Journal of Memory and Language **42,** 481–496 (2000)
doi:10.1006/jmla.1999.2692, available online at http://www.idealibrary.com on IDEAL®

Perception of Wordlikeness: Effects of Segment Probability and Length on the Processing of Nonwords

Stefan A. Frisch, Nathan R. Large, and David B. Pisoni

Indiana University

A probabilistic phonotactic grammar based on the probabilities of the constituents contained in a dictionary of English was used to generate multisyllabic nonwords. English-speaking listeners evaluated the wordlikeness of these patterns. Wordlikeness ratings were higher for nonwords containing high-probability constituents and were also higher for nonwords with fewer syllables. Differences in the processing of these same nonwords that partially reflected their perceived wordlikeness were also found in a recognition memory task. Nonwords with higher probability constituents yielded better recognition memory performance, suggesting that participants were able to use their knowledge of frequently occurring lexical patterns to improve recognition. These results suggest that lexical patterns provide the foundation of an emergent phonological competence used to process nonwords in both linguistic and metalinguistic tasks.

Key Words: phonotactics; wordlikeness; recognition memory; nonwords.

In studying the ability to recognize spoken words, investigators have described many factors that affect the way linguistic stimuli are perceived and processed. Some of these are based on lexical properties of a particular word, including its semantic relationship with other words (Swinney, 1979), its frequency of use in text or speech (Treisman, 1978), its familiarity to the listener (Chalmers, Humphreys, & Dennis, 1997), or its confusability with similar-sounding words in the lexicon (Luce & Pisoni, 1998). Other relevant properties are due to the linguistic structure of the word itself, including morphological considerations, such as the presence of bound grammatical morphemes, embedded words, or compounding, and phonological features, such as length, stress pattern, and segmental composition. In order to remove most lexical and morphological factors from linguistic stimuli, studies commonly employ nonwords. For example, nonwords have been used to examine word learning in young children (Dollaghan, 1985; Gathercole, 1989), the effects of lexical status on memory span (Hulme, Roodenrys, Brown, & Mercer, 1995), and the process of lexical access (Forster, 1978; Marslen-Wilson, 1984; Rubenstein, Garfield, & Milliken, 1970).

Nonword stimuli vary in their similarity to actual words in a listener's native language, which is referred to as *wordlikeness.* Wordlikeness has a strong influence on how quickly and accurately nonwords are processed in a variety of tasks. There have been two general approaches to explaining and quantifying wordlikeness effects that reflect two different factors that could influence wordlikeness judgments and the processing of nonwords. Some studies have investigated the influence of similarity between a nonword and one or more particular words in the lexicon. Other studies have examined the influence of the phonotactic structure of the nonword itself. Despite the importance of wordlikeness to psycholinguistic research, relatively few studies have attempted to examine the factors that listeners use when they make

This work was supported by NIH Training Grant No. DC00012 and NIH Research Grant DC00111 to Indiana University. We thank Janna Carlson for her patient assistance in recording the stimuli used in this study. We also thank Janet Pierrehumbert, Richard Shiffrin, and Mike Vitevitch for comments on an earlier version of this article and the anonymous reviewers for their many helpful suggestions that were too numerous to acknowledge in the text.

Address correspondence and reprint requests to Stefan Frisch, University of Michigan, Program in Linguistics, 1076 Frieze Building, 105 S. State St., Ann Arbor, MI 48109-1285. E-mail: sfrisch@umich.edu.

0749-596X/00 $35.00

judgments of wordlikeness or how metalinguistic judgments are related to differences in language processing.

Greenberg and Jenkins (1964) carried out one of the earliest investigations of the influence of similarity of a nonword to particular words. They examined the "distance" of four-phoneme (CCVC) monosyllabic nonsense words from the words of English using phoneme substitution. Their distance score correlated well with the judgments of distance from English that were elicited from participants. Greenberg and Jenkins proposed that there is a phonological space in which words are assembled in memory, according to their sound structure (see also Treisman, 1978). Words differ in the number and frequency of other words that occupy the *similarity neighborhood* around them in this space. Combinations of segments that are more common reside in *dense* neighborhoods in the lexical space, while uncommon combinations occupy *sparse* neighborhoods. Luce and Pisoni (1998) reviewed a considerable body of evidence showing that the size of the lexical neighborhood of a word influences its processing.

Similarity to actual words also influences the processing of nonwords. For example, Wurm and Samuel (1997) used nonword stimuli that varied in similarity to words in the lexicon to study processing demands in the phoneme-monitoring task. They generated multisyllabic nonwords by replacing one phoneme, two phonemes, or one phoneme per syllable from actual words and found that phoneme monitoring latency was influenced by the number of phoneme changes. Actual words had the smallest response latency, with latency increasing for one phoneme, two phoneme, and many phoneme substitutions. Such studies suggest that the similarity between nonwords and actual words influences both language processing and wordlikeness judgments.

In related research, Vitz and Winkler (1973) sought to describe the information people use to make similarity comparisons between pairs of words. They analyzed similarity judgments based on a metric that considered the amount of phonemic overlap between words. They found a relationship between degree of overlap and similarity ratings that was analogous to the findings of Greenberg and Jenkins (1964), suggesting that two-item comparison tasks may involve the same processes as wordlikeness judgments. Using similar methodology, Sendlmeier (1987) compared listeners' abilities to classify nonwords based on similarity. He found that naïve participants made judgments based only on the most salient characteristics of the nonwords, such as the word onset or the stressed syllable. Given the parallels between similarity judgments for words and wordlikeness judgments for nonwords, particular salient constituents may also play an important role in influencing wordlikeness judgments.

In linguistic theory, the phonotactic constraints of a language specify its well-formed words and morphemes. Nonwords which are composed of phonemes in legal combinations, but which have no lexical entry, have been termed *pseudowords*, in contrast to impossible nonword patterns (Brown & Hildum, 1956). It has also been noted that certain sound patterns may not be completely absent from a language, but are rare enough that their acceptability is questionable. For example, Pierrehumbert (1994) examined restrictions between word-internal (medial) consonant clusters in monomorphemic English words. A large number of these clusters do not occur, but are not subject to any particular phonological constraint of English. She proposed that these clusters are highly improbable combinations that are psychologically equivalent to illegal clusters. Kessler and Treiman (1997) provided support for the descriptive utility of probabilistic phonotactic constraints in a study of onset–vowel and vowel–coda co-occurrences in English monosyllables. They found that many vowel–coda pairs occurred more or less frequently than would be expected by chance, while onset–vowel combinations were statistically random. For example, the rime /æl/ occurs less frequently than would be expected based on the high probabilities of /æ/ and /l/ independently appearing as the vowel and coda in a syllable. They argued that probabilistic constraints between vowel and coda but not onset and vowel provide evidence for the rime as a syllable-internal constituent in

English. Their findings converge with a number of studies of processing that show onset–rime structure for English (see Treiman, 1988; Treiman, Kessler, Knewasser, Tincoff, & Bowman, 2000).

Probabilistic phonotactic patterns have been shown to influence language processing. For example, Vitevitch and Luce (1998) demonstrated that higher probability patterns facilitate repetition of nonwords by adults. Pitt and McQueen (1998) showed that the identification of a phoneme can be influenced by transitional probability from the preceding segment. Knowledge of probabilistic phonotactic patterns develops at an early age. Jusczyk, Luce, and Charles-Luce (1994) found that infants become sensitive to differences in segment probability and transitional probability of CVC sequences during their first year.

Landauer and Streeter (1973) carried out a computational analysis of a dictionary (as an approximation of a native speaker's mental lexicon) to demonstrate that word frequency effects in language processing are confounded with phonotactic probability. They demonstrated that high-frequency words are composed of different, more common sounds than low-frequency words. This finding was replicated and extended by Frauenfelder, Baayen, Hellwig, and Schreuder (1993), who showed that words with more common segments are more frequent than words with less common segments in both English and Dutch. Frauenfelder et al. further demonstrated that word frequency and phonotactic probability are also confounded with neighborhood density. Thus, it is not clear whether the influences of neighborhood density and of phonotactic probability discussed above are separate influences or whether they can be derived from a common source.

Experiments that manipulate the phonotactic probability of words or nonwords typically vary the frequency of a few target segments or the transitional probability of a particular segment pair. Recently, Coleman and Pierrehumbert (1997) examined the influence of probabilistic phonotactics on wordlikeness judgments using a more sophisticated probability measure. They analyzed acceptability judgments by 12 participants for 116 multisyllabic nonwords, half of which contained illegal sequences. Although judgments were influenced by the presence of an illegal sequence, a great deal of variation in responses could not be accounted for by categorical phonotactic restrictions. Coleman and Pierrehumbert proposed that some of the variation could be explained using a *stochastic grammar* that generates expected probabilities for nonwords by taking the product of the nonwords' constituent probabilities (i.e., assuming independent combination of constituents). The logarithm of the expected probability for each nonword was compared to the number of participants who accepted it, and the two measures were significantly correlated ($r = .50$, $p < .001$). Coleman and Pierrehumbert proposed that acceptability judgments for nonwords involve an evaluation of the entire composition of the nonword (i.e., the cumulative phonotactic probability). In their stochastic grammar, the presence of an illegal sequence in one part of the nonword can be ameliorated by a high-probability sequence elsewhere in the word, in contrast to the traditional linguistic view of acceptability (e.g., Hawkins, 1979, p. 50). The findings of Coleman and Pierrehumbert suggest that probabilistic phonotactics are an important influence on wordlikeness judgments.

The stochastic grammar may also provide a means of unifying the influences of phonotactic probability and lexical neighborhood effects in a single measure. Since the expected probability metric takes into account the entire composition of the nonword, the stochastic grammar reflects the number of words in the lexicon that are similar to the nonword to some degree. A nonword with several low-probability segments will have low aggregate probability, and since it contains several infrequent segments it will likely be in a sparse region of the lexical similarity space.

Coleman and Pierrehumbert (1997) based their analyses on stimuli that were not specifically designed to examine probabilistic phonotactics. In order to replicate and extend their findings with more carefully controlled stimuli, we generated a corpus of nonwords that varied in their length and the probability of their constituents and that covered a wide range of expected probabilities according to the stochastic grammar. These nonwords were used in three experiments using two different behavioral tasks: subjective metalinguistic ratings and recognition memory. While the focus of this paper is primarily on metalinguistic judgments for nonwords, we also examined recognition memory for nonwords in a preliminary attempt to integrate these lines of research. No study that we know of has investigated metalinguistic and linguistic tasks together, for the same set of stimuli. The primary goal of our study was to evaluate the usefulness and psycholinguistic relevance of Coleman and Pierrehumbert's probabilistic grammar as a predictor of wordlikeness.

EXPERIMENT 1

Method

Stimuli. The Coleman and Pierrehumbert (1997) grammar considers the likelihood of each constituent (onset and rime) of each syllable in terms of its prosodic position within the word. Although there are alternative ways of describing the internal division of the syllable, we refer to the onset and rime structures for descriptive convenience (see Pierrehumbert & Nair, 1995; Treiman & Kessler, 1995, for discussion). In the grammar, the prosodic positions of the constituents differentiate the stress of the syllable (stressed or unstressed) and the constituent's position within the word (initial, final, or medial). Since only onsets can be initial and only rimes can be final, there are eight different probability distributions of constituents-by-position: stressed initial onsets, stressed medial onsets, stressed medial rimes, stressed final rimes, and unstressed counterparts to each of these. Each constituent distribution consists of the inventory of potential constituents occurring in a given position, as well as the probability of that item's occurrence in that position, computed from an on-line dictionary of English (Nusbaum, Pisoni, & Davis, 1984). The probability of each constituent is equal to the number of words that contain that particular constituent

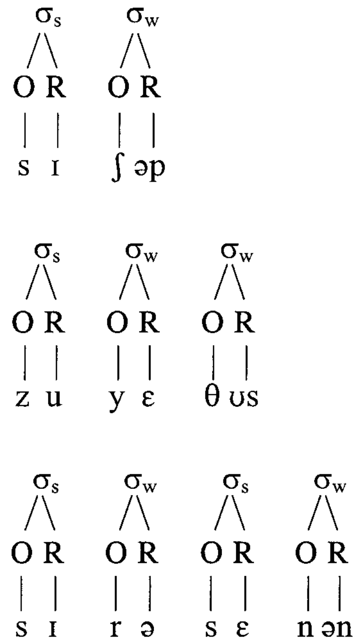

FIG. 1. Structured phonological representation of example nonwords created from the eight constituent distributions, one for each length. The two- and four-syllable nonwords have high constituent probability and the three-syllable nonword has low constituent probability. σ_s, strong syllable (stressed); σ_w, weak syllable (unstressed); O, onset; R, rime.

at the specified position divided by the total number of words containing constituents at that position.

To construct our groups of nonword stimuli, we selected two sets of constituents, a high-probability group and a low-probability group, for each prosodic position. We used only single-consonant onsets (C), all medial rimes consisted only of a single vowel or diphthong (V), and all final rimes were a vowel–consonant (VC) pair, so that all of our nonwords had an alternating CV pattern plus a final consonant. We used these constituents to construct groups of nonwords of two, three, and four syllables in length, using the most common stress pattern of that word length in the dictionary (see Fig. 1 for three examples). Each nonword consisted entirely of high-probability constituents or low-probability constituents. Coleman and Pierrehumbert multiplied the probabilities of the constituents of each syllable in a nonword to yield that nonword's expected probability. By the product probability metric, our nonwords with high- and low-probability constituents covered a wide range of overall probabilities. Note that in longer nonwords more probabilities are multiplied together, and so some of the longest nonwords with high-probability constituents actually had a lower product probability than the shortest nonwords with low probability constituents. The complete set of stimuli is given in Appendix 1.

Within each constituent probability group, the segment composition was equated as much as possible across the different-length nonwords. To reduce the amount of similarity between the stimuli no two syllables occurred in the same word more than once within a group. Each syllable in the stimulus set occurred six times across the 144 stimuli. In the stimulus set, no nonword appeared as a contiguous part of any other nonword. Monosyllabic nonwords were avoided for this reason, as the repeated use of syllables across stimuli of different lengths would have led to the monosyllabic nonwords appearing as parts of the longer nonwords. To avoid morphological confounds, the highest probability final rimes (e.g., /-əs/, /-əd/, which might be interpreted as inflectional suffixes) were avoided. However, it is possible that some of our stimuli were perceived as morphologically complex, as discussed below.

A female talker produced spoken versions of the 144 nonwords in a randomized order, recorded over several sessions using a digital signal recording program. The talker was a research assistant with some phonetic training but was not aware of the methods used to generate the nonwords or the purposes of the experiment. The talker sat in a sound-attenuated booth and recordings were made using a Shure SM98 microphone. Phonemic transcriptions of the nonwords were presented to the talker individually on a computer monitor. Each stimulus was digitally leveled to 64 dB SPL, and the recorded stimuli were screened for accuracy and fluency by the first and second authors. The resulting stimuli reflected normal phonetic processes of English: unstressed, intervocalic /t/ and /d/ were flapped, final /t/ was glottalized, other final stops were unreleased, and unstressed vowels were reduced and centralized. The productions were typical of laboratory speech or spelling pronunciation and more carefully articulated than casual speech. In particular, we believe that the unstressed full vowels (e.g., /e/, /ɛ/, /ʊ/) were articulated distinctly from /ə/, and that the voicing contrast for unstressed /t/ and /d/ was maintained in the length of the preceding vowel.

Participants. Twenty-four undergraduate students earning experimental credit for introductory psychology courses participated. The participants in this and all following experiments were native speakers of American English and reported no recent history of speech or hearing disorders.

Procedure. Participants were instructed to rate the nonwords for their wordlikeness. The instructions included descriptors for a 7-point scale to be used for rating the nonwords. A rating of 1 was described as "Low–Impossible—this word could never be a word of English," rating of 4 as "Medium–Neutral—this word is equally good and bad as a word of English," and 7 as "High–Possible—this word could easily be a word of English." The other numbers were to be used for nonwords between these categories. Ratings of 2 or 3 represented "unlikely" and 5 or 6 were "likely." We instructed participants to work as quickly as possible without sacrificing accuracy.

The stimuli were presented by computer over Byerdynamic DT100 headphones at 74 dB SPL. Participants sat in individual sound absorbent testing booths, facing a computer monitor. Participants responded using a 7-button response box. The rating response and latency were recorded for each test item. The response boxes were covered by printed cards indicating the digits 1–7 and the guidelines "Low Wordlikeness," "Medium Wordlikeness," and "High Wordlikeness" above the digits 1, 4, and 7, respectively. Stimuli were randomized for each listener. There were 2 s between successive trials.

Results and Discussion

Ratings. Analysis of variance (ANOVA) was performed with constituent probability group

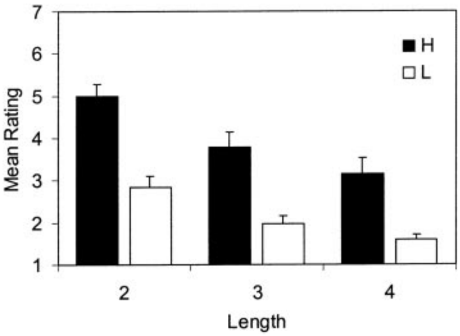

FIG. 2. Mean subjective ratings for the nonword stimuli in Experiment 1 for each constituent probability group (High vs Low) and length in syllables (two, three, or four). Error bars show 99% confidence interval for the mean.

and length in syllables as factors across subjects and across stimuli. Figure 2 displays the mean rating for each length and probability group. Stimuli with high-probability constituents were rated more wordlike than stimuli with low-probability constituents [$F1(1, 138) = 206$, $p < .001$, $F2(1, 138) = 426$, $p < .001$]. Shorter stimuli were rated more wordlike than longer stimuli [$F1(2, 138) = 49.7$, $p < .001$; $F2(2, 138) = 103$, $p < .001$]. There was no clear interaction between probability and length [$F1(2, 138) = 3.6$, $p < .05$, $F2(2, 138) = 1.8$, $p = .18$].

Mean ratings for each stimulus, as a function of log product probability, are shown in Fig. 3. Expected log probability and average ratings were highly correlated, $r = .87$, $p < .001$, replicating Coleman and Pierrehumbert's (1997) findings. Note that there is overlap between the stimuli with high-probability constituents and the stimuli with low-probability constituents at expected probability 10^{-10}. These stimuli are four-syllable nonwords with high-probability constituents and two-syllable nonwords with low-probability constituents. The high-constituent probability nonwords and low-constituent probability nonwords have similar ratings at this point, as predicted by their similar product probabilities in the stochastic grammar. This suggests that the log product probability metric correctly predicts the integration of segmental probability and length factors in determining wordlikeness.

While the log product probability provides an extremely good prediction of wordlikeness judgments, it is possible that some other metric would provide a better prediction or that the rating patterns can be attributed to idiosyncrasies of the stimuli. We examine a wide range of predictors for the rating data in the "Other Grammars" section below. Potential confounds in our stimuli that could account for the ratings come in two forms. First, the low constituent probability nonwords contained lax vowels /ɛ/, /æ/, and /ʊ/ that might be considered ungrammatical in English in unstressed syllables. Thus, many of our nonwords could be considered illegal according to the traditional descriptive phonology of English, despite the fact that these constituents are attested in the lexicon. This could potentially account for the constituent probability and length differences, as nonwords with illegal vowels would be rated less wordlike, and the longer low-constituent probability nonwords contained more of these unstressed vowels. However, upon closer examination this does not appear to be the case. Table 1 shows mean ratings for low-constituent probability nonwords for each unstressed rime, along with the log probability of that constituent. Note that there were unstressed medial rimes only in nonwords of three or four syllables in length. Both the unstressed medial rimes and unstressed final rimes contained potentially illegal vowels; how-

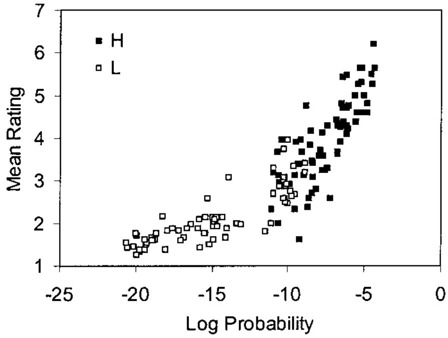

FIG. 3. Mean subjective ratings for each nonword in Experiment 1 as a function of the log product of onset and rime probabilities for the nonword.

TABLE 1

Mean Rating and Unstressed Rime Probability of Low-Constituent Probability Nonwords

Rime	Log *P*	*M*	*SD*
Medial rimes			
æ	−2.32	1.85	.38
e	−2.50	1.77	.38
ε	−2.66	1.74	.26
Final rimes			
ʊs	−4.03	2.32	.71
εd	−4.03	2.19	.66
oˡl	−4.03	2.14	.70
əʃ	−4.03	1.91	.42

ever, there were no significant differences in the ratings for the different medial rimes [$F1(2, 69) < 1$; $F2(2, 45) < 1$] or final rimes [$F1(3, 92) = 1.61$, $p = .19$; $F2(3, 68) = 1.32$, $p = .27$]. Since there are no large differences in probability between the different rimes and there is no evidence for a difference in ratings between vowels, the data are completely consistent with the stochastic grammar.

The second potential confound is that the high constituent probability nonwords had final rimes that could be interpreted as suffixes (/ən/ as in *golden,* or as in *dancin'* in casual speech), as unstressed pronouns (/ət/ as *it* and /əm/ as *them*), or as particles (/əp/ as in *write up* and /ən/ as in *break in*). As a result, some of the high probability nonwords could have been perceived as phrases consisting of real words when the rime is treated as a word (e.g., /midət/ as *meet it* and /sæʃəm/ as *sash them*). This confound could account for the constituent probability and length effects, as none of the low constituent probability nonwords can be parsed as word sequences, and fewer of the long high constituent probability nonwords can be parsed as word sequences. Table 2 shows mean ratings for two- and three-syllable nonwords that can and cannot be parsed as word sequences (or as words with suffixes). None of the four-syllable nonwords can be parsed as containing a word sequence. There is no effect of a potential word parse for two-syllable nonwords, but the increase in ratings for three-syllable nonwords that can be parsed as word sequence is significant [$t(23) = 2.09$, $p < .05$]. While an account based on parsing cannot account for the different findings for two- and three-syllable nonwords, the difference is easily accounted for by probability. In our stimuli, the log-product probabilities of the initial portions of the nonwords are nearly identical for two-syllable nonwords. Log probabilities happen to be higher for the initial portions of the three-syllable nonwords with a potential word parse than those without.

Post hoc analysis: Implicit rejections. Figure 3 shows that nonwords with expected probability below 10^{-12} were given very low wordlikeness ratings, with mean ratings approaching the lower limit of 1. Coleman and Pierrehumbert (1997) found no such floor effect, but our experiment differs from theirs in two important ways. First, nearly all of their stimuli had probabilities above 10^{-12}, so it may be that our nonwords exceeded some lower limit to acceptable probability. Second, they used a two-alternative forced-choice accept/reject task rather than a 7-point wordlikeness rating scale. The difference in the type of judgment asked for may affect how the judgment is made. With these possibilities in mind, we carried out a post hoc analysis to determine whether participants' ratings really did reach a floor level or whether they still discriminated the differences in probability even for very low-probability nonwords. We analyzed the number of ratings equal to "1," either for each nonword across subjects or by each subject across nonword groups. Our in-

TABLE 2

Mean Rating and Log Probability of the Initial Fragment of High-Constituent Probability Nonwords

Group	*n*	Log *P*	*M*	*SD*
Two-syllable				
w/parse	16	−3.59	4.96	.55
w/o parse	8	−3.61	5.03	.62
Three-syllable				
w/parse	7	−5.16	4.13	.42
w/o parse	17	−5.62	3.64	.72

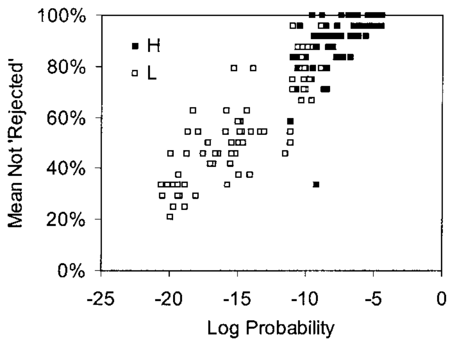

FIG. 4. Percentage of participants responding with a rating higher than an implicit rejection for each nonword in Experiment 1 as a function of the log product probability.

structions described a rating of "1" as equivalent to "Low–Impossible—this word could never be a word of English," making such a rating potentially analogous to an "unacceptable" response in a two-choice task. We call this response an "implicit rejection." An ANOVA using mean number of implicit rejections per constituent probability and length group as the dependent measure showed main effects of constituent probability [$F1(1, 138) = 65.8$, $p < .001$; $F2(1, 138) = 339$, $p < .001$] and length [$F1(2, 138) = 15.45$, $p < .001$; $F2(2, 138) = 77.7$, $p < .001$]. There was also a significant divergent interaction between constituent probability and length [$F1(2, 138) = 4.7$, $p < .05$; $F2(2, 138) = 23.5$, $p < .001$]. The difference between high- and low-constituent probability groups was larger for longer stimuli than for shorter stimuli.

Figure 4 shows the percentage of ratings greater than 1 (ratings that were not implicit rejections) across subjects for each stimulus as a function of expected log probability. The percentage of ratings greater than 1 is shown for ease of comparison to the original rating data. There is a high correlation between the number of ratings greater than 1 and expected log probability, $r = .91$, $p < .001$. Comparing Figs. 3 and 4, the number of implicit rejections appears to better differentiate the low-probability stimuli, while the original mean rating appears to better differentiate the high-probability stimuli.

The present findings show that a stochastic grammar can be successfully employed to generate nonword stimuli of relatively high or low wordlikeness as determined by subjective ratings. Native speakers' metalinguistic knowledge of phonotactics appears to include the probabilistic information encoded in the stochastic grammar. Wordlikeness ratings approached the low asymptote for very low expected probability strings using a rating scale, but these stimuli are still clearly differentiated in the number of floor-level ratings they receive. To determine if these implicit rejections are substantially different from rejections on a two-choice accept/reject task, as well as to more closely replicate Coleman and Pierrehumbert's study, we conducted a second experiment with the same stimuli using the accept/reject task. The number of rejections in this task can be compared to our measure of implicit rejections, and to the wordlikeness rating data, to see if differences in metalinguistic judgments of wordlikeness arise under different response conditions.

EXPERIMENT 2

Methods

Participants. Twenty-four undergraduate students earning experimental credit for introductory psychology courses participated.

Stimuli. The stimuli were the same as in Experiment 1.

Procedure. The procedure and presentation of stimuli was identical to Experiment 1, except that the participants were told to accept or reject each sound pattern as a possible word of English. The 7-button boxes had labels reading "Yes" and "No" placed above the "2" and "6" buttons, respectively.

Results and Discussion

The percentage of responses that indicated a stimulus was acceptable is shown in Fig. 5 for each constituent probability and length group. Nonwords with high-probability constituents were more acceptable than those with low-probability constituents [$F1(1, 138) = 343$, $p < .001$; $F2(1, 138) = 506$, $p < .001$], and shorter nonwords were more acceptable than longer nonwords [$F1(2, 138) = 49.1$, $p <$

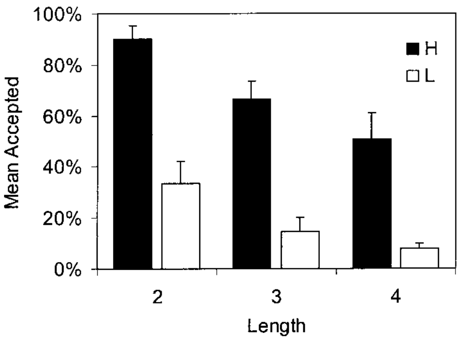

FIG. 5. Percentage of acceptances for each nonword in Experiment 2 for each constituent probability group (High vs Low) and length in syllables (two, three, or four). Error bars show 99% confidence interval for the mean.

.001; $F2(2, 138) = 72.5$, $p < .001$]. The interaction was not significant [$F1(2, 138) = 3.3$, $p < .05$; $F2(2, 138) = 2.2$, $p = .11$].

Figure 6 shows the percentage of participants that judged each nonword acceptable as a function of the log product probability of the nonword. Judgments begin to asymptote near probability 10^{-12} as in Experiment 1. Despite this leveling, acceptability and log probability for the stimuli correlated highly, $r = .86$, $p < .001$. Thus the two-choice acceptability task and the 7-point wordlikeness task produce essentially the same results. By contrast, a comparison between the acceptability judgments and the implicit rejections measure from Experiment 1 shows that these are not equivalent rejections. With only two choices available in the acceptability task, participants used the "unacceptable" choice more often than the "impossible" rating of "1" on the 7-point wordlikeness task.

Taken together, Experiments 1 and 2 replicate and extend the findings of Coleman and Pierrehumbert (1997), providing support for the onset/rime grammar and the log product probability metric as a method of generating and describing nonwords. The very high correlation we obtained between metalinguistic wordlikeness judgments and log product probability also provides strong evidence that phonotactic knowledge emerges from the patterns of segments in the mental lexicon (Bybee, 1988; Ohala & Ohala, 1986; Plaut & Kello, 1999).

EXPERIMENT 3

Experiment 3 employed a recognition memory task to study the representation and processing of our nonwords in memory. If participants use knowledge about the words in their language to perceive and encode nonwords, as the metalinguistic judgments suggest, then items that are more wordlike should have stronger activation or more associations with real words in the language. The more wordlike items should thus be more distinctive in memory. Thus, we predicted that participants would recognize the nonword stimuli with high-probability constituents more accurately than the stimuli with low-probability constituents.

This prediction might be viewed as the reverse of the pattern found for real words. Low-frequency words are more easily discriminated (i.e., recognized) than high-frequency words in recognition memory experiments. The word-frequency effect in recognition memory is apparently due to listeners' difficulty in separating recent presentations of high-frequency words from their previous presentations. Studies comparing low-frequency words with very low-frequency words have reported that low-frequency words actually show better recognition memory performance than very low-frequency items. This finding is analogous to the predictions for nonword patterns (see Chalmers & Humphreys, 1998; Guttentag & Carroll, 1997). In related work, Vitevitch and Luce (1998) presented data from a naming task that demonstrated a similar reversal of patterns for words and nonwords.

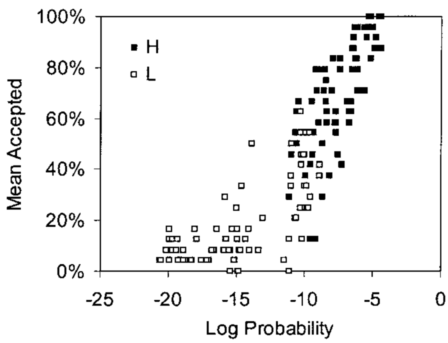

FIG. 6. Percentage acceptances for each nonword in Experiment 2 as a function of the log product probability.

They found that nonwords with high-probability segments and high transitional probability between pairs of segments were repeated more quickly than low-probability nonwords. For real words, however, those containing high-probability segments were repeated more slowly than those containing low-probability segments.

Predicting an effect of length on recognition memory for our nonword stimuli is more problematic. In spoken-word recognition, longer words are more easily identified than shorter words, presumably due to the redundant acoustic-phonetic cues for lexical access contained in the longer words (Miller, Wiener, & Stevens, 1946). An identification advantage for longer words has also been demonstrated for visually presented words (Zechmeister, 1972). Since longer nonwords contain additional cues for recognition and recall, they might be easier to discriminate in recognition memory than shorter nonwords. However, the rating data and the product probability measure indicate that the longer nonwords are less word-like. Presumably, they would invoke less activation of familiar lexical items during the study phase and this might put longer nonwords at a disadvantage in a recognition memory task.

Method

Participants. Thirty undergraduate students participated in exchange for credit in an introductory psychology course.

Stimuli. The nonword stimuli used were the same as in Experiment 1.

Procedure. The first part of this experiment, the study phase, was identical to the nonsense word-rating task used in Experiment 1, with the exception that each participant rated 72 items, or half of the complete stimulus set. The stimuli in each probability and length group were randomly divided into four quarters (a, b, c, d) that were combined into six different pairings for the rating task (ab, ac, ad, bc, bd, and cd). Thus, half of the participants rated each nonword and each participant rated half of the stimuli.

The ratings were collected in the same manner as Experiment 1, with one small difference. On each trial, the test item was presented twice for every participant before rating. The second presentation was provided to make the recognition memory task easier. We verified that this manipulation had no effect on the ratings task in a pilot study (henceforth called Experiment 3a) using 20 participants where the number of presentations (one or two) was manipulated as a between-subjects variable. There were no differences in the effect of constituent probability or length on the ratings as a function of number of repetitions.

The rating task served as the study phase for the recognition memory experiment. However, the participants were not informed that there would be a recognition test at the end of the rating task. Between the study phase and test phase, we gave participants a paper-and-pencil arithmetic activity with 20 simple addition and subtraction questions. For the recognition memory test, participants were asked to determine if each of the 144 stimuli was old or new. Responses were recorded using the same 7-button boxes, but with the label “Old” placed over the second button and “New” over the sixth button.

Results and Discussion

Ratings. The results of Experiments 1 and 2 were replicated. Nonwords with high-probability constituents were judged more word-like than nonwords with low-probability constituents [$F1(1, 174) = 149$, $p < .001$; $F2(1, 138) = 119$, $p < .001$]. Shorter nonwords were judged more word-like than longer nonwords [$F1(2, 174) = 51.9$, $p < .001$; $F2(2, 138) = 81.6$, $p < .001$]. There was no interaction of constituent probability and length. Mean rating and log product probability were highly correlated ($r = .87$; $p < .001$).

Recognition memory. Three measures of recognition memory performance were examined: d' and its two components, hit rate and false-alarm rate. Figure 7 displays the mean d' for stimuli in each constituent probability and length group. An ANOVA on mean d' revealed significantly higher d' for nonwords with high-probability constituents [$F1(1, 174) = 16.8$, $p < .001$; $F2(1, 138) = 13.4$, $p < .001$], but there was no effect of length in either the subject or item analysis. There were also no interactions. Separate analysis of the components of

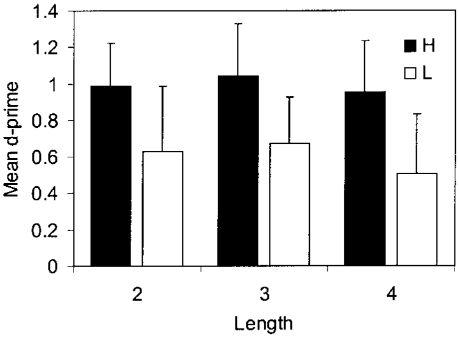

FIG. 7. Mean d' within each constituent probability group (High vs Low) and length in syllables (two, three, or four) in Experiment 3. Error bars show 99% confidence interval for the mean.

d' shows that the difference over probability groups can be attributed entirely to a difference in hit rate. Hit rate was higher for nonwords with high-probability constituents [$F1(1, 174) = 18.0$, $p < .001$; $F2(1, 138) = 16.0$, $p < .001$], but there was no effect of length [$F1(2, 174) < 1$; $F2(2, 138) < 1$]. There was no interaction. Analysis of the false-alarm rate showed no effects.

Our hypothesis concerning the effects of constituent probability on recognition memory performance was supported. Nonword patterns with high-probability constituents were recognized more accurately than nonword patterns with low-probability constituents. Unlike in the rating task, there was no length effect. We have two plausible accounts of the lack of a length effect. First, recall that Sendlmeier (1987) found that between-word similarity comparisons appeared to depend on a few salient constituents. In the recognition task, the comparison of a stimulus to the memory traces from the study phase may have involved analogous processes. Thus, only a few constituents may have been utilized regardless of the length of the nonword. Alternately, the lack of any length effect in our experiment may have been due to the particular nonword stimuli we used. Recall that a limited inventory of constituents was used to construct all of the stimuli, so the words in the stimulus set shared the same constituents in equal proportions. In addition, many of the possible combinations of onset and rime appeared as syllables in multiple nonwords, and groupings of syllables were similarly combinatorial. Thus, unlike in real words, there was very little redundancy in the groupings of different constituents in the nonwords, regardless of their length. These artifacts of the design process for constructing the nonwords may have reduced the advantage of additional cues for longer stimuli that is found with real words.

OTHER GRAMMARS

The consistently high correlation between wordlikeness judgments for nonwords and their expected probability in the log product probability grammar suggests that the stochastic grammar based on onset and rime constituents captures a significant portion of the information that participants use in making their judgments. However, there are a number of other possible stochastic grammars that are closely related to the grammar presented here. In addition, the influence of emergent lexical patterns could come directly from lexical exemplars, as in lexical neighborhood effects, rather than indirectly through a stochastic grammar. Finally, the stimuli might contain particularly salient good or bad components, and so predictors based more closely on traditional linguistic notions of constraint violation may account for the rating data equally well. To assess these alternatives, a series of correlations were carried out.

Methods

Correlations between mean wordlikeness or acceptability for each stimulus and a variety of predictors were generated for the rating data in Experiments 1, 2, 3, and 3a. The predictors were selected to explore the variety of approaches to wordlikeness discussed in the introduction.

Other stochastic grammars. The use of a stochastic grammar built from onset and rime constituents was inspired in part by Kessler and Treiman's (1997) analysis of the combinations of segments in CVC words in English. It is equally plausible, however, that combinations of larger units (syllables) or smaller units (segments) would provide a more accurate prediction of wordlikeness and acceptability judgments. Stochastic grammars based on syllable and segment probability (retaining the prosodic categories of stressed and unstressed and initial, medial, and final) were created from the same lexicon used for the onset–rime grammar. The log products of the syllable probabilities and segment probabilities were generated for each stimulus. These stochastic grammars share the property that all constituents contribute to the overall probability measure equally. Linear product probabilities were also evaluated as predictors of wordlikeness judgments for all three stochastic grammars and all of the other probability measures given below.

Violation grammars. Since all of the stimuli used in this study had alternating CV patterns and were created using constituents that appeared in at least one lexical item (and usually many more, even for the low-probability constituents), there were no truly impossible constituents or combinations of constituents in the stimulus set. However, some of our low-probability constituents appear in very few words in the lexicon and thus might be considered illegal. Coleman and Pierrehumbert (1997) found some influence of illegal clusters on acceptability judgments and they examined the probability of the least probable element as a predictor of judgments. Extending their analysis, the lowest probability segment, constituent, and syllable were considered as predictors. The probability of the lowest probability onset and lowest probability rime were also considered independently as predictors.

Between-word similarity. The size of a word's lexical neighborhood is an important factor in performance on a variety of psycholinguistic tasks (Luce & Pisoni, 1998). For a CVC word, a successful measure of the neighborhood of similar words is the number of words that share two of the three phonemes with the target. Extending this similarity metric, we considered the neighborhood of a multisyllabic nonword to be number of real words that shared 66% of the segments in the nonword. In determining whether a word was a neighbor to a nonword, the word and nonword were optimally aligned (by syllables) to maximize the count of shared segments. Shared segments between the word and nonword in each aligned onset and rime were counted. If the match was greater than or equal to 66%, the word was considered part of the similarity neighborhood of the nonword. For example, the stimulus /hɪ.tə.nət/ and the word *infinite* /ɪn.fə.nət/ share the vowel in the first two syllables and all three segments in the final syllable. The words are best aligned at their left edges. In this case, five of the seven segments in the nonword are found in matching positions in the word, so *infinite* is counted as a neighbor. The number of words in the nonwords' neighborhood was examined as a predictor of wordlikeness judgments.

The log product probability metric includes the probabilities of all constituents in the nonword. However, the word onset and the syllable with primary stress in a word are likely to be especially salient positions that might have the most influence over wordlikeness judgments (Pierrehumbert & Nair, 1995; Sendlmeier, 1987). Accordingly, the probability of the initial syllable, onset, rime, vowel, and primary stressed syllable, onset, rime, and vowel were also considered as separate predictors of wordlikeness.

Results and Discussion

Correlations for the best four predictors from each category are shown in Table 3. Correlations across the experiments are quite consistent. For all four sets of rating data, the log product onset and rime probability provided the best correlation with the participants' judgments. In almost all cases, the differences between the log product onset and rime grammar and the other predictors were significant using a t test for correlations on the same population (141 df). All correlations were significantly below the onset and rime grammar at $p < .01$ unless otherwise marked. Note that the significance test takes into account the correlation between the predictors, and so there is no particular r level above which all correlations are not different from the stochastic onset and rime grammar.

Overall, the best predictors of participants' judgments were the stochastic grammars. The logarithm of the number of real-word neighbors was also a good predictor, and in all but Exper-

TABLE 3

Correlation between Wordlikeness Judgments and Most Effective Predictors across All Experiments

Predictor	Expt. 1 rating	Expt. 2 accept	Expt. 3 rating	Expt. 3a rating
Stochastic grammar				
LogP onset–rime	.87	.86	.87	.87
LogP syllable	.84	.82	.83	.83*
LogP segment	.83	.81	.82	.83
P segment	.55	.48	.49	.57
Violation-based				
LogP worst syllable	.76	.77	.72	.73
LogP worst onset–rime	.75	.78	.72	.71
P worst segment	1 -muli low-prob-LogPtwo,E39.2(DT6-13.47839337E8913l32egmenuaob-7007d5se)(r3			

iment 3 it was not significantly worse than the log product of onset and rime probabilities. All of these predictors share the property that they consider the composition of the entire nonword, so both high- and low-probability constituents as well as length influence the predictor. They can therefore account for the wide range of ratings over this set of nonwords, which covers a large range of constituent probabilities and lengths.

Similarities and differences between the stochastic grammars and the neighborhood metric can be found when the data are examined in greater detail. When just the nonwords with high-probability constituents are considered, there are no significant differences between any of the three log product stochastic grammars and the log number of neighbors, with all correlations around .78. Over the high-constituent probability nonwords, the violation-based and other similarity-based predictors that consider just a single constituent perform much worse ($r < .38$). Conversely, over just the nonwords with low-probability constituents, the violation-based and similarity-based predictors do somewhat better, the best being log P worst syllable ($r \approx .48$). By contrast, the log number of neighbors does not fare well over the low-constituent probability stimuli ($r < .34$), as very few of these stimuli have any neighbors at all. The stochastic grammars still perform well for these stimuli ($r \approx .73$). Thus, overlap with the lexicon may be more important for high-constituent probability stimuli, while the presence of an extremely unlikely constituent may be more important for low-probability stimuli. The stochastic grammar is the best predictor, as it is able to capture the influence of both effects in a single, composite measure. The high-probability constituents occur in many lexical items, while the lowest probability constituents will tend to dominate the overall probability computation and therefore have the greatest effect on the product probability.

Separate correlations for each length group (two, three, and four syllables) once again reveal that the stochastic grammars provide a consistently good fit to the rating data ($r \approx .91$, .89, and .85, respectively). Analogous to the difference between nonwords with high-probability constituents and nonwords with low-probability constituents, the log number of neighbors performs well for the two-syllable nonwords ($r \approx .86$) but not for the four-syllable nonwords ($r \approx .56$), with the three-syllable nonwords in between ($r \approx .82$). With the length differences removed, the violation-based and other similarity-based predictors perform much better overall. This is especially the case for the longer nonwords. For four-syllable nonwords, log *P* worst syllable performs as well as the stochastic grammars ($r \approx .85$). *P* stressed rime and *P* initial syllable also perform well ($r \approx .80$ and $r \approx .76$ respectively). Thus, there is evidence that particularly salient constituents play an important role determining wordlikeness for some types of nonwords.

GENERAL DISCUSSION

Across all three experiments we found a strong relationship between expected phonotactic probability in a stochastic grammar and subjective judgments of wordlikeness. The pattern of responses for wordlikeness and acceptability judgments were quite similar, suggesting that listeners were performing the same task regardless of the response options. The Coleman and Pierrehumbert (1997) grammar incorporates prosodic features such as stress pattern, syllable structure, and syllable position into the probability of constituents. This grammar is a more detailed model of lexical patterns than is usually considered in computations of phonotactic probability or between-word similarity. We see their analysis of the lexicon as a useful experimental tool, both for the creation of novel nonword stimuli with specific properties and for the examination of how those properties may affect nonword perception and processing in a range of psycholinguistic tasks.

The results of Experiment 3, comparing these same nonword stimuli in a recognition memory task, provided further evidence that the perception of nonword stimuli makes use of knowledge about familiar word patterns. Nonwords with high-probability constituents were more easily recognized than nonwords with low-probability constituents. The fact that there was no effect of pattern length on recognition memory in this study may have been an artifact of our stimulus-construction method. It may also indicate that processes that are involved in the recognition memory task beyond those that are involved in metalinguistic judgments are sensitive to different aspects of wordlikeness. Alternately, it may be that recognition memory for nonwords involves different processes than recognition memory for words (see Vitevitch & Luce, 1998, for discussion).

In short, the present findings demonstrate that nonwords are processed against a background of probabilistic knowledge about the sound patterns of words stored in long-term memory. While this finding is incompatible with the traditional symbolic view of phonotactic constraints and grammatical competence, our results are compatible with the lexical representations used in a number of models of speech perception and production. For example, a distributed connectionist network model of the phonological lexicon would contain a segmental level that implicitly encodes probabilistic information about constituents through connections to the lexical items that contain those constituents (Plaut & Kello, 1999). These same probabilistic patterns could also be viewed as an emergent property of the collection of word tokens in an exemplar model of lexical memory (Goldinger, 1998).

Our results show that judgments of wordlikeness are heavily influenced by the relative frequency of occurrence of sounds and sound patterns in the lexicon. We also found that phonotactic patterns influence processing in other tasks such as recognition memory. Our findings are consistent with a growing body of research on probabilistic phonotactics in language processing (Aslin, Saffran, & Newport, 1998; Frisch, 1996, 2000; Jusczyk et al., 1994; Pitt & McQueen, 1998; Treiman et al., 2000; Vitevitch, Luce, Charles-Luce, & Kemmerer, 1997; Wurm & Samuel, 1997). This literature demonstrates that phonotactic information is accessed and used not only in psycholinguistic tasks employing familiar real words but also in processing novel nonwords. Thus, it appears that phonotactic knowledge is best viewed as an emergent property of the encoding and processing of lexical information that is an integral part of language use.

APPENDIX

Phonemic Transcriptions of Nonword Stimuli Used in the Experiments

Two-syllable (CˈV.CVC)		Three-syllable (CˈV.CV.CVC)		Four-syllable (CˌV.CV.CˈV.CVC)	
High	Low	High	Low	High	Low
midət	zo^{I}ʒəʃ	særɪʃəp	guyæðəʃ	sætətɛnən	gufɛʃɚgʊs
kinəp	yʊgo^{I}l	hesɚləm	vɔweθɛd	herɪsaʃəp	vɔyætʃo^{w}ðəʃ
henət	vɔθəʃ	kitɪdəp	yʊfæʒo^{I}l	kisɚtaləm	yʊweʃo^{w}θɛd
sæʃəm	guðʊs	mirɚnəm	zo^{I}wɛgɛd	mitɪsɛdəp	zo^{I}fætʃɚʒo^{I}l
sinən	vugʊs	hæsələp	yɔfeðo^{I}l	særɚtɛnəp	guwɛʃɚgo^{I}l
mæləp	gɔðəʃ	kɪtɚlən	zuyɛθʊs	hesəsanət	vɔfetʃo^{w}θəʃ
kɪnəm	zuθɛd	mætədət	gɔwæʒəʃ	kitɚtaʃəm	yʊyɛʃo^{w}θʊs
hæləm	yɔʒo^{I}l	sɪrənəp	vufɛgo^{I}l	mirəsɛdət	zo^{I}wætʃɚʒɛd
hɪdəp	go^{I}gɛd	kerɪnət	yo^{I}yæθəʃ	hætətɛnəm	yɔfɛʃɚgɛd
sænət	vʊðo^{I}l	mesɚʃəm	zʊweðʊs	kɪrɪsaləp	zuyætʃo^{w}ðo^{I}l
meləp	zʊθʊs	sitɪdən	go^{I}fæʒɛd	mæsɚtalən	gɔweʃo^{w}θʊs
kɪdət	yo^{I}ʒəʃ	hɪrɚnən	vʊwɛgʊs	sɪtɪsɛdən	vufætʃɚʒəʃ
sɪdət	guʒɛd	sæsənəm	gufegɛd	hærɚtɛdən	yɔwɛʃɚʒɛd
heʃəm	vɔðʊs	hetɚdəp	vɔyɛʒo^{I}l	kɪsəsaʃəm	zufetʃo^{w}ðʊs
kelən	yʊθəʃ	kirəʃəp	yʊwæθɛd	mætɚtanət	gɔyɛʃo^{w}θəʃ
minəp	zo^{I}go^{I}l	mitələm	zo^{I}fɛðəʃ	sɪrəsɛnəp	vuwætʃɚgo^{I}l
hɪʃəp	yɔʒɛd	hærɪnəp	yɔyægo^{I}l	ketətɛdət	yo^{I}fɛʃɚʒəʃ
kidən	zʊθʊs	kɪsɚdət	zuweʒəʃ	merɪsalən	zʊyætʃo^{w}θʊs
mænən	gɔðo^{I}l	mætɪləp	gɔfæθʊs	sisɚtaləp	guweʃo^{w}ðo^{I}l
sɪʃəp	vugɛd	sɪrɚlən	vuwɛðo^{I}l	hɪtɪsɛnəm	vɔfætʃɚgɛd
keləm	yo^{I}ʒo^{I}l	kesənən	yo^{I}fegʊs	kerɚtɛdəp	yo^{I}wɛʃɚʒo^{I}l
menəm	zuθɛd	metɚdən	zʊyɛʒɛd	mesəsaləm	zʊfetʃo^{w}θɛd
sɪləp	vʊðəʃ	sirəʃəm	vʊwæðʊs	sitɚtaʃəp	guyɛʃo^{w}ðəʃ
hælən	go^{I}gʊs	hɪtənət	go^{I}fɛθəʃ	hɪrəsɛnən	vɔwætʃɚgʊs

REFERENCES

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science,* **9,** 321–324.

Brown, R. W., & Hildum, D. C. (1956). Expectancy and the perception of syllables. *Language,* **32,** 411–419.

Bybee, J. L. (1988). Morphology as lexical organization. In M. Hammond & M. Noonan (Eds.), *Theoretical morphology* (pp. 119–141). San Diego: Academic Press.

Chalmers, K. A., & Humphreys, M. S. (1998). Role of generalized and episode specific memories in the word frequency effect in recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* **24,** 610–632.

Chalmers, K. A., Humphreys, M. S., & Dennis, S. (1997). A naturalistic study of the word frequency effect in episodic recognition. *Memory & Cognition,* **25,** 780–784.

Coleman, J., & Pierrehumbert, J. B. (1997). Stochastic phonological grammars and acceptability. In *Computational phonology: Third meeting of the ACL special interest group in computational phonology* (pp. 49–56). Somerset, NJ: Association for Computational Linguistics.

Dollaghan, C. (1985). Child meets word: "Fast mapping" in preschool children. *Journal of Speech and Hearing Research,* **28,** 449–454.

Forster, K. I. (1978). Assessing the mental lexicon. In E. Walker (Ed.), *Explorations in the biology of language* (pp. 139–174). Montgomery, VT: Bradford.

Frauenfelder, U. H., Baayen, R. H., Hellwig, F. M., & Schreuder, R. (1993). Neighborhood density and frequency across languages and modalities. *Journal of Memory and Language,* **32,** 781–804.

Frisch, S. A. (1996). *Similarity and frequency in phonology.* Unpublished doctoral dissertation, Northwestern University.

Frisch, S. A. (2000). Temporally organized representations as phonological units. In M. Broe & J. B. Pierrehumbert (Eds.), *Papers in laboratory phonology V: Language acquisition and the lexicon.* (pp. 283–298). Cambridge: Cambridge Univ. Press.

Gathercole, S. E. (1995). Is nonword repetition a test of phonological memory or long-term knowledge? It all

depends on the nonwords. *Memory and Cognition,* **23,** 83–94.
Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review,* **105,** 251–279.
Greenberg, J. H., & Jenkins, J. J. (1964). Studies in the psychological correlates of the sound system of American English. *Word,* **20,** 157–177.
Guttentag, R. E., & Carroll, D. (1997). Recollection-based recognition: Word frequency effects. *Journal of Memory and Language,* **37,** 502–516.
Hawkins, P. (1979). *Introducing phonology*. London: Hutchinson.
Hulme, C., Roodenrys, S., Brown, G., & Mercer, R. (1995). The role of long-term memory mechanisms in memory span. *British Journal of Psychology,* **86,** 527–536.
Jusczyk, P. W., Luce, P. A., & Charles-Luce, J. (1994). Infants' sensitivity to phonotactic patterns in the native language. *Journal of Memory and Language,* **33,** 630–645.
Kessler, B., & Treiman, R. (1997). Syllable structure and the distribution of phonemes in English syllables. *Journal of Memory and Language,* **37,** 295–311.
Landauer, T. K., & Streeter, L. A. (1973). Structural differences between common and rare words: Failure of equivalence assumptions for theories of word recognition. *Journal of Verbal Learning and Verbal Behavior,* **12,** 119–131.
Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing,* **19,** 1–36.
Marslen-Wilson, W. D. (1984). Function and process in spoken word recognition. A tutorial review. In H. Bouma & D. G. Bouwhis (Eds.), *Attention and performance X: Control of language processes* (pp. 125–150). Hillsdale, NJ: Erlbaum.
Miller, G. A., Wiener, F. M., & Stevens, S. S. (1946). *Transmission and reception of sounds under combat conditions* (Summary Technical Report of D1 T6.5nA.,n1Tf 1 D (20,)Tj4ac034Tmodel.