In order to determine the relationship between vocabulary size and repetition accuracy, we correlated accuracy scores for the low and high frequency sequences with our vocabulary measures. These correlations were significant and were greater for low frequency sequences, as compared to high frequency sequences (for low frequency sequences, \( r^2 = .28, \ p < .001 \) for PPVT-III and \( r^2 = .33, \ p < .001 \) for EVT; for high frequency sequences, \( r^2 = .19, \ p < .001 \) for PPVT-III and \( r^2 = .22, \ p < .001 \) for EVT). When the adults were excluded from the analysis, the correlations were somewhat smaller, but still significant (for low frequency sequences, \( r^2 = .21, \ p < .001 \) for PPVT-III and \( r^2 = .21, \ p < .001 \) for EVT; for high frequency sequences, \( r^2 = .14, \ p < .001 \) for PPVT-III and \( r^2 = .13, \ p < .001 \) for EVT). The familiarity effect was also correlated with vocabulary size (\( r^2 = .15, \ p < .001 \) for PPVT-III and \( r^2 = .17, \ p < .001 \) for EVT with adults included; \( r^2 = .06, \ p < .05 \) for PPVT-III and \( r^2 = .08, \ p < .005 \) for EVT with adults excluded). Figure 4 shows accuracy scores plotted against each of our two measures of vocabulary size.

Accuracy and the familiarity effect are correlated both with vocabulary size and age. Furthermore, vocabulary size and age are highly correlated with each other. To tease apart the influence of these two factors, we performed two stepwise multiple regressions. For both analyses, the independent variables were age, the natural log of the EVT raw score, and the natural log of the
PPVT-III raw score. The familiarity effect was the dependent variable for one analysis and overall accuracy (averaged across all items) was the dependent variable for the other analysis. When the dependent variable was the familiarity effect, the only significant predictor was EVT raw score, accounting for 17 percent of the variance. When the dependent variable was overall accuracy, the only significant predictor was PPVT-III raw score, accounting for 31 percent of the variance. These analyses were also performed excluding the adult subjects. Again, when the dependent variable is the familiarity effect, the only significant predictor was EVT raw score, accounting for 8 percent of the variance. When the dependent variable was overall accuracy, the only significant predictor was PPVT-III raw score, accounting for 19 percent of the variance. The results of these regression analyses suggest that it is vocabulary size, rather than age per se, that accounts for the higher accuracy and the smaller effect of transitional probability on accuracy for older children and adults.

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 some actual words containing them. That is, they are more likely to have incorporated some particular sequences into their phonotactics (their general knowledge of what can be a well-formed word of a language), and are more likely to have practiced the auditory and motor representations necessary for perceiving and for fluently producing each of the two sounds in a sequence in the context of the other sound. 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 robust phonotactic knowledge (contributing to fluency) and more robustly abstracted representations of individual phonemes (contributing to flexibility).

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 sequences. The first explanation (i.e., greater likelihood of being familiar with the specific sequence) predicts that we should see an interaction, with the larger-vocabulary children being more accurate than the smaller-vocabulary children just on the attested sequences. Moreover, if this is the sole explanation, a regression function should be able to predict the advantage of a larger vocabulary for the low-frequency attested sequences from the advantage of a larger vocabulary for the high-frequency sequences, but there should be no advantage of having a larger vocabulary for the unattested sequences. If the second explanation is correct, on the other hand, then the larger-vocabulary children should be more accurate than the smaller-vocabulary children on both the unattested and on the attested sequences. Moreover, if this is the sole explanation, a regression function should be able to predict the advantage of a larger vocabulary for the attested low-frequency sequences from the advantage of a larger vocabulary for either the unattested sequences or from the attested high-frequency sequences.

For this analysis, then, we first divided the children into a larger-vocabulary group (the 52 children with the highest EVT raw scores) and a smaller-vocabulary group (the 52 children with the
lowest EVT raw scores). The two groups were well separated by vocabulary size, as measured by EVT raw score (for high vocabulary group, \( M = 73.63, SD = 10.15 \); for low vocabulary group, \( M = 47.58, SD = 5.05 \)). We computed the mean accuracy score for each word for the two vocabulary groups, and then used the means of the smaller-vocabulary children to predict the means of the larger-vocabulary children across three different groups of sequences, as determined by the sequence frequency based on the HML. That is, most of the low-frequency sequences in our corpus were completely unattested in the MHR, but there were five low-frequency sequences that were attested in two to six words in the HML. (Two of these sequences were used in both two- and three-syllable nonwords, so this resulted in seven words with low-frequency but attested sequences.) This gave us three types: unattested (zero-frequency) sequences, attested low-frequency sequences, and high-frequency sequences. We regressed mean accuracy for the target sequences for the larger-vocabulary children against mean accuracy for the smaller-vocabulary children for each of these sequence types. We reasoned that if there was no interaction between vocabulary size and the effect of frequency, then the regression lines for the three functions should be parallel and close together. As can be observed in Figure 5, the regression curve for the unattested sequences overlaps almost completely with the regression curve for the high-frequency sequences, but the curve for the attested low-frequency sequences lies above these two curves and is not as steep. That is, the group of children with larger vocabularies is more accurate relative to the smaller-vocabulary group for the attested low-frequency sequences than would be predicted by the difference between the two groups for either the high-frequency or the unattested sequences.

Discussion

In this study, we found that children and adults repeated low-frequency sequences less accurately than high-frequency sequences, even when they were able to produce all of the
component sounds in the sequence. The same sound was subject to error more often when it was in a less familiar low-frequency phonetic context. Participants also repeated low-frequency sequences with less fluency than high-frequency sequences. The same sound was produced with a longer duration when it appeared in a less familiar phonetic context, as compared to when it appeared in a more familiar context. We interpret these results in terms of the greater opportunity for practicing a phonetic generalization afforded by the more frequent context. With more encounters with a particular sequence, representations in memory become more robustly abstracted away from individual instances. Relevant levels of representation to consider here are the detailed acoustic and articulatory representations of each different prior episode and a higher-level representation that reflects the common properties and relations among prior episodes and that can index both acoustics and articulation.

A second set of results involved the measures of the participants’ receptive and expressive vocabularies. Participants with larger vocabularies showed less of a frequency effect than participants with smaller vocabularies. This was true not only for the sample as a whole (which included 22 adults), but also within the group of children. The 104 children showed a wide range of lexicon sizes as gauged by their EVT and PPVT scores, and the effect of frequency on production accuracy was mediated by vocabulary size. We can think of two possible interpretations of this result. First, a child with a larger vocabulary could simply have had more opportunity to practice the perception and the production of a sound in more different attested contexts. Thus, having a larger vocabulary reduces the likelihood that a particular low-frequency context is completely novel. Alternatively, a child with a larger vocabulary could have elaborated a more robustly abstracted general representation of the individual phonemes or other components internal to the sequence. The latter interpretation would be in keeping with results showing that repeated
practice with a difficult non-native phoneme contrast does not increase long-term facility with the contrast for adult speakers unless there is sufficient variability in the prosodic and segmental contexts for the phonemes (e.g., Logan, Lively, & Pisoni, 1991; Bradlow, Pisoni, Akahane-Yamada, & Tohkura., 1997).

To differentiate between these interpretations, we divided the children into a larger-vocabulary group and a smaller-vocabulary group to determine whether this influence of vocabulary size was due to the larger-vocabulary children simply being familiar with more of the target sequences, or was due to their making a more robust, abstract phonological generalization. The larger-vocabulary children out-performed the smaller-vocabulary children both on low-frequency attested sequences and on unattested sequences. This result strongly supports the claim that, as their vocabularies increase in size, children make more robust, flexibly extensible phonological generalizations. That is, the effect of having a larger vocabulary is not simply that the child knows more phoneme sequences. Rather, the larger-vocabulary children were better at extending phonological generalizations from known patterns to parallel but unknown patterns. They showed less of an effect of frequency not just on the accuracy of the attested low-frequency sequences, but also on the completely unattested, zero-frequency sequences.

The fact that the influence of frequency on production accuracy was mediated by vocabulary size is in keeping with the considerably less robust or inconsistent effects on production accuracy in adults. For example, Vitevitch et al. (1997) found a difference in latency and in accuracy in speeded repetitions of CVCCVC nonsense words with low- versus high-frequency phoneme sequences, but Vitevitch and Luce (1999) were able to replicate only the effect on latency. Munson (in press) found that phonotactic probability predicted diphone durations for children in the same age range as the children we studied, but not for adults. In the current study,
we found no effect of age once vocabulary size was partialled out. Thus, the smaller effect (or lack of an effect) of sequence frequency on nonword repetition accuracy in adults in the earlier studies is probably an effect of the adult’s typically much larger vocabulary, since we know that vocabulary size in this population of American English speakers continues to increase dramatically throughout the school-age years (Nagy & Herman, 1987).

What does this decline in the frequency effect on accuracy (and the lack of interaction for the unattested forms for the children) tell us about the relationship between lexicon and phonology in general? First and foremost, it suggests that learning the phonology cannot be separated from learning individual forms. As Ferguson and Farwell (1975, p. 36) put it, “A phonic core of remembered lexical items and the articulations that produce 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.” A specific prediction that emerges from this view of phonological acquisition is that, other things being equal (i.e. barring other motor or neural problems), young children with larger lexicons should be better at making phonological generalizations and, conversely, children who are better at making phonological generalizations will probably also be better at learning new words. This prediction is supported by research on specific language impairment (SLI). A consistent finding in this literature is that children with SLI have difficulties both with nonword repetition (e.g., Gathercole & Baddeley, 1990; Dollaghan & Campbell, 1998; Edwards & Lahey, 1998; Ellis-Weismer, Tomblin, Zhang, Buckwalter, Chynoweth, & Jones, 2000) and with novel word learning tasks (e.g., Dollaghan, 1987; Oetting, Rice, & Swank, 1995; Rice, Buhr, & Nemeth, 1990; and Rice, Buhr, & Oetting, 1992). Preliminary results of a parallel study of the performance of children with phonological disorder on the same nonword repetition task (Edwards, Beckman, Munson, Draper, & Katagiri, 2000) also
supports this hypothesis. A subset of children with phonological disorder had smaller receptive vocabularies than their typically developing age peers, and these children also showed a greater influence of phoneme sequence frequency on nonword repetition accuracy.

More generally, our results support a particular view of the relationship between grammatical knowledge and processing skills in general. Knowledge of more wordforms is associated with more robustly generalized knowledge of how to learn to hear and say new wordforms. This is in keeping with the emerging view of grammar as an emergent property of the history of interactions between the language user and the language events in the world (see, e.g., Allen & Seidenberg, 1999; Bates & Goodman, 1999; Beckman & Edwards, 2000b; Pierrehumbert, in press; Werker, et al., 2000). 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 two are intimately linked because one is the synoptic long-term cognitive representation and the other is the dynamic immediate cognitive representation of the same encountered events. 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, and can recognize the word produced by unfamiliar speakers — even when the speaker has an unfamiliar accent or is a machine. 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.

Under this view, several seeming paradoxes disappear. For example, the research on cross-language speech perception and development shows that adults are simultaneously better than infants at parsing fine details of native language sound patterns even when presented in ways that
are completely novel, and worse at parsing non-native sound patterns. For example, whereas at 3 years of age, English-acquiring children are quite variable in their ability to attend to spectral variation along a synthetic /s/-/ʃ/ fricative continuum, English-speaking adults show a fine-tuned response to the synthesized fricative, with a sharp category boundary that shifts subtly but consistently to accommodate to different following CV formant transitions (Nittrouer, 1992). At the same time, English-speaking adults who have not been exposed to Hindi in childhood often cannot reliably attend to the similarly subtle differences in natural burst frication spectrum that differentiate Hindi retroflex /ɭ/ from dental /t/ even after a year of studying Hindi (Werker & Tees, 1984). At first glance, these two results might seem contradictory. Adult perceptual processing seems to tap a representation that is at once less abstract (less schematic, more attentive to contextually relevant detail) and more abstract (less attentive to contextually irrelevant detail) than that suggested by the behavior of the younger subject. However, if we think of the effect of experience as one of elaboration, one which robustly encodes memories at more relevant different levels of representation, then this paradox disappears.

The effects of having a larger vocabulary, then, could reflect a similar process of elaboration. As the child gains more experience with more words, and 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. To recast Ferguson and Farwell’s (1975) idea of a “lexical core” in this view, then, it is not so much that a “pre-grammatical” foundation of knowledge of how to produce a small core of words is overlaid by phonological knowledge, but that 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 wordform.
ACKNOWLEDGEMENTS

This work was supported by NIDCD grant 02932 to Jan Edwards and by NIH training grant T32 DC0051 to Robert A. Fox. 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


delayed preschoolers. *Journal of Speech and Hearing Disorders, 55*, 33-42.

word learning (QUIL) of words: The effect of a pause. *Journal of Speech and Hearing
Research, 35*, 1040-1048.

Rose, R. C., Schroeter, J., & Sondhi, M. M. (1996). The potential role of speech production models
in automatic speech recognition. *Journal of the Acoustical Society of America, 99*, 1699-
1709.

Schaadt, L. (1997). Phonological representations and phonological awareness in early elementary
school-age children. Unpublished senior honor’s thesis, Ohio State University, Columbus,
OH.

word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition,
13*, 64-75.

acquisition: Swedish and American /t/. *Phonetica, 51*, 146-158.

sensitivity to phonotactic patterns. In M. Broe & J. Pierrehumbert (Eds.), *Papers in
Laboratory Phonology V* (pp. 269-282). Cambridge, UK: Cambridge University Press.

Tsushima, T., Takizawa, O., Sasaki, M., Shiraki, S., Nishi, K., Kohno, M., Menyuk, P., & Best, C.
(1994). Discrimination of English /r/-l/ and /w-y/ by Japanese infants at 6-12 months:
Language-specific developmental changes in speech perception abilities. *Proceedings of the
Vocabulary size and phonotactic probability 37


Table 1. Wordlikeness and transitional probabilities for nonword stimuli. Segments from pairs for which we measured the duration of one of the target phonemes are indicated.

<table>
<thead>
<tr>
<th>Phonetic form</th>
<th>Measured segments</th>
<th>Wordlikeness</th>
<th>MHR</th>
<th>HML</th>
</tr>
</thead>
<tbody>
<tr>
<td>low freq high freq</td>
<td>low freq high freq</td>
<td>low freq high freq</td>
<td>low freq high freq</td>
<td>low freq high freq</td>
</tr>
<tr>
<td>/jugoin/ /bogib/</td>
<td>3.06 3.30</td>
<td>-12.42 -9.71</td>
<td>-12.92 -10.84</td>
<td></td>
</tr>
<tr>
<td>/moipød/ /mæbeıp/ [m]</td>
<td>2.96 2.76</td>
<td>-13.11 -8.09</td>
<td>-12.00 -7.81</td>
<td></td>
</tr>
<tr>
<td>/vugim/ /vidæg/ [v]</td>
<td>3.19 2.91</td>
<td>-13.11 -8.73</td>
<td>-12.92 -8.53</td>
<td></td>
</tr>
<tr>
<td>/bodøjau/ /medøjau/</td>
<td>2.35 2.96</td>
<td>-13.11 -8.37</td>
<td>-14.30 -7.56</td>
<td></td>
</tr>
<tr>
<td>/vukatem/ /vitægap/ [v]</td>
<td>2.96 2.65</td>
<td>-13.11 -8.73</td>
<td>-12.92 -8.53</td>
<td></td>
</tr>
<tr>
<td>/gaunøpek/ /giømok/</td>
<td>2.78 2.64</td>
<td>-12.42 -9.71</td>
<td>-11.82 -10.84</td>
<td></td>
</tr>
<tr>
<td>/nubømøn/ /nødøbip/ [n]</td>
<td>1.68 1.88</td>
<td>-13.11 -8.26</td>
<td>-10.84 -7.79</td>
<td></td>
</tr>
<tr>
<td>/tedaum/ /podaud/</td>
<td>2.90 3.11</td>
<td>-13.31 -10.67</td>
<td>-14.59 -11.81</td>
<td></td>
</tr>
<tr>
<td>/dugnødæd/ /tægnødæit/ [g]</td>
<td>2.68 3.03</td>
<td>-13.31 -9.98</td>
<td>-14.59 -10.53</td>
<td></td>
</tr>
<tr>
<td>/aukpøde/ /ikbøni/</td>
<td>2.41 2.06</td>
<td>-13.31 -9.48</td>
<td>-14.59 -9.77</td>
<td></td>
</tr>
<tr>
<td>/auftøga/ /auntøko/ [au]</td>
<td>2.43 3.11</td>
<td>-13.31 -8.56</td>
<td>-14.59 -8.96</td>
<td></td>
</tr>
<tr>
<td>/nøfæmb/ /minæmp/</td>
<td>2.49 3.03</td>
<td>-13.57 -9.32</td>
<td>-15.73 -11.08</td>
<td></td>
</tr>
<tr>
<td>/pwagəb/</td>
<td>/tweket/</td>
<td>1.69</td>
<td>2.28</td>
<td>-13.88</td>
</tr>
<tr>
<td>----------</td>
<td>---------------</td>
<td>------</td>
<td>------</td>
<td>--------</td>
</tr>
<tr>
<td>/buflkit/</td>
<td>/kiften/</td>
<td>2.61</td>
<td>3.68</td>
<td>-14.00</td>
</tr>
<tr>
<td></td>
<td>[f]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/dogdet/</td>
<td>/tæktut/</td>
<td>2.76</td>
<td>3.38</td>
<td>-14.00</td>
</tr>
<tr>
<td>/næfkaτu/</td>
<td>/gæftədai/</td>
<td>2.73</td>
<td>2.44</td>
<td>-14.00</td>
</tr>
<tr>
<td></td>
<td>[f]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/degdəne/</td>
<td>/tikɔpo/</td>
<td>2.43</td>
<td>2.54</td>
<td>-14.00</td>
</tr>
</tbody>
</table>

\(^a\) Mean wordlikeness rating (ranging from 1 to 5); \(^b\) transitional probabilities based on the MHR database; \(^c\) transitional probabilities based on the HML database.
Figure 1. Mean accuracy for target sequence plotted against its transitional probability calculated from the MHR database (Fig. 1a) and from the HML database (Fig. 1b), for all 44 nonwords.

Figure 2. Mean accuracy over all low frequency sequences (filled circles) and over all high frequency sequences (open circles) plotted against age in months for all participants (Fig. 2a) and for child participants only (Fig. 2b). Familiarity effect plotted against age in months for all participants (Fig. 2c).

Figure 3. Vocabulary size plotted against age in months, with size measured by EVT raw score (Fig. 3a) and by PPVT-III raw score (Fig. 3b).

Figure 4. Mean accuracy of low frequency sequences (filled circles) and of high frequency sequences (open circles) plotted against vocabulary size, as measured by natural log of EVT raw score (Fig. 4a) and natural log of PPVT-III raw score (Fig. 4b).

Figure 5. Mean accuracy scores for the larger vocabulary group of child participants plotted against the mean accuracy scores for the smaller vocabulary group of child participants for the high frequency target sequences (open circles), low frequency attested sequences (filled circles), and unattested sequences (asterisks).
Figures 1a and 1b
Figures 2a and 2b
Figure 2c

A scatter plot showing the relationship between age in months and familiarity effect. The x-axis represents age in months, ranging from 0.0 to 500.0, and the y-axis represents familiarity effect, ranging from -10.0 to 40.0. The data points are plotted, and a trend line is drawn through them, indicating a negative correlation.
Figures 3a and 3b

![Graphs showing EVT and PPVT-III raw scores against age in months.](image-url)
Figures 4a and 4b

1. ln (EVT raw score)
2. ln (PPVT-III raw score)

- Low freq sequences
- High freq sequences

Mean accuracy score
Figure 5

Mean accuracy: smaller-vocabulary group

Mean accuracy: larger-vocabulary group

High freq sequences
Low freq sequences
Unattested sequences