Chapter 2: A Three-Tiered Theory of the Syllable

A universal theory of the syllable has, in our view, three specific tasks. First, it must specify the well-formed expressions of the theory. Thus, it provides an alphabet out of which syllable units are constructed together with a characterization of the permissable arrays of alphabetic units. Second, it must specify the parameters along which individual languages vary in their choice of syllable types. Third, it must characterize the class of language-particular rules which modify or extend the underlying syllable representations ("syllabification rules") and state how these rules are integrated into the general organization of the phonological component. We discuss each of these in turn in the following sections.

### 2.1 Well-formed Expressions

Let us consider, then, the first task. As we saw in Chapter 1, syllable trees consist of three-tiered representations, in which each tier has a certain vocabulary associated with it. The vocabulary of the first, or σ-tier, consists of the single element σ. The vocabulary of the second, or CV-tier, consists of the two elements C,V; and the vocabulary of the third, or segmental tier, consists of single-column phonetic matrices characterizing consonants and vowels in the usual manner. Well-formed strings on each tier consist of concatenations of the members of the alphabet defined on that tier.

Elements of neighboring tiers may be related in much the same way that syntactic elements are related in tree structures. In syntactic theory these relations are specified in terms of lines which are called "branches" while in multi-tiered phonological representations they are specified in terms of "association lines". The notion of "immediate constituent" holds in multi-tiered phonological representations just as it does in syntactic theory. Consider, for example, a tree of the following form:

<u>Canterbury Tales</u>, namely <u>cinamome: to me</u> (with primary stress on the penultimate syllable of each member), or such Modern English pairs as <u>sinister: minister</u> or <u>higgledy: piggledy</u>. In all such cases the rhyme depends upon the identity of the final stressed vowel and the entire string to its right within the word or verse line, including, crucially, all intervocalic consonants. In such pairs, there is clearly no single constituent which uniquely defines the notion rhyme.

The theory which we present in the following chapters represents an attempt to achieve maximal theoretical simplicity in the face of data of considerable intricacy and variety. Wherever possible, we have tried to avoid unnecessary additions to the theoretical apparatus of phonological theory by making maximal use of the notational distinctions provided by three-tiered syllable structures. Our strategy will be to demonstrate the descriptive power of an otherwise highly constrained phonological theory which incorporates the level of the CV-tier. We hope to show that, given this level, all of the phonological generalizations motivating the recognition of the syllable can be readily captured without the need for further notational apparatus.

Figure 1.

A
B   C
D   E
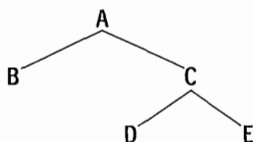
In structures like that in (1), B and C are said to be the immediate constituents of A because A immediately dominates each member of the string BC and nothing else.   Similarly, D and E are the immediate constituents of C. Furthermore, if a node A exhaustively dominates a string S and nothing else, S is said to be a member of the category A. Thus, the consonant /n/ and the sequence /d_3/ in (5) of Chapter 1 are both members of the category C and the string /nɪf/ is a member of the category syllable.

While the similarity between syntactic trees and syllable trees is instructive, there are several differences which should be kept in mind. First, the notion of tier plays no significant role in current syntactic theory. Thus, in the tiered representations presented here, the number of levels between the root and the terminals of a given structure is fixed at three. In syntactic trees no such fixed number is characterized. Second, while in syllable theory the elements of the alphabet are exhaustively partitioned among the three tiers of syllable representation (i.e. each tier has its own alphabet and shares it with no other tier), in syntactic theory the non-terminal symbols may appear at any non-terminal level of the tree. A third difference concerns the nature of the inter-tier associations. In phrase structure trees, any non-root node must be immediately dominated by one and only one node. In multi-tiered phonological representations, however, non-root nodes may be dominated by two or more elements.

2.2 Core Syllables

We now turn to the second task of syllable theory, which involves the characterization of the set of syllable types encountered at the earliest level of phonological derivations. It is our view that words are fully syllabified at the level of lexical representation: that is, syllable trees are not built up in the course of phonological derivations but are already present, fully formed, in the lexical representations that constitute the input to the phonological component.[1] This is the strongest possible claim we can make with respect to syllable representation since it suggests that syllable structure is assigned at a single level, uniquely specifiable for all languages. A theory in which this is true is also the simplest possible theory from the point of view of acquisition since it entails that the syllable structures encountered in surface representation will be similar or identical to those found in underlying representation.

There are two types of evidence in favor of this view. First, there are languages in which the postulation of syllable structure in the lexicon makes it possible to achieve a significant simplification of the phonological component. Efik, which we discuss in greater detail below, is one such language. A second form of evidence is psycholinguistic in character and involves lexical recall tasks. For example, the so-called tip-of-the-tongue phenomenon reported on by Brown and McNeill (1966) and subsequent writers can be best understood in terms of a lexical entry that is fully syllabified. The tip-of-the-tongue phenomenon arises when subjects recall suprasegmental properties of a given lexical item such as stress placement and the number of syllables, but cannot recall properties characteristic of the segmental level of representation. Since these suprasegmental properties presuppose syllabification, such data suggests that words are

---

1. We return in Section 2.4 to a discussion of the algorithm whereby syllable structure is supplied to lexical representations.

stored in fully syllabified form.

We propose that the primary set of core syllable types comprises the following sequences:[2]

Figure 2

 a. CV
 b. V
 c. CVC
 d. VC

These syllable types are not equal in status, however. Notice first that all languages (to the best of our knowledge) have the syllable type CV, while some languages lack each of the other three types. Furthermore, type (2d) is the most highly marked in the sense that any language that has (2d) must also have (2a-c). It is possible to derive these systematic relationships in the following way. We propose that the syllable type CV belongs to the grammar of all languages. This syllable type may be operated on to yield one or more of the other core syllable types by the following two operations:

Figure 3

 a. delete syllable initial C.
 b. insert syllable final C.

Any language may choose either, both, or neither of these two rules to expand its inventory of primary core syllable types. This system thus gives rise to the following types of languages:

---

2. Proposals similar in spirit to those presented below are set out in Abercrombie (1967).

Figure 4

Type I:	CV
Type II:	CV, V
Type III:	CV, CVC
Type IV:	CV, V, CVC, VC

Type I involves neither rule in (3). Type II involves Rule (3a) alone. Type III involves Rule (3b) alone. Type IV involves both (3a) and (3b). Notice, in particular, that the following hypothetical language types cannot be characterized by the rules of (3):

Figure 5

V, VC	CV, V, CVC → VC
CVC, VC	CV, VC → CVC, V
CV, V, VC	V, CVC → CV, VL
CV, CVC, VC	V, VC, CVC → CV

As far as we have been able to determine, each of the language types in (4) have been instantiated. For example, Type I is represented by Senufo, Type II by Maori, Type III by Klamath and Type IV by English. On the other hand, none of the types in (5) are instantiated.[3]

In addition to the parameters given in (3), languages may select among certain further options. First, some languages allow core syllable types to include sequences of consecutive V-elements:

---

3. This prediction of our theory clarifies and generalizes a claim made by Jakobson: "There are languages lacking syllables with initial vowels and/or syllables with final consonants, but there are no languages devoid of syllables with initial consonants or of syllables with final vowels." (Jakobson 1962, 526)

Figure 6
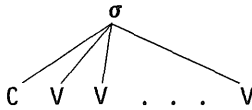
In such languages well-formed core syllable types may consist not only of CV and CVC, for example, but also of CVV and CVVC, and perhaps of CVVV and CVVVC and so on. Accordingly, we allow languages to freely select CV* as the representative of the primary core syllable (where V* represents one or more V-elements). If a language selects CV*, the operations of (3) apply as before to yield a set of derived core syllable types like those of (4) except for the fact that V* everywhere replaces V.
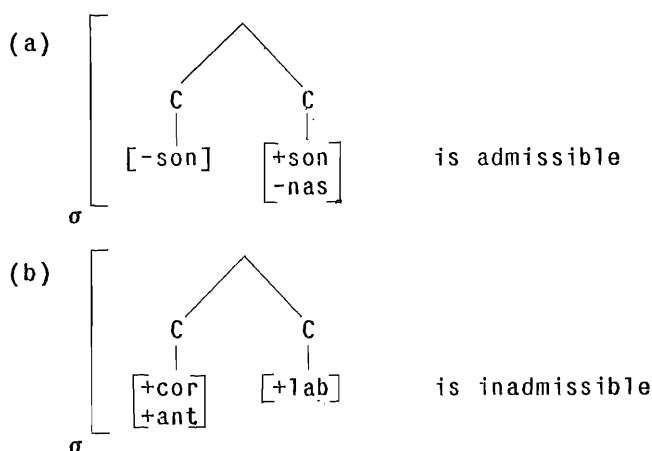
Similarly, some languages allow more than one C-element in initial or final position in the syllable. We represent this condition by C*. In order to expand our inventory of core syllable types to allow for the possibility of C* and V*, then, we will additionally define each language in terms of its maximal syllable, stated as a single expansion of the general schema C(*)V(*)(C(*)), where any occurrence of * may be replaced by an integer greater than 1. Thus, for example, the formula $C*V^2$ designates a language allowing syllable-initial clusters of any length and up to two vowels. The formula CV^2C^2 characterizes a language allowing two member consonant clusters in final position, but not in initial position in the syllable, and two member vowel clusters in the nucleus. Using this notation, we may characterize the English core syllable in terms of the maximal syllable $C*VC^2$, realized fully by the English word sprint.[4] The symbol C* will be used in place of the more specific C^2, C^3, etc. when the upper bounds on the length of a cluster follow from language specific constraints on

4. We assume, following arguments presented in Kiparsky (1981), that longer syllables may be created by a rule adjoining extrasyllabic coronal segments to the end of a preceding syllable to form such words as next and sixth.

sequences of vowels and consonants within the syllable, and need not be independently stipulated.

Constraints on cooccurrence within the syllable are represented, in the present theory, by positive and negative syllable structure conditions which, taken together, generate the set of well-formed core syllables for each language. The positive syllable structure conditions (PSSCs) state the general canonic form of well-formed consonant or vowel clusters in terms of sequences of natural classes. For example, the PSSC in (8a) states that initial clusters may contain, as their first two (or only) members, any obstruent followed by any oral sonorant:

Figure 7

(a)

```
              /\
             /  \
            C    C
            |    |
         [-son] [+son]    is admissible
                [-nas]
     σ
```

(b)

```
              /\
             /  \
            C    C
            |    |
         [+cor] [+lab]    is inadmissible
         [+ant]
     σ
```

In case there are no constraints on clusters (a situation unknown to us in the case of consonant clusters, but not uncommon in the case of vowel sequences), no PSSC need be stated. The negative syllable structure conditions (NSSCs), applying to the output of the PSSCs, specify certain subsequences within the syllable as ill-formed, thus performing a filtering operation. For example, the NSSC (7b) excludes such sequences as /tw,

dw, sw, zw/ from the set of consonant clusters generated by PSSC (7a).
NSSCs range not only over subsequences dominated by Cs alone and
subsequences dominated by Vs alone, but also subsequences dominated by
both C and V elements. The empirical justification for the distinction
between PSSCs and NSSCs will become apparent in our discussion of
syllable initial clusters in English, in Section 2.5.

2.3 Core Syllable Associations

Permissible core syllable associations between elements of the CV-tier
and elements of the segmental tier are determined in part by universal
principles. Unless otherwise stipulated in the grammar or lexicon of a given
language, V-elements of syllable structure are freely allowed to dominate
[-consonantal] segments, and C-elements are freely allowed to dominate
both [+ consonantal] segments and [+ high, -consonantal] segments. Other
associations are possible only when admitted by language specific rules.
For example, some languages allow post-vocalic consonants to be
dominated by V if they are tautosyllabic with the preceding vowel.
Languages of this type are those in which the "mora" is a unit of prosodic
organization capable of bearing pitch or tone contrasts, and include
Lithuanian, Japanese and Akan. As a further example, some languages
allow tautosyllabic VC sequences to dominate single consonant or vowel
segments. English is such a language in our view, since it has core syllables
of the following type:

Figure 8

(8a) gives the underlying representation for the word hide which acquires its

surface quality through the operation of rules of vowel shift and
diphthongization. In core syllable representation, the vowel of this word is
represented as a single segment assigned to two adjacent positions on the
CV-tier, indicating that the vowel is long.   (8b), giving the underlying
representation of bird, reflects the phonetic fact that the syllabic nucleus of
this word does not contain a vowel such as we find in similar words like
breed and beard. Rather, the portion of the utterance which occurs between
the initial and final consonant presents a steady-state, r-colored segment
occupying the syllable peak and consisting of two units of timing.[5]
Accordingly, we represent this element as a single phonetic matrix
associated with two elements of the CV-tier.

As we observed in Chapter 1, the "bipositional" representation of long
vowels and consonants given in (8) allows us to formulate a unitary
characterization of syllable weight.   Further motivation for such
"bipositional" representation in English comes from an observation due to
Selkirk (1978) regarding constraints on syllable final clusters. In the first
place, Selkirk notes that in syllable final clusters of the form VCCC, where
the vowel is short, the final member must be [+coronal]. Thus, while next
/nekst/ and glimpse /glɪmps/ are well-formed, words like like *nexp
/neksp/ and *glimpf /glɪmpf/ are not. In the case of V̄CC clusters, where
the V̄ is a long vowel or diphthong, the final consonant is subject to the
same constraint; namely, it must be [+coronal]. Thus, we find pint /paynt/
and fiend /fiynd/ but not /paynk/ or /fiymp/. It is clear that long vowels
and diphthongs are functioning equivalently to VC sequences. If we express
this equivalence by representing long vowels and diphthongs as suggested
in (8), we may formulate a single constraint to the effect that the third
member of a syllable final C cluster must dominate [+coronal] segments

5. Cf. Klatt (1975) for measurements showing that stressed and unstressed syllabic /r/ in
English have twice the length of stressed and unstressed non-low lax vowels, respectively, for

only.[6]

In core syllable representations not all associations are one-to-one. We have just given examples of many-to-one associations in Figure (8) above. Additional one-to-many and many-to-one configurations are illustrated below:

Figure 9

a.
```
    C
   / \
  t   ʃ
```

b.
```
  C   C
   \ /
    s
```
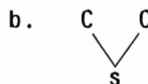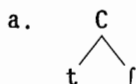
As remarked earlier, the elements of the CV-tier are interpreted as corresponding to the timing units of speech production at the sub-syllabic level. Thus, a single C represents a single unit of timing, while a sequence CC represents a double timing unit. Accordingly, (9a) is interpreted as an affricate (i.e. a single, internally complex segment as in English church). We now have a natural way of distinguishing such minimal pairs as Polish czy 'whether' and trzy 'three', in which the affricate [tʃ] of the first example is acoustically and perceptually distinct from the otherwise identical stop-fricative sequence [tʃ] of the second (Brooks 1965):

---

6. In terms of the analysis suggested in footnote 4, in which the maximal core syllable of English is C^*VC^2, this amounts to the requirement that extrasyllabic C elements are adjoined to the preceding syllable only if they dominate [+coronal] segments.

Figure 10

'whether'   'three'

(9b) represents a geminate continuant such as is found in the core syllables of some languages. In English, geminate consonants are found in derived syllable structure in the casual pronunciation of phrases such as of cour[ss]ey do (i.e. of course they do). (See Shockey 1977 for further discussion.)

The availability of syllabifications involving many-one relationships provides a ready account for certain consonantal distributions which are otherwise puzzling. Consider, for example, the account of Efik given in Welmers (1973, 74-6). Welmers notes that this language has three sets of consonants which are differentiated by their distribution. Set 1 appears in both word-initial and intervocalic position; set 2 appears only in word-final position; and set 3 appears only intervocalically. Set 1, for example, is represented by the consonant [kp], set 2 by the consonant [p], and set 3 by the bilabial flap [B]. Thus we find, in word-initial position, examples like kpók 'cut up' (set 1); in word final position examples like dép 'buy' (set 2); but in intervocalic position examples from both sets 1 and 3, such as èkpàt 'bag' (set 1) and sɨBé 'cut down' (set 3).

Welmers notes that the consonants [kp] and [p] are in complementary distribution. Since Efik has the voiced phonemes /b/ and /d/ beside the voiceless phoneme /t/, he suggests, on grounds of pattern congruity, that [kp] is reasonably interpreted as representing a basic labial phoneme /p/ characterized by velar coarticulation. A phonetic rule removes the velar articulation in syllable-final position, thereby accounting for the final sound

But what of the bilabial flap [B]? Welmers notes that this flap is in complementary distribution with both [kp] and [p]. In particular, while it occurs intervocalically like [kp], it occurs only after vowels whose quality is elsewhere characteristic of vowels in closed syllables. To further illustrate this complementary distribution, Welmers provides the example <u>dwȯp ė bà</u> 'twelve (ten plus two)' which varies with the fast speech form <u>dwòBė bà</u>, where the closed syllable allophone of /o/ occurs in both cases.

In order not to proliferate the vocalic register of Efik, Welmers proposes, instead, to recognize the syllable as an entity in phonological description.[7] He then proposes that the flapped consonants that follow vowels showing closed syllable variants in words like <u>sɨBé</u>   are,   in   fact, underlyingly ambisyllabic, as in (11a). Furthermore, Welmers proposes that the syllable structure of <u>dwòBė bà</u> in rapid speech be represented as in (11b), where the dashed line indicates the effect of a rule affiliating a word-final consonant to a vowel-initial syllable:

Figure 11
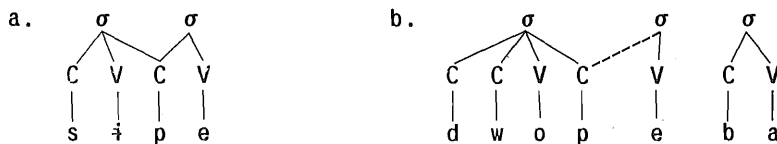
In these representations the ambisyllabic character of the medial /p/ is indicated by the fact that it is dominated by two syllable nodes. This /p/ undergoes the flap rule that applies to all ambisyllabic consonants, becoming [B].

---

7. In particular, he posits a unit to mark syllable division, although he does not make clear precisely how this unit is to be formally introduced. In what follows we adopt Welmers' basic insight but reformulate his analysis in terms of tree structures.

There are two consequences of Welmers' solution which argue overwhelmingly in its favor. First, the ambisyllabic representation of the medial /p/ causes the preceding syllable to be closed and thereby provides the requisite syllabic environment for the allophonic variation observed in the vowel system. Without this, it would be necessary, as Welmers notes, to virtually double the vowel inventory of Efik. Secondly, the representations in (11) provide the requisite syllabic environment to account for the allophonic variation observed between [kp], [p] and [B], assuming that the flap rule applies only to ambisyllabic consonants. Hence, the consonantal segments need not proliferate in the inventory of Efik phonemes.

2.4 Core Syllable Division

Elements of the CV-tier are grouped into core syllables corresponding to the core syllable inventory selected by the language in question. This grouping is constrained by the following principle:

Figure 12

The Onset First Principle

a. Syllable-initial consonants are maximized to the extent consistent with the syllable structure conditions of the language in question.

b. Subsequently, syllable-final consonants are maximized to the extent consistent with the syllable structure conditions of the language in question.

These two principles, which are adapted from rules formulated for English by Kahn (1976), apply in the order given. Hence, given a string of the form VCV, where both VC and CV are well-formed syllables, the syllable division is V-CV. Similarly, given a string of the form VCCV, where CCV is a

well-formed syllable, the syllable division is V-CCV, even though VC (or VCC) may be well-formed syllables as well. Thus, the Onset First Principle implies that given alternative syllable divisions, languages will select that which maximizes syllable-initial consonant sequences.

Principle (12) might be interpreted in any number of empirically equivalent ways; for example, as conditions holding of fully formed core syllable representations, or of partly formed core representations which are then built up with reference to the language-particular syllable structure conditions to create fully formed core syllables. We adopt the latter interpretation here. In particular, we assume that fully formed core syllables are constructed in the following manner.

Figure 13

a. V-elements are prelinked to σ's.

b. C-elements to the left are adjoined one by one as long as the configuration resulting at each step satisfies all relevant syllable structure conditions.

c. Subsequently, C-elements to the right are adjoined in the manner described in (b) above.

To illustrate this algorithm, which builds up syllables in onionlike fashion from the center outward, we offer the following example:

Figure 14
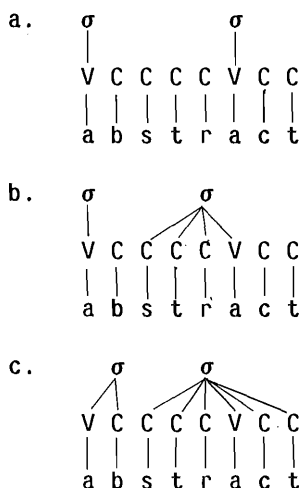
(14a) represents the initial configuration as defined by (13a). (14b) conflates three steps in accordance with (13b). The first of these creates the syllable /ra/, the second the syllable /tra/, and the third the syllable /stra/. As we see in Section 2.5, each of these intermediate configurations satisfies the syllable structure conditions of English pertaining to initial clusters. /b/ cannot be adjoined to the resulting configuration, however, since the sequence /bstr/ does not satisfy the conditions on syllable-initial position in English. The final configuration (14c) is then created in a similar manner in accordance with (13c).

According to the principles given so far, long clusters of consonants may not be exhaustively parsable into core syllables in some cases. Thus, some consonants in such clusters may remain extrasyllabic. An extrasyllabic consonant is one which is not a member of any syllable. Typically, such consonants are separated from neighboring consonants by short neutral or voiceless vowels and are historically susceptible to processes which either eliminate them or incorporate them into well-formed syllables by means of processes such as vowel epenthesis, sonorant vocalization and metathesis. English has several such examples. The usual

pronunciation of knish in Cambridge, Massachusetts, for example, inserts a short, voiceless schwa after the k. However, this is not a full schwa as is evidenced by its near-minimal contrast with the word canoe. Other examples of extrasyllabic consonants in English include the initial consonants of the names Pnin, Knievel, Zbigniew, Khmer Rouge, Dvořák, Phnom Penh, Dmitri and Gdansk, in usual pronunciations, not to speak of the b in common renderings of the name of the former Iranian minister Ghotbzadeh. We represent extrasyllabic consonants as follows:

Figure 15
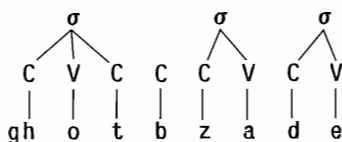
We claim that the segment b in an example such as this is a member of neither the first nor the second syllable, either underlyingly or on the surface. Extrasyllabic segments occurring in core syllable representations often play an important role in the syllable-conditioned phonology of languages. We return to this topic in Chapters 3 and 4.[8]

---

8. While our theory allows extrasyllabic C-elements, it does not permit extrasyllabic V-elements. This restriction follows from the logic of our theory. Given, first, that elements dominated by V are interpreted as belonging to the syllable peak, and second, that the notion of syllable peak is interpretable only with respect to a syllable of which the peak is a constituent, it follows that any notion of "extrasyllabic peak" would be conceptually anomalous

2.5 A Case Study: Initial Consonant Clusters in English

We illustrate and develop the theory presented so far with a partial account of the core syllable in English. Here we consider the constraints on syllable-initial consonant sequences.

In terms of the set of parameters presented in Section 2.2, English is a Type IV language which permits consonant clusters (C*) in syllable-initial (and syllable-final) position. Specifically, as noted above, English conforms to the maximal syllable formula C^*VC^2. Initial consonant clusters are subject to a number of constraints. In the following figure we display in columnar format the well-formed and ill-formed initial clusters of English core syllable representations. The rows specify the first members of such clusters and the columns specify the second members. A " + " indicates that the row/column pair is a well-formed syllable-initial cluster, while a "-" indicates that it is not:

Figure 16

		w	l	r	p	t	k	m	n	f	θ
a.	p	-	+	+	-	-	-	-	-	-	-
	b	-	+	+	-	-	-	-	-	-	-
	f	-	+	+	-	-	-	-	-	-	-
b.	t	+	-	+	-	-	-	-	-	-	-
	d	+	-	+	-	-	-	-	-	-	-
	θ	+	-	+	-	-	-	-	-	-	-
c.	k	+	+	+	-	-	-	-	-	-	-
	g	+	+	+	-	-	-	-	-	-	-
d.	s	+	+	-	+	+	+	+	+	?	-
	ʃ	+	+	+	?	?	-	?	?	-	-

The following consonants do not occur as the first member of a cluster: /w, l, r, y, h, ð, z, ʒ, tʃ, dʒ, n, m, v/.

It will be noticed that no clusters containing /y/ have been included in the above chart. Following the analysis given by Levin (1981), we assume that /y/ in words such as pure or cue is not present underlyingly but is inserted by rule before the vowel /ɨ/, which becomes the surface /uw/.[9] This analysis receives support from two independent considerations. First, except for a few words of foreign orgin (for example, proper names such as Tokyo), /y/ appears in clusters before no other vowels than /uw/. This anomaly is explained under Levin's proposal by the fact that Cy clusters have no other source than the insertion rule. Secondly, /v/ and /m/, which do not otherwise appear in word-initial clusters in English, appear before /y/ in such words as muse and view. Under Levin's analysis, this /y/ is not underlying and thus these words do not constitute exceptions to the generalization that voiced fricatives and nasals do not occur as the first members of syllable-initial clusters in core syllable representations.

Let us examine more closely the cooccurrence restrictions given in (16). If we exclude for the moment the clusters of (16d), we observe that only obstruents occur as the initial member of these clusters, and only oral sonorants occur as the second member. On the basis of these observations we may formulate a positive syllable structure condition for two-member syllable-initial clusters as follows:
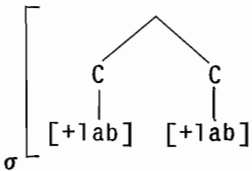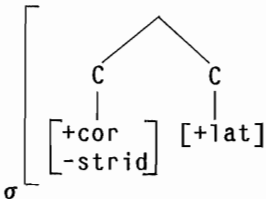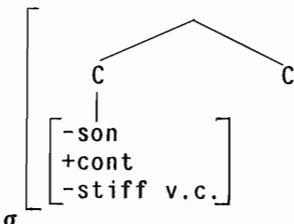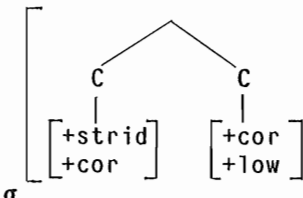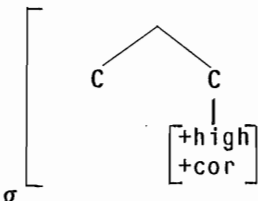
Figure 17

---
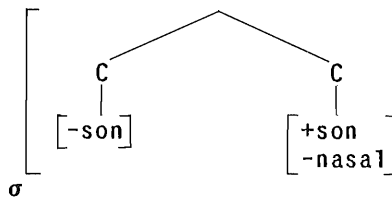
9. The rule of y-insertion does not apply in words like poor since there is no /ɨ/ at any point in their representation.

This condition is, needless to say, specific to English since many languages allow initial clusters that do not conform to it. It therefore forms a part of the set of rules constituting the phonological component of English.

(17) will correctly generate not only the clusters of (16a-c) but also the clusters in the first three columns of (16d). On the other hand, (17) is overly general in that it admits a number of clusters that are systematically excluded in English, as indicated by the minus entries in the table. Further examination shows that the exclusions are not random but fall into phonologically well-defined classes: labials may not be followed by /w/; nonstrident coronals are not followed by /l/; voiced fricatives are everywhere excluded as the first member of clusters; the sequence /sr/ is excluded; and finally, the second member of a cluster cannot be /y/. These observations motivate the following set of <u>negative syllable structure conditions</u> for English syllable-initial clusters: [10]

---

10. It should be noted that, following a suggestion by Halle and Vergnaud (1979), (18d) below can be eliminated in favor of a rule of the following sort: s $\rightarrow$ $\int$ / ___ r

Figure 18 - Negative Syllable Structure Conditions

condition: excludes:

a.   *

[+lab]   [+lab]   bw, pw, fw

b.   *

$\begin{bmatrix}+cor\\-strid\end{bmatrix}$ [+lat]   dl, tl, θl

c.   *

$\begin{bmatrix}-son\\+cont\\-stiff\ v.c.\end{bmatrix}$ vw, zl, ...

d.   *

$\begin{bmatrix}+strid\\+cor\end{bmatrix}$ $\begin{bmatrix}+cor\\+low\end{bmatrix}$ sr

e.   *

$\begin{bmatrix}+high\\+cor\end{bmatrix}$ py, ty, gy, ...

Let us now consider how the remaining s-initial clusters of (16d) are to be accounted for. Excluding for the moment the marginal cluster /sf/, which occurs only in a few items of the learned vocabulary, we find two further groups of s-clusters: /sp, st, sk/ and /sm, sn/. It should be clear that there is no obvious way to extend the positive syllable structure condition given in (17) to accommodate these clusters that does not involve the use of rather powerful notational devices permitting the expression of discontinuous dependencies.

As a first hypothesis, we might propose to account for these clusters in terms of the following additional positive conditions:
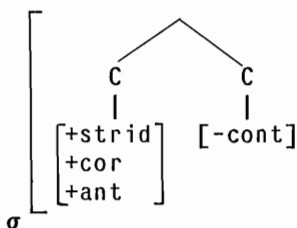
Figure 19
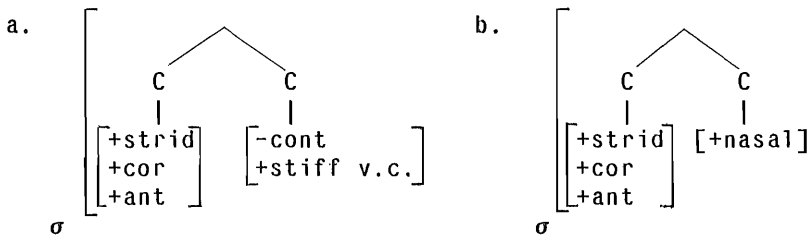
The first of these conditions admits /sp, st, sk/ while the second admits /sm, sn/. Notice, however, that these two conditions are partly similar, a fact which raises the question whether they may be simplified further. It appears that it is unnecessary to qualify condition (19a) by the feature [+ stiff v.c.]. Consider alternations such as the following:

Figure 20

```
believe / disbelieve
direct / misdirect
governor / ex-governor
hardball / softball
hug David / kiss David
lime green / grass green
```

In the first member of each pair, the initial stop of believe, direct, governor, ball, David and green can be pronounced with voicing throughout the articulation. The corresponding stop in the second member of each pair, however, is obligatorily devoiced (Davidsen-Nielsen 1974), except in the case of overly precise pronunciations in which a pause is inserted between the two constituents. This rule of devoicing, which devoices an oral stop after an unvoiced sound both within the same word and across word boundaries, is fully general and productive in English, and independently accounts for the absence of [sb, sd, sg], on the assumption that this rule is a non-cyclic rule in the sense of Mascaró (1976). There is, therefore, no need to incorporate this restriction in the set of syllable structure conditions. We may, accordingly, remove the specification [+ stiff v.c.] from (19a). Observing now that /p, t, k, m, n/ all share the feature [-continuant], we may collapse the two statements of (19) into the following:

Figure 21 - Positive Syllable Structure Condition
(s-clusters)

$$
\sigma \left[\begin{array}{c} \overset{\displaystyle\frown}{\underset{\begin{bmatrix}+\text{strid}\\ +\text{cor}\\ +\text{ant}\end{bmatrix}}{\overset{|}{C}}} \quad \underset{[-\text{cont}]}{\overset{|}{C}} \end{array} \right]
$$

Turning now to three-member syllable-initial clusters, the important observation is that all such clusters are made up of overlapping sub-sequences of well-formed two-member clusters.   The following three-member clusters are attested. [11]

Figure 22 - Three-member Syllable-initial Clusters

	w	l	r
s p	-	+	+
s t	-	-	+
s k	+	+	·+·

Thus, for example, /spl/ is made up of the (overlapping) well-formed syllable-initial sequences /sp/ + /pl/. All sequences of three consonants that <u>cannot</u> be so parsed are excluded in English. Accordingly, /snr/ is not a possible three-member cluster, since /nr/ is not a possible two-member cluster.

It should be clear at this point that the set of occurring three-member clusters (and only these) is generated by the rules and principles given so far. Any three-member cluster, A B C, is analyzable into two two-member clusters, A B and B C, each of which necessarily satisfies either (17) or (21) and neither of which satisfies any of the negative syllable structure conditions (18). No other three-member clusters can be created under the algorithm stated in (13).

---

11. The cluster /skl/ is marginal, occurring in the learned vocabulary only: <u>sclerosis</u>, <u>sclaff</u>, <u>sklent</u> and in certain proper names such as <u>Sklar</u>. We consider its marginality accidental rather than systematic, although our decision here has no incidence upon the rest of our analysis.

Notice further, however, that while all well-formed three-member clusters can be analyzed into pairs of well-formed two-member clusters, it is not necessarily the case that any pair of well-formed two-member clusters, AB and BC, combine to form a well-formed three-member cluster, ABC. This is because of the requirement in the algorithm (13) that not only positive but also negative syllable structure conditions be satisfied. Thus, while both /st/ and /tw/ are well-formed two-member clusters in English, they cannot combine into a well-formed three-member cluster /stw/ because that cluster is ruled out by a negative syllable structure condition specific to this sequence.[12]

This system also makes the prediction that syllable-initial clusters in English consist of no more than three members. To see this, consider any arbitrary string of four consonants, ABCD. If consonants A and B satisfy condition (17), B and C satisfy neither (17) nor (21). Hence the string cannot be syllabified. Suppose, on the other hand, A and B satisfy condition (21). In this case consonants B and C can only satisfy condition (17). C and D, now, can satisfy neither condition (17) nor (21) and the string cannot be syllabified. Thus, it is an automatic consequence of our analysis that English contains initial clusters that are at most three segments long. This requires no independent statement in the grammar of English.[13]

To complete our analysis of initial consonant clusters in English core syllables, we consider first the observation that a few English words begin with /sf/ clusters. These include sphere, sphynx, sphragistics and a few proper names of foreign origin such as Svedberg, Sforza and Sfravara. This is most simply stated as a separate positive condition permitting the voiceless labial fricative /f/ after /s/;   /sfr/ clusters will then be automatically generated due to the independent acceptability of /fr/

---

12. In Clements and Keyser (1981) use was made of a Parsing Convention to account for the relation between two and three-member clusters in English. The present account eliminates the need for this convention.

13. We are indebted to Wendy Lewis for pointing out this consequence of our analysis to us

clusters in core syllable representations. Secondly, the analysis so far does not accommodate certain /ʃ/-initial clusters which are fully integrated into the phonological systems of many monolingual English speakers, not only in words such as schwa, shrub and Schlesinger (as is predicted by (17)), but also in s[h]piel, shtick, schmalz and schnapps, which our analysis does not so far generate. The latter group of clusters may be admitted by a minimal generalization of (21) eliminating the feature [+anterior]. Note that this generalization predicts the presence of three-member /ʃ/-initial clusters. The apparent absence of such clusters is probably best to be regarded as accidental in view of the relatively small number of words beginning with /ʃ/-initial clusters in the first place.[14]

Summarizing our analysis of initial consonant clusters in English core syllables, we have postulated two positive syllable structure conditions (17) and (21) (as well as a third, not formalized, for /sf/ clusters) and five negative syllable structure conditions (18a-e). These conditions, taken together, define the set of systematically possible (or well-formed) underlying initial consonant clusters in English core syllables.

As a further consequence of our analysis, we can explain the dual status of affricates with respect to syllable structure constraints. The affricate /tʃ/ behaves as a single segment in its ability to begin a syllable on its own; if we had treated it as a phoneme sequence, /t + ʃ/, it would have violated the positive syllable structure conditions (17) and (21). On the other hand, this segment behaves as a cluster with respect to the fact that it may not adjoin to a following liquid or glide to form any of the sequences /tʃr, tʃl, tʃw/. This fact follows automatically from the principles stated so far. Thus, note that a representation such as the following satisfies neither of the positive syllable structure conditions (17) or (21):

---

14. These clusters may not be totally absent, as witnessed by those speakers who allow initial /ʃ/ in such words as spritz, strudel, Strauss or Springer.

Figure 23
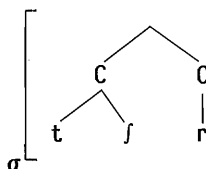
The representation of affricates as <u>single</u> segments on the CV-tier and as <u>dual</u> segments on the segmental tier enables us to account for their ambiguous status with respect to syllabification rules.

Our distinction between positive and negative conditions has been founded on the so far tacit premise that the language learner will learn the simplest set of conditions that characterizes the well-formed clusters of his or her language. This criterion is satisfied by the analysis given above. As the reader may verify, we cannot reformulate any of our existing positive conditions as negative conditions, or any of our negative conditions as positive conditions, without increasing the formal complexity of the analysis as determined by the usual evaluation metrics. In addition, our analysis has been carried out under the further assumption that syllable structure conditions are properly stated in terms of a highly restricted vocabulary which does not include the abbreviatory conventions or other notational devices that have been proposed elsewhere in phonology. Thus, while we could have collapsed our two positive conditions or any two or more of our negative conditions by the use of angled brackets subject to Boolean conditions, we have not done so due to our belief that such abbreviations do not reflect genuine linguistic generalizations concerning core syllable structure.

These methodological premises lead us to an interesting empirical consequence. Our analysis has recognized two classes of "systematically excluded" clusters: those that are not generated by the positive conditions, and those that are excluded by the negative conditions. These two classes of clusters are of a strikingly different character. The clusters not generated

by the positive conditions are totally deviant sequences. These include such sequences as /zb, bd, dv, bz/, which, as we have already seen in our discussion of extrasyllabicity, cannot be pronounced as tautosyllabic sequences by phonetically untrained English speakers. Further such sequences as /rb, wt, lr, nk/, etc., which are equally deviant, are also excluded by the positive conditions. Exceptions to our generalization, we believe, are only apparent. Thus, the initial sound of the name Nkrumah is pronounced as a separate syllable by most monolingual English speakers, while the initial sequences of words such as tsetse and tsar are probably to be considered as affricates (which are not, as we have seen, excluded by the syllable structure conditions). If we consider, on the other hand, the further class of "systematically excluded" syllables which is excluded by the negative syllable structure conditions, we find a very different state of affairs. These sequences, in large part, are not only easily pronounceable by the phonetically untrained speaker but occur with some frequency in words of foreign origin. Examples include fjord, Kyoto, bwana, pueblo, Tlingit, Vladimir, zweiback, vroom, voyeur, foie gras, svelte and so on. We believe that the traditional confusion regarding whether clusters such as these should be accounted for in a general account of the English syllable[15] reflects their intermediate status between fully acceptable and fully deviant clusters. This difference is formally characterized in the present approach by the fact that they are excluded by negative, rather than positive

---

15. See Algeo (1978) for a useful review of the literature.

conditions on syllable structure.[16]

2.5.1 Excursus: The Acquisition of Consonant Clusters

There are very few detailed longitudinal studies of the acquisition of phonology and only one, to our knowledge, which bears directly on the acquisition of syllable-initial clusters in English, that of Smith (1973). Smith found that the acquisition of initial clusters was characterized by a highly systematic progression of stages, each exhibiting fewer restrictions on consonant cooccurrence than the previous stage. Interestingly, three-member clusters were reported to appear at the same time as the two-member clusters of which they were constituted; thus, /spl/ appeared at the point when /pl/ was already present and /sp/ had just been acquired (p. 169).

This result is of theoretical interest for language acquisition. Consider, in particular, the view of acquisition expressed in Jakobson (1968). Jakobson observed that, all else being equal, syntagmatically simple sequences are acquired earlier than syntagmatically more complex ones. Specifically, Jakobson claimed that within a word, the maximum number of

---

16. In an earlier study Greenberg and Jenkins (1964) examined the linguistic distance of monosyllables from well-formed English along a four-point scale as follows:

  1. existing word: well-formed existing cluster - /strʌk/
  2. nonexisting word: well-formed existing cluster - /strib/
  3. nonexisting word: well-formed nonexisting cluster - /stwip/
  4. nonexisting word: ill-formed nonexisting cluster - /gvsurs/

An implication of their findings was that "...the perceptual sensitivity of the subjects to [linguistic distance from English] is uniform across the entire length of the scale as opposed to being highly sensitive for small departures from English and less and less sensitive for differences between very distant 'words'" (p. 167). Note that Greenberg and Jenkins did not test the validity of their four-point scale directly, but rather the validity of a scale of distance from English as measured by the number of ways we can perform phoneme substitutions in four-phoneme "words" and arrive at existing words (the more ways, the closer the word is to English).

  Our results motivate the recognition of a subdivision in the fourth point of this scale; namely, a subdivision into ill-formed clusters that are excluded by negative conditions (pwin) and those that are excluded by positive conditions (pkin).

phonemes, the number of their possibilities of distribution, and the maximum number of phonemic distinctions all increase by degrees in child language acquisition (op.cit. pp.85-86). While this claim is probably correct as a broad and general one, it does not allow us to accommodate Smith's results in any straightforward way, since it would predict that /sp/ and /pl/ clusters would be acquired earlier than /spl/ clusters. On the assumption that a child's knowledge of English phonology includes knowledge of the positive syllable structure conditions (17) and (21), however, Smith's result is what we expect to find. Once a child has acquired two-member clusters like /sp/ and /pl/, this condition predicts that the child has concurrently acquired the competence to produce three-member clusters such as /spl/.

Due to the lack of relevant studies, we do not know whether all children show the same pattern of behavior as Smith's subject. In fact, we would not be surprised to find a lag in the production of three-member clusters in some subjects caused by performance difficulties due, perhaps, to differing rates of physical maturation. Until such time as more data is available, we can do no more than note that Smith's otherwise anomalous observation finds a natural explanation within our framework.

2.6 Syllable Transformations

Up to this point we have described how languages construct a core syllable inventory by selecting among the limited number of parameters provided by the theory. We have so far said nothing about the third task of syllable theory, that of characterizing the class of processes which transform core syllable representations into the frequently distinct set of surface syllable trees. It is obvious that such processes play an important role in the phonological systems of many languages; see, for example, the discussion of Klamath syllabification given in Chapters 4 and 5. Some of these processes affect elements of the segmental tier exclusively, while some affect higher levels of representation. We number among these operations

the following: the insertion and deletion of association lines, and the insertion, deletion, substitution and metathesis of segments on the CV-tier.

Operations affecting syllable structure are governed by certain general conventions that apply automatically to their output. Their general effect is to preserve phonological well-formedness throughout derivations. Thus, phonological rules which apply so as to create ill-formed syllables will generally induce the operation of applicable well-formedness conditions. We shall not illustrate all of these conventions here as we shall have ample opportunity to do so in the discussion to follow. However, we shall pause to mention one particularly pervasive convention, that which deals with resyllabification.

In many languages we find that the constraints on syllable structure hold not only at the level of core syllable representation, but also after the application of phonological rules. We may express this fact by assuming that in such languages the rules of syllabification continue to apply throughout phonological derivations in accordance with the Resyllabification Convention given below:

Figure 24 - Resyllabification Convention

The output of every rule is resyllabified according to
the syllable structure rules examined up to that point in
the derivation.

We interpret this convention to mean that all association lines between C-elements and σ's, as well as all floating σ's, are erased, and that the resulting configuration is resyllabified according to the algorithm stated in (13).

what of
domesticity ?

Whenever, in the course of a derivation, resyllabification is called for, we assume that all syllabification rules encountered previously in that derivation will reapply. This includes not only those syllabification rules belonging to the core syllable component, but those belonging to the phonological component as well, and includes affiliation rules of the type to be examined just below and in Chapter 4. We add two further comments on the Resyllabification Convention. First, this convention may cease to operate late in derivations in some languages, where, for example, vowel elision may create consonant clusters, extrasyllabic segments or other such sequences that are ill-formed in earlier stages of derivation. We propose, then, that individual grammars may specify a point in the set of ordered rules at which the Resyllabification Convention becomes inoperative; indeed, some languages may not make use of the Resyllabification Convention at all. Second, there is evidence that in some languages at least, syllabification takes place initially at the domain of the morpheme and subsequently at the level of the stem or word. Resyllabification across word boundaries, on the other hand, is normally optional, and may differ in some respects from initial syllabification. For some discussion see Kiparsky (1979), Harris (1983) and Odden (forthcoming).

We turn finally to a class of rules which effect association lines by explicitly introducing, deleting or respecifying them. Such is the effect, for example, of the rule of "Right Capture" (Kahn 1976), which creates ambisyllabic segments in words such as <u>Monica</u>:

Figure 25
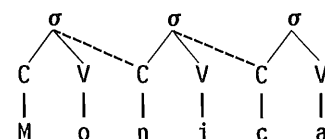
This rule, which affects intervocalic consonants in unstressed syllables, has

the sole effect of introducing the association lines indicated by the dashed lines in (25). It accounts for the phonetic fact that the first and second syllables of <u>Monica</u> are closed and that the second and third consonants are ambisyllabic. We may assume, harmlessly, that the Resyllabification Convention (24) applies to the result of rules of this type just as it does to all others. Since it requires that the structure be resyllabified according to the syllable structure rules examined up to that point in the derivation (thus, including the affiliation rule illustrated in (25) itself), its effect is vacuous, and (25) remains unchanged. We refer the reader to Kahn (1976) for a fuller account of the role of affiliation rules in English, including rules not mentioned in the above discussion.[17]

---

17. The rules of English given in this chapter predict that /s/ is syllabified rightward in sequences of the form: VsC_1V. This syllabification seems correct for most examples. Kahn (1976) points out that the location of main stress in nouns such as <u>amnesty</u>, <u>orchestra</u> and <u>pedestal</u> indicates that the penultimate syllable is light, and hence that the /s/ initial cluster is tautosyllabic with the following vowel. Other examples, however, seem to suggest that /s/ clusters are heterosyllabic in some cases; for example, <u>asbestos</u>, <u>manifesto</u>, <u>clandestine</u> and <u>fiesta</u>, which receive penultimate stress. However, numerous further examples like <u>allegro</u>, <u>vanilla</u>, <u>regatta</u> and <u>Kentucky</u> exhibit penultimate stress even though their penultimate syllables are light. To attract penultimate stress to these syllables, we may provide these words with idiosyncratically marked closed syllables in the lexicon, or, alternatively, provide them with lexically marked stress. The same solution can then be extended to words like asbestos.

CV PHONOLOGY
A Generative Theory of the Syllable

George N. Clements
Samuel Jay Keyser

This book was printed and bound in the United States of America.