

Appeared in *Chinese Phonology* [音韻論叢] 11: 97-120, 2002.

Two Theories of Onset Clusters

San Duanmu (端木三)

June 2002

中文摘要

中文和英文音節首的輔音組(如 [pl], [pr]等)和輔音介音組(如[kw], [pj]等), 哪些可用哪些不可用, 一般都用音響級別(sonority scale)來解釋。本文提出, 這些音組實際上都是複合單音(同時有兩個或三個發音動作的單音)。本文先指出音響理論的缺點, 然後根據發音理論, 討論複合單音的類型, 其次再逐一討論中文和英文的輔音組和輔音介音組。本文認為, 中文的輔音介音組全部都能分析為複合單音。英文除了音節首的舌尖擦音(如[s]), 剩下的輔音組和輔音介音組也都全部能分析為複合單音。英文音節首的[s]等音對於任何理論都是例外, 對本文也如此。本文的理論比現有的理論更簡單, 可以解釋更多的現象, 同時還反映出了中文和英文的共性。

關鍵詞: 音節; 輔音組; 複合單音

English Abstract

Onset clusters refer to consonant clusters, such as [pl] and [pr], and consonant-glide clusters, such as [kw] and [pj], that occur at the beginning of a syllable. Which clusters are used in a language is usually explained in terms of sonority scale. I argue instead that the occurring clusters are in fact complex single sounds. I first point out the shortcomings of the sonority approach. Then I introduce articulator-based feature theory, which defines possible complex sounds. Next I survey each onset cluster in Chinese and English and argue that it can be represented as a single complex sound. An initial coronal fricative, such as [s] or [ʃ], can also be added to an English syllable, which every theory must accommodate. The present proposal is simpler, accounts for more facts, and shows a similarity between Chinese and English.

Key words: syllable; consonant cluster; complex sound

1. Introduction

Many languages have what appear to be clusters of two (or more) sounds before the nuclear vowel. Some examples from English and Chinese are shown in (1) and (2) (see Appendix 1 for English clusters and Appendix 2 for Chinese clusters). I assume that, before the nuclear vowel, there is no difference between [i] and [j] or between [u] and [w], and I will transcribe them as glides.

- (1) Onset clusters in English
- | Cluster | Example |
|---------|---------|
| [pl] | place |
| [pr] | price |
| [pj] | pure |
| [kw] | quick |
- (2) Onset clusters in Chinese
- | Cluster | Example |
|---------|---------------|
| [mj] | [mjæn] ‘face’ |
| [nw] | [nwæn] ‘warm’ |

Let us call them onset clusters. In languages that allow onset clusters, not every cluster is good. For example, neither English nor Chinese allows those in (3).

- (3) Bad onset clusters for English and Chinese
[tm], [kf], [nt], etc.

The goal of this article is to explain which clusters are good and which are bad.

A standard analysis of permissible and non-permissible onset clusters is based on the sonority scale, according to which the second sound should normally be more sonorous than the first. However, the sonority-based theory cannot fully explain the distribution of good and bad clusters. For example, [pl] and [tl] have the same sonority slope (that is, how much louder the second sound is than the first), yet in English [pl] is good and [tl] is bad. The usual solution is to add another constraint (a place dissimilation constraint), namely, clusters in which the two sounds have the same place of articulation are not allowed. But [dr] and [tr] are both good in English, where the sounds in each pair have the same place of articulation (retroflex coronals).

In this paper I propose a different analysis, in which there is only one onset slot in each English or Chinese syllable (the single-slot analysis). What appears to be a cluster of two sounds is in fact a single complex sound. The analysis assumes articulator-based feature theory, which determines what are possible single sounds and what are not possible single sounds. For lack of space, restrictions on other sequences, such as the onset-nucleus sequence or the nucleus-coda sequence, are not discussed (but some discussion will be given on coda clusters). In addition, I limit the discussion to English and Chinese and will say little about languages that have more complex clusters, such as German, Croatian, or Polish.

In section 2 I introduce the sonority-based analysis. In section 3 I introduce articulator-based feature theory. In section 4 I offer the single-slot

analysis of onset clusters. In section 5 I discuss some further issues. In section 6 I offer concluding remarks.

2. The sonority-based analysis

The sonority-based analysis makes two basic assumptions. First, sounds of a language are divided into several degrees on a scale according to their sonority (roughly, how loud a sound is). In English four degrees have been proposed for consonants and glides (Kenstowicz 1994, Roca & Johnson 1999, and references therein), shown in (4).

(4) Sonority scale for English

| Sound Class | Example | Sonority Scale |
|-------------|---------|----------------|
| Glides | [j w] | 4 |
| Liquids | [l r] | 3 |
| Nasals | [m n] | 2 |
| Obstruents | [f t k] | 1 |

The second assumption is that there is a requirement on the minimal sonority distance (MSD) between the two sounds in an onset cluster. In English the MSD is assumed to be two, i.e. the sonority of the second sound must be at least two degrees higher than that of the first. Some examples are shown in (5).

(5) English onset clusters (Minimal Sonority Distance = 2)

| Example | Sonority distance | Prediction |
|---------|-------------------|------------|
| [fn] | $1 < \text{MSD}$ | bad |
| [fl] | $2 = \text{MSD}$ | good |
| [kw] | $3 > \text{MSD}$ | good |

The sonority of [f] is 1 and that of [n] is 2, so the sonority distance of [fn] is $2 - 1 = 1$, which is lower than the MSD, so the cluster is bad (that is, no English words start with [fn]). In [fl], the sonority distance is $3 - 1 = 2$, which meets the MSD, so the cluster is good (such as *fly*). In [kw], the sonority distance is $4 - 1 = 3$, which exceeds to the MSD, so the cluster is also good (such as *quick*).

The sonority analysis must address two other problems. First, some clusters satisfy the MSD but are not found in English, such as [tl] and [fw]. Second, some clusters violate the MSD but are found in English, such as [st] and [sn]. The solution is to add two more assumptions. First, the two sounds of a cluster should not have the same place of articulation. For example, in [tl] both sounds are alveolar, and in [fw] both sounds are labial. Second, an initial [s], and sometimes [ʃ], is allowed to precede almost any consonant, regardless of the sonority distance or places of articulation. For example, [st] is good (such as *stop*,

step), even though the sonority difference is zero and both sounds are alveolar. Similarly, [sn] is good (such as *snack*, *snow*), even though the sonority difference is only one and both sounds are alveolar. The same is true for a small number of words that start with [ʃ], such as *schnapps*, *schnauzer*, *schnorrer*, and so on. There is no other solution except saying that [s] (and [ʃ]) is an exception. The additional assumptions are summarized in (6) and (7).

- (6) Additional assumptions for English
- a. Sounds in a cluster must have different places of articulation
 - b. Initial [s] (and [ʃ]) can be added
- (7)
- | Cluster | Problem | Explanation |
|--------------|----------------------|-----------------|
| *[tl], *[pw] | meet MSD but missing | Violating (6a) |
| [st], [sn] | violate MSD but used | Allowed by (6b) |

The sonority analysis is the only major theory for English onset clusters (an analysis in government phonology, such as Harris 1994, may be considered a variation of the sonority approach). However, even with the additional assumptions, there are still problems. Consider the examples in (8).

- (8) Additional problems for the sonority analysis
- | Cluster | Sonority | Comment |
|------------|----------|---------------------------------------|
| [lj], [rw] | 1 | violates MSD but used |
| [tr] | 2 | violates (6a) but used |
| [nw] | 2 | no violation but missing |
| [tʃ] | 0 | violates MSD but used (if two sounds) |
| [tʃw] | 3 | should be good (if [tʃ] is one sound) |

First, in some English dialects [lj] is a good cluster (such as *lewd*), yet it violates the MSD. Similarly, in all English dialects [r] in the onset is rounded, i.e. [rw] (thus, some children pronounce *rabbit* as *wabbit*). However, the sonority distance in [rw] is also below the MSD. Second, although clusters like [tl] and [fw] can be ruled out by disallowing sounds of the same place of articulation, English does have [tr] and [dr] (such as *try*, *dry*), where both sounds of a cluster are retroflex coronals. Third, there is a large number of clusters that satisfy the sonority distance and the place of articulation requirement yet are still not found in English: [nw, zr, ʒw, ʒr, ʒl, ðw, ðr, ðl, θl]. Fourth, consider the affricate [tʃ] as in *chip*. If it is two sounds, their sonority distance is 0, which is below the MSD, yet the cluster is good. If [tʃ] is one sound, whose sonority is 1, then it should combine with [w], with a resulting sonority distance of 3, which satisfies the

MSD. But [tʃw] is not recognized in traditional analysis (I will return to [tʃw] below).

In summary, the sonority analysis can account for many English onset clusters in terms of (a) a minimal sonority distance and (b) a constraint against two sounds with the same place of articulation, and (c) a special exception for initial [s] and [ʃ]. On the other hand, there remain serious problems: some clusters that violate (a) and (b) are still good, such as [lj, rw, tr, dr], and some clusters that meet all the requirements are still missing, such as [nw, zr, ʒw, ʒr, ʒl, ðw, ðr, ðl, θl], or unaccounted for, such as [tʃw, dʒw].

In Chinese the only onset clusters are made of a consonant and a glide. If we use the same sonority scale as in English, there will be many problems. For example, consider the examples in (9).

(9) Chinese onset clusters (Minimal Sonority Distance = 2)

| Example | Sonority Distance | Prediction | Actual |
|---------|-------------------|------------|--------|
| [fn] | 1 | bad | bad |
| [lj] | 1 | bad | good |
| [fl] | 2 | good | bad |

In [fn] the sonority distance is $2-1=1$, below the threshold of 2, so we correctly predict the cluster to be bad. On the other hand, in [lj] the sonority distance is $4-3=1$, also below the threshold of 2, yet the cluster is good, as in *lian* ‘connect’. Moreover, in [fl] the sonority distance is $3-1=2$, meeting the threshold of 2, so we predict the cluster to be good, yet it is bad.

A possible solution to the problem is to use a different sonority analysis for Chinese, shown in (10) and (11), according to which Chinese ignores the fine-grain differences among the consonants.

(10) New sonority scale for Chinese

| Sound Class | Example | Sonority Scale |
|-------------|-----------|----------------|
| Vowels | [a o u i] | 3 |
| Glides | [j w] | 2 |
| Consonants | [f m l r] | 1 |

(11) Chinese onset clusters (Minimal Sonority Distance = 1)

| Example | Sonority Distance | Prediction | Actual |
|---------|-------------------|------------|--------|
| [fn] | 0 | bad | bad |
| [fl] | 0 | bad | bad |
| [lj] | 1 | good | good |

If we assume the sonority scale in (10) and the MSD to be one for Chinese, then good clusters are again those that satisfy the MSD. On the other hand, proposing a different sonority scale and a different MSD for Chinese makes the theory less general. For example, why is Chinese less sensitive to sonority differences among consonants? And if the sonority difference in [lw] is big enough for Chinese (that is, MSD = 1), why is MSD = 1 not big enough for English (that is, why must MSD = 2 in English)?

3. Articulator-based feature theory

I follow a standard assumption that words are made of consonants and vowels and consonants and vowels in turn are made of distinctive features. Some traditional features are shown in (12).

| | | |
|------|-----------|------------------------------------|
| (12) | FEATURE | EXPLANATION |
| | sonorant | airflow is not obstructed |
| | stop | complete closure in the oral tract |
| | fricative | oral tract has a narrow opening |
| | aspirated | vocal cords are open |
| | voice | vocal cords are vibrating |
| | nasal | nasal tract is open |
| | high | tongue body is high |
| | back | tongue body is back |
| | round | lips are rounded |
| | labial | closure at the lips |
| | alveolar | closure at the alveolar area |

We can usually refer to a sound unambiguously by two or three features. Some examples are shown in (13).

| | | |
|------|-----|----------------------------|
| (13) | [m] | labial nasal |
| | [p] | labial voiceless stop |
| | [b] | labial voiced stop |
| | [f] | labial voiceless fricative |
| | [t] | alveolar voiceless stop |
| | [u] | rounded high back vowel |

The articulator-based feature theory differs from traditional feature theory in that a distinction is made between articulators and features. Articulators are movable parts in the vocal tract that participate in speech production. A feature is a gesture made by an articulator. For example, in the traditional theory both ‘labial’ and ‘round’ are features, but in the articulator-based theory ‘labial’ is an articulator

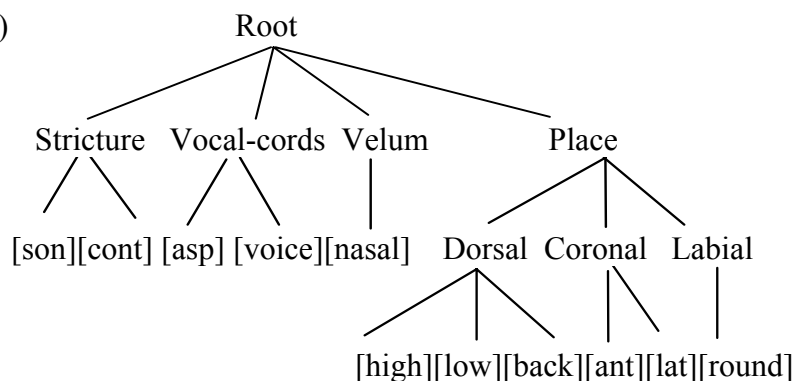
and ‘round’ is a gesture made by ‘labial’. Most features are made only by a given articulator; for example, [round] is made only by labial, and [anterior] is made only by coronal. The stricture (or manner) features [sonorant] and [continuant] can be made by different articulators (more below). The specific feature structure (also called feature geometry) I will assume is given in (14) and diagrammed in (15), based on the works of Clements (1985), Sagey (1986), Ladefoged and Halle (1988), McCarthy (1988), Steriade (1989), Kenstowicz (1994), and Halle (1995). For ease of discussion, some details are omitted, such as the Tongue-root articulator. In addition, customary abbreviations of features (the first three or four letters of a feature) are used.

(14) Articulator-based feature structure

Stricture features [sonorant][continuant]

| Articulators | Their features |
|--------------|---------------------|
| Vocal-cords | [aspirated][voice] |
| Velum | [nasal] |
| Dorsal | [high][low][back] |
| Coronal | [anterior][lateral] |
| Labial | [round] |

(15)



Features are shown in square brackets; each feature can take a value of either ‘+’ or ‘-’. Articulators include Vocal-cords, Velum, Dorsal (tongue body), Coronal (tongue tip), and Labial. For example, the articulator for [t] in English is Coronal, instead of the traditional ‘alveolar’, since it is the tongue tip that initiates the closure. In (16) I list some traditional terms and their translations in feature geometry (see Chomsky and Halle 1968, Halle and Clements 1983, Steriade 1989, Clements and Hume 1995, and Halle 1995).

| | | |
|------|-------------------|----------------------------|
| (16) | Traditional terms | Feature geometry |
| | stop | [-cont] |
| | obstruent | [-son] |
| | fricative | [-son, +cont] |
| | retroflex | Coronal-[-ant] |
| | palatal | Coronal and Dorsal-[-back] |

Palatals are represented as a combination of Coronal and Dorsal-[-back]. The Coronal part accounts for alternations between palatals and coronals, such as [ç] and [sʲ] in Chinese. The feature [-back] accounts for close interaction between front vowels and palatals. Other translations between traditional terms and feature geometry are largely transparent.

A sound need not use all the articulators; unused articulators and features are not shown in the representation. However, a sound can use more than one articulator at the same time. The articulators Dorsal, Coronal, and Labial are often grouped together. Let us call them place articulators. A sound that uses one place articulator can be called a simple sound. A sound that uses two or three place articulators can be called a complex sound. Some examples are given in (17), where * indicates a major articulator (following the notation of Kenstowicz 1994, instead of the pointer notation of Sagey 1986).

| | | | |
|------|--------------------|------------|---------------------------|
| (17) | Sounds | Complexity | Place articulators |
| | [p] | simple | *Labial |
| | [t] | simple | *Coronal |
| | [k] | simple | *Dorsal |
| | [k ^w] | complex | *Dorsal, Labial |
| | [tç] | complex | *Dorsal, *Coronal |
| | [tsʲ] | complex | Dorsal, *Coronal |
| | [kʲ] | complex | *Dorsal, Coronal |
| | [tç ^w] | complex | *Dorsal, *Coronal, Labial |

When a sound has one articulator, it is the major articulator. When a sound has two (or more) place articulators, they may have different degrees of closure. The one that has greater constriction is the ‘major’ articulator, and the one that has less constriction is the ‘minor’ articulator. The stricture (or the manner) features specify the degree of closure of the major articulator. The closure of the minor articulator is left unspecified; it is always assume to be that of an approximant, i.e. [+son, -cont]. Thus, the [-cont] (stop closure) of [p] is executed by *Labial. The [-cont] of [k^w] is executed by *Dorsal, while the minor articulator Labial executes an approximant action, adding [ʷ]. When two articulators have the same degree of

closure, they are both major articulators. Thus, the [-cont] of the palatal [tç] is executed by both *Dorsal and *Coronal. Like [tç], [ts^j] is [-cont], but now the feature is executed by *Coronal alone, while the minor articulator Dorsal executes an approximant [j]. The stricture feature of [k^j] is also [-cont], executed by *Dorsal, and Coronal is the minor articulator.

The articulatory actions of major and minor articulators occur at the same time, although their results can differ in time acoustically. For example, in [k^w] the velar closure for [k] occurs at the same time as the lip rounding for [w], although acoustically the effect of [w] (formant transition) is seen after the release of [k]. Similarly, in [p^h] the glottal opening for [h] occurs at the same time as the labial closure for [p], although acoustically aspiration comes after the release of [p]. Thus, as far as features are concerned, there is no need to assume a time sequence between major and minor articulators.

Although a sound can use several articulators and features at the same time, there is a crucial restriction: a sound can only use a given articulator or feature once. Duanmu (1994) calls the constraint the No Contour Principle, which I repeat in (18).

- (18) No Contour Principle (for a single sound)
- | | |
|------------|------------------------------|
| *N | N = any node |
| /\ | α = any value (i.e., + or -) |
| [αF] [-αF] | F = any feature |

For example, no single sound can have the feature [+round] and then [-round]. The No Contour Principle is the most important criterion in deciding what is a possible single sound and what is not. I will return to this point below.

4. Single-slot analysis of onset clusters

In the present proposal, what appears to be an onset cluster is in fact a single sound. Compared with the cluster analysis, which the sonority analysis assumes, the single-slot analysis is the null-hypothesis, because it offers a simpler syllable structure. So unless there is compelling evidence for the cluster analysis, the single-slot analysis is preferred.

The single-slot proposal assumes that both sounds of the cluster are in the onset, including a consonant-glide cluster such as [kw] or [mj], which nearly all researchers agree to be the case in English. The only potential confusion comes from words like *mew* and *cute*. Davis and Hammond (1995) argue that some dialects use [ju], where [j] is in the onset, but some dialects, including American English, use [iu], where both vowels are in the nucleus. On the other hand,

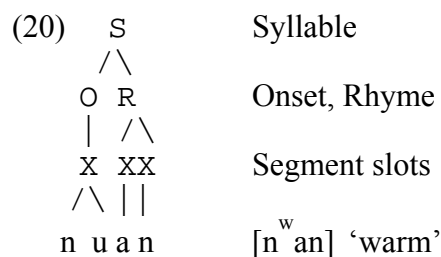
according to Harris (1994), what the two dialects use are [ju] and [iw] respectively. In other words, there is no glide in the nucleus.

In Chinese there are various views on the position of the prenuclear glide. I will mention two arguments that it is also in the onset (both were given in Duanmu 2000). First, there is no compelling evidence that the glide is in the nucleus. Second, if the glide is in the nucleus, sharing the same position with the main vowel, as proposed by Goh (2000), then we get a cluster that is impossible to represent as a single sound. For example, according to Goh (2000), [u] and [a] of *kuai* ‘fast’ share a single nucleus position. This means that [ua] is a single sound. However, if we assume that each feature value can only occur once in a sound (Duanmu 1994), then it is impossible to represent [ua] as a single sound. The reason is that [u] is [+round] and [a] is [-round], and [u] is [+high] and [a] is [-high]. Thus, to represent [ua] fully, we need to use two values for [round] and two values for [high]. A feature with two values of within one sound is called a contour feature. Although many linguists in autosegmental phonology (Goldsmith 1976) freely use such representations, there is a serious theoretical problem. As Duanmu (1994) points out, if each feature can take two values within a sound, we predict about 1,000,000 possible sounds, which far exceeds the total number of sounds in world’s languages (as estimated by Maddieson 1984). In contrast, if each feature can take just one value within a sound, we predict about 1,000 possible sounds, which is close to the total number of sounds in world’s languages.

A potential objection to the single-slot analysis is that it may assume more phonemes than the cluster analysis. Compare the two analyses in (19).

- (19) Phonemes in two analyses
- | | |
|----------------------|---|
| Onsets | [s t k w sw tw kw] |
| Cluster analysis | [s t k w] |
| Single-slot analysis | [s t k w s ^w t ^w k ^w] |

Suppose a language has the phonemes [s w t k] and the clusters [sw], [tw], and [kw]. If a cluster is two sounds, we do not need to assume new phonemes beyond [s w t k]. However, if [sw], [tw], and [kw] are single sounds, we seem to need three new phonemes [s^w], [t^w], and [k^w], in addition to [s w t k]. An answer to this objection is offered in Duanmu (2000). The idea is to separate an underlying level and a surface level. Consider the syllable [nwan] ‘warm’. Underlyingly, there are four phonemes [n u a n]. However, because there is only one onset slot, [nu] must share the slot, giving a surface form [n^w]. This is shown in (20).



In fact, the single slot analysis assumes fewer phonemes than the traditional analysis. For example, the traditional analysis assumes that the palatals [tç tç^h ç] are independent phonemes in Standard Chinese, whereas the single-slot analysis assumes that they are not underlying phonemes but derive from underlying clusters [tsj ts^hj sj] (Duanmu 2000).

In Chinese, the only onset clusters are CG (consonant-glide) clusters. What the single-slot theory needs to show is that good CG clusters in Chinese can be represented as single sounds. In English, if we ignore the initial [s], all clusters are made of an obstruent and an approximant. What the single-slot theory needs to show, minimally, is that good obstruent-approximant clusters in English can be represented as single sounds. Ideally, perhaps, the single-slot theory should make the claims in (21).

- (21) A strong claim:
- a. If a cluster is not a single sound, it is bad in English and Chinese
 - b. If a cluster is a single sound, it is good in English and Chinese

It is possible to maintain (21a), provided we make some provisions for the initial [s], and perhaps initial [ʃ] and [θ] as well (more below). It is hard to maintain (21b), which claims that all possible single sounds are used in both English and Chinese. For example, [nw] is a possible single sound (see below), which is used in Chinese (e.g. *nuan* ‘warm’) but not in English. Similarly, [pl] is a possible single sound (see below), which is used in English but not in Chinese. Moreover, [mw] is a possible single sound (see below), which is not used in either English or Chinese. In view of this, I will assume a moderate claim, given in (22).

- (22) A moderate claim:
All clusters used in English and Chinese are single sound.

What (22) implies is that there are clusters that can be represented as single sound but are not used in English or Chinese. I will return to this issue in section 5. In what follows I discuss Chinese and English separately.

4.1. Chinese onset clusters

All onset clusters in Chinese are made of a consonant and a glide (CG) (see Appendix 2). There are three glides [j w ɥ]. All CG clusters can be represented as single sounds. Some examples are shown in (23), where Vocal-cords features are omitted.

| | | | |
|------|-------------------|-------------------------|----------------------------------|
| (23) | Sounds | Example | Articulator representation |
| | [t] | <i>ta</i> ‘he’ | *Coronal |
| | [k] | <i>kan</i> ‘see’ | *Dorsal |
| | [t ^w] | <i>tuan</i> ‘league’ | *Coronal, Labial |
| | [t ^j] | <i>tian</i> ‘sky’ | *Coronal, Dorsal-[-back] |
| | [t ^ɥ] | <i>lue</i> ‘abbreviate’ | *Coronal, Labial, Dorsal-[-back] |
| | [k ^w] | <i>kuai</i> ‘fast’ | *Dorsal, Labial |
| | [k ^j] | | *Dorsal-[-back], Coronal |
| | [k ^ɥ] | | *Dorsal-[-back], Coronal, Labial |

The last two sounds in (23) are not used in Standard Chinese; their absence needs a separate explanation in any theory. It is worth noting, however, that [k^j] is used in classic Chinese and some Min dialects. Next consider the sounds in (24).

| | | | |
|------|---------------------|----------------------|------------------------------------|
| (24) | Sounds | Example | Articulator representation |
| | [s] | <i>san</i> ‘three’ | *Coronal |
| | [s ^w] | <i>suan</i> ‘garlic’ | *Coronal, Labial |
| | [s ^j] | <i>xia</i> ‘down’ | *Coronal, Dorsal-[-back] |
| | [s ^ɥ] | <i>xue</i> ‘study’ | *Coronal, Dorsal-[-back], Labial |
| | ([ç]) | <i>xia</i> ‘down’ | *Coronal, *Dorsal-[-back]) |
| | ([ç ^w]) | <i>xue</i> ‘study’ | *Coronal, *Dorsal-[-back], Labial) |

The pronunciation [s^j] and [s^ɥ] is used by some Beijing speakers, especially among females (Cao 1987, Hu 1991). Other Beijing speakers use [ç] and [ç^w] instead (shown in parentheses), in which case both Coronal and Dorsal are major articulators. This analysis shows that [s^j] and [ç] have the same articulators, so do [s^ɥ] and [ç^w]. It is possible, therefore, not to count [ç] as an underlying phoneme in Standard Chinese. The same applies to [tç] and [tç^h] (Duanmu 2000). In fact, one can go a step further and argue that the affricates [ts] and [ts^h] are not underlying phonemes either but are derived from [t]+[s] and [t^h]+[s] respectively when the cluster shares a single onset slot.

A reviewer points out that evidence from speech errors suggests that affricates behave like single sounds, but clusters like [pl] and [pr] behave like two sounds. I will return to this issue in section 5.

4.2. English onset clusters

Let us now consider English. As mentioned above, the initial [s] must be treated separately in any theory. Excluding three-sound clusters, where the initial sound is always [s], there are 56 two-sound clusters, shown in (25), including 15 consonant-[j] clusters (see Appendix 1). Some clusters are rare but are still included, such as [fw], [bw], [zw], and [zl]. The reason is that if they do occur, even infrequently, a theory should account for them. Similarly, not all dialect has all 15 Cj clusters, but some do, so all of them are included. The cluster [hw] is not included, although some speakers use it in such words as *why* and *what*. However, [hw] can be represented as a single sound, so even if it is included, the conclusion does not change. Affricates are not included but will be discussed later.

(25) English onset cluster

fw, fr, fl, sw, sv, sr, sl, sn, sm, sf, sp, st, sk, ʃw, ʃr, ʃl, ʃn, ʃm, θw, θr, vw, vr, vl, zw, zl, bw, br, bl, dw, dr, gw, gr, gl, pw, pr, pl, tw, tr, kw, kr, kl, lj, nj, mj, vj, hj, fj, sj, θj, zj, bj, dj, gj, pj, tj, kj

Example

fuate, fry, fly, swing, svelte, Sri Lanka, slot, snow, smoke, sphere, spot, stop, sky, schwa, shrink, schlock, schnauzer, schmooze, thwart, three, voile, vroom, Vladimir, zwieback, zloty, Buenos, bring, blow, dwarf, dry, Guam, green, glass, pueblo, pray, play, twin, try, quick, cry, class, lewd, news, mew, view, huge, few, assume, enthuse, presume, beauty, due, guile, pure, tube, cute

First, consider the analysis of labials in (26).

| (26) | Sounds | Example | Articulator representation |
|------|--------------------|---------|---------------------------------------|
| | [p] | pay | *Labial |
| | [p ^w] | pueblo | *Labial-[+round] |
| | [p ^l] | pure | *Labial, Coronal, Dorsal-[-back] |
| | [p ^l] | play | *Labial, Coronal-[+lateral] |
| | [p ^{rw}] | pray | *Labial-[+round], Coronal-[-anterior] |

Both [p] and [p^w] have the articulator Labial. Their difference is that [p^w] has an additional feature [+round] associated with Labial. The liquid [l] is Coronal. The liquid [l] is in fact [l^{rw}], so it is Coronal and Labial. Thus, the two liquids can be

distinguished even without using the feature [lateral] or [anterior]. The glide [j] is Dorsal-[-back] and Coronal, and so it is distinct from both [l] and [r^w]. The same analysis applies to [fr, fl, vw, vr, vl, bw, br, bl, fj, bj]. Next consider velars, shown in (27).

| | | | |
|------|-------------------------------|---------|---------------------------------------|
| (27) | Sounds | Example | Articulator representation |
| | [k] | car | *Dorsal |
| | [k ^w] | quick | *Dorsal, Labial |
| | [k ^l] | class | *Dorsal, Coronal-[+lateral] |
| | [k ^j] | cute | *Dorsal-[-back], Coronal |
| | [k ^{r^w}] | cry | *Dorsal, Coronal-[-anterior], Labial, |

Again, the approximants [l r w j] can be distinguished as different minor articulations. Next consider coronals, shown in (28).

| | | | |
|------|-------------------|---------|----------------------------|
| (28) | Sounds | Example | Articulator representation |
| | [t] | tin | *Coronal |
| | [t ^w] | twin | *Coronal, Labial |
| | [t ^l] | tune | *Coronal, Dorsal-[-back] |
| | [t ^l] | | *Coronal-[+lateral] |

There is no problem representing [t] and [t^w], nor is there any problem with [t^l], which is used in some dialects. The sound [t^l] can also be represented, a lateral stop perhaps, although it is not used in English. The cluster [tr] will be discussed below. Next, consider some palatal sounds, shown in (29).

| | | | |
|------|--------|---------|---|
| (29) | Sounds | Example | Articulator representation |
| | [ʃ] | show | *Coronal-[-ant], *Dorsal-[-back] Stricture: [-son, +cont] |
| | [tʃ] | chip | *Coronal-[-ant], *Dorsal-[-back] Stricture: [+strident, -cont, -son] |

As discussed above, a palatal has Coronal and Dorsal-[-back]. The fricative [ʃ] differs from [ç] by having [-anterior] (i.e. retroflex) associated with Coronal. Following Steriade (1989), affricates are represented as strident stops, so the stricture features of [tʃ] are [+strident, -continuant, -sonorant]. Next, consider [tr], analyzed in (30).

| | | | |
|------|------------------------------|-----------------|---|
| (30) | Sounds [tr ^w] | Example tree | Articulator representation Stricture: [+strident, -son, -cont] *Coronal[-ant], Labial (*Coronal[-ant], *Dorsal[-back], Labial) |
|------|------------------------------|-----------------|---|

This sound is an affricate and rounded, so its stricture features are [+strident, -son, -cont]. Its articulators are Coronal and Labial. The feature [-anterior] indicates that the tongue tip is curled back (retroflex). It is possible that [tr^w] also has a dorsal component, similar to that in [tʃ], shown in parentheses. If so, the sound can also be transcribed as [tʃ^w] or [tʃ^{rw}]. A similar analysis applies to [dr].

As discussed in the previous section, in the single-slot analysis, it is possible not to include affricates in the underlying phonemic inventory. Instead, [tʃ dʒ tr dr] can be derived from stop-fricative clusters [t]+[ʃ], [d]+[ʒ], [t]+[r], and [d]+[r] respectively, which merge into an affricate when they share a single onset slot.

I have shown clusters that can be represented as single sounds. There are also clusters that cannot be represented as single sounds. Consider the examples in (31).

| | | |
|------|--------|------------------------------------|
| (31) | Sounds | Problems |
| | [tm] | [t] is [-nasal] [m] is [+nasal] |
| | [ms] | [m] is [+nasal] [s] is [-nasal] |
| | [mt] | [t] is [-nasal] [m] is [+nasal] |

Such clusters do not occur in English. In the sonority analysis, it is because the sonority distance is below 2 (see section 2). In the present analysis, it is because they cannot be represented as single sounds. In particular, in a single sound each feature can only take one value, but in the above clusters the feature [nasal] must take two values. The same analysis applies to other obstruent-nasal or nasal-obstruent clusters. Next, consider another kind of clusters that do not occur, shown in (32).

| | | |
|------|--------|---|
| (32) | Sounds | Problems |
| | [bz] | [b] is the major articulator [z] is not a major or minor articulator |

In the sonority analysis the cluster is bad because the sonority distance is too small (zero). In the single-slot analysis the reason is different. Here [b] has greater

closure, so it is the major articulator; [z] cannot be a major articulator because its stricture ([-son, +cont]) is different from that of [b] ([-son, -cont]). But [z] cannot be a minor articulator either, because the stricture of a minor articulator is an approximant [+son, +cont], whereas the stricture of [z] is [-son, +cont]. Thus, [bz] cannot be a single sound. The same analysis applies to a large number of stop-fricative and fricative-stop clusters. Next consider the cluster in (33), which does not occur.

| | | |
|------|--------|--------------|
| (33) | Sounds | Problems |
| | [rb] | same as [br] |

In the sonority analysis the cluster is bad because the sonority distance is too small (negative 2). In the single-slot analysis it is possible to represent [rb] as a single sound, because [b] and [r] involve different articulators, so they can be made at the same time. But this means that [rb] is the same as [br] (or [b^r]). Thus, we cannot have [br] and [rb] as two different clusters. The same analysis applies to other approximant-obstruent clusters. Next consider the cases in (34), which do occur in English.

| | | | |
|------|--------|-----------|--|
| (34) | Sounds | Example | Problems |
| | [sm] | smooth | [s] is [-nasal] [m] is [+nasal] |
| | [θr] | three | [θ] is [+anterior] [r] is [-anterior] |
| | [ʃn] | schnauzer | [ʃ] is [-nasal] [n] is [+nasal] |

In each cluster, a feature must take two values, thus the cluster cannot be represented as a single sound. Nevertheless, such clusters are found in English. A solution is to say that English sometimes tolerates an initial fricative outside a syllable, in particular a voiceless fricative with the Coronal articulator (but see also [zw] ‘zwieback’ and [zl] ‘zloty’, where a cluster starts with a voiced Coronal fricative). The same analysis applies to the clusters [sw, sv, sr, sl, sn, sp, st, sk, ʃw, ʃr, ʃl, ʃm, θw]. It is worth noting that [sw, ʃw, θw] can be represented as single sounds, because the labial gesture [w] and the coronal gesture [s, ʃ, θ] can be made at the same time. However, English [sw, ʃw, θw] are not pronounced as single sounds, because the labial gesture does not start at the beginning of [s, ʃ, θ] but after them.

Finally, consider the clusters [nw, zr, ʒw, ʒr, ʒl, ɸw, ɸr, ɸl, θl]. As mentioned earlier, they satisfy the sonority distance and the place of articulation requirement, but they are not found in English. In the sonority analysis their absence has no explanation. In the present analysis, most of them can be explained, as shown in (35).

- (35) Not single sounds: [zr, ʒr, ʒl, ɸr, ɸl, θl]
 Fricative-[w]: [ʒw, ɸw]
 Unexplained: [nw]

Six of the clusters cannot be represented as single sounds, so their absence is expected. The lack of [ʒw, ɸw] seems to be related to the fact that, except [hw], English fricatives do not take [w] as a minor articulator in general (in [fw] and [vw], [w] is not a minor articulator; see the discussion of [pw] above). Even in [sw, ʃw, θw], which do occur, the fricatives are pronounced separately from [w], as just mentioned. The lack of [nw] remains unexplained (it does occur in Chinese though).

4.3. Summary

I have shown that all onset clusters in Chinese can be represented as a single sound. In addition, excluding an initial voiceless coronal fricative, all English onset clusters can be represented as a single sound. On the other hand, there are some clusters that can be represented as single sounds but are not used in English or Chinese. I address the problem in the next section.

5. Further issues

In this section I discuss some issues that are related to the present proposal and topics for further research.

5.1. Evidence from speech errors

An anonymous reviewer points out that in English speech errors, consonant clusters such as [br] and [fl] can often be split. An example is shown in (36), from Fromkin (1973: 249).

- | | | |
|------|-------------|-------------|
| (36) | Intended | Error |
| | brake fluid | blake fruid |

In the example, the [r] from [br] has switched position with the [l] from [fl]. Such examples indicate that each cluster is made of two separate sounds. In contrast,

there is no evidence that affricates such as [tʃ] and [dʒ] can split, indicating that they are single sounds.

In the present analysis, [br, fl, tʃ, dʒ] are all complex single sounds, so their difference needs an explanation. The answer, I suggest, lies in the difference between major and minor articulators. Consider the articulator structure of [br] and [tʃ], shown in (37).

| | | |
|------|--------|----------------------------|
| (37) | Sounds | Articulator representation |
| | [br] | *Labial, Coronal |
| | [tʃ] | *Coronal, *Dorsal |

In [br], Labial is the major articulator (for [b]) and Coronal is the minor articulator (for [r]). In [tʃ], both articulators are major articulators. The speech error evidence shows that it is possible for errors to occur only with the minor articulator, without affecting the major articulator. It is also possible for errors to occur with both articulators, or with true clusters, as shown in (38), from Fromkin (1973: 247).

| | | |
|------|----------------|-----------------|
| (38) | Intended | Error |
| | drop a bomb | bop a dromb |
| | western Mexico | weksern mestico |

In the first example, the complex sound [dr] has switched with [b]. In the second example, a true cluster [st] has switched with another cluster [ks].

In summary, evidence from speech errors does not contradict the present proposal that some consonant clusters are complex single sounds.

5.2. Unused complex sounds

There are some clusters that can be represented as single sounds but are not used. For example, [kj] can be represented as a single sound, but it is not used in Standard Chinese, and [nw] is used in Standard Chinese but not in English, although both English and Chinese use [rw]. The answer, I suggest, is that a language can choose which sounds to use and which not to use. For example, English uses [θ] and [ʃ] but not [ʒ] or [ç], whereas Standard Chinese uses [ʒ] and [ç] but not [θ] or [ʃ]. The same may apply to clusters (or complex sounds), namely, each language can choose which complex sounds to use and which not to use. On the other hand, it remains an open question whether the chosen sounds of a language should show some balance or cohesion. The answer is beyond the scope of the present discussion.

5.3. Difference between English and Chinese

There is a striking difference between onset clusters in English and Chinese. In English a cluster is usually made of a consonant and a liquid ([l] or [r]), whereas in Chinese it is usually made of a consonant and a glide. It would be desirable to find an explanation for the difference, but that is again beyond the present paper.

5.4. Coda consonant

I have not discussed the coda. For example, why is [kɛpl] a bad English syllable? In the sonority approach, the answer is as follows. The sonority slope should rise in the onset and fall in the coda. In [kɛpl], the sonority is rising in [pl], so the syllable is bad. In the present analysis, the answer seems less obvious: since English allows CVC syllables, and since [pl] can be analyzed as a single complex sound [p^l], shouldn't we expect [kɛp^l] to be good?

The real answer, I suggest, lies in the fact that there are more restrictions on codas than on onsets. For example, there are languages that do not allow any coda consonants at all. There are also many languages in which coda consonants are more restricted than onset consonants. For example, Cantonese has about 20 consonants, but only six can occur in the coda: [p t k m n ŋ]. Similarly, Standard Chinese has about 20 consonants, but only two can occur in the coda: [n ŋ]. Therefore, it is natural that English also has restrictions on the coda; in particular, it cannot be a complex sound.

5.5. Coda clusters

English has monosyllabic words that have consonant clusters in the coda, such as [græsp] 'grasp', [sɪks] 'six' or [frɪlz] *frills*. The clusters are pronounced in sequence, and so they cannot be analyzed as complex single sounds. Indeed, an English word can have up to four consonants after the vowel, such as [sɪksθs] 'sixths'.

Coda clusters pose two problems for the sonority theory. First, the sonority distance in clusters like [sp] of 'grasp' or [ks] of 'six' is zero, which differs from the minimal sonority distance for onset clusters (which is two). The sonority theory must assume that onsets and codas have different sonority requirements. Second, even if we assume that the sonority distance for English codas can be zero, many other coda clusters with zero sonority distance are still bad, such as [kp], [tp], [mn], [nm], etc.

It has been noted that the extra coda consonants in English are usually [t, d, θ, s, z], which are all coronal obstruents (stops and fricatives). However, not all coronal obstruents are allowed; in particular, [ð] is not allowed as an extra

consonant. It is relevant to observe that [t, d, θ, s, z] are suffixes in English but [ð] is not. A better generalization, therefore, is to say that extra coda consonants in English must be suffixes or suffix-like consonants. Obviously, suffixes must be used regardless what the regular syllable requirement is. But leaving aside suffix-like consonants, the coda cluster [sp] in ‘grasp’ still remains a problem for the sonority approach.

There is an alternative analysis of English coda consonants (Borowsky 1989, Harris 1994, and Hammond 1999), according to which the English syllable only allows one coda consonant. The extra [p] in ‘grasp’ is not inside the first syllable, but in a separate syllable whose nucleus is an empty vowel [græs.pØ] (Burzio 1994, Harris 1994). Suffix consonants can be attached to the word final syllable, but they need not be an integral part of it (such attachment has been called ‘appendixes’). In other words, just as there is no onset cluster in English, there is no coda cluster in English either. The maximal English syllable, therefore, is CVX, as it is in Chinese (Duanmu 1990, 2000).

5.6. Phonetic evidence

An anonymous reviewer points out that if clusters like [pl] and [pr] are complex single sounds, they should be phonetically similar in duration as other single sounds, such as [p]. Is there such evidence for the present analysis?

The segmentation of speech sounds is a complicated matter and is subject to various interpretations. For example, an English vowel is considerably longer before a voiced consonant. Did the consonant lose some duration to the vowel, or did the vowel gain duration at no cost to the consonant? Also, if a consonant cluster is 70% of the sum of two separate consonants, does it mean that the cluster counts as one sound, or do we always expect a cluster to be shorter than the sum of two separate consonants?

With the above cautionary note, what kind of phonetic evidence would bear on the present proposal? I suggest the following. In the syllable CCVX or CGVX, if CC or CG are single sounds, the duration of CCVX and CGVX should be similar to that of CVX. On the other hand, if CC and CG are clusters of two sounds, then CC and CG should perhaps be consistently longer than CVX.

I am not aware of any evidence in Chinese that suggests that CGVX is consistently longer than CVX. I am not aware of any evidence in English either that suggests that CCVX and CGVX are consistently longer than CVX. In the absence of counter-evidence, it is better to maintain the simpler theory or the null hypothesis, according to which both Chinese and English have a single slot for the syllable onset.

5.7. Other languages

The present analysis is based on English and Chinese. It is natural to ask whether it can extend to languages that have more complex onset clusters, such as German, Polish, and Croatian. I am unable to fully address the issue here but will offer some speculations. The central assumption is that sounds around a syllable need not be a proper part of it. For example, in English an extra [s] or [ʃ] is allowed to precede a syllable. In addition, English consonant suffixes can follow the word final syllable. In other languages the allowance may extend to other consonants. It is important to note that, in languages like English and Polish, complicated onset and coda clusters occur mostly at the beginning and the end of a word, but not in the middle of a morpheme (Borowsky 1989, Rubach and Booij 1990). This means that the core syllable is generally quite simple. The extra sounds at word-edges may be extrasyllabic (Borowsky 1989), or in a separate syllable whose nucleus is an empty vowel (Burzio 1994, Harris 1994). Thus, there is no compelling reason to assume that the basic syllable structure in other languages is more complicated than that in English and Chinese.

6. Concluding remarks

I have pointed out problems in the sonority-based analysis of onset clusters. In particular, there are clusters that satisfy the sonority requirement (and the place dissimilation requirement) but are not used, and there are clusters that violate the sonority requirement (or the place dissimilation requirement) but are used. In addition, the sonority approach has to assume different analyses for Chinese and English.

I have offered an alternative in which there is a single onset slot in both Chinese and English. Using articulator-based feature theory, and excluding initial coronal fricatives in English (as other theories must do), all good onset clusters are shown to be single sounds. The present analysis keeps the null hypothesis of syllabic structure, in which there is only one onset slot (and one coda slot). In other words, there is no need to assume more complicated syllable structures. In addition, there is no need to assume different syllable structures for Chinese and English.

An anonymous reviewer suggests that, although the present proposal has outlined an alternative to the cluster theory, more arguments against the latter are needed, in order for people to give up the established view. I hope to have shown that, although the cluster theory was the only choice before the single-slot theory became available, there have always been problems with the sonority analysis, which comes with the cluster theory. Since the single-slot theory assumes a simpler syllable structure and avoids the problems of the sonority analysis, there is no motivation for the cluster theory anymore.

*Acknowledgments:

The idea that English has a single onset slot was originally suggested by Peter Kipka (class discussions, 1987-1988), although the details of the present proposal are worked out by myself. Some ideas in this article were discussed in my phonology classes at the University of Michigan. I thank the students for their input. An earlier and shorter version of this work was presented at the Seventh International and the Nineteenth National Conferences on Chinese Phonology, Taipei, May 26, 2001, and was included in the proceedings. I thank Yuchau Hsiao for organizing the conference and the audience, in particular Moira Yip, for their comments. I also thank an anonymous reviewer for some useful suggestions.

Reference

- Borowsky, Toni. 1989. Structure preservation and the syllable coda in English. *Natural Language and Linguistic Theory*. 7: 145-166.
- Burzio, Luigi. 1994. *Principles of English Stress*. Cambridge, UK: Cambridge University Press.
- Cao, Yun. 1987. Beijinghua tç zhu shengmu de qianhua xianxiang [The fronting of the tç series of onsets in Beijing Mandarin]. *Yuyan Jiaoxue Yu Yanjiu* [Language Teaching and Research] 1987.3: 84-91.
- Chomsky, Noam and Morris Halle. 1968. *The Sound Pattern of English*. N.Y.: Harper and Row.
- Clements, G. N. 1985. The geometry of phonological features. *Phonology Yearbook* 2: 225-252.
- _____, and Elizabeth Hume. 1995. The internal organization of speech sounds. *The Handbook of Phonological Theory*, ed. by John Goldsmith, 245-306. Cambridge, MA: Blackwell.
- Davis, Stuart and Michael Hammond. 1995. On the status of onglides in American English. *Phonology* 12: 159-182.
- Duanmu, San. 1990. *A Formal Study of Syllable, Tone, Stress and Domain in Chinese Languages*. Cambridge: MIT dissertation.
- _____. 1994. Against contour tone units. *Linguistic Inquiry* 25.4: 555-608.
- _____. 2000. *The Phonology of Standard Chinese*. Oxford: Oxford University Press.
- Fromkin, Victoria A. 1973. *Speech Errors as Linguistic Evidence*. The Hague: Mouton.
- Goh, Yeng-Seng. 2000. Beijinghua shi danyinjie yuyan de zhiyi [Is Beijing Mandarin a monosyllabic language?]. *Dangdai Yuyanxue* [Contemporary Linguistics] 2.4: 231-247.
- Goldsmith, John. 1976. *Autosegmental Phonology*. Cambridge: MIT dissertation. Reproduced by the Indiana University Linguistics Club, Bloomington, Indiana.
- Halle, Morris. 1995. Feature geometry and feature spreading. *Linguistic Inquiry* 26: 1-46.
- _____, and G. N. Clements. 1983. *Problem Book in Phonology*. Cambridge: MIT Press.
- Harris, John. 1994. *English Sound Structure*. Oxford, UK: Blackwell.
- Hu, Mingyang. 1991. *Yuyanxue Lunwenxuan* [Selected writings in linguistics]. Beijing: Renmin Daxue Chubanshe.
- Kenstowicz, Michael. 1994. *Phonology in Generative Grammar*. Oxford, UK: Blackwell.

- Ladefoged, Peter and Morris Halle. 1988. Some major features of the International Phonetic Alphabet. *Language* 64.3: 577-82.
- Maddieson, Ian. 1984. *Patterns of Sounds*. London: Cambridge University Press.
- McCarthy, John. 1988. Feature geometry and dependency: A review. *Phonetica* 43: 1988; 45: 84-108.
- Roca, Iggy and Wyn Johnson. 1999. *A Course in Phonology*. Oxford, UK: Blackwell.
- Rubach, Jerzy and Geert Booij. 1990. Syllable structure assignment in Polish. *Phonology* 7: 121-158.
- Sagey, Elizabeth. 1986. *The Representation of Features and Relations in Nonlinear Phonology*. Cambridge: MIT dissertation.
- Steriade, Donca. 1989. Affricates are Stops. Paper presented at the Conference on Features and Underspecification Theories. Cambridge: MIT.