Speculations about the Representations of Words in Memory

Morris Halle 1985

1. Introduction

The purpose of this chapter is to explore one aspect of a fluent speaker's knowledge of a language. I present reasons for believing that information about the phonic shape of the words is stored in a fluent speaker's memory in the form of a three-dimensional object that for concreteness one might picture as a spiral-bound notebook. I realize that this is a fairly radical claim and that it is likely to be met with considerable skepticism on the part of working phoneticians who have, of course, never encountered spiral-bound notebooks in any of their studies of the sounds produced and the articulations executed by hundreds of speakers of dozens of languages.

2. What is Memorized

I begin with the negative assertion that it is unlikely that the information about the phonic shape of words is stored in the memory of speakers in acoustic form resembling, for instance, an oscillogram or a sound spectrogram. One reason that this is improbable is that when we learn a new word we practically never remember most of the salient acoustic properties that must have been present in the signal that struck our ears; for example, we do not remember the voice quality of the person who taught us the word or the rate at which the word was pronounced. Not only voice quality, speed of utterance, and other properties directly linked to the unique circumstances surrounding every utterance are discarded in the course of learning a new word. The omissions are much more radical. For example, there is reason to believe that English speakers do not store in memory such a salient property as, for example, the stress contour of the word. It is known that for a large class of English words stress is determined by Rule 1:

Reprinted with permission from V. Fromkin, (ed.), *Phonetic Linguistics* Academic Press, Orlando, Florida, pp. 101–114, 1985

Rule 1:

a. Stress the antepenultimate syllable if the penultimate syllable is "light"; that is, ends with a short vowel; otherwise,
b. stress the penultimate syllable.

Examples of words stressed in accordance with this rule are given in (1):

(1)	límerick	jávelin	América	Cánada
	addéndum	veránda	Augústa	Aláska
	decórum	ultimátum	marína	rutabága

Not all English words have stress contours that follow directly from (1). Examples of such deviant stress contours are given in (2).

(2) statuétte devélop órchestra

It should be noted that the exceptions are not random; rather, they are stressed in accordance with rules that differ somewhat from (1). Moreover, the existence of exceptions to (1) does not undermine its status: Rule 1 expresses a true regularity that is manifested in a large class of words of the language.

Having established that Rule 1 is true of a large class of English words, one might next inquire whether regularities such as those expressed in Rule 1 are of any significance for speakers of English. There are two diametrically opposed answers that might be offered here. On the one hand, one might argue that these regularities play no role as far as English speakers are concerned; they are discoveries that have been made by professional linguists, and, like many other such discoveries, they have no bearing on the way ordinary speakers of English produce and understand utterances in their language. For example, linguists have established that English is derived from the same protolanguage as Sanskrit, Armenian, Greek, Latin, and Irish, and that the stems of the English words *brotherly* and *fraternal* derive from the same Indo-European root, but these facts surely have no effect on the ability of an ordinary English speaker, say a bright ten-year-old, to speak and understand her teacher, her class-mates, or others about her.

By contrast with the preceding, one might respond to our question by suggesting that, unlike the historical or archeological facts noted above, the regularities in Rule 1 are a crucial component of the knowledge that English speakers must have in order to speak and to understand each other. For example, one might speculate that for various psychobiological reasons speakers find it difficult or impossible to memorize the stress contours of words, but they find it easy to compute the stress contours by means of rules such as 1.

This is obviously a fairly bold speculation that will be rejected out of hand unless evidence in support of it is immediately provided.

Some evidence that speakers do not memorize the stress contours of words directly comes from the treatment of foreign words when these are borrowed into English. It is well known that, when borrowed into English, foreign words are frequently anglicized. Thus, for example, the three words in (3) which are fairly recent borrowings from Russian, are stressed in accordance with Rule 1, whereas their original Russian stress contours are as shown in (4).

(3) bólshevik Rómanov babúshka

(4) bolshevík Románov bábushka

It is commonly said that words are anglicized when they have features that English speakers might find difficult or impossible to pronounce. In the present instance, however, this is hardly a plausible explanation, for, as shown in (2), there are English words that have precisely the same stress contours as our three words have in Russian.

A more plausible explanation for the anglicized stress contours of the borrowed words is that the speakers who introduced these words into English did not memorize their original Russian stress contours, because, as suggested above, they find memorizing stress contours difficult or impossible. They, therefore, remembered the words without their original stress contours and supplied them with stress contours by utilizing Rule 1, the standard stress rule for English nouns. In addition to explaining why words borrowed into English from Russian (and other foreign languages) are stressed as illustrated in (3) and not as in (4), our suggestion also provides a rationale for the existence of the regularities in stress contour that we have illustrated in (1). These regularities exist—as already suggested—because speakers find it difficult to memorize the stress contours of each word separately but find it easy to compute the stress contours by means of rules such as Rule 1. The assertion that speakers do not memorize the stress contour of words implies further that rules such as 1 are part of their knowledge of English, that is, that Rule 1 is of interest not just to linguists but also to speakers of English and that words are represented in memory in an abstract form in which many characteristics found in the physical signal are systematically omitted. This is, of course, not a trivial proposal, for it implies that words are represented in memory in a form that is quite abstract in that it omits many characteristics that can be observed in the acoustic signal and the articulatory gymnastics. Moreover, it implies that the process of speaking involves computations of the type made necessary by rules such as Rule 1. While we are still far from being able to demonstrate all this, the evidence accumulated by generative phonology since the late 1950s suggests that the proposal might not be far off the mark.

For some additional evidence supporting the view that words are represented in memory in a quite abstract form, that is, in a form that is indirectly related to the observable articulatory behavior and acoustic signal by means of special rules, consider the following experiment that can be conducted without elaborate equipment and preparations. One can present a list of nonsense syllables such as the one in (5) to fluent speakers of English and ask them to indicate which syllables in the list might have been taken from an unabridged dictionary of the language and which might not.

(5) flib slin vlim smid fnit vrig plit trit brid blim tnig bnin

It has been my experience that speakers have very clear intuitions about which of these nonsense words are or are not part of English. And on the basis of this experience I have little doubt that most people would regard *vlim, fnit, vrig, tnig, bnin* as unlikely candidates for words of English, but that they are likely to accept the others.

This fact raises two questions: (1) on what basis do speakers make these judgments? and (2) how do speakers acquire the knowledge that underlies these judgments? Since the nonsense syllables in (5) have never been seen by our experimental subjects, it is not possible that they arrive at their judgments simply by checking the list of all the words they know and by discovering that the non-English items are not in the list. We must rather assume that speakers know some general principle that allows them to determine whether any arbitrary sequence of sounds is or is not a well-formed syllable of English. The principle involved in the judgments under discussion is given in Rule 2.

Rule 2:

> English syllable onsets containing two consonants are composed of {p t k b d g f θ} followed by {l r w} or of {s} followed by {p t k m n l w}.

It may be somewhat puzzling to some that we should know such complicated principles as those in Rule 2 in addition to those in Rule 1, especially since practically none of us is likely ever to have been consciously aware of their existence. It takes, however, but a moment's reflection to convince oneself that there are many things that people know without being conscious of this fact. For instance, major league ball players must surely have knowledge of parabolic trajectories, for each time they catch a ball they must somehow calculate such a trajectory. But no one is likely to want to conclude that baseball players have explicit knowledge of Newton's Laws of Motion, that they can solve differential equations, or even that they can do simple sums. Like the knowledge of parabolic trajectories possessed by ball players, knowledge of syllable onsets is largely implicit knowledge, but that, of course, does not make it any less real.

If lack of awareness on our part is, thus, no bar to the assumption that our knowledge of English includes knowledge of such abstract principles as those in Rules 1 and 2, there remains the question as to how we could have ever come into possession of this knowledge. It obviously could not have been taught to us by our parents or teachers, for they are as little aware of this information as we are. We must, therefore, assume that we somehow acquired it on our own in the course of learning English. Given the conditions under which young children ordinarily learn their mother tongue, the only plausible assumption is that we are so constructed that when we store the words in our memory we simultaneously abstract the distributional regularities in the phoneme sequences that make up these words and that in the course of this procedure we establish that the words obey principles such as those in Rules 1 and 2.

Like our knowledge of what phonetic attributes should be memorized and what attributes should be discarded, the knowledge that we must establish all sequential regularities in words could not have been plausibly learned and must, therefore, be assumed to be innate. It must be part of the genetic equipment by virtue of which humans, unlike members of other species, are capable of learning natural languages. To motivate this story one may speculate that space in our memory is at a premium and that we must, therefore, store in our memory as little information as possible about the phonetic shape of each word, eliminating as many redundancies as possible and placing maximum reliance on our ability to compute the omitted information. For example, as observed above, if Rule 1 is available, information about stress need not be stored in memory. Similarly, given the redundancies noted in Rule 2 we can omit in the representation of English onset clusters beginning with /s/ information about such phonetic features as voicing and continuancy, for these are totally predictable.[1]

In order to realize these economies in memory storage, however, we must be able to compute the omitted features. Thus, we arrive once again at the conclusion that the process of speaking involves rules and computations and that words are represented in memory in a form that omits many of the characteristics directly observable in the acoustic signal and vocal tract gymnastics.

3. How it is Represented

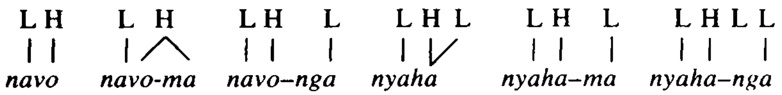
If Rule 2 is an integral part of a speaker's knowledge of English, then it must be the case that English speakers represent words in their memory in a form that is compatible with that rule. Since Rule 2 makes reference to discrete speech sounds and to such features as voicing and continuancy, we are led to infer that speakers represent words in memory as sequences of discrete speech sounds or 'phonemes', that is, in a form that resembles transcriptions of language in familiar alphabetic writing systems. The proposition that words are stored in speakers' memories as sequences of phonemes or phoneme-like units has been generally accepted by linguists and phoneticians for a long time. There is a considerable body of facts, however, that cannot be readily dealt with by means of representations composed of a single linear sequence of units. For example, all languages utilize variations in the fundamental pitch of the voice in their utterances. Thus, in English, utterances are pronounced with quite different melodies when they are used as a response to a neutral question and when they are intended to express surprise. In response to the neutral question *What are you studying*? the response might be *linguistics*, with a melody composed of the tones M(id)–H(igh)–L(ow); the response to a comment occasioning great surprise would be *linguistics* with a LHH melody. Since speakers normally produce such melodies in their utterances, they must possess a means for representing melodies.

The problem that arises here is that a given tone sequence may be spread over an arbitrary number of syllables. For example, the same two melodies that we encountered above in our little story would have been produced if, instead of the trisyllabic *linguistics*, the response had been the bisyllabic *Brasil* or the decasyllabic *antidisestablishmentaria-nism*. What this shows is that the tone sequences are independent of the syllable sequences. It should not come as a great surprise that notations like our normal alphabetic writing system or the standard phonetic or phonological transcriptions, which are modeled on our alphabetic writing system and which represent utterances by means of a single linear sequence of symbols, are in principle incapable of handling signals composed of two independent sequences of elements: tones and syllables. Since there are two independent sequences of elements encoded in the signal, the notation must have two independent sequences of symbols and that is what in fact we find in the various notations, some quite ancient, especially designed to record both the melody and the words of a chant or song. In these notations the words were recorded on one line and the melody was recorded by a system of diacritic marks written above the line of letters on what is in effect a second line. (There is even a technical term for such diacritics *neumes*, which Webster defines as "symbols used in the notation of Gregorian chant.") Various informal adaptations of this idea have been employed in phonetic studies of tonal phenomena. It is only quite recently that these tonal notations have been formally investigated. As a result of work by Goldsmith (1979), Williams (1976), Liberman (1975), Pulleyblank (1983), and others on what has been called "autosegmental phonology," great advances have been made in our understanding of such representations. To convey some idea as to what has been learned we examine below an actual example.

4. Autosegmental Phonology: Some Examples

In many languages, tonal melodies serve to distinguish different words. In such languages, for example, two otherwise identical words are differentiated by the fact that one has the melody LH and the other, the melody HL. The tonal differences function, therefore, exactly like other phonetic distinctions, for example, like differences in the quality of the vowels /i/ and /e/ in the words *bit–bet*. And like information about vowel quality, information about the tonal melodies of the different words must be memorized by the speaker of a tone language. Moreover, there is reason to believe that speakers store this information on a separate autosegmental tier; that is, the words in these languages are stored in the form of two parallel sequences of units: the phonemes and the tones. To see why this might be so, consider the facts in (6), where we reproduce three forms of the Mende words *navo* 'money' and *nyaha* 'woman' (data from Leben 1978). The first of the three forms is the bare word, the second gives the word with a suffix meaning 'on', and the third, with a suffix signaling the indefinite plural.

(6)



The word *navo* has the melody LH; the word *nyaha* has the melody LHL. The suffix *ma* is toneless, whereas the suffix *nga* has a melody consisting of a single L tone. Most of the work linking tones to syllables is accomplished by the universal linking convention (see Pulleyblank 1983), which states that tones are linked to vowels one to one and from left to right. Any deviations from this simple correspondence between tones and vowels must be licensed by language-particular rules.

In the examples in (6), the universal linking convention accounts for the fact that *ma*, though inherently toneless, surfaces with a L tone in *nyahama*. The convention fails to account for the cases where the linking between tones and vowels is other than one to one. These are the domain of the language-particular rules, and Mende has two of these. On the one hand, Mende has a rule that spreads the last tone of one morpheme onto the following toneless syllable. It is by virtue of this rule that *ma* has H tone in *navoma*. A second rule links an unlinked tone to the word-final syllable. In consequence of this rule, we find two tones on the last syllable of the bare stem *nyaha* but not elsewhere. There is a third deviation from one-to-one correspondence in (6): In *nyahanga*, the second L tone remains unlinked. This fact is accounted for by what has been said above. Since neither the universal linking convention nor the tone rules of Mende provide a way for linking the second L tone in *nyahanga* to a vowel of the word, the tone remains unlinked. Since tones can only be pronounced when they are linked to phonemes, this L tone is not pronounced. It is worth noting that the phenomena we have just discussed are totally opaque if tones are viewed as attributes of the individual vowels rather than as an autosegmental sequence parallel to and separate from the sequences of phonemes.

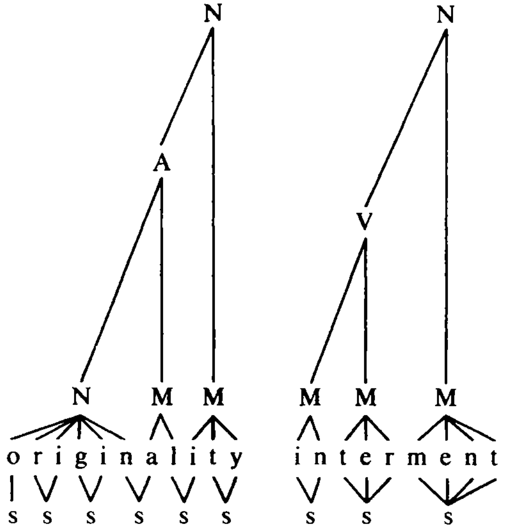
Since two parallel lines define a plane, we can say that the parallel sequences of tones and phonemes in (6) constitute a plane, the tone plane. One of the major insights gained by linguistic research since the early 1970s is that the tone plane is not the only property of phonological sequences that must be represented on its own separate plane. Other entities requiring such treatment are the two major types of constituents that simultaneously make up each word: syllables and morphemes.[2]

Until the 1970s the only means utilized by linguists for delimiting subsequences in the phonetic string have been boundary markers or junctures that are intercalated among the phonemes of the sequence at appropriate points. The problem with this device is that it introduces into the representation all sorts of symbols that, if taken seriously, tend to make the statement of various phonological regularities all but impenetrable, as shown in (7).

(7) # + /o/ri/gi/n + a/l + i/ty# + /   # + /in + /ter + /ment# + /
where # = word boundary, + = morpheme boundary,
/ = syllable boundary

It was suggested by Kahn (1976) that syllables should be represented on a separate autosegmental tier, like the tones in (6), and a similar suggestion was made with respect to morphemes by Rotenberg (1978). We illustrate these suggestions in (8) (N, noun; A, adjective; V, verb; M, morpheme; s, syllable).

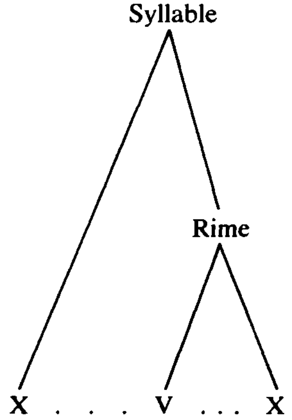
(8)



We can say that in (8) there are two planes intersecting in the phoneme sequence or 'skeleton'. We refer to the plane above the phoneme sequence as the morpheme plane and to the plane below the phoneme sequence as the syllable plane.

A noteworthy feature of the plane on which the morphological structure of the word is represented is that the linking lines reflect nested constituent structure of the sort widely encountered in syntax. This kind of constituent structure is not restricted to the morphological plane. In fact, some sort of constituent structure in syllables was tacitly assumed in our discussion of the distributional restrictions on English words. It will be recalled that the restrictions exemplified in (5) apply not just to any subsequence in the syllable but only to what was referred to above as the onset of the syllable. It may have been noticed that in the discussion of the onset above a detailed characterization of the onset itself was not included. This omission may now be repaired. Recent work by Levin (1983), Steriade (1982), and others appears to lead to the conclusion that syllables universally have the constituent structure given in (9). 'Onset' is, therefore, the name given to the phoneme subsequence in the syllable that precedes the rime.

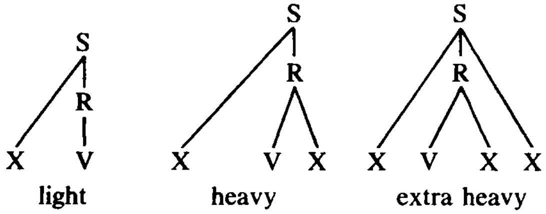
(9)



The syllable thus consists of a left-headed constituent, the rime, nested inside of a right-headed constituent. The V(owel), which is the head of the rime, is the only obligatory constituent of the syllable. The presence of any other constituents in the syllable is governed by language-particular rules of the type given in Rule 2 above.[3]

Since English syllable structure is too complex to discuss here, we examine here the syllables of Classical Arabic. Unlike English, which has a very rich syllable repertoire, Classical Arabic permits only the three types of syllable illustrated in (8), with the further limitation that extra-heavy syllables can occur only word finally.

(10)



Arabic shares an interesting property with all Semitic languages as well as with a number of non-Semitic languages, such as the African language Hausa and the American Indian language Yokuts. In these languages the syllabification of a word is determined not by the phonemes that compose it, but rather by its morphological structure. To illustrate, I cite in (11) some forms from McCarthy (1984) epitomizing the formation of the socalled broken plurals.

(11)	*jundab*	*janaadib*	'locust'
	sultaan	*salaatiin*	'sultan'
	duktar	*dakaatir*	'doctor'
	safarjal	*safaarij*	'quince'
	maktab	*makaatib*	'office'
	miftaaḥ	*mafaatiiḥ*	'key'
	nuwwar	*nawaawir*	'white flower'
	9andaliib	*9anaadil*	'nightingale'

The first thing to observe about these examples is that while the singular forms may be either bi- or trisyllabic, the plural forms are uniformly trisyllabic. Moreover, the syllable structure in the plural is also fixed. The first syllable is light, the second syllable is heavy, and the third syllable, which always ends with a consonant, has a vowel that is identical in length with that of the second syllable in the singular. This is a typical instance of morphology-driven syllabification, that is, of syllable structure being imposed not by the phonological composition of the word but by the fact that the word belongs to a particular grammatical category, the plural in the case under discussion.

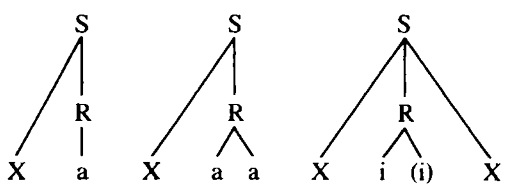
Nor is this all that is determined by the fact that the form is plural. Note that the vowel pattern in all plural forms is the same: /i/ in the last syllable, /a/ in the first and second. The vowel pattern in the plural is, thus, totally unrelated to that of the singular. Only the consonants are not determined by the morphology, by the fact that the word is plural. These convey the lexical meaning of the form. And even their distribution is severely restricted: They occur only in specific positions in the word, and there must be precisely four consonants in every plural form. If the word has more than four consonants in the singular, the extra consonants are omitted, as shown by the words for *quince* and *nightingale* in (11). It would appear, therefore, that the plural form of Arabic nouns consists of the skeleton given in (12) in which the empty slots, represented by X in (12), are filled in by the consonants representing the lexical meaning of the noun.

(12) X a X aa X i(i) X

It was pointed out by McCarthy (1979) that if the consonants are represented on a separate autosegmental tier, the linking of the consonants to the empty slots in the skeleton (12) is effected by the same left-to-right one-to-one convention that is regularly encountered in tone languages; compare this with our discussion of Mende tones in (6) above. Notice that this convention predicts that if the noun has more than four consonants the extra consonants will remain unlinked and that it will not be any four, but precisely the first four consonants that will be linked to the empty slots. And as we have seen in the last line of (11), this prediction is fully borne out.

We now recall the proposal made above that syllable structure must be represented on a separate autosegmental plane. Formally this means that, instead of as in (12), Arabic noun plurals should be represented as in (13):

(13)



It turns out that skeleta having the syllabic structure given in (13) are found in other parts of Arabic morphology. For instance, in the verb inflection there are several forms with the syllable structure in (13) but differing in the vowels that appear in it. Thus, in the perfective active the vowels are all /a/, whereas in the perfective passive the vowels are /u/ in the first two syllables and /i/ in the third. In the light of these facts, McCarthy suggested that the vowels should also be represented on a separate autosegmental plane and that the skeleton, that is, the line of intersection of the different autosegmental planes, should, therefore, consist exclusively of empty slots. If this is done, then we can further postulate that in the plurals of (11) the vowel melody consists of the sequence /a i/, of which the latter is linked by a morphological rule to the head of the last rime. The universal linking convention will then link /a/ to the head of the first rime, and a language-particular spreading rule will link /a/ to the empty rime slots in the first syllable.

The different autosegmental planes—the vowels, the consonants, the syllabic structure, the morphological constituency, and so on—all intersect in the skeleton core, which can be viewed as being the counterpart of the spine of the spiral-bound notebook referred to at the beginning of this essay. I have tried to present a picture of this type of representation in Figure 1.

The three autosegmental planes shown in the figure are, of course, not the only ones encountered. As indicated above, tone must be represented on an autosegmental tier of its own. Hayes (1980) and Halle and Vergnaud (1987) have shown that stress should also be represented autosegmentally. Moreover, Poser (1982), Yip (1982), and others have provided telling arguments for representing still other phonological properties of words on separate autosegmental tiers. As noted above, to the best of my knowledge there are no promising alternative notations to the multi-tiered autosegmental representation that has been described here. Since there can be no doubt that speakers must have knowledge of the complicated and varied facts that autosegmental representations

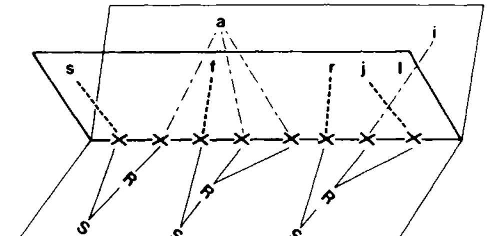


Fig. 1. A representation of the Arabic word *safaarij* 'quince' (pl.).

permit us to capture, I conclude that this multi-tiered three-dimensional representation properly reflects an aspect of linguistic reality, in particular the form in which speakers store information about the shape of words in memory.

Notes

I am grateful to Sylvain Bromberger for many improvements in the content and form of this paper, to Maggie Carracino for technical assistance, and to Rosamond T. Halle for help with the graphics. Shortcomings remaining in the paper are my responsibility exclusively. This work was supported in part by the Center for Cognitive Science, MIT.

1. The feature of voicing is the cue that distinguishes /p t k/ from /b d g/. The fact that the former but not the latter set can appear after /s/ might be captured by a special rule that redundantly supplies the feature of voicelessness to stops in position after /s/. Given such a rule, the language would not need to specify voicelessness for stops after /s/, since that information would be supplied by the rule.
2. The ordinary commonsense notion of syllable serves adequately for present purposes, so nothing further needs to be said about it at this point. We assume for present purposes that a morpheme is the smallest phoneme subsequence that has independent semantic or grammatical function. As shown in (7), the words *originality* and *interment* are each composed of three morphemes.
3. It is a canard put out, no doubt, by graduate students in Slavic who failed their qualifying examination that Polish has a particularly complicated syllable structure. While Polish syllable structure is far from simple, it is not significantly more complicated than that of English where syllables may end in strings such as [ɪksθs] or [ɛŋkθs] as in *sixths, lengths*.

References

Goldsmith, J. A.
1979 *Autosegmental phonology*. New York: Garland Press.
Halle, M., & J.-R. Vergnaud
1980 Three dimensional phonology. *Journal of Linguistic Research, 1,* 83–105.
Halle, M., & J.-R. Vergnaud
1987 *An essay on stress*. Cambridge, MA: MIT Press.
Hayes, B.
1980 *A metrical theory of stress rules*. Unpublished MIT doctoral dissertation distributed by Indiana University Linguistics Club.
Kahn, D.
1976 *Syllable-based generalizations in English phonology*. Unpublished doctoral dissertation, MIT.

Leben, W. R.
1978 The representation of tone. In V. Fromkin (Ed.), *Tone: a linguistic survey* (pp. 179–219). New York: Academic Press.
Levin, J.
1983 *Reduplication and prosodic structure*. Unpublished manuscript, MIT.
Liberman, M. Y.
1975 *The intonational system of English*. Unpublished doctoral dissertation, MIT.
McCarthy, J.
1979 *Formal problems in Semitic phonology and morphology*. Unpublished doctoral dissertation, MIT.
McCarthy, J.
1984 A prosodic account of Arabic broken plurals. In I. R. Dihoff (Ed.), *Current approaches to African linguistics*. Dordrecht: Foris Publications.
Pierrehumbert, J. B.
1982 *The phonology and phonetics of English intonation*. Unpublished doctoral dissertation, MIT.
Poser, W. J.
1982 Phonological representations and action-at-a-distance. In H. van der Hulst & N. Smith (Eds.), *The structure of phonological representation* II (pp. 121–158). Dordrecht: Foris Publications.
Pulleyblank, D. G.
1983 *Tone in lexical phonology*. Unpublished doctoral dissertation, MIT.
Rotenberg, J.
1978 *The syntax of phonology*. Unpublished doctoral dissertation, MIT.
Steriade, D.
1982 *Greek prosodies and the nature of syllabification*. Unpublished doctoral dissertation, MIT.
Williams, E.
1976 Underlying tone in Margi and Igbo. *Linguistic Inquiry, 7,* 463–484.
Yip, M.
1982 Reduplication and C-V skeleta in Chinese secret languages. *Linguistic Inquiry, 13,* 637–661.