The CVX Theory of Syllable Structure*

San Duanmu

University of Michigan

1. Introduction

This chapter is an outline of the CVX theory of syllable structure, which is discussed in detail in Duanmu (2009). The CVX theory claims that in all languages the maximal syllable size is CVX, which can be CVV, such as [hau] *how* and [bi:] *bee*, or CVC, such as [bɛt] *bet*.

Many words larger than CVC may spring to mind, such as *smile* [smail], which is CCVVC, *texts* [tɛksts], which is CVCCCC, and *sprints* [sprɪnts], which is CCCVCCC. Before we turn to them, let us consider the full structure of CVX, which is shown in (1).

(1) The maximal syllable: CVX

```
  σ          Syllable
 /\
O  R         Onset/Rime
|  /\
X XX         Timing slots (X-slots)
```

For familiarity, I call the first slot C or the onset, the second slot V or the nucleus, and the third slot X or the coda. I also assume that VX forms a constituent called the rime, although this assumption is not crucial for what the maximal syllable size is (see Davis 1988, who questions the rime). Some examples are shown in (2), where I follow a standard assumption that a lax vowel fills one timing slot but a tense vowel can fill two. (However, I shall argue in section 6 that a tense vowel can fill one timing slot when it is in a VC rime. For example, the American English [ɒ] fills two timing slots in *law* [lɒ:] but one in *almost* [ɒlmost].)

(2) Sample syllables

Transcription	C	V	X	
[bæt]	[b	æ	t]	*bat*
[bai]	[b	a	i]	*buy*
[wi:]	[u	i	i]	*we*
[jæm]	[i	æ	m]	*yam*
[æn]	[	æ	n]	*Ann*
[ðə]	[ð	ə	]	*the*
[ə]	[	ə	]	*a*
[n:]	[	n	n]	'fish' (Shanghai Chinese)
[sz:]	[s	z	z]	'four' (Standard Chinese)

Several comments can be made. First, the C slot can be filled by a consonant, as in *bat*, or a high vowel (transcribed as a glide), as in *we* and *yam*. Second, the V slot can be filled by a syllabic consonant, as in [n:] 'fish' in Shanghai Chinese and [sz:] 'four' in Standard Chinese. Third, a syllable can be smaller than CVX, because the onset can be missing, as in *Ann*, so can the coda, as in *the*, or both the onset and the coda, as in *a*; however, I shall not focus on such syllables here. Finally, [ii], [nn], and [zz] are not repetitions of the same sound but a sound doubly linked to two timing slots, represented in (3). A doubly linked sound has one articulatory configuration that lasts for two units of time.

(3)

```
CVX        VX        CVX
| \/        \/        | \/
b  i         n        s  z

[bi:]      [n:]      [sz:]
```

There are many words that exceed CVX. For example, *smile* [smail] is CCVVC, with an extra C on each side of [mai]; *texts* [tɛksts] has three extra Cs at the end; and [sprɪnts] *sprints* has two extra Cs at each end of the word. The analysis of such words involves additional issues, detailed in the sections below. In particular, I argue that not every C is syllabified, although every C should be accounted for. For example, I argue that a consonant affix can be kept, even if it is not in a syllable. Similarly, I argue that an extra word-final (or stem-final) C can be kept only if a language has V-initial suffixes, so that the extra C can potentially serve as the onset of the following V. I also argue that if the articulatory gestures of two sounds can overlap, then they can form a 'complex sound'. For example, the gesture of [p] (Labial) and that of [l] (Coronal) are independent and can overlap. Therefore, [pl] can form a complex sound and fit into the onset slot. The proposal will mostly be illustrated with data from English, whose maximal syllable is often thought to be larger than CVX. Other languages will be mentioned only briefly.

2. Not every consonant is syllabified

It is well known that an unsyllabified C is often either deleted or given an epenthetic vowel to form a syllable with. This can be seen in Chinese loans of English words, exemplified in (4).

(4) Extra C in loan words: deletion and vowel epenthesis

	Deletion of [d]	Epenthetic [u]
English:	[aiɚlənd]	[ʤi:p]
Chinese:	[ai ɚ lan]	[ʤi: p^hu]
	'Ireland'	'Jeep'

Syllables in Standard Chinese cannot end in [d] or [p]. In the loan for *Ireland*, there is an extra [d], which is deleted, and in the loan for *Jeep*, there is an extra [p], for which a vowel [u] is added. Whether an extra C is deleted or given an epenthetic V may depend on the length of the source word (e.g. V-epenthesis for monosyllables and deletion for long words), but either way the effect is to eliminate extra Cs.

Many linguists, such as Jones (1950), Abercrombie (1967), Haugen (1956a, b), Fudge (1969), Hoard (1971), Kahn (1976), Clements and Keyser (1983), Cairns (1988), Hammond (1999), Hall (2002a), and Blevins (2003), believe that every consonant must be syllabified (i.e. belonging to a syllable). Under this 'all-in' assumption, English allows very large syllables, such as CVCCCC in *texts* [tɛksts], CCVVCC in *smiles* [smailz], and CCCVCCC in *sprints* [sprɪnts].

However, there are two problems in the all-in analysis. First, it is well known that word-medial syllables are generally quite simple, although extra consonants can occur at word edges. This is the case, for example, in Greek (Steriade 1982), English (Borowsky 1986, 1989), German (Giegerich 1985, 1989), Bella Coola (Bagemihl 1991), Spokane Salish (Bates and Carlson 1992), Polish (Bethin 1992), Georgian (Butskhrikidze 2002), and Jiarong (Lin 1993). The all-in analysis must explain why extra-large syllables are not found in word-medial positions. Second, there is no explanation why all consonants must be in a syllable. One might think that it is not possible to pronounce a consonant without a syllable, but English does have consonant interjections, such as *shh* [ʃ], *pff* [pf], and *psst* [ps], and English speakers have no trouble pronouncing [s] or [f] alone. If such utterances are not syllables, then consonants can be pronounced without being in a syllable, and there is no need for the all-in analysis. If such utterances are syllables, there is another problem: a word like *text* need not be one syllable but can be three [tɛk][s][t], because [s] and [t] can be syllables themselves. Unless there are answers to these problems (of which I am not aware), we should reject the all-in analysis.

If some consonants are not syllabified, why can they still be kept, instead of being deleted or given an epenthetic vowel? I shall argue that being in a syllable is not the only reason for keeping a consonant. Other reasons include the morphology of a language, to be discussed next.

3. Morphological factor: Consonant affixes

Most word-final consonant clusters in English involve [t, d, s, z, θ]. If we exclude these sounds, the final syllable will be smaller. For example, *tenths* [tɛnθs] is CVCCC, but if we exclude [θ, s], the syllable is CVC. Similarly, *texts* [tɛksts] is CVCCCC, but if we exclude [s, t, s], the syllable is again CVC. The question is: what is the reason to exclude word-final [t, d, s, z, θ]?

Phonetically, [t, d, s, z, θ] are made with the articulator Coronal. One might suggest, therefore, that Coronal sounds can be absorbed into a syllable as a special case. On the other hand, [t, d, s, z, θ] also occur as suffixes in English; indeed they are the only consonant suffixes in English. Therefore, it is possible that final [t, d, s, z, θ] are there not because they are in a syllable but because they can serve as suffixes. The point has been made by Goldsmith (1990:

127), who proposes that consonant morphemes are 'licensed' to occur even if they are not in a syllable. I shall call it the affix rule and state it in (5).

(5) The affix rule:

Affix or affix-like sounds can be pronounced, even if they cannot fit into a syllable.

The affix rule is intuitively natural. In the CELEX lexicon of English (Baayen et al 1993), among the 54,447 basic words, 41,911 (or 77%) end in C. This means that in most words the final syllable is already full (needs no more C). If consonant suffixes could not be added when the preceding syllable is full, then they would rarely surface and we would not be able to tell whether there is a suffix or not. Also, since the affix rule already accounts for word-final [t, d, s, z, θ], it is redundant to assume that they must also be part of a syllable.

In English, no all final [t, d, s, z, θ] are suffixes. For example, the final [s] and [t] in *text* [tɛkst] are not suffixes. Such sounds are covered by the term 'affix-like' in the affix rule. The idea has been proposed before by Fujimura (1979) and Pierrehumbert (1994), who use the notion 'perceived suffixes' to refer to final [t, d, s, z, θ] that are not suffixes, such as [s] in [æks] *ax*.

The lack of distinction between real affixes and affix-like sounds suggests that phonetic judgment is independent of semantics. I shall call it 'the independence of phonetic judgment' and state it in (6).

(6) The independence of phonetic judgment:

Judgment on phonetic well-formedness is independent of meaning.

For example, if the sequence [æks] sounds good in *backs*, it should sound good in *ax*, even though [s] is a suffix in the former but not in the latter. Similarly, if the sequence [ɛkst] sounds good in *indexed*, it should sound good in *text*, even though [t] is a suffix in the former but not in the latter.

In summary, while it is in principle possible that all word-final consonants are in the preceding syllable, there is no evidence for it. In addition, the all-in analysis is more complicated in that it cannot explain why non-final syllables are smaller. Finally, the all-in analysis is redundant, because word-final [t, d, s, z, θ] can be explained independently by morphology and there is no need to assume that they have to be in a syllable.

4. Morphological factor: Potential V and anti-allomorphy

If we exclude word-final [t, d, s, z, θ], the largest final rime in English is VXC, which can be either VVC, as in [baik] *bike*, or VCC, as in [hɛlp] *help*. There are three analyses of VXC, shown in (7), where square brackets indicate syllable boundaries.

(7) Analyses of final VXC

[…VXC]	Large rime
[…VX]<C>	Extrasyllabic <C>
[…VX][C(V)]	Onset (of a potential V)

In the first analysis the entire VXC is in a large rime (e.g. Kahn 1976, Kiparsky 1981, Giegerich 1992, Harris 1994, Blevins 1995, and Hall 2001). The problem is that nonfinal rimes are limited to VX (Giegerich 1985, Borowsky 1989), instead of VXC. To say that the maximal rime is VX non-finally and VXC finally, as Hall (2001) does, is a restatement of the fact, not an explanation. In addition, should there be VXC rimes, one would expect them to be more prominent than VX rimes (e.g. more likely to attract stress), but there is no such evidence.

In the second analysis a word-final C is extrasyllabic (e.g. McCarthy 1979, Hayes 1982, Borowsky 1989, Giegerich 1989, Goldsmith 1990, and Gussmann 2002), and there is a consistent maximal rime VX for both final and nonfinal syllables. The question though is why the final C should be excluded from syllabification and why it is allowed to stay.

In the third analysis the final C is the onset of a following V. There are two views of what the V is. In the first, the V can be completely abstract (Burzio 1994). In the second, the V can be real—it is provided by a V-initial suffix (Giegerich 1985, Borowsky 1986). Since a word does not always have a suffix, I refer to the V as a 'potential V'. For example, the [p] in *help* is an extra C when the word occurs alone but it is the onset of the following V in *helper* and *helping*. The question for this analysis is why the final C is kept when there is no following V, such as the [p] in *help*, *helpful*, and *helpless*. Here I appeal to a requirement known as paradigm uniformity or anti-allomorphy (Burzio 1996), according to which one aims to keep a morpheme in the same shape regardless of the environment. In (8) I state the conditions that support the final C, which are illustrated in (9).

(8) Conditions for an extra word-edge C

a. Potential V: A word-final C can serve as the onset of a potential V, which may come with a V-initial suffix. Similarly, a word-initial C can serve as the coda of a potential V, which may come with a V-final prefix.

b. Anti-allomorphy: Keep a morpheme in the same shape regardless of the environment.

(9)

Final C	Supported by potential V	Supported by anti-allomorphy
[hɛl]p	[hɛl][pɪŋ], [hɛl][pɚ]	[hɛl]p, [hɛl]p[fʊl], [hɛl]p[lɪs]
help	*helping*, *helper*	*help, helpful, helpless*

[rɪs]k	[rɪs][kɪŋ], [rɪs][ki]	[rɪs]k, [rɪs]k[fri]
risk	*risking*, *risky*	*risk, risk-free*

Without the anti-allomorphy requirement, unsyllabified consonants would be deleted and many lexical contrasts would be lost. For example, if we delete the unsyllabified [p], *help, helpful,* and *helpless* would be pronounced as [hɛl], [hɛl][fʊl], and [hɛl][lɪs] respectively, confusing with *hell, hellful,* and *hellless*.

Charles Cairns (personal communications) points out that, while the final [p] in *help* can be attributed to a potential V (e.g. from a V-initial suffix *–ing* or *–er*), the [p] in the noun *kelp* cannot be accounted for this way, because it does not take such suffixes. There are two answers to the problem. First, there are V-initial suffixes for nouns, such as *–y*. Second, such words can be explained by a proper interpretation of (6): because phonetic judgment is independent of meaning (or of the morphological pattern of a word), if final [-lp] does not sound bad for words like *help*, then it does not sound bad for any word.

One might point out that in a word like *file*, the [l] is velarized (a 'dark' [l]), which is an indication that it is in the rime; therefore, the syllabification should be [fail]. In the CVX analysis, [ail] is not a possible rime; instead, there are two other reasons why [l] is dark. First, while [l] is dark in the rime (for all speakers) and clear in the onset (for some speakers), it is unclear whether [l] is dark or clear when it is unsyllabified. Second, it is possible that *file* has two syllables [fai][ɫ], where [ɫ] is syllabic, in the rime, and hence dark.

The potential-V analysis also predicts that, if a language has CV prefixes, an extra C may occur as a 'potential coda' in word-initial position. This is schematically shown in (10).

(10) Initial C supported by potential V and anti-allomorphy

Root	Supported by potential V	Supported by anti-allomorphy
CCVC	[CV-C][CVC]	C[CVC]

An example of (10) can be seen in (11), from the Tibeto-Burman language Jiarong, where [tɕʰ] is an affricate (Lin 1993: 36).

(11)

Root	Supported by potential V	Supported by anti-allomorphy
ntɕʰok	[kɐ-n][tɕʰok]	
'dip'	'dip'	
ʒba	[tə-ʒ][ba]	ʒ[ba-n][tɕʰok]
'face'	'face'	'face dip (dimple)'

The root for 'dip' is CCVC. When it follows a CV prefix (or a vowel-final word), the root-initial [n] can serve as the coda of the preceding V. Similarly, the root for 'face' is CCV. When there is a CV prefix, the root-initial C can serve as the coda of the preceding V. When there is no prefix, as in 'face dip (dimple)', the root-initial C is supported by anti-allomorphy.

The potential-V analysis makes specific predictions for whether and where an extra C may occur in a given language. In a language that has V-initial suffixes, an extra C may occur in root-final position. In a language that has V-final prefixes, an extra C may occur in root-initial position. If the predictions are correct, we have already explained the extra C and there is no reason to assume that it has to be in a syllable, too.

5. Gestural overlap and complex sounds

English has many onsets that are made of a CC cluster, such as [pr] in *pray*, [pl] in *play*, [br] in *bring*, [kw] in *quick*, [mj] in *mute*, and so on. Similar onsets are found in Chinese, too, such as [kwan] 'wide', [twan] 'group', [njan] 'year', [nwan] 'warm', [ljan] 'connect', and so on. We cannot exclude the first C by appealing to the affix rule or a potential V. For example, Chinese has no prefixes. Therefore, the first C in a CC cluster cannot be excluded by the affix rule or explained in terms of a potential V.

It is tempting to suggest that we extend the maximal syllable from CVX to CCVX. However, I shall argue that there is a crucial property of CC onsets that has been overlooked: all CC onsets are possible complex sounds in that their articulatory gestures can overlap. Because overlapping gestures can be made simultaneously, a complex sound takes just one timing slot, as a single sound does. Let me explain it in detail.

I assume that a speech sound (consonant or vowel) is made of one or more articulatory gestures or 'features' (e.g. Ladefoged and Halle 1988, Browman and Goldstein 1989, Halle 2003). For example, [k] is made of the gesture [+stop] by the tongue body (the Dorsal articulator) and [p] is made of the gesture [+stop] by the lips (the Labial articulator) (ignoring laryngeal features). In addition, I assume that speech sounds are made in sequence over time. For example, in [it] *eat* [i] is made before [t] whereas in [ti] *tea* [t] is made before [i]. Moreover, I assume that different gestures can overlap in time, creating a complex sound. The notion of gestural overlap is not controversial. Some well-known examples are shown in (12), where I use [͡] to indicate gestural overlap (a complex sound).

(12) Gestural sequence vs. gestural overlap (complex sound)

Sound	Example	Language
[kp]	[bækpæk] *backpack*	English
[k͡p]	[k͡pu] ‘die’	Eggon
[sw]	[swei] *sway*	English
[s͡w]	[s͡wei] ‘year’	Chinese

In *backpack*, [k] and [p] are pronounced in sequence as two sounds. In [k͡pu] ‘die’, a word in the African language Eggon (Ladefoged and Maddieson 1996: 334), [k] and [p] are pronounced simultaneously as one sound. Similarly, as pointed out by Chao (1934), [sw] is pronounced in sequence as two sounds in the English word [swei] *sway*, where the lip rounding of [w] starts after [s], but simultaneously as one sound in the Chinese word [s͡wei] ‘year’, where the lip rounding of [w] starts at the beginning of [s]. Chao’s observation can be demonstrated by spectrographic data, as shown in Figure 1.

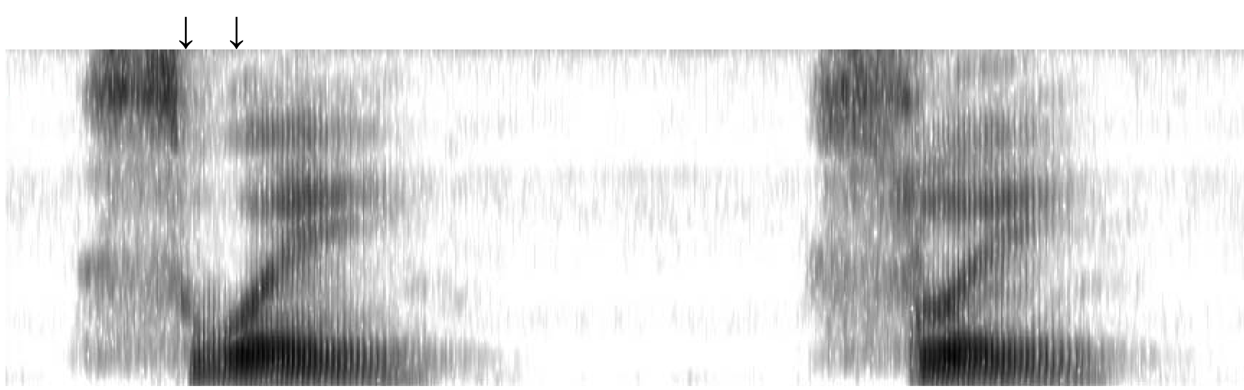

Figure 1: Spectrograms of the English word *sway* [swei] on the left and the Chinese word [s͡wei] ‘year’ on the right. The independent duration of [w] (between arrows) is visible in the English word but not in the Chinese word.

The notion of complex sounds has been proposed before in a rather loose way; it usually refers to two sounds that share one timing slot. Some examples are shown in (13).

(13) Previous proposals of complex sounds

Author	Proposal	Example
Selkirk (1982)	[sC]	[sp, st, sk, …]
Sagey (1986)	‘contour features’	[st, ts, nt, tn, …]
Borowsky (1989)	NC	[mp, nt, ns, …]

Selkirk (1982: 347) proposes that [s] can pair with any obstruent C to form a single sound, a view shared by Lamontagne (1993). Sagey (1986) proposes that opposite gestures can occur in sequence within a sound, such as [-stop, +stop] in [st] and [+stop, -stop] in [ts]; she calls such sequential gestures ‘contour features’. Borowsky (1989) proposes that NC clusters with the same place of articulation can count as one sound, a view shared by Hall (2001). Clearly, such views of complex sounds radically increase the inventory of possible sounds in the world’s languages.

Unlike previous proposals, I assume a highly restricted version of complex sounds. In my analysis, two sounds cannot form a complex sound if they have conflicting gestures, such as [-nasal] and [+nasal], or [+round] and [-round]. The reason is that conflicting gestures cannot overlap (i.e. they cannot be made simultaneously) but must be made in sequence, and therefore it would require more than one timing slot to do so.

A pair of conflicting gestures, such as [+round, -round] by the articulator Labial, [+nasal, -nasal] by the articulator Velum, and [+anterior, -anterior] by the articulator Coronal, have been called ‘contour features’. They can be ruled out by the No Contour Principle, first proposed by Duanmu (1994) and given in (14).

(14) No Contour Principle:

An articulator cannot make opposite values of the same feature (F) within a sound (i.e. within one timing slot).

```
*Articulator        *Articulator
     /\                  /\
  [+F][-F]            [-F][+F]
```

The principle assumes that each gesture takes up a unit of time, and to perform opposite gestures we need two units of time, or the duration of two sounds. It also assumes that all gestures within a sound are simultaneous. For example, [k͡p] is the same as [p͡k].

Although the No Contour Principle allows fewer possible complex sounds than previous analyses do, it also allows many that are not thought to be complex sounds before. For example, onset clusters [pl, pr, kl, kr] can form complex sounds [p͡l, p͡r, k͡l, k͡r]. In addition, the No Contour Principle can distinguish certain clusters that are thought to be the same. For example, consider [fr] and [θr], shown in (15) and (16), where [+anterior] means that the Coronal articulator (tongue tip) is front and [-anterior] means that it is back.

(15)	[f]	Labial—[+fricative]
	[r]	Coronal—[-fricative, -anterior]
	[f͡r]	Labial—[+fricative], Coronal—[-fricative]
(16)	[θ]	Coronal—[+fricative, +anterior]
	[r]	Coronal—[-fricative, -anterior]
	*[θ͡r]	Coronal—[+fricative, -fricative, +anterior, -anterior]

In [f͡r], there are no conflicting gestures; therefore, it is a possible complex sound. In [θ͡r], there are conflicting gestures [+anterior, -anterior]; therefore, it is not a possible complex sound.

It is worth noting that, in the theory of underspecification (e.g. Steriade 1987, Archangeli 1988, Keating 1988), non-contrastive features or articulators are not specified. For example, [f] is unspecified for Coronal and [r] is unspecified for Labial. In addition, while [-voice] is specified for [f] (in contrast to [+voice] in [v]), [voice] is not contrastive for [r] and so [r] is unspecified for [voice]. Therefore, in [f͡r] there are no conflicting gestures [-voice, +voice], but [-voice] only.

In previous analyses, onset clusters are governed by sonority (e.g. Jespersen 1904, Selkirk 1982, Steriade 1982, Kenstowicz 1994, Zec 1988, and others), which I rephrase in (17) and illustrate with English examples in (18).

(17) The sonority analysis of onset clusters:

The sonority in an onset cluster must show a sufficient rise.

(18) Examples of the sonority analysis of onset clusters

Cluster	Prediction
pr	good: enough sonority rise
fr	good: enough sonority rise
θr	good: enough sonority rise
sr	good: enough sonority rise
*fn	bad: not enough sonority rise

In the CVX theory, an onset cluster is possible only if it can form a complex sound. Some examples are shown in (19).

(19) The complex-sound analysis of onset clusters:

a. Possible onset clusters (those that can form complex sounds): pr, fr, pl, kr, …

b. Impossible onset clusters (those that cannot form complex sounds): θr, ʃr, ʃm, ʃn, sr, sl, sm, sn, st, sp, sk, sf, …

In the sonority analysis, [θr] and [sr] are both good, because they have the same sonority rise as [fr]. In contrast, *[θr] and *[sr] are bad in the complex-sound analysis, even though [fr] is good. The reason is that [f] and [r] do not have conflicting gestures, whereas [θ] and [r] do.

The sonority analysis and the complex-sound analysis make different predictions. The sonority analysis predicts that clusters like [θr, ʃr] are possible word-medial onsets, because they have a proper sonority rise. In addition, clusters like [ʃm, ʃn, sr, sl, sm, sn, st, sp, sk, sf] are possible word-medial onsets, because in the sonority analysis [s] (and [ʃ]) is exempt from the sonority requirement. In contrast, the complex-sound analysis predicts that such clusters are not possible onsets and will not be found word medially. As shown in Duanmu (2009), evidence from an exhaustive search of the English lexicon supports the complex-sound analysis. Word initially, an extra C can be supported by a 'potential V', because English has V-final prefixes.

Charles Cairns (personal communications) points out that in *asparagus* and *astronomy*, the [p] and [t] are not aspirated, which seems to indicate that [sp] and [str] are medial onsets, which contradict my claim that [sp] and [str] are not possible medial onsets. There are two possible solutions to this problem. First, aspiration may not be a reliable indicator of syllable boundary, a point made by Wells (1990). For example, in *gastronomy*, the [t] is unaspirated, yet it would be better to place the syllable boundary between [s] and [t], because [æ] cannot end a regular English word and the syllable [gæ] would violate the Law of Finals (Vennemann 1988). Similarly, in *gestation*, the [t] is unaspirated, yet it would be better to place the syllable boundary after [s], because [ɛ] never ends an English word. Second, there could be a perceived word boundary in words like *asparagus*. For example, *asparagus* is thought to be 'a sparrowgrass' by some English speakers.

Besides onset clusters, there are three other cases where a complex sound can be formed. The first case is affricates, which are formed by a stop plus a fricative, such as [t] + [s] → [t͡s]. The formation of an affricate is not controversial, but the analysis is not obvious. For example, if [t] is [+stop] and [s] is [-stop], [t͡s] would have conflicting gestures [+stop, -stop]. Therefore, [t͡s] should not be a well-formed complex sound. A solution is not to treat affricates as [+stop, -stop], but as [+stop, +strident], where [strident] refers to strong frication (e.g. Steriade 1989, Clements 1999). This analysis has two problems. First, it is hard to interpret the feature [strident] as an articulatory gesture. Second, there are still feature conflicts between a stop, which is [+stop, -strident], and a fricative, which is [-stop, +strident]. We must explain why when a stop and a fricative form an affricate, the features [-stop] and [-strident] are lost.

A better solution is to represent stops and fricatives with independent gestures, instead of opposite ones (Lombardi 1990), so that their gestures are compatible. Specifically, I propose that [+fricative] is a gesture that closes the edges (left and right sides) of an articulator, whereas [+stop] is a gesture that closes the middle of an articulator. In this analysis, [t͡s] is a well-formed complex sound, although it is rarely used as an onset in English (but see *Tswana*, *scherzo*, and *Alzheimer*). Similarly, [p͡s] and [k͡s] are possible complex sounds, although they are more common in German than in English. The notion of affricates can explain some words that appear to be problems. For example, the first syllable in [dɛks][trəʊz] *dextrose* seems to be CVCC, which exceeds CVX. In the present analysis, the syllabification is [dɛk͡s][t͡rəʊ]z, where no syllable exceeds CVX.

Another case of complex sounds involve VNC, where NC has the same place of articulation. Such a cluster often becomes ṼC, where Ṽ is a complex sound. The feature analysis of *tent* [tɛnt] → [tɛ̃t] is given in (20), where only relevant features are shown.

(20)	[ɛ	n	t]	→	[ɛ̃	t]
Coronal		(+stop)	+stop			+stop
Dorsal	-back				-back	
Velum		+nasal	-nasal		+nasal	-nasal

I assume that [nt] share the feature Coronal-[+stop], indicated under [t]. When [ɛnt] merge into [ɛ̃t], all features are preserved, and there are no conflicting features in the complex sound [ɛ̃].

Borowsky (1986) observes that non-final rimes in English are mostly VX. However, there are some exceptions, most of them involving VNC rimes. For example, the first syllable in *symptom* [sɪmp.təm] seems to be VNC, which exceeds VX. In the present analysis, VNC can be ṼC and *symptom* is [sĩptəm], where no rime exceeds VX. The present analysis agrees with independent judgments that VNC is indeed often realized as ṼC. For example, a number of linguists have given similar transcriptions, such as *simple* [sĩpɫ̩], *sinker* [sĩkə], *symptom* [sĩptəm], and *council* [kãũsɫ̩] (e.g. Malécot 1960, Bailey 1978, Fujimura 1979, and Cohn 1993).

Yet another case of complex sounds involves V+[r], where V is typically a back vowel. If we represent [r] as Coronal-[-anterior], then [Vr] can form a complex sound. The analysis of [or] is shown in (21).

(21)	[o	r]	→	[o͡r]
Coronal		-anterior		-anterior
Dorsal	+back			+back
Labial	+round			+round

In this analysis, the syllabification of *ordnance* could be [o͡rd][nəns], where the first syllable is VX. Similarly, *arctic* could be [a͡rk][tɪk], where the first syllable is again VX.

6. [V:C] rimes

Another kind of rimes that occur in non-final positions is [V:C], where [V:] is a tense vowel, such as [i:s] in *aesthetic* [i:sθɛtɪk] in British English and [ɒ:l] in *also* [ɒ:lso] in American English. The [V:C] rime is often treated as VXC, because a tense vowel can take two timing slots phonologically. For example, the rime of a stressed monosyllable must be either [VC] (as in *bit* [bɪt] and *bet* [bɛt]) or [V:] (as in *bee* [bi:] and *law* [lɒ:]), but not just [V] (e.g. *[bɪ] or *[bɛ]). This shows that a tense vowel is equivalent to a lax vowel plus a consonant.

If we treat a [V:C] rime as VXC, we face two problems. First, there is a general lack of medial VCC rimes. Second, there is a general lack of medial VVC rimes, where VV is a diphthong. Therefore, [V:C] does not represent a general case, but a special one.

It has been proposed that while tense vowels can be long, they need not in all environments (Pike 1947, Jones 1950, Abercrombie 1967, Giegerich 1985, and Alcantara 1998). In particular, a tense V can be long when there is no following C in the syllable but short when there is. If so [V:C] is in fact VC, where V is tense and short but still distinct from a lax vowel. For example, *aesthetic* is [isθɛtɪk] in British English, where [i] is short (compared with [i:] in *bee*) yet distinct from [ɪ]. Similarly, *also* is [ɒlso] in American English, where [ɒ] is short compared with [ɒ:] in *author*. Under this proposal, apparent [V:C] rime are still VX.

In some languages, such as Cantonese and Thai, there is a contrast between [CVC] and [CV:C] syllables. For example, Cantonese has a contrast between [sam] 'heart' and [sa:m] 'shirt'. If the contrast lies in vowel length, [CV:C] has four timing slots and therefore poses a challenge to the CVX theory. There are several possible solutions though. First, what appears to be a monosyllabic word can in fact be more than a syllable. For example, while [sam] is CVX, [sa:m] could be [sa:]<m>, where <m> is either extra-syllabic or a separate syllable. Second, a long vowel is shortened in Cantonese (Wang 1999) and Thai (Leben 1971) when the syllable is not final, and it remains to be seen what the representation of non-final syllables should be. Third, there is often a quality difference between reported long and short vowels, and it is sometimes possible to represent their difference without appealing to vowel length. For example, Huang (1971) represents Cantonese [sam] 'heart' as [sʌm] and [sa:m] 'shirt' as [sam]. For lack of space, I leave it open what the correct analysis should be of languages like Cantonese and Thai.

7. Multiple solutions to the CVX requirement

In the CVX theory, the only restriction on syllable size is that it does not exceed CVX. Therefore, for a given a string of sounds, it is possible to satisfy the requirement in more than one way. Two examples are shown in (22).

(22) Multiple solutions to the CVX requirement

Word	Sounds	Syllabification options
pumpkin	[pʌmpkən]	[pʌ̃p][kən], [pʌm][p͡kən], [pʌŋ][kən]
arctic	[arktɪk]	[a͡rk][tɪk], [ar][tɪk]
prints	[prɪnts]	[p͡rɪn]ts, [p͡rɪn]t͡s, [p͡rɪ̃t]s, [p͡rɪ̃t͡s]

The options in (22) are not exhaustive. For example, the second syllable of *pumpkin* can be a syllabic [n], which is not shown. All the options satisfy the CVX requirement. The availability of such options could explain possible variations among speakers.

8. Can [lp] be an onset or [pl] be a coda?

I proposed above that [pl] can form a complex sound [p͡l]. Several questions can be raised. First, if [p] + [l] can form a complex sound, so should [l] + [p]. Why then is there no word such as *lpum*? The answer is that all articulatory features in a complex sound are simultaneous, so that [p͡l] and [l͡p] are identical. Therefore, *lpum* would be the same as *plum*.

One might also observe that *help* is not realized as [hɛl͡p], where [p] is pronounced simultaneously as [l] (same as [p͡l] in *plum*). Instead, *help* is pronounced as [hɛlp], where [p] is pronounced after [l]. If [l] and [p] can form a complex sound, why is [hɛ͡lp] not used? The answer is that there is no need to incorporate [p] into the preceding syllable, because [p] is already supported by 'anti-allomorphy' (see above). Besides, the sequence [lp] is more similar to that in *hel.per*, and hence is a better way to satisfy anti-allomorphy.

Next, one might ask why there are no such words as *tikl* or *tepl* in English, where [k͡l] and [p͡l] are complex sounds in the coda. A possible answer is that, as is common in the world's languages, there are more restrictions on codas than on onsets. For example, in Standard Chinese, almost any consonant can be the onset, but the coda is limited to [n] and [ŋ]. Similarly, English may simply disfavor complex sounds in the coda. In addition, a final [l] can be syllabic in English, and *tikl* and *tepl* would be disyllabic. Indeed, there are such words in English, such as *tickle* and *nipple*. This point was made by Lamontagne (1993).

Finally, one might ask why there is no medial coda [k͡l] or [p͡l] in English, which could result from a hypothetical word *neplsa* or *tiklny*. There are again two possible answers. First, English may disfavor complex codas. Second, because [l] can be syllabic, *neplsa* or *tiklny* can be trisyllabic, which would usually be spelled as *neppelsa* or *tickelny*. Unlike *neplsa* and *tiklny*, which do not seem to be possible English words, *neppelsa* and *tickelny* seem a lot more like possible English words.

9. The CV-only analysis

Lowenstamm (1996) and Scheer (2004) propose that even CVX is too large. Instead, there is only one syllable size for all languages, which is CV. Let us call it the 'CV-only' analysis. Two examples are shown in (23), where Ø is an empty C or V.

(23) The CV-only analysis

mix [mɪ][kØ][sØ]

spiked [sØ][pa][Øi][kØ][tØ]

The CV-only analysis has three problems. First, it makes extensive use of empty sounds that have no phonetic content. In contrast, the CVX theory assumes no such empty sounds. Second, the CV-only analysis must explain why only some word-edge Cs can occur with an empty V but others cannot. For example, while initial [sØ] is good in English, [fØ] is not (e.g. *stop* is good but **ftop* is not). Similarly, it must explain why extra word-final Cs are mostly [s, z, t, d, θ] in English (e.g. *text* is good but **texp* is not). The CV-only analysis must say, as other theories do, that [s] is special initially and [s, z, t, d, θ] are special word finally. But once we agree that these sounds are special, it is unnecessary to assume that they must also occur with an empty V, or be syllabified at all. A third problem with the CV-only theory is that it opens up alternative solutions that are hard to choose from. For example, assuming exhaustive syllabification, there are at least three analyses of *spiked*, shown in (24).

(24) Alternative analyses that are hard to distinguish

CVX-only [sØØ][pai][kØt]

CV-only [sØ][pa][Øi][kØ][tØ]

VC-only [Øs][Øp][aØ][ik][Øt]

The CVX-only analysis (a version of the CV-only analysis, not to be confused with the CVX theory) uses a larger syllable size but fewer syllables and fewer empty elements. The CV-only and VC-only analyses use the same syllable size, the same number of empty elements, and the same number of syllables. All the solutions can represent any word without problem. It is not obvious how one should choose among the alternatives.

10. Other languages

Although I have focused on English, there are reasons to believe that the CVX theory has a general nature. First, English has long been thought to have very large syllables. If they turn out to be no larger than CVX, one would like to reexamine other languages that are thought to have very large syllables. Second, there is evidence for the CVX theory from major language families. In the Indo-European family, I have discussed English and German in Duanmu (2009), and others have discussed Polish (Bethin 1992), Georgian (Butskhrikidze 2002), and Hindi (Kumar 2005). In the Sino-Tibetan family, I have discussed Chinese and Jiarong in Duanmu (2009). In the Amerindian family, there is supporting evidence from Bella Coola (Bagemihl 1991) and Spokane Salish (Bates and Carlson 1992). Finally, African languages are known to have simple syllable structures. Thus, evidence for the CVX theory goes beyond a few random languages.

Robert Vago (personal communications) suggests that Hungarian might be a problem for the CVX theory. Let us look at a few examples that involve medial consonant clusters. According to Olaszy (2007), the largest medial consonant cluster in Hungarian is CCCCC. There are five examples, shown in (25), following the original transcription.

(25) Medial CCCCC clusters (underlined) in Hungarian

akordstruktúra	[akortStruktu:ra]	‘accord structure’
angsztrőm	[ankstrO:m]	‘angstrom’
marketingstratégia	[mArketinkStratE:gia]	‘marketing strategy’
platformstratégia	[platformStratE:gia]	‘platform strategy’

sportstratégia [SportStratE:gia] 'sports strategy'

Of the five words, four are compounds, where the CCCCC cluster is not truly medial. The remaining example [ankstrO:m] can be syllabified as [ãk͡s][t͡rOm], where [ã], [k͡s], and [t͡r] are well-formed complex sounds. Therefore, all the words are compatible with the CVX theory.

Dell and Elmedlaoui (2002: 254) report that Berber syllables are usually limited to CVC, but VCC rimes can arise when the CC is a geminate. Two trisyllabic examples are shown in (26).

(26) VCC rimes in Berber

/m-qiddš-a/ → [m.qidd.ša] 'mischieveous girl'

/t-nššr-i/ → [t.nəšš.ri] 'you (fem sing) spread'

Two remarks can be made here. First, we may need to examine the phonetic properties of the geminate CC coda and see if they are shortened or if they contrast with a single C in this environment. Second, the VCC rimes are stem final and so may be subject to the anti-allomorphy requirement. In particular, if there is no suffix, or if the suffix is a consonant, the stem are possibly syllabified as [qid.dš] and [nəš.šr], where no syllable exceeds CVX. In the suffixed forms, the syllabification could be [m.qid<d>.ša] and [t.nəš<š>.ri], where < > indicate an unsyllabified consonant. The free consonant is kept (rather than deleted) because of the anti-allomorphy requirement. This is similar to the case in English. For example, in *helpful*, the [p] is unsyllabified, but it is kept by anti-allomorphy, because *help* is otherwise a good word, where the [p] is supported by a potential V (see section 4).

11. The metrical basis of CVX

If the maximal syllable is indeed CVX, we would like to know why. Without an answer, CVX may still seem to be an arbitrary size.

McCarthy (1979) and Kiparsky (1979, 1981) proposes that a syllable structure is in fact a metrical structure. For example, a CCCVCC syllable is represented in (27). Following the metrical theory of Liberman and Prince (1977), S is a strong node and W is a weak node, and the taller the x-column, the stronger a position is.

(27) Metrical representation of a CCCVCC syllable (Kiparsky 1981: 250)

```
                   x
              x    x
         x    x    x   x
[[[  W   S ]  S ][ S [ S  W  ]]]

     C   C    C    V   C   C
```

Kiparsky argues that, if we interpret W has having less sonority and S as having more, the metrical representation yields an ideal sonority contour, whereby sonority rises towards the nucleus and falls from the nucleus (Jespersen 1904). In other words, sonority 'is simply the intrasyllabic counterpart of stress'.

Kiparsky's proposal contains an important insight, but there are two problems. First, the metrical structure is ill formed, because it contains adjacent strong beats, or 'stress clash', which violate a fundamental property of rhythm, which is the alternation between strong and weak beats (Hayes 1995). Second, the syllables Kiparsky assumes are too big. If the CVX theory is correct, no language has syllables larger than CVX.

The CVX theory can solve both problems in Kiparsky's analysis, yet keep his insight that syllable structure reflects metrical structure. In particular, let us assume that the simplest rhythm is the alternation between S and W beats, or …SWSW…, where there is no ‘clash’ SS or ‘lapse’ WW. In addition, let us assume that each syllable has only one peak. Now consider the structures in (28), where V is a strong beat and C a weak one.

(28)	String	Rhythm	Analysis
	*CCV	WWS	Bad rhythm: lapse of WW
	*VCC	SWW	Bad rhythm: lapse of WW
	*VCV	SWS	good rhythm but bad syllable: two peaks
	CVC	WSW	good rhythm and good syllable
	VC	SW	good rhythm and good syllable
	CV	WS	good rhythm and good syllable

The analysis predicts that CCV and VCC are not good syllables (unless CC can form a complex sound), nor is VCV, but CVC, VC, and CV are.

The metrical analysis I just offered is only a preliminary one. A full analysis may require both moraic feet and syllabic feet in the same language (Duanmu et al 2005). For example, a heavy syllable is a moraic foot, and two syllables can make syllabic foot. For lack of space, such issues cannot be covered here.

12. Summary and implications

There is no doubt that many languages have CVX syllables (CVC or CVV). What I have proposed is that CVX is also the upper limit on syllable size. Extra consonants do occur but they can be explained by morphology. In particular, a word has the schematic structure C_mCSCC_m, where C_m is one or more affix or affix-like consonants, C is a consonant supported by a potential V from an affix, and S is one or more syllables whose maximal size is CVX.

If the CVX theory is correct, there are some interesting implications. First, while some linguists believe that the notion of the syllable is central to phonological analysis (e.g. Ladefoged 2001), other linguists doubt whether the syllable is a real phonological entity (e.g. Chomsky and Halle 1968, Gimson 1970, Lamontagne 1993, Steriade 1999, Blevins 2003, and many participants at the 2007 CUNY Syllable Conference). If the maximal syllable size is CVX, it is a real phonological entity to be accounted for, and the best term for it seems to be the syllable.

On the other hand, if the CVX limit can be derived from metrical structure, as I have suggested, there is a more fundamental aspect of language to be investigated, which is the nature of rhythm. Indeed, the manifestation of rhythm is not limited to language alone.

The CVX theory also has implications for how to approach language variation. In generative phonology, the standard way to account for language variation is to use 'parameters' (Chomsky 1981). For example, parameters for stress patterns have been proposed by Halle and Vergnaud (1987) and Hayes (1995). Likewise, Blevins (1995) proposes six such parameters for the maximal syllable size, rephrased in (29), along with the values (or 'settings') for English. Similar parameters have been proposed before by Clements and Keyser (1983: 28-30).

(29) Binary parameters for the maximal syllable size (Blevins 1995: 219)

Parameters	English settings
Can the onset contain two sounds?	Yes
Can the nucleus contain two sounds?	Yes
Is the coda allowed?	Yes
Can the coda contain two sounds?	Yes

Can extra C occur initially?	No (Yes)
Can extra C occur finally?	Yes

If all parameters are set to 'yes', a syllable can be CCVVCC in non-edge positions and CCCVVCCC in monosyllabic words. In fact, Clements and Keyser (1983: 32) assume that C and V can each repeat at least three times, so that a maximal possible syllable is at least CCCVVVCCC.

Like previous analyses, the CVX theory recognizes cross-linguistic variations in the maximal size of a monosyllabic word. However, unlike previous analysis, which attributes the variations to syllable parameters, the CVX theory derives the variations from morphology, which must be recognized independently. In addition, the CVX theory recognizes the fact that complex sounds can result from gestural overlap. Excluding such factors, the maximal syllable is found to be CVX, much smaller than has been proposed before, such as CCCVCCC (Cairns 1988), CCCCVCCC (Hooper 1976a: 229), or CCCVVVCCC (Clements and Keyser 1983: 32).

When one looks at the world's languages, it is easy to get the impression that there is a wide range of patterns. People have often wondered about the question: How different can human languages be and what are the limits of variation? With regard to syllables, I have proposed that there is no variation in the maximal size. The range of possible syllables is therefore far smaller than has been conceived before. My study shows that in at least some parts of language there is no structural variation, despite apparent diversity at first sight.

*Acknowledgments: I thank the audience at the CUNY Syllable Conference for their comments on my talk and Chuck Cairns and François Dell for their comments on an earlier draft.

References:

Abercrombie, David. 1967. Elements of general phonetics. Chicago: Aldine.

Alcantara, Jonathan Brett. 1998. The Architecture of the English Lexicon. Doctoral Dissertation, Cornell University.

Archangeli, Diana. 1988. Aspects of underspecification theory. *Phonology* 5: 183-207.

Baayen, R. Harald, Richard Piepenbrock, and L. Gulikers. 1993. The CELEX lexical database (CD-ROM). Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania.

Bagemihl, Bruce. 1991. Syllable structure in Bella Coola. Linguistic Inquiry 22.4: 589-646.

Bailey, Charles-James Nice. 1978. Gradience in English syllabification and a revised concept of unmarked syllabification. Indiana University Linguistics Club, Bloomington.

Bates, Dawn, and Barry Carlson. 1992. Simple syllables in Spokane Salish. Linguistic Inquiry 23.4: 653-659.

Bethin, Christina Y. 1992. *Polish syllables: the role of prosody in phonology and morphology*. Columbus: Slavica Publishers.

Blevins, Juliette. 1995. The syllable in phonological theory. In The handbook of phonological theory, ed. John Goldsmith, 206-244. Cambridge, MA: Blackwell.

Blevins, Juliette. 2003. Evolutionary phonology: the emergence of sound patterns. Cambridge: Cambridge University Press.

Borowsky, Toni. 1986. Topics in the lexical phonology of English. Doctoral dissertation, University of Massachusetts, Amherst.

Borowsky, Toni. 1989. Structure preservation and the syllable coda in English. Natural Language and Linguistic Theory 7: 145-166.

Browman, Catherine P. and Louis M. Goldstein. 1989. Articulatory gestures as phonological units. Phonology 6: 201-251.

Burzio, Luigi. 1994. Principles of English stress. Cambridge, UK: Cambridge University Press.

Burzio, Luigi. 1996. Surface constraints versus underlying representation. In Current trends in phonology: models and methods, vol. 1, ed. Jacques Durand and Bernard Laks, 123-141. Salford: European Studies Research Institute, University of Salford Publications.

Butskhrikidze, Marika. 2002. The consonant phonotactics of Georgian. Doctoral dissertation, Universiteit Leiden. [Distributed by Netherlands Graduate School of Linguistics, Utrecht.]

Cairns. Charles E. 1988. Phonotactics, markedness and lexical representation. *Phonology* 5.2: 209-236.

Chao, Yuen-Ren. 1934. The non-uniqueness of phonemic solutions of phonetic systems. Bulletin of the Institute of History and Philology, Academia Sinica 4.4: 363-397.

Chomsky, Noam. 1981. Lectures on government and binding. Dordrecht: Foris.

Chomsky, Noam, and Morris Halle. 1968. The sound pattern of English. N.Y.: Harper and Row.

Clements, G. N. 1999. Affricates as noncontoured stops. In Osamu Fujimura, Brian D Joseph, and Bohumil Palek, eds., Proceedings of LP'98: Item Order in Language and Speech (Columbus, the Ohio State University, September 15-20, 1998), Vol. I, 271-299. Prague: Karolinum Press (Charles University in Prague).

Clements, G. N., and Samuel Jay Keyser. 1983. CV phonology: a generative theory of the syllable. Cambridge, Mass.: MIT Press.

Cohn, Abigail C. 1993. Nasalisation in English: phonology or phonetics. Phonology 10.1:43-81.

Davis, Stuart. 1988. Topics in syllable geometry. New York: Garland.

Dell, Francois, and Mohamed Elmedlaoui. 2002. Syllables in Tashlhiyt Berber and in Moroccan Arabic. (Kluwer International Handbooks in Linguistics, Volume 2) Dordrecht and Boston: Kluwer Academic Publishers.

Duanmu, San. 1994. Against contour tone units. Linguistic Inquiry 25.4: 555-608.

Duanmu, San. 2009. Syllable structure: The limits of variation. Oxford: Oxford University Press.

Duanmu, San, Hyo-Young Kim, and Nathan Stiennon. 2005. Stress and syllable structure in English: approaches to phonological variations. Taiwan Journal of Linguistics 3.2: 45-77.

Fudge, Erik C. 1969. Syllables. Journal of Linguistics 5:253-286.

Fujimura, Osama. 1979. An analysis of English syllables as cores and affixes. Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung 471-476.

Giegerich, Heinz. 1985. Metrical phonology and phonological structure: German and English. Cambridge, UK: Cambridge University Press.

Giegerich, Heinz. 1989. Syllable structure and lexical derivation in German. Bloomington, Indiana: Indiana Linguistic Club.

Giegerich, Heinz. 1992. English phonology. Cambridge, UK: Cambridge University Press.

Gimson, A. C. 1970. An introduction to the pronunciation of English. 2nd edition. New York: St. Martin's Press.

Goldsmith, John. 1990. Autosegmental and metrical phonology. Oxford, UK and Cambridge, MA: Basil Blackwell.

Gouskova, Maria. 2004. Relational hierarchies in Optimality Theory: the case of syllable contact. Phonology 21.2: 201-250.

Gussmann, Edmund. 2002. Phonology: analysis and theory. Cambridge: Cambridge University Press.

Hall, Tracy Alan. 2001. The distribution of superheavy syllables in Modern English. *Folia Linguistica* XXXV/3-4: 399-442.

Hall, Tracy Alan. 2002a. Against extrasyllabic consonants in German and English. *Phonology* 19.1: 33-75.

Halle, Morris. 2003. Phonological features. In International encyclopedia of linguistics, ed. by William J. Frawley. 2nd ed. Oxford: Oxford University Press.

Halle, Morris, and Jean-Roger Vergnaud. 1987. An essay on stress. Cambridge, Mass.: MIT Press.

Hammond, Michael. 1999. The phonology of English: a prosodic Optimality Theoretic approach. Oxford: Oxford University Press.

Harris, John. 1994. English sound structure. Oxford, UK: Blackwell.

Haugen, Einar. 1956a. Syllabification in Kutenai. International Journal of American Linguistics 22.3: 196-201.

Haugen, Einar. 1956b. The syllable in linguistic description. In For Roman Jakobson, edited by Morris Halle, Horace Lunt, Hugh MacLean, and Cornelis van Schooneveld, 213-221. The Hague: Mouton.

Hayes, Bruce. 1982. Extrametricality and English stress. Linguistic Inquiry 13: 227-276.

Hayes, Bruce. 1995. Metrical stress theory: Principles and case studies. Chicago: University of Chicago Press.

Hoard, James E. 1971. Aspiration, tenseness, and syllabification in English. Language 47.1: 133-140.

Hooper, Joan Bybee. 1976a. An introduction to natural generative phonology. New York: Academic Press.

Huang, Parker Po-fei. 1970. Cantonese dictionary: Cantonese-English, English-Cantonese, Yale University Press, New Heaven and London.

Jespersen, Otto. 1904. Lehrbuch der Phonetik. Leipzig & Berlin: Teubner.

Jones, Daniel. 1950. The pronunciation of English. 3rd edition. Cambridge, England: Cambridge University Press.

Kahn, Daniel. 1976. Syllable-based generalizations in English phonology. Doctoral dissertation, MIT, Cambridge, Mass.

Keating, Patricia. 1988. Underspecification in phonetics. Phonology 5:275-292.

Kenstowicz, Michael. 1994. Phonology in Generative Grammar. Oxford: Blackwell.

Kiparsky, Paul. 1979. Metrical structure assignment is cyclic. Linguistic Inquiry 10.3: 421-441.

Kiparsky, Paul. 1981. Remarks on the metrical structure of the syllable. Phonologica 1980: Akten der Vierten Internationalen Phonologie-Tagung, Wien, 29, Juni-2. Juli 1980, ed. Wolfgang U. Dressler, Oskar E. Pfeiffer, and John R. Rennison, 245-256. Innsbruck: Institute für Sprachwissenschaft der Universität Innsbruck.

Kumar, Aman. 2005. Aspects of Hindi Syllable Structure. Doctoral dissertation, University of Michigan, Ann Arbor.

Ladefoged, Peter. 2001. Vowels and consonants: an introduction to the sounds of languages. Malden, MA: Blackwell.

Ladefoged, Peter, and Morris Halle. 1988. Some major features of the International Phonetic Alphabet. Language 64.3:577-82.

Ladefoged, Peter, and Ian Maddieson. 1996. The sounds of the world's languages. Cambridge, MA: Blackwell.

Lamontagne, Gregory A. 1993. Syllabification and consonant cooccurrence conditions. Doctoral dissertation, University of Massachusetts, Amherst.

Leben, William. 1971. On the segmental nature of tone in Thai. Quarterly Progress Report No. 101: 221-224. Research Laboratory of Electronics, Massachusetts Institute of Technology.

Liberman, Mark, and Alan Prince. 1977. On stress and linguistic rhythm. Linguistic Inquiry 8.2:249-336.

Lin, Xiangrong. 1993. Jiarongyu yanjiu [Studies on the Jiarong (rGyalrong) language]. Chengdu: Sichuan Minzu Chubanshe.

Lombardi, Linda. 1990. The nonlinear organization of the affricate. Natural Language and Linguistic Theory 8:374-425.

Lowenstamm, Jean. 1996. CV as the only syllable type. In Current trends in phonology, models and methods, edited by Jacques Durand & Bernard Laks, 419–443. Salford: European Studies Research Institute, University of Salford Publications.

Malécot, André. 1960. Vowel nasality as a distinctive feature in American English. Language 36.2: 222-229.

McCarthy, John. 1979. On stress and syllabification. Linguistic Inquiry 10.3: 443-465.

Olaszy, Gábor. 2007. Consonant clusters in Hungarian speech: Spoken word database for the presentation of acoustic structure and coarticulation in consonant clusters. http://fonetika.nytud.hu/cccc/adatbazisrol_en.html

Pierrehumbert, Janet. 1994. Syllable structure and word structure: a study of triconsonantal clusters in English. In Phonological structure and phonetic form, Papers in Laboratory Phonology III, ed. Patricia A. Keating, 168-188. Cambridge, England and New York, NY: Cambridge University Press.

Pike, Kenneth. 1947. On the phonemic status of English diphthongs. Language 23:151-159.

Sagey, Elizabeth. 1986. The representation of features and relations in nonlinear phonology. Doctoral dissertation, MIT, Cambridge, Mass.

Scheer, Tobias. 2004. A lateral theory of phonology: what is CVCV, and why should it be? [Studies in Generative Grammar 68.1] Berlin and New York: Mouton de Gruyter.

Selkirk, Elisabeth. 1982. The syllable. In The structure of phonological representations (Part II), ed. Harry van der Hulst and Norval Smith, 337-83. Linguistic Models 2. Dordrecht: Foris.

Steriade, Donca. 1982. Greek prosodies and the nature of syllabification. Doctoral dissertation, MIT, Cambridge, Mass.

Steriade, Donca. 1987. Redundant values. In Papers from the 23rd Annual Regional Meeting of the Chicago Linguistic Society, Part 2: Parasession on Autosegmental and Metrical Phonology, 339-362.

Steriade, Donca. 1989. Affricates are Stops. Paper presented at Conference on Features and Underspecification Theories, October 7-9, MIT.

Steriade, Donca. 1999. Alternatives to syllable-based accounts of consonantal phonotactics. In Osamu Fujimura, Brian D Joseph, and Bohumil Palek, eds., Proceedings of LP'98: Item Order in Language and Speech (Columbus, the Ohio State University, September 15-20, 1998), Vol. I, 205-245. Prague: Karolinum Press (Charles University in Prague).

Vennemann, Theo. 1988. Preference laws for syllable structure and the explanation of sound change. Berlin and New York: Mouton de Gruyter.

Wang, Jenny Zhijie. 1999. Vowel Length in Cantonese. Paper presented at the 11th North American Conference on Chinese Linguistics, Harvard University.

Wells, John Christopher. 1990. Syllabification and allophony. In Studies in the pronunciation of English, A commemorative volume in honour of A.C. Gimson, ed. Susan Ramsaran, 76-86. London and New York: Routledge.

Zec, Draga. 1988. Sonority constraints on prosodic structure. Doctoral dissertation, Stanford University.