Running head: SYLLABLE STRUCTURE AND PHONEME DISTRIBUTION

Syllable Structure and the Distribution of Phonemes in English Syllables

Brett Kessler — Stanford University

Rebecca Treiman — Wayne State University

Correspondence concerning this article should be addressed to:

Brett Kessler
Department of Linguistics
Stanford University
Stanford CA 94305-2150

Electronic mail may be sent to the Internet address kessler@csli.stanford.edu.

Word count: 8,302 (incl. Refs, excl. Abstract and Tables)

Abstract

In describing the phonotactics (patterning of phonemes) of English syllables, linguists have focused on absolute restrictions concerning which phonemes may occupy which slots of the syllable. To determine whether probabilistic patterns also exist, we analyzed the distributions of phonemes in a reasonably comprehensive list of uninflected English CVC (consonant-vowel-consonant) words, some 2,001 words in all. The results showed that there is a significant connection between the vowel and the following consonant (coda), with certain vowel-coda combinations being more frequent than expected by chance. In contrast, we did not find significant associations between the initial consonant (onset) and the vowel. These findings support the idea that English CVC syllables are composed of an onset and a vowel-coda rime. Implications for lexical processing are discussed.

Syllable Structure and the Distribution of Phonemes in English Syllables

Linguists have often observed absolute restrictions in the patterning of phonemes in syllables. For example, it is often noted that /h/ can occur only at the start of an English syllable and that /ŋ/ can occur only at the end. (See Tables 1 and 2 for an explanation of the phonetic symbols.) By the same token, certain combinations of phonemes occur in the language, whereas others do not. In the General American English accent, /ɑ/ can occur before /r/ at the end of a syllable (car), but /æ/ cannot, and this rule has no exceptions. In some languages, there are many more such gaps or restrictions at the end of the syllable than at the beginning, and such asymmetry has been accepted as evidence that the syllable has a particular kind of internal structure (Goldsmith, 1990, p. 123–127). In English, however, it is not so obvious that the end of the syllable has many more restrictions than the beginning. As a result, there has been some debate as to whether there is enough imbalance in phonotactic constraints to suggest internal syllable structure.

This impasse can be broken, in our view, by abandoning the idea that only absolute, inviolable restrictions are worthy of note. We take a probabilistic approach in the present study by asking whether some consonants occur in certain positions of the syllable more or less often than expected by chance, and whether some legal combinations occur more or less often than expected by chance. For example, among the words we consider, we find that the sequence /ʌf/ occurs much more often than one would expect from the frequency of /ʌ/ and /f/ considered independently, and /æl/ occurs much less often than expected. Even though there are several words that end in /æl/, their unusually low frequency suggests a phonotactic restriction of a nonabsolute, violable kind. Broadening our purview to include restrictions of this kind gives a clearer picture of the phonotactic patterns of English. In the present study, we use statistical techniques to examine the patterns of phoneme co-occurrences in the CVC words of English—single morphemes that consist of a vowel flanked by exactly one consonant on each side.

Relevance of Syllable Phonotactics

A statistical study of syllabic phonotactics is motivated by several considerations. On one level, we are interested in exploring the validity of quantitative approaches to language in general. Some important schools of linguistics hold that language is best described in terms of symbols and categories, and that quantitative tendencies that cannot be reduced to a system of categorical rules are merely accidents that are irrelevant to language as a systematic entity. Davis (1985), for example, rejected all arguments for internal syllable structure, on the grounds that he was able to find exceptions, however rare, to all properties that had been proposed for specific subsyllabic entities. But many other language researchers find quantitative approaches, such as connectionist models, to be very useful. One of our motivations is to show that a quantitative approach to English phonotactics can uncover the same types of patterns that have been noted as absolute rules in other languages. Researchers interested in universal properties of language might take this as evidence that statistical patterns are not necessarily accidental and deserve a closer look.

A second motivation for our study is to contribute toward research in lexical processing. Recognition and production of a word can be affected by how many other words are similar to it in pronunciation or spelling (e.g., Goldinger, Luce, & Pisoni, 1989; Grainger, 1992). The magnitude of such effects is associated with which end of the word one is considering. For example, the beginning of a word is the most salient part for identification (Cutler, 1982, p. 19). Also, word production is slower and more erratic when there are other words with a similar beginning; interference is not so pronounced when there are other words that have a similar ending (Sevald & Dell, 1994). If we were to find that phonotactic patterns in English show a tendency to constrain combinations of elements at the end of the syllable, thus contributing toward making words more similar at their ends than at their beginnings, then researchers will want to take such facts into account when exploring the relation between processing asymmetries and the structure of the English vocabulary.

Our final reason for studying syllable phonotactics is the evidence it may bring to bear on syllable structure. Linguists have adduced every possible configuration for the internal structure of syllables. For CVC syllables, the main concern is whether the vowel is grouped with the prior consonant (called the onset), with the posterior consonant (the coda), or with neither. Figure 1 illustrates those three basic theories. The leftmost tree illustrates the theory of the flat syllable, where the vowel groups with neither the onset nor the coda (Clements & Keyser, 1983; Davis, 1985; Hockett, 1955). The second tree illustrates the onset–rime theory, where the vowel groups with the coda to form a constituent called the rime (Fudge, 1969; Goldsmith, 1990; Kuryłowicz, 1973; Selkirk, 1982). The last tree illustrates the theory of body–coda organization, where the vowel is grouped with the onset to form a constituent called the body (McCarthy, 1979, p. 455; Iverson & Wheeler, 1989). More recently, some phonologists have claimed that the components of the syllable are units of weight called moras (Hayes, 1989; Hyman, 1985). As the trees in Figure 2 illustrate, basic moric theory always has the vowel in the first mora and the coda in the second, with the complication that a long vowel is considered to be simultaneously in both moras.

***** PLACE FIGURE 1 ABOUT HERE ****

***** PLACE FIGURE 2 ABOUT HERE ****

Of all these theories, the onset–rime is perhaps the most widely accepted, in our opinion rightly so. Many linguistic phenomena are easily described in terms of properties of the vowel and the coda (i.e., the rime), with the onset consonant being irrelevant. These phenomena include verse metrics, word stress, and compensatory lengthening (see Halle & Vergnaud, 1982; Kuryłowicz, 1973; and Hayes, 1989 for descriptions). It is much harder to find important processes that depend on the body to the exclusion of the coda. Rhyming traditions, language games such as Pig Latin, and speech errors often treat rimes as units but rarely treat bodies as units (e.g., MacKay, 1973; Stemberger, 1983). Furthermore, the bulk of experimental evidence favors the onset–rime theory (see Treiman & Kessler, 1995). On the other hand, Davis (1985) argued for the flat syllable; experiments with speakers of Korean suggest a body–coda organization (Derwing, Yoon, & Cho, 1993; Yoon, 1995); and Pierrehumbert and Nair (1995) claimed that moric theory can account for much of the experimental evidence. The question of internal syllable structure is not yet settled.

We believe that a statistical study of syllable phonotactics can bring light to the issue of syllable structure. Although the theoretical concept of linguistic structure is hard to pin down, many would agree that a structure is the natural domain for a constraint or process. If, for example, different types of consonants may appear before the vowel than after it, then that suggests that those are not undifferentiated consonant slots, but rather that those two elements belong to different structures. We explore this issue in Study 1. Of course, most theories put the onset and coda in different structures, but such information would argue against some versions of flat syllable theory (e.g., Clements & Keyser, 1983). Another approach is to test whether proposed constituent structures like the head or rime constitute a natural domain for phonotactic constraints. If there are many constraints against combining certain vowels with certain codas (such as the aforementioned illegal combination */ær/) but few if any against combining vowels with onsets, then that suggests that the vowel and coda form a structure, the rime. We investigate this matter in Study 2.

The idea of considering phonotactic constraints as evidence for syllable structure is not new. Fudge (1969) argued for an onset-rime structure, claiming that most or all phonotactic constraints in English involve the vowel and coda. But Clements and Keyser (1983, p. 20) favored the flat syllable, stating that "cooccurrence restrictions holding between the nucleus [i.e., vowel] and preceding elements of the syllable appear to be just as common as cooccurrence restrictions holding between the nucleus and following elements." Fudge (1987) responded with extensive tables that showed that vowel-coda pairs in English have many more gaps than do onset-vowel pairs. Part of the reason that the matter remains unsettled is statistical. The problem is twofold, involving both false zeroes and false positives. Some phonemes are fairly uncommon in English, and the number of morphemes is finite, so some possible combinations may fail to exist just because they do not have a reasonable chance to occur. A count of zero co-occurrences does not mean there is a principled constraint against a sequence. On the other hand, finding a few co-occurrences does not mean that the phonemes combine freely. Some phonemes may be so common that one would expect them to appear together dozens or hundreds of times. Previous researchers were by no means naïve on these points: Clements and Keyser admitted that their claim that voiced fricatives do not appear before /ʊ/ may be accidental, and they counted /vu/ as a circumscribed sequence even though they knew about the words voodoo and rendezvous. But statistical tests have rarely been applied. A notable exception is Randolph (1989), who used the likelihood ratio statistic to test the significance of collocational constraints within the syllable, all of which he rejected. But the statistics he reported were so remarkably low that one suspects that scaling factors threw them off by several orders of magnitude.

In this study, we explored co-occurrence patterns by examining a reasonably comprehensive list of uninflected English CVC words. We readily admit that a study of CVC words will not answer all questions about syllable structure. Some of the patterns we uncover may reflect properties of word edges rather than syllable edges; indeed, some phonologists declare that some or all word-final consonants should not even be considered part of a syllable (Kenstowicz, 1994, p. 260–261). Some patterns may have more to do with whether a consonant is prevocalic or postvocalic, or whether it precedes or follows sentence stress, than with whether it is in the onset or coda of a syllable. Despite these warnings, we believe that CVC words are an ideal place to begin study. The word list is very homogeneous: All single-coda consonants are paired with a single onset consonant, and there are no confounding factors such as extra phonemes or difference in stress or morphemic composition. Nor is there any doubt as to which syllable a consonant belongs to, which is a matter of no small controversy for English intervocalic consonants (Lass, 1984, p. 262–268). Our use of CVC words should not only make the statistical analyses more straightforward and interpretable, but should also facilitate follow-up studies. If in the future different patterns are found in other carefully constructed databases, the homogeneous nature of our database should help researchers formulate hypotheses as to what factors in the word lists account for those differences.

Study 1

In Study 1, we ask whether there are differences in the frequency of occurrence of the different consonants depending on whether they are in the onset or the coda.

Method

We analyzed the 2,001 monomorphemic CVC words found in the unabridged Random House Dictionary (Flexner, 1987). We were fervid in our zeal to eradicate polymorphemic words: A word was rejected if any part of it is used in the same sense in some other word, so that even words like this and then were omitted on the grounds that th may be a demonstrative morpheme. We omitted all words which the dictionary gave any reason to believe were not in current general use throughout America. Thus we omitted words with foreign phonemes or accented letters, foreign measures, and place and ethnic names that were not obviously Anglicized. We did include given names such as *Dave*.

We used the first pronunciation listed in the dictionary. The dictionary phonemes were transcribed on a unit-by-unit basis, except that /ɝ/, which is treated as the sequence *ûr* by Random House (Flexner, 1987), was here treated as a vowel. As a result, words like *bird* were included in our list of CVCs. We did this because /ɝ/ is a phonetically unitary sound and we wished to avoid any bias from prejudging its underlying properties. For the other vowels before /r/, which are variously treated in different accents, we followed the usage of the Dictionary in recognizing diphthongs before /r/, as well as the vowels /i/ (beer), /ʊ/ (boor), /ɛ/ (bear), /ɔ/ (bore), and /ɑ/ (bar). Tables 1 and 2 list the phonemes recognized in this study, and the phonetic classes to which they belong. Note that we treated diphthongs and affricates as units.

The Random House scheme distinguishes /w/ as in wine from /ʍ/ as in whine. It also draws a distinction between the vowels /ɒ/ cot, /ɔ/ caught, and /ɑ/ khat, spa. Although these are more distinctions than are commonly made in America, we observe the full set of contrasts because they are dialectically neutral: All those vowel distinctions are made in parts of New England and throughout England. By the same token, we count /ɔ/ as a mid round vowel even though for most Americans it is low and perhaps even unrounded.

***** PLACE TABLE 1 ABOUT HERE ****

***** PLACE TABLE 2 ABOUT HERE ****

Table 2 lists the number of times each consonant occurs in the word list. Only word types were considered, unweighted by their frequency. A word type may contain two occurrences of consonants: Thus bib contributes 2 toward the count of /b/. To determine whether the frequencies are affected by syllable position, we performed for each consonant type separate two-cell goodness-of-fit tests with Pearson's χ^2, computing the expected frequencies under the null hypothesis that consonants would be evenly distributed between onset and coda. Because all words had exactly one onset and one coda consonant, this means that each consonant should occur half the time in an onset, and half the time in a coda. To correct for the fact that the size of the χ^2 statistic depends in part on the total number of times each consonant occurs, we also computed the ϕ coefficient of association for the contingency tables. This statistic includes the total number of consonant tokens as a divisor, and so scales from 0 to 1. In order to determine whether there is an overall association between consonant type and syllable slot across the consonantal system as a whole, we also computed the χ^2 statistic across all 25 consonant types. Finally, we performed a χ^2 decomposition by phonetic feature class to help understand effects intermediate between those of the entire table and of individual phonemes. Here we used $\underline{G}^2$, the likelihood-ratio version of χ^2, because it is additive across decompositions. To guard against the danger of finding significant results simply because we made huge numbers of comparisons, we restricted ourselves throughout this study to only making comparisons between feature sets that are immediate children of the same node in the trees presented in Figures 3 and 4. In Study 1, we only looked at the place of articulation, contrasting two classes of phonemes only when they were immediate children of the same node in the topmost tree of Figure 3. For example, anterior coronal consonants as a group were only compared against nonanterior coronal consonants, but not against labial consonants.

***** PLACE FIGURE 3 ABOUT HERE ****

Results

Table 3 shows how often each consonant occurs in onset and coda. The table is arranged by the strength of association between consonant type and syllable position (ϕ). As one can see by the number of starred χ^2 statistics, most of the consonants appear in either the onset or the coda more often than one would expect. A χ^2 test for the consonantal system as a whole gives a significant result ($\chi^2(24) = 496.52$, $p < .05$). The strength of the association, as measured by Cramér's coefficient, is .35.

***** PLACE TABLE 3 ABOUT HERE ****

In the G^2 decomposition, the total G^2 was 601.26. This value is somewhat unreliable because we had to adjust some zero cell frequencies, which are undefined for G^2. Our first step, therefore, was to partition off any consonants that had structural zeroes, including the glides /j/ and /w/, which cannot occur in codas because we define them to be parts of diphthongs (the vowel slot), and /ʒ/ and /ð/, which arguably occur in the onset in only special circumstances (loan words and function words, respectively). Effects among those consonants, or between them as a group and the other consonants as a group, were each significant, and accounted for two thirds of the deviation ($\underline{G}^2$ of 408.83). But when we decomposed the remaining $\underline{G}^2$ by place of articulation according to the scheme of Figure 3, we found several other significant deviations having nothing to do with structural zeroes; these are summarized in Table 4. The table shows, for example, that coronal consonants prefer the coda significantly more than noncoronals. Among coronals, anterior consonants have a more marked preference for the coda than do nonanterior coronals. In contrast to Table 3, this table shows the direction of the skew within the contrast, not the absolute direction of the preference. Thus /d/ is listed as favoring the onset more than other anterior coronals, even though in absolute numbers (i.e., when compared to <u>all</u> other consonants) it is found in the coda somewhat more than in the onset.

***** PLACE TABLE 4 ABOUT HERE ****

<u>Discussion</u>

Our results show that there is an association between consonant type and syllable position. In the phonemic analysis used by our source dictionary, glides (/h/, /j/, /w/, /ʍ/) can only occur in the onset and /ŋ/ can only occur in the coda. This much is common knowledge. What is not commonly recognized is the skew that is present for several other consonants as well. In particular, /z/, /θ/, /n/, /t/, /l/ and /k/ show a significant preference for the coda, and /b/, /ǰ/, /ʃ/, and /r/ show a preference for the onset. The fact that consonants as a group prefer particular syllable slots is strong enough to be significant even if one factors out the consonants for which there is an absolute, inviolable restriction as to which slot they can go in. About half of the $\underline{G}^2$ among the remaining consonants can be accounted for by significant differences between contrasting places of articulation, with the strongest effect being the contrast between anterior and nonanterior coronal consonants.

Because the words in our study are all monomorphemic, the prevalence of /z/, /θ/, /n/, and /t/ in codas cannot be explained by their use as inflectional or derivational endings. The fact that coronals, especially anterior coronals, appear disproportionately often in codas echoes absolute constraints found in other languages. When languages restrict codas or word endings to consonants of a particular place of articulation, anterior coronals are the least likely to be excluded. The core, native vocabulary of Spanish, for example, has many words ending in the anterior coronals /ð/, /θ/, /s/, /l/, /n/, and /r/, but almost no words ending in other consonants, even though such consonants are frequent at the beginning of words.

Study 2

As we pointed out earlier, one test of structural constituency is whether the items within a proposed constituent are more strongly associated with each other than they are with items outside of the structure. Put another way, items within a constituent should vary less freely with respect to each other than they do with respect to other items. Study 2 was designed to determine whether the vowel and the coda are more strongly associated with each other than are the vowel and the onset. Such a finding would suggest that the vowel and the coda form a constituent, the rime. In addition, the existence of a rime constituent would allow us to treat the distributional patterns seen in Study 1 as properties of that constituent. That approach would simplify our account by obviating the need to hypothesize that the onset or the coda, or both, are distinguished constituents in their own right.

Method

The most straightforward way of investigating association patterns between the three syllable slots would be to consider the data as a three-variable contingency table, with the levels of each variable being the different phoneme types. A first logical step would be to use log-linear models to test whether the phonemes in each slot are associated at all with the phonemes in the other slots. Unfortunately, a three-dimensional contingency table for 24 onset consonant types $\times$ 15 vowel types $\times$ 21 coda consonant types would require 7,560 cells to be populated by 2,001 observations. That would leave an undesirable number of empty cells, to say the least. Performing three-variable contingency table analyses on these data requires a collapsing into broader categories. There is, however, no obvious optimal way to group consonants or vowels into a small number of classes. We decided to classify the consonants in three different ways (by place, manner, and voice; see Table 2 and Figure 3) and the vowels in three different ways (by height, backness, and tenseness; see Table 1 and Figure 4), choosing categories a priori according to our informal judgments as to what categories tend to play the biggest role in phonological processes. For CVC words, this results in 27 different comparisons, for which we compensate by changing the critical significance level p from .05 to .001.

***** PLACE FIGURE 4 ABOUT HERE ****

When looking at the three-way interactions, we first asked which associations are significant at the $p < .001$ level, by looking at the standard hierarchical log-linear models. That is, we first tested for complete independence of all three slots; at the next higher level, we tested for each slot the hypothesis that it is independent of the other two slots; at the next level, we tested for each pair of slots the hypothesis that they are conditionally independent given the value of the third slot; and finally, we tested for homogeneous association. When more than one model adequately fit the data (that is, the G^2 was lower than the significance level for .001), we selected the model at the lowest level, unless a higher-level (i.e., more complicated) model provided a significantly better fit (i.e., the difference in G^2 exceeded the .001 significance level at the differential degrees of freedom).

Of course, the mere fact of association does not say anything about the degree of association. One way to assess the strength of association is to look at the percentage of total G^2 that a pair association can account for by itself. For each pair of slots, we computed the difference between the G^2 that obtains under a model that assumes that all three pairs of slots are independent, and one that assumes only that the other two pairs of slots are independent. Because we are comparing the triplets along many different features, each of which has different patterns of variation, we standardized this difference as a percentage of total G^2. Of primary interest are the percentages for onset–vowel and vowel–coda pairs. We computed the ratio of the vowel–coda percentage to the onset–vowel percentage and averaged that across all the feature systems for which statistically significant pairwise associations were found.

It could be that any high association that turned up in in vowel–coda pairs is due to the absolute co-occurrence restrictions that are already well known (e.g., the ban against /ær/). In order to confirm that nonabsolute restrictions point in the same direction, we ran the above tests again, this time omitting all words that end in /r/ or /ŋ/. These are the two sounds for which all phonologists agree that there are absolute co-occurrence restrictions for preceding vowels, in that tense and lax vowels do not contrast in that environment. If the pattern of associations still holds when those 170 words are excluded, then our results cannot be explained by those absolute co-occurrence restrictions.

Although the foregoing is sufficient to test our hypothesis that vowel–coda association is much more important than onset–vowel association, we also want to describe where the interactions lie. For example, if we were to find that vowel tenseness interacts with coda voicing, it would be useful to know whether it is the tense or lax vowels that are associated with voiced codas. When one of the syllable slots varies independently of the other two, this question can be pursued with some rigor by ignoring that slot and doing a $\underline{G}^2$ decomposition on the remaining two-way table. When doing so, we applied Bonferroni corrections for 9 comparisons (3 ways of collapsing each of the two slots by feature) so as to maintain a familywise significance level of .05. When none of the slots was independent, we proceeded less rigorously. With the understanding that we may be missing some indirect interactions, we decomposed two-way tables for each of the associated pairs of slots, so as to get a feel for each of the pairwise interactions. We then informally compared that to the three-way table, noting any obvious interactions. This part of the analysis is exploratory, intended to provide hypotheses that later can be tested more rigorously over a larger body of data.

Finally, we performed some two-dimensional phoneme-level analyses over each pair of slots taken independently. Although this introduces the danger of misinterpreting indirect associations, it is useful to look at a less abstract level of analysis (phonemes rather than features) to avoid losing sight of the broader picture. To take a theoretical example, results that generalize over labiodentals in general may obscure the fact that English has only two labiodental consonant types, /f/ and /v/, and that indeed they may even behave differently. So for each pair of slots, we performed a goodness-of-fit χ^2 on each pair of phonemes. For example, for the onset–vowel pair /lu/, we compared the number of times those two sounds occurred together (15) with the marginal totals for onset /l/ (135) and vowel /u/ (117). There being 2,001 words, one would expect them to occur together 7.89 times. Whenever the expected frequency was 5 or higher, we tested whether the difference was significant, using Pearson's χ^2 to compute a 2×2 goodness-of-fit statistic. We report which pairs occur significantly more or less often than expected by chance, and how the number of significant pairs compares to the count we would expect to see by chance at our significance level ($\underline{p} < .05$). We also ask whether there is a significant association between each pair of slots, computed over all phonemes that appear in those slots. The usual approach would be to perform a χ^2 test over a large table of which one dimension would be, for example, all possible vowel phonemes, and the other dimension all possible coda consonants. Unfortunately, the tables for any of the three pairs of slots would be too sparse for one to put much faith in a standard χ^2 test: Vowel–coda pairs have 44% of their cells with expectation lower than 5, many of which are even lower than 1. However, one can overcome these limitations by calculating exact probabilities using a multilevel extension of Fisher's exact test, determining the distribution by random sampling of many permutations of the tables. We did so over 50,000 samples, and report the Fisher statistic and the Monte Carlo estimate of the $\underline{p}$ value. In each case, we looked for probabilities of .05 or less. We also calculated Cramér's association statistic for these tables.

We also ran all of the above tests on a subset of the dictionary that only contained words having lax vowels. Comparing those two sets of results can help to show which vowel features are responsible for what association. The lax vowel subset is also of interest because virtually all linguists agree that a word-final consonant after a short, stressed vowel is part of the syllable, rather than extrasyllabic.

Results

Table 5 shows the results of the log-linear analysis of the three syllable slots, one analysis for each of three ways of collapsing each of the three slots. The models presented in the table can be read as saying that a significant association is found between two slots when they are enclosed within the same set of square brackets. In all cases, the hypothesis of three-way association is rejected. Because of the low $\underline{p}$ threshold used for the individual tests, the finding of an association in any one comparison is sufficient for concluding the association in CVC words as a whole at $\underline{p} < .05$. Thus the results as a whole are [OC][VC], i.e., there are associations between onset and coda and between vowel and coda, but there is insufficient evidence for assuming any direct association between onset and vowel. The patterns are the same whether we look at the entire set of words or whether we omit those that end in /r/ or /ŋ/. On the other hand, as the rows labeled Lax show, the results change in particular cases if all the words with tense vowels and diphthongs are omitted. But even then, the overall pattern is [OC][VC].

***** PLACE TABLE 5 ABOUT HERE ****

When we looked at the strength of the associations, we found that the percentage of G^2 explained by vowel–coda associations was always higher than that explained by onset–vowel associations: The ratio of the first over the second averaged 22.7 (range 1.3 to 104.8) over the 21 feature comparisons for which the null hypothesis of no pairwise association had to be rejected. When we omitted from consideration all words that end in /r/ or /ŋ/, the average ratio of vowel–coda strength over onset–vowel was almost the same, 25.2 (range 1.6 to 119.4). The average ratio was 12.9 (range 1.9 to 54.4) when only words with lax vowels were considered.

Having verified our hypothesis that the vowel associates with the coda much more strongly than with the onset, we turn now to the more exploratory issues. One question concerns the direction of the associations in each comparison. Such directionality is easier to visualize in two dimensions than in three. Fortunately, because most comparisons show an interaction of only one pair of slots, we can in most cases collapse the tables down to a comparison between vowel and coda on the one hand, and onset and coda on the other. Whenever the log-linear analysis (Table 5) shows that a pairwise association between two syllable slots is significant when collapsing along a particular pair of features, we did a two-way $\underline{G}^2$ analysis over those variables, and report the significant subtables obtained by a $\underline{G}^2$ decomposition. For example, Table 5 shows that onset and coda are associated when the phonemes in both slots are collapsed by place of articulation. Table 6 shows the subtables that are significant for either the entire word set or for the words with lax vowels, when a $\underline{G}^2$ decomposition is performed for onset $\times$ coda, collapsing both sets of phonemes by place. Table 7 gives analogous information for vowel–coda associations.

***** PLACE TABLE 6 ABOUT HERE ****

***** PLACE TABLE 7 ABOUT HERE ****

Caution is needed when collapsing consonants by place, because there are both onset–coda and vowel–coda effects. However, informal inspection of the data suggests that the interactions are not complicated. The associations between vowel height and coda place do not conflict with those between onset place and coda place, and indeed the patterns of association shown in Table 7 hold regardless of the place of the onset consonant. On the other hand, associations between onset place and coda place win out over the associations between vowel tenseness and coda place. When the onset is labial, lax vowels lose their antipathy for coronals, with the dispreference for labial codas showing up much more strongly. And when the onset is postcoronal, lax vowels reverse their normal pattern, preferring labial codas to velar ones.

Finally, we report the results of our phoneme-level analyses over each pair of syllable slots. Table 8 lists the pairs that occur significantly more or less often than expected, based on the frequencies of the individual phonemes and the assumption of no association. The percentage in the last column tells what percent of the testable pairs (i.e., those with expected frequencies of 5 or higher) the given list constitutes. That figure can be compared to the approximate figure of 5% that would be expected by chance at a significance level of .05 even if there is no association between slots.

***** PLACE TABLE 8 ABOUT HERE ****

Finally, we looked at each pair of slots as a whole, and asked whether, for example, vowels and codas as a set are significantly associated at the phoneme level. The onset–vowel association was not significant (Fisher statistic 366.1, with a Monte Carlo estimate of the p value of .13), but the vowel–coda association was easily significant (Fisher statistic 855.9, with an estimated $p < .0001$). The onset–coda association was also significant (Fisher 521.9, $p = .0067$). Scaling by table size to find the associated Cramér association statistic, onset–vowel pairings and onset–coda pairings each have a value .11, but the vowel–coda statistic is .18. The comparable figures for the lax-voweled subset is .16 for onset–vowel pairs, .25 for vowel–coda pairs, and .14 for onset–coda pairs.

Discussion

The results of Study 2 demonstrate that there is a stronger connection between vowel and coda than between vowel and onset. Indeed, none of our tests showed significance for onset–vowel phoneme-level association. In contrast, every approach we took to estimating the association between vowel and coda proved significant. Similar results obtain when we compare various indices of the strength of the two associations, showing that the effect is real. It is also encouraging that the same results hold if we omit words with final /r/ and /ŋ/, the major source of absolute co-occurrence restrictions, and that the basic relationships remain the same if we look only at words with lax vowels. Assuming that co-occurrence restrictions are functions of local structure, these results strongly suggest that vowel and coda form a constituent, the rime.

In addition, the more exploratory phase of our investigation revealed several interesting co-occurrence patterns worthy of being investigated in future studies with larger word sets. Among onset–vowel phoneme pairs, for example, the fact that /ki/ is less frequent than expected but /si/ is more frequent than expected may prove to be a real case of onset–vowel association. This statistical pattern has been noted for other languages (Maddieson & Precoda, 1992), and in general the replacement of velars by coronals before front vowels (palatalization) is absolute in many languages, including Old French, a major source of English vocabulary. Such meaningful patterning cautions us against concluding there is no association at all between onset and vowel, even though that association may not be significant at our critical level, and even though it is manifestly much weaker than the vowel–coda association.

Among onset–coda pairings, all patterns point to favoring of discords. This conclusion is very strong with respect to place of articulation. In particular, coronal onsets prefer noncoronal codas, and vice versa. We also found clear evidence for discord on the basis of manner of articulation. The list of statistically significant onset–coda phoneme pairs almost entirely shows discord in place and especially manner. Of the pairs that are more frequent than expected, only one of them (/č-k/) fails to match obstruents with sonorants, and it at least has a place contrast. Of the pairs that are less frequent than expected, all are either two sonorants (/l-l, r-l, r-r, h-n, r-n/) or two obstruents (/p-d, b-p, f-k/), and some of these lack place contrasts as well. (For the last pair, reasons of tabu probably also contribute to its rarity.) That the manner contrasts do not turn up when we look only at words with lax vowels suggests that there may really be a three-way interaction here: Perhaps the constraint is against having both consonants be sonorant when the vowel is tense, or long, because that would make the syllable too sonorous overall. As for the discord in consonant place, similar constraints have been noted in more complex English syllables (e.g., */spip/ is impossible), and constraints against identical onset and coda consonants occur in several other languages (Davis, 1985, pp. 23–30). But only through the kind of statistical analysis carried out here can one determine that tendencies toward onset–coda dissimilarity, although violable, nevertheless show up in English CVC words. Very similar results were obtained by Berkley (1994) through statistical methods.

Vowel–coda pairs show many interesting effects, as itemized in Table 7. We can quickly pass over the sparse rightmost three columns. The association effects listed there, while real, may simply reflect a few local effects such as the lack of short central vowels before /r/, and an association of the round vowel /ɒ/ with the labial stops. In contrast, large effects show up repeatedly in the first two columns. However, many of those effects may reflect the fact that the phonetic features within a given slot do not co-occur in balanced numbers. For example, the voicing and manner of articulation of a coda consonant are directly associated with its coronality, and most low vowels are lax. Taking such factors into account, the tantalizing associations between tenseness and either voicing or manner, and between low vowels and voice or place, may turn out to be indirect. The most reliable associations in the table, from a causal point of view, may be those between low vowels and less sonorous codas on the one hand, and between tense vowels and coronal codas on the other. This latter finding agrees with the conclusions of Berg (1994), who performed a statistical study of VC sequences in standard British English. Both our study and Berg's show that vowels that are phonetically longer (tense vowels, including diphthongs) prefer to be associated with coda consonants that are phonetically shorter (coronals: Crystal & House, 1988). The tendency for phonetically longer vowels to pair with phonetically shorter consonants may reflect a drive toward isochrony, in Berg's terminology. In informal terms, if one has to add a consonant to an already long rime, it is best that that additional consonant be as short as possible. This is often seen on a grosser scale in other languages, where long vowels must be followed in the same syllable by fewer consonants, or indeed none at all.

General Discussion

Our statistical study of English monomorphemic CVC words sheds light on the substructure of the syllable. If we apply the oft stated criterion that co-occurrence constraints can be taken as evidence of structure, and broaden the search criteria so that we accept quantitative tendencies in addition to absolute restrictions, we find that co-occurrence constraints between vowel and coda are significant. That is, we have adduced further evidence for a rime structure. The fact that consonants have a different distribution in the coda than in the onset, quite apart from any association with the vowel (Study 1) can also be referred to the same structure: Consonants have a different distribution within the rime than outside it.

It must be acknowledged, however, that this interpretation is not definitive. It is conceivable that the vowel only associates with the last consonant of the word, not the last consonant of all syllables, or to the contrary that the vowel associates with any immediately following consonant, whether or not it belongs to the same syllable as the vowel. Studies with different word lists will be needed to tease apart these possibilities, although some research already tends to corroborate the idea that the associations are syllable-based. Berg (1994), for example, found almost exactly the same type of VC associations as we did, and his database included polysyllabic words. Randolph (1989) also used polysyllabic words, and although one must keep in mind his report that associations were not significant, it is striking that the strength of the associations he reports between vowels and consonants are two to three times larger for coda consonants than for onset consonants.

The other main caveat is that the idea of hierarchical structure in linguistics is a vaguely defined concept that can easily be overworked. We believe that a connectionist approach to language leads to a consistent and flexible view of structure. If all components of a syllable are associated by connections of varying weights, then a structure can be seen as a reification of those associations. In this sense it is meaningful to speak of an element as simultaneously being a member of many structures, which vary in strength. Although such views have been rejected by a number of linguists (e.g., Davis, 1985, p. 76), Vennemann (1988a, 1988b) advocates precisely the sort of syllable structure that Davis rejected: A syllable has a body and also a rime, and even a discontiguous shell (onset-coda structure), and different patterns of association may exist within any of these structures (see also Derwing, Dow, & Nearey, 1988). That vowel–coda patterns have significant phonotactic associations justifies talking about a rime structure. So does the bulk of other linguistic and experimental evidence. But we have also uncovered significant onset-coda associations, which may justify talking about shells.

The implications of this study to lexical processing are also obvious, though not unequivocal. If our results generalize to the vocabulary as a whole, then from the standpoint of temporal processing, the first two thirds of the syllable (onset and vowel) are largely unpredictable, that is, informative and distinctive, and the last third (the coda) is largely predictable, that is, redundant. Such asymmetries dovetail with experimental evidence showing that English language processing appears to be optimized for recognizing words from their beginnings (e.g., Cutler, 1982, p. 19), and for producing words that have distinctive beginnings (Sevald & Dell, 1994). These processing asymmetries may, at least in part, be learned adaptations to preexisting patterns in the vocabulary, patterns that may have their origin in physical facts of articulation and acoustics. Of course the opposite conclusion is also supported by these statistics: that the vocabulary itself is shaped because speakers tend to reject neologisms and sound changes that conflict with the most natural patterns of temporal processing. The data raise interesting issues that are not obvious when one looks at English phonotactics only from the standpoint of absolute rules and inviolable co-occurrence restrictions.

Finally, we believe that our results satisfy our goal of demonstrating that quantitative methods can pick up patterns that are not unfamiliar to linguists who advocate strictly categorical approaches to language. We have shown that English codas and onsets have different consonant inventories; that the selection of coda consonants depends on vowels more than the selection of onsets does; that bodies show palatalization, shells show constraints against identity, and rimes show isochrony. These patterns are very common cross-linguistically. In fact, the results suggest that languages may be more similar to each other at a statistical level than they are at a categorical level. This has important implications for the search for linguistic universals.

But one may wonder whether language users are sensitive to the patterns uncovered here. On one view, the partial constraints that we have documented are the detritus of historical inviolable rules. For instance, the statistical tendency toward rime isochrony could have resulted from rules that applied without exception at some point in the history of some of the source languages of English but that no longer apply. On this view, present-day users of English may not judge newly coined words that follow the statistical patterns as sounding more natural than words that do not. On the other hand, language users may have developed a sensitivity to the distributional characteristics of English phonemes regardless of their historical causes. Work reviewed by Kelly (1992) suggests that adults and even children show a remarkable sensitivity to subtle patterns in the language. Research that is presently being pursued (Treiman, Kessler, Knewasser, Tincoff, & Bowman, 1996) shows that speakers are indeed sensitive to the frequencies of rimes. For example, they rate nonsense words with more frequent rimes as better than those with less frequent rimes. In those experiments, stimuli were selected so as to cancel out the effects of the frequency of the individual vowel and coda, so that observed results were necessarily due to the frequency of the rime as a whole. Such findings suggest that, whatever the origins of the distributional patterns in present-day English, language users pick up these patterns.

Conclusion

We have shown that statistical techniques reveal phonotactic patterns in English that are very similar to the inviolable phonotactic restrictions that have been noted in other languages. In particular, the fact that the association between vowel and coda is so much greater than that between vowel and onset buttresses analyses that posit rimes as syllable constituents. Those unequal associations mean greater distinctiveness at the beginning of the syllable, which may be connected to findings that people produce and recognize words more efficiently the more distinctive the beginnings of those words are.

Together with other recent studies that have analyzed large lexical databases in order to elucidate the patterns in languages (e.g., Berkley, 1994; Frauenfelder, Baayen, Heelwig, & Schreuder, 1993; Krupa, 1982; Treiman, Mullennix, Bijeljac-Babic, & Richmond-Welty, 1995), the present study represents a departure from traditional linguistic and psycholinguistic work. We believe that there is much to be learned from detailed statistical studies of language and that the results of such studies will have important implications for the search for language universals and for our understanding of language processing.

References

Berg, T. (1994). The sensitivity of phonological rimes to phonetic length. Arbeiten aus Anglistik und Amerikanistik, 19, 63–81.

Berkley, D. M. (1994). Variability in Obligatory Contour Principle effects. In K. Beals et al. (Eds.), Papers from the 30th Regional Meeting of the Chicago Linguistic Society, Vol. 2 (pp. 1–12). Chicago: Chicago Linguistic Society.

Clements, G. N. & Keyser, S. J. (1983). CV phonology : A generative theory of the syllable. Cambridge, MA: MIT Press.

Crystal, T. H. & House, A. S. (1988). The duration of American-English stop consonants: An overview. Journal of Phonetics, 16, 285–294.

Cutler, A. (1982). The reliability of speech error data. In A. Cutler (Ed.), Slips of the tongue and language production (pp. 7–28). Amsterdam: Mouton.

Davis, S. (1985). Topics in syllable geometry. Unpublished doctoral dissertation, University of Arizona, Tucson.

Derwing, B. L., Dow, M. L., & Nearey, T. M. (1988). Experimenting with syllable structure. In J. Powers & K. de Jong (Eds.), Proceedings of the Fifth Eastern States Conference on Linguistics (pp. 83–94). Columbus, OH: Ohio State University.

Derwing, B. L., Yoon, Y. B., & Cho, S. W. (1993). The organization of the Korean syllable: Experimental evidence. In P. M. Clancy (Ed.), Japanese/Korean Linguistics, vol. 2 (pp. 223–238). Stanford, CA: Center for the Study of Language and Information.

Flexner, S. B. (Ed.) (1987). The Random House dictionary of the English language (2nd ed.) New York: Random House.

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.

Fudge, E. (1969). Syllables. Journal of Linguistics, 5, 253–287.

Fudge, E. (1987). Branching structure within the syllable. Journal of Linguistics, 23, 359–377.

Goldinger, S. D., Luce, P. A., & Pisoni, D. B. (1989). Priming lexical neighbors of spoken words: Effects of competition and inhibition. Journal of Memory and Language, 28, 501–518.

Goldsmith, J. A. (1990). Autosegmental and Metrical Phonology. Oxford, England: Blackwell.

Grainger, J. (1992). Orthographic neighborhoods and visual word recognition. In R. Frost & L. Katz (Eds.), Orthography, phonology, morphology, and meaning (pp. 131–146). Amsterdam: Elsevier.

Halle, M. & Vergnaud, J.-R. (1982). Three dimensional phonology. Journal of Linguistic Research, 1, 83–105.

Hayes, B. (1989). Compensatory lengthening in moraic phonology. Linguistic Inquiry, 20, 253–306.

Hockett, C. (1955). A manual of phonology. Baltimore: Waverly Press.

Hyman, L. M. (1985). A theory of phonological weight. Dordrecht, Netherlands: Foris.

Iverson, G. K. & Wheeler, D. W. (1989). Phonological categories and constituents. In Linguistic Categorization (pp. 93–114). Amsterdam: Benjamins.

Kelly, M. H. (1992). Using sound to solve syntactic problems: The role of phonology in grammatical category assignments. Psychological Review, 99, 349–364.

Kenstowicz, M. (1994). Phonology in generative grammar. Cambridge, MA: Blackwell.

Krupa, V. (1982). The Polynesian languages. London: Routledge & Kegan Paul.

Kuryłowicz, J. (1973). Contribution à la théorie de la syllabe [Contribution to the theory of the syllable]. In his Esquisses linguistiques, vol. 1 (2nd ed., pp. 193–220). München, Germany: Wilhelm Fink.

Lass, R. (1984). Phonology. Cambridge, MA: Cambridge University Press.

MacKay, D. G. (1973). Spoonerisms: The structure of errors in the serial order of speech. In V. A. Fromkin (ed.), Speech errors as linguistic evidence (pp. 164–194). The Hague, Netherlands: Mouton.

McCarthy, J. J. (1979). On stress and syllabification. Linguistic Inquiry, 10, 443–466.

Maddieson, I. & Precoda, K. (1992). Syllable structure and phonetic models. Phonology, 9, 45–60.

Pierrehumbert, J. & Nair, R. (1995). Word games and syllable structure. Language and Speech, 38, 77–114.

Randolph, M. A. (1989). Syllable-based constraints on properties of English sounds. Unpublished doctoral dissertation, MIT, Cambridge, MA.

Selkirk, E. O. (1982). The syllable. In H. van der Hulst & N. Smith (Eds.), The structure of phonological representations, vol. 2 (pp. 337–383). Dordrecht, Netherlands: Foris.

Sevald, C. A. & Dell, G. S. (1994). The sequential cuing effect in speech production. Cognition, 53, 91–127.

Stemberger, J. P. (1983). Speech errors and theoretical phonology: A review. Bloomington, IN: Indiana University Linguistics Club.

Treiman, R. & Kessler, B. (1995). In defense of an onset–rhyme syllable structure for English. Language and Speech, 38, 127–142.

Treiman, R., Kessler, B., Knewasser, S., Tincoff, R., & Bowman, M. (1996, July). Phonotactic patterns in English words and speakers' sensitivity to these patterns. Paper presented at the Fifth Conference on Laboratory Phonology, Northwestern University, Evanston, IL.

Treiman, R., Mullennix, J., Bijeljac-Babic, R., & Richmond-Welty, E. D. (1995). The special role of rimes in the description, use, and acquisition of English orthography. Journal of Experimental Psychology: General, 124, 107–136.

Vennemann, T. (1988a). Preference laws for syllable structure and the explanation of sound change. Berlin, Germany: Mouton de Gruyter.

Vennemann, T. (1988b). The rule dependence of syllable structure. In

C. Duncan-Rose & T. Vennemann (Eds.), On language: Rhetorica, phonologica, syntactica: A festschrift for Robert P. Stockwell from his friends and colleagues (pp. 257–283). London: Routledge.

Yoon, Y. B. (1995). Experimental studies of the syllable and the segment in Korean. Unpublished doctoral dissertation, University of Alberta, Edmonton.

Author Note

Brett Kessler, Department of Linguistics; Rebecca Treiman, Psychology Department.

The research was supported, in part, by NSF grants SBR-9020956 and SBR-9408456. Some of the results were presented at the meeting of the Psychonomic Society, St. Louis, Nov. 1994, and at the Conference on Laboratory Phonology, Northwestern University, July 1996. The authors thank Jarrett Rosenberg for assistance with the exact statistical tests, and Thomas Berg, Edward Flemming, and William Poser for extensive comments.

Correspondence concerning this article should be addressed to Brett Kessler, Department of Linguistics, Stanford University, Stanford, California 94305-2150; or to Rebecca Treiman, Psychology Department, Wayne State University, 71 W. Warren Ave., Detroit, Michigan 48202. Electronic mail may be sent via Internet to kessler@csli.stanford.edu or to treiman@math.wayne.edu, respectively.

Table 1

Frequency and Features of Vowels

Vowel	Example	Frequency	Height	Backness	Tenseness
ɑ	alms	30	low	central	tense
æ	ax	198	low	front	lax
ɒ	odd	128	low	back	lax
ai	ides	136	—	—	tense
ɑu	out	44	—	—	tense
e	ape	183	mid	front	tense
ɛ	ebb	159	mid	front	lax
ɝ	erg	115	mid	central	tense
i	eat	210	high	front	tense
ɪ	if	207	high	front	lax
o	oats	146	mid	back	tense
ɔ	ought	110	mid	back	tense
ɔi	oink	25	—	—	tense
u	ooze	117	high	back	tense
ʊ	push	38	high	back	lax
ʌ	up	155	mid	central	lax

Table 2

Frequency and Features of Consonants

Phone	Example	Frequency	Place	Manner	Voice
b	boy	216	bilabial	interrupted	voiced
č	chin	115	nonanterior	interrupted	unvoiced
d	dog	268	anterior	interrupted	voiced
ð	this	15	anterior	fricative	voiced
f	fox	160	labiodental	fricative	unvoiced
g	girl	155	postcoronal	interrupted	voiced
h	hop	105	postcoronal	approximant	unvoiced
j	young	30	nonanterior	approximant	voiced
ǰ	jump	115	nonanterior	interrupted	voiced
k	kiss	324	postcoronal	interrupted	unvoiced
l	love	365	anterior	approximant	voiced
m	maid	243	bilabial	nasal	voiced
n	new	306	anterior	nasal	voiced
ŋ	sang	46	postcoronal	nasal	voiced
p	pad	240	bilabial	interrupted	unvoiced
r	read	287	nonanterior	approximant	voiced
s	sing	242	anterior	fricative	unvoiced

(table continues)

Phone	Example	Frequency	Place	Manner	Voice
ʃ	sheep	109	nonanterior	fricative	unvoiced
t	tongue	323	anterior	interrupted	unvoiced
θ	thin	56	anterior	fricative	unvoiced
v	vase	99	labiodental	fricative	voiced
w	win	82	labial	approximant	voiced
ʍ	whip	24	labial	approximant	unvoiced
z	zoo	71	anterior	fricative	voiced
ʒ	rouge	6	nonanterior	fricative	voiced

Table 3

Distribution of Consonants Within Onset and Coda

Phone	Onset	Coda	χ^2	ϕ
j	30	0	30.00*	1.000
ʍ	24	0	24.00*	1.000
w	82	0	82.00*	1.000
ŋ	0	46	46.00*	1.000
h	105	0	105.00*	1.000
ð	1	14	11.27*	.867
ʒ	1	5	—	.667
z	13	58	28.52*	.634
b	154	62	39.19*	.426
θ	17	39	8.64*	.393
n	99	207	38.12*	.353
ǰ	74	41	9.47*	.287
t	119	204	22.37*	.263
l	135	230	24.73*	.260
ʃ	65	44	4.05*	.193
f	92	68	3.60	.150
r	163	124	5.30*	.136

(table continues)

Phone	Onset	Coda	χ^2	ϕ
g	88	67	2.85	.135
k	142	182	4.94*	.123
v	45	54	0.82	.091
p	128	112	1.07	.067
d	126	142	0.96	.060
m	116	127	0.50	.045
s	126	116	0.41	.041
č	56	59	0.08	.026

Note. Statistics examine the difference between the frequency of each consonant in the onset and its frequency in the coda.

*$p < .05$, 1 df.

Table 4

$\underline{G}^2$ Decomposition of Consonant Distributions By Place of Articulation, Onset vs. Coda

Contrasting	Favoring onset	Favoring coda	$\underline{G}^2$	Cramér
Consonants	noncoronal	coronal	30.06	.09
coronal	nonanterior	anterior	60.93	.16
anterior	/d, s/	/l, n, θ, t, z/	47.11	.17
noncoronal	labial	velar	7.85	.07
labial	/b/	/p, m/	28.53	.20
velar	/g/	/k/	7.05	.12

Note. Omits consonants largely restricted to either onset (/h, j, w, ʍ/) or coda (/ŋ, ð, ʒ/). First column shows the phonetic category within which a significant skew between onset and coda is found. The members of that category are differentiated by whether they appear more often in onset than do other members of the same category. The nonanterior consonants /ǰ, ʃ, r, č/ did not vary significantly among themselves.

Table 5

Significant Associations Between Onset, Vowel, and Coda

Vowel–Coda	Onset	
	Place	Manner or Voice
Coda place		
Vowel height	[OC][VC]	[O][VC]
Lax	[OC][V]	[O][V][C]
Vowel backness	[OC][VC]	[O][VC]
Lax	[OC][V]	[O][V][C]
Vowel tenseness	[OC][VC]	[O][VC]
Coda manner		
Vowel height	[O][VC]	[O][VC]
Lax	[O][VC]	[O][VC]
Vowel backness	[O][V][C]	[O][V][C]
Lax	[O][VC]	[O][VC]
Vowel tenseness	[O][VC]	[O][VC]
Coda voice		
Vowel height	[O][VC]	[O][VC]
Lax	[O][V][C]	[O][V][C]

(table continues)

Vowel–Coda	Onset	
	Place	Manner or Voice
Vowel backness	[O][V][C]	[O][V][C]
Lax	[O][V][C]	[O][V][C]
Vowel tenseness	[O][VC]	[O][VC]

Note. Table shows which log-linear models are significant at $p < .001$ when the phonemes in the three syllable slots are collapsed by the indicated features. [O][V][C] means that no associations are accepted; [O][VC] means that vowel and coda are associated, but onset is independent; [OC][V] means that the onset and coda are associated, but the vowel is independent; [OC][VC] means that the coda is associated with both onset and vowel, but that the latter two are conditionally independent. Rows labeled Lax tell how the preceding row changes if only words with lax vowels are included.

Table 6

Significant Tables in $\underline{G}^2$ Decompositions of Pairwise Onset–Coda Interactions, By Place

Coda	Onset	
	Noncoronal	Labial
Coronal	21.99	9.30
Lax	18.56	n.s.
Velar	n.s.	9.79
Lax	n.s.	n.s.

Note. Numbers are $\underline{G}^2(1)$ values for significant 2×2 associations, or n.s. when decomposition is not significant at $\underline{p} < .05$ with Bonferroni correction for 9 comparisons. Features are contrasted with their sibling features in Figure 3; e.g., Labial is contrasted with Postcoronal. Main figures are for the full word set; subrows labeled Lax exclude words with tense vowels or diphthongs.

Table 7

Significant Pairwise Vowel–Coda Interactions, By Phonetic Feature

Coda	Vowel				
	Tenseness	Height		Backness	
	Lax	Low	High	Round	Central
Voicing					
Unvoiced	17.35	10.47	9.79	n.s.	n.s.
Manner					
Obstruent	26.24	15.17	n.s.	n.s.	n.s.
Lax	NA	15.25	n.s.	9.04	n.s.
Interrupted	21.53	14.59	n.s.	n.s.	n.s.
Lax	NA	n.s.	n.s.	n.s.	n.s.
Nasal	13.84	11.85	n.s.	n.s.	n.s.
Lax	NA	14.07	n.s.	n.s.	15.37
Place					
Noncoronal	51.39	27.00	n.s.	n.s.	n.s.
Velar	15.98	n.s.	n.s.	n.s.	n.s.
Bilabial	n.s.	11.68	n.s.	14.06	n.s.
Nonanterior	n.s.	9.72	n.s.	n.s.	n.s.

Note. Numbers are $G^2(1)$ values for significant 2×2 associations, or n.s. when decomposition is not significant at $\underline{p} < .05$ with Bonferroni correction for 9 comparisons or when Table 5 indicates no significant association along the relevant dimension. Features are contrasted with their sibling features in Figures 3 and 4. Main figures are for the full word set; subrows labeled Lax exclude words with tense vowels or diphthongs. Lax/Tense contrasts are not applicable (NA) to that lax-vowel subset.

Table 8

Pairs of Phonemes that Occur at Frequencies Significantly Different from Chance, $\underline{p} < .05$

Pair	More frequent than chance	Less frequent	Count
Onset–Vowel	lu, ru, we, fe, si, wai, nɒ	rɜ˞, ki	9 (3.6%)
Lax	gæ, ǰɒ	wæ	3 (3.3%)
Vowel–Coda	ɔr, ɛr, ir, ʌf, ʌg, en, ol, ez, ɒp, æg, el, ait, ɔl, up, ɒt	(ɪr, ær, er, ʌr), æl, (or, ɒr, ur, ɜ˞r), ʌl, (ib, eč), os, ɒl, ɔp, æz, ɔt, ɔm	33 (17.7%)
Lax	ɛr, ɒb, ʌf, ɒp, ɪl, ɒt, ɛl, ʌm, ɛs, ɪŋ	(ɪr, ær, ɛŋ, ʌr), ɛʃ, ɛb, ɛp, ɒf, ʌp, æl	20 (25.6%)
Onset–Coda	f-n, ǰ-n, r-d, č-k, n-t	l-l, r-l, p-d, r-r, h-n, r-n, b-p, f-k	13 (4.7%)
Lax	—	—	0 (0.0%)

Note. Only pairs whose expected frequencies are 5 or more are considered. Last column tells how many pairs are here listed, and what percentage that is of all pairs that have sufficiently high expectation. Lists are in descending order of the association statistic ϕ. Pairs in parentheses are not found at all. Subrows labeled Lax omit from consideration all words that have tense vowels or diphthongs.

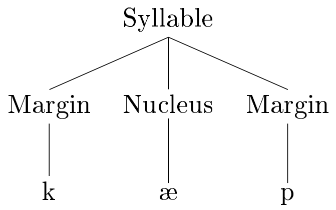
Figure Captions

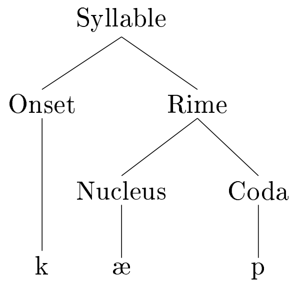
Figure 1. Flat, onset-rime, and body-coda theories of syllable structure, illustrated with the word cap.

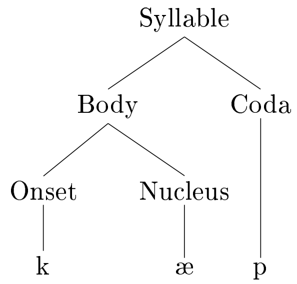
Figure 2. Moric theory of syllable structure, illustrated with cap (short vowel) and keep (long vowel).

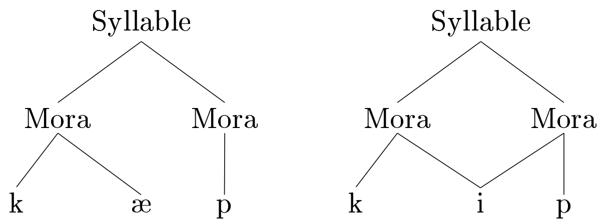
Figure 3. Hierarchical organization of the three classes of consonant features. This study compares the distribution of two consonant feature classes only when they are immediate children of the same node. Table 2 lists the features of each of the consonant phonemes.

Figure 4. Hierarchical organization of the three classes of vowel features. This study compares the distribution of two vowel feature classes only when they are immediate children of the same node. Table 1 lists the features of each of the vowel phonemes.


Syllable
Margin
Nucleus
Margin
k
æ
p


Syllable
Onset
Rime
Nucleus
Coda
k
æ
p


Syllable
Body
Coda
Onset
Nucleus
k
æ
p


Syllable
Mora
Mora
k
æ
p
Syllable
Mora
Mora
k
i
p

- Place
  - coronal
    - anterior
    - nonanterior
  - noncoronal
    - labial
      - bilabial
      - labiodental
    - postcoronal

- Manner
  - obstruent
    - interrupted
    - fricative
  - sonorant
    - nasal
    - approximant

- Voice
  - voiced
  - unvoiced

- Height
  - low
  - nonlow
    - high
    - mid

- Backness
  - back (round)
  - nonround
    - central
    - front

- Tenseness
  - tense
  - lax