Phonology: Sound Structure

Abigail C. Cohn*

1 Introduction

Consider the "words" shown in (1):

(1)	I	II	III	
a.	*xoda [x] as in German **ch**.	poda	poda[z]	(cf. coda, codas)
b.	*ngatus =[ŋ]	matus	matus[ɪz]	(cf. mattress, mattresses)
	*rudih	hurid	hurid[z]	(cf. herd, herds)
c.	*bnick	blick	blick[s]	(cf. block, blocks)

Fluent speakers of English would agree that none of these are actual words of English, yet most speakers would also agree that those in column I are not possible words (we use an * to indicate an impossible or "ungrammatical" form); while those in column II are. In addition, most speakers would agree that the plurals of the would-be words in column II would be pronounced as indicated in column III. How do we know this? Our knowledge of the sound patterns of our native language(s) comes not through memorizing a list of words, but rather by internalizing information about the allowed and disallowed sound patterns of that language. As fluent speakers of English, we know which sounds occur in our language and which don't. For example, in (1a), the [x] sound of German (written **ch** in borrowings from German, as in the German pronunciation of *Ba**ch***) just doesn't occur in English. In addition, some occurring sounds of English are nevertheless restricted in the position where they occur within the word. As shown in (1b), the sound represented by the spelling sequence **ng** [ŋ] cannot occur at the beginning of a word (though it occurs in the middle (*si**ng**er*) or end (*si**ng***)), while **h** cannot occur at the end of a word (but it occurs at the beginning (***h**ot*) or middle (*a**h**ead*)). We also know which sounds can be combined into a sequence. Thus in (1c), **bl** is an allowable sequence at the beginning of a word (***bl**ue*), while **bn** is not. Finally, we also know how sound patterns alternate. For example, in the regular plural formation in English, what is written as **s** or **es** is pronounced [s], [z], or [ɪz] depending on certain properties of the last sound of the word. As native speakers, without thinking we produce the expected forms (*block[s], herd[z], mattress[ɪz]*). It is this knowledge about sound structure—which sounds occur, what their distribution is, how they can be combined, and how they might be realized differently in different positions in a word or phrase—that constitutes the study of **phonology**.

Central to research in phonology is documenting and characterizing the full range of attested sound structures and patterns across the languages of the world.[1] In this chapter, we explore some of the central generalizations about sounds, using theories and tools that allow us to insightfully analyze these patterns. We will focus on three areas: sound inventories and contrasts (Section 2), structure above the level of the sound unit or segment, that is prosodic organization (Section 3), and structure internal to the segment (Section 4). The general approach followed here is **generative phonology** (see Chomsky and Halle 1968, also Kenstowicz and Kisseberth 1979) where the

---

*This chapter is to appear in M. Aronoff and J. Rees-Miller (eds.) *Handbook of Linguistics, 2nd edition*. Oxford: Blackwell. Thanks to Beverley Goodman, Lisa Lavoie, Peggy Renwick, Ayako Tsuchida, and Draga Zec for providing helpful input on earlier drafts of this chapter and special thanks for Emma Lantz for assisting in the formatting and editing of this version and Linda Heimisdóttir for her careful editing.

[1] We also talk about the organization of gestures of sign languages as phonology, but we won't develop the parallels here.

goal is to develop a theory that accurately models a speaker's knowledge of his or her language. In Section 5, we consider phonology in a broader context, considering alternative views and identifying emerging trends.

2 Inventories and Contrasts

2.1 Inventories

All languages have consonants and vowels. Consonants are sounds with a constriction in the vocal tract, while vowels lack such a constriction. Vowels can serve as the core of a syllable (see below in Section 3), while consonants generally cannot. Consonants must cooccur with vowels to produce forms that are pronounceable. Both consonants and vowels can be defined in terms of where in the mouth and how they are produced. Consonants are characterized in terms of **place** and **manner** of articulation. Place of articulation (e.g. labial, coronal, velar) indicates where the obstruction occurs created by the movement of an "active" (e.g. the tongue) and "passive" (e.g. the soft palate) articulator. The manner of articulation indicates the degree of constriction: complete closure (stops), noticeable obstruction (fricatives) or a combination of closure and obstruction (affricates), closure in the mouth with air escaping through the nose (nasals), or only approximation (liquids and glides). Vowels are generally characterized in terms of the height of the tongue or jaw (high, mid, low), the relative backness of the tongue (front, central, back), and whether the lips are rounded or unrounded. Other properties play a role, such as whether the vocal cords are close together and vibrating (voiced) or farther apart, allowing freer passage of air from the lungs (voiceless).[2]

So far we have presented examples using English spelling, with some additional pronunciation information provided in square brackets ([ ]). English spelling is sorely inadequate for accurately describing the sounds currently used in English. The 26 symbols of the Roman alphabet are not sufficient to represent all of the consonant and vowel sounds of English (as we'll see there are about 39, depending on the dialect). But this isn't the only issue. In order to describe sounds reliably, we need a fully systematic relationship between sound and symbol, something that English spelling doesn't provide, since there are many-to-many correspondences of sound to symbol. For example, the sound [k] corresponds to several different symbols or symbol combinations—***c**at, **k**ite, **kh**an, **q**uite* (**qu** = [kw]), *e**ch**o, pa**ck**, bo**x*** (**x** = [ks]), whereas the letter **c** represents various sounds— [k]: ***c**at*, [s]: ***c**ite*, [tʃ]: ***c**ello* (not including two-symbol combinations, such as **ch**).

Additionally, we often need to be able to include more pronunciation detail. This need for greater detail is true even of languages with much more transparent spelling systems than English. We need what is called **phonetic transcription**. The International Phonetic Alphabet (IPA) is a system of phonetic transcription that allows us to systematically represent the sounds of any language. This system, developed by the International Phonetic Association (founded in 1886) is periodically updated, to reflect changes in thinking on transcription and to include new speech sounds that have been "discovered".

A sound inventory is the set of sounds occurring in a particular language. Looking across the inventories of the languages of the world, we find that the number of consonants and vowels, as well as the specific selection of sounds, varies greatly from one language to another. In his study of the sound inventories of 317 languages, Maddieson (1984, updated Maddieson and Precoda 1990) found that the number of consonants in a language ranged from 6 to 95, with a mean of 22.8; while the number of vowels ranged from 3 to 46 with a mean of 8.7.

---

[2]The description of possible human speech sounds falls within the purview of (linguistic) phonetics. For an introduction, see Ladefoged and Johnson 2011, and Ladefoged and Disner 2012, also Chapter 10 of the *Handbook of Linguistics (2nd edition),* Phonetics.

Considering this range of sound inventory size, let's see how the inventory of American English compares, as shown in (2). For the consonants (C), the places of articulation are the column headings and the manners of articulation are the row headings. When two sounds appear within a single cell, the one on the left is voiceless (without vocal cord vibration) and the one on the right is voiced (with vocal cord vibration). For the vowels (V), in addition to tongue backness (marking the columns) and height (marking the rows), adjacent pairs within a category differ in "tenseness" (longer and more peripheral in the vowel space, e.g. [i]) vs. "laxness" (shorter and more centralized, e.g. [ɪ]). There are also three diphthongs (vowel-glide combinations that function as a single vocalic unit).

(2) Sound inventory of English

C's	labial	dental	alveolar	palato-alv	palatal	velar	glottal
stop	p b		t d			k g	
fricative	f v	θ ð	s z	ʃ ʒ			h
affricate				tʃ dʒ			
nasal	m		n			ŋ	
liquid			r l				
glide					j	w	

V's	front	central	back
high	i ɪ		ʊ u
mid	e ɛ	ɝ ʌ	ɔ o
low	æ	a	

Diphthongs: aj, aw, oj

There is some variation in the number of sounds argued to occur in English. This is due in part to dialect variation; for example, [a] ***cot*** vs. [ɔ] ***caught*** contrast in some dialects but not others. There are also analytic questions, such as whether the affricates, [tʃ] (***ch**ur**ch***) and [dʒ] (***j**udg**e***), should be treated as single units or as sequences of sounds. Nevertheless, the characterization of American English in (2) with 24 consonants, 12 vowels and 3 diphthongs is fairly common. Thus, English has an average-sized consonant inventory, though notable in its rich array of fricatives. There are whole classes of consonants that English doesn't exemplify, such as clicks, found in languages of Southern Africa. With 12 vowels, English has a relatively rich vowel inventory, especially considering that the distinctions are all made using only the two dimensions of tongue height and backness. (In the inventory above, we haven't included schwa [ə], which occurs only in unstressed position.)

We can compare the English inventory with a language that has a rich consonant inventory such as that found in Arabic. In Modern Literary Arabic, we find a very small vowel inventory, only three distinct vowel qualities (though length differences also result in differences in meaning, e.g. [d**u**r] 'turn!' vs. [d**u:**r] 'houses'), but a very rich consonant inventory. Not only are most of the consonants seen in English found here, but there are additional places of articulation, notably at the back of the mouth (uvular and pharyngeal). In addition, there is a contrast between plain consonants and those with a superimposition of a back tongue position (pharyngealization) and finally consonants also contrast for length ([ba**r**a] 'sharpen' vs. [ba**rr**a] 'acquit'). Including all these contrasting dimensions, there are 48 consonants in this variety of Arabic.

While the inventories of English and Arabic might suggest that there is a tendency for languages with large consonant inventories to have correspondingly small vowel inventories and vice versa, this is not necessarily the case. Consider for example Rotokas, spoken in Papua New Guinea (the smallest inventory in Maddieson's 1984 database), with a very common five vowel inventory [i, e, a, o, u], but only six consonants [p, t, k, g, β, ɾ] for a total of only 11 segments.

While there is great variation in the segments that occur in particular languages—Maddieson and Precoda (1990) identify roughly 900—strong predictions can nevertheless be made about which sounds will occur. Some sounds and categories of sounds are just more common than others. For example, all languages have stops, but not all languages have fricatives. Beyond these basic observations, in many cases the presence of one property implies the presence of something else in the same system; such generalizations are called implicational language universals. For example, if a language has the mid vowels [e, o] (as in English, *bait* [bet] and *boat* [bot]), it is predicted that it will also have the high vowels [i, u] (English *beat* [bit] and *boot* [but]) and the low vowel [a] (English *pot* [pat]); but the converse doesn't hold, as we've seen in Arabic with [i, u, a], but lacking [e, o]. (See Hyman (2008) for a discussion of such phonological universals.) An ongoing area of investigation is identifying principles (both phonological and phonetic) that play a role in defining attested as well as potentially well-formed sound inventories, such as the principle of "economy" (Clements 2003), whereby contrasting phonological classes tend to be used maximally.

2.2 Contrast

When we characterize the inventory of sounds of a language, we draw an important distinction between those sounds that can be used to make meaningful contrasts in a language, vs. those that occur only in predictable contexts. The above discussion of the inventories of English, Arabic, and Rotokas reflects those sounds argued to be **distinctive** or **in contrast** in the language (though, as we discuss below in Section 4, the status of [ŋ] in English is debatable).

In order to determine the status of a sound, we use a simple test for **minimal pairs**. Minimal pairs (or sets) are words with distinct meanings differing only in one sound. Thus we can show that [m] and [n] (differing only in place of articulation) are distinct sounds in English, since the substitution of these sounds alone is enough to change the meaning of a word: ***m****eat* vs. ***n****eat, si****mm****er* vs. *si****nn****er, ra****m*** vs. *ra****n***. The presence of [m] vs. [n] at the beginning, middle, or end of a word results in words with distinct meanings.

If a sound is used distinctively in a particular language, it is a **phoneme** in that language (and is transcribed in / /'s). Phonemes are argued to be the units encoded in lexical entries (the forms in our mental dictionaries), and upon which speakers judge "sameness" and "differentness". However, phonemes can vary in their actual pronunciation, depending on the context of the neighboring sounds, the structure of the utterance, and so forth.

Two languages may have the same sounds or **phones** (the actual phonetic events, transcribed in [ ]'s), but their grouping into phonemes might be different. In English, the sounds [b, p, p^h] all occur (that is, voiced, voiceless unaspirated, and voiceless aspirated). While [p^h] and [b] contrast, the selection of [p] or [p^h] is determined by the phonological context, as schematized in (3a). *Buy* [baj] contrasts with *pie* [p^haj], but the realization of a voiceless stop as aspirated word initially (*pie* [p^haj]) or unaspirated following [s] (*spy* [spaj]) is predictable and there are no minimal pairs for [p] and [p^h]. These three phonetic categories are mapped to only two abstract phonological categories. Yet in Thai, all three sounds occur and can produce differences in meaning, as shown by the minimal set in (3b).

(3) a. English b. Thai

	phonemes	phones	phonemes	
buy [baj]	/b/	**[b]**	/b/	[baa] ‘crazy’
pie [p^haj]	/p/		/p/	[paa] ‘aunt’
spy [spaj]		**[p^h]**	/p^h/	[p^haa] ‘cloth’

but no *[paj] or *[sp^{h}aj]

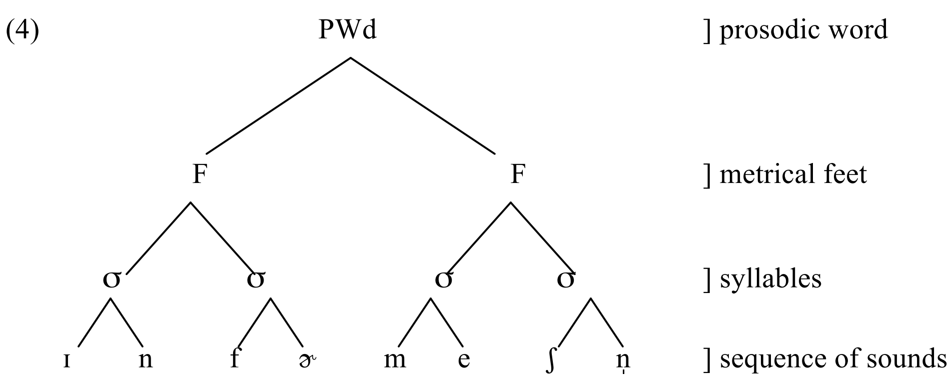
To summarize, these three phones [b, p, p^h] constitute three separate phonemes in Thai, but only two in English.

In English [p, p^h] stand in a special relationship to each other, since they are part of the same phoneme (usually taken to be /p/). Such sounds are called **allophones**. We can capture this relationship by describing the distribution, e.g., [p^h] occurs at the beginning of words and [p] occurs after [s]. (There is more to this pattern, but we won't pursue it here.) Or we can go a step further and argue that the phoneme /p/ occurs at an abstract or **underlying** level and account for the observed surface distribution with a phonological rule. We return to the issue of rules in Section 4.

3 Structure Above the Level of the Segment: Prosodic Organization

The sound structure of a word (a unit that can be defined on several linguistic levels, including morphologically and prosodically) includes not only the sequence of sounds (made up in turn of bundles of distinctive features, as discussed in Section 4), but also entails the hierarchical grouping of these sounds. Let's take the English word *information* as an example:

(4)



This word consists of a sequence of sounds **ɪ-n-f-ɚ-m-e-ʃ-n̩**. Most speakers of English would agree that this form consists of four syllables (σ) broken up as **ɪn-fɚ-me-ʃn̩.** Consonants and vowels are grouped into syllables in non-arbitrary ways, with a vowel forming the core or **nucleus** optionally flanked by consonants. In the final syllable [ʃn̩], the nucleus is n̩, which is a syllabic nasal, serving the role of a vowel. These syllables are in turn organized into stress groupings (**ɪ̀n-fɚ**)(**mé-ʃn̩**). The third syllable is the most prominent (primary stress, indicated with an acute accent ´) and the first also has some prominence (secondary stress, indicated with a grave accent `). These patterns of prominence can be accounted for by grouping the syllables together into units known as metrical feet (F). (For an introduction to metrical theory, see Hayes 1995). Finally the feet are grouped together into a Prosodic Word (PWd). The Prosodic Word often has the same shape as a morphologically defined word, but not necessarily. There are, for example, so-called function words (grammatical words), which we take to be words morphologically, but that can't stand on their own phonologically, such as *a,* or *the*. The syllables, feet, and prosodic words together constitute the prosodic structure of a word. Words in turn can be grouped into higher levels of prosodic structure as well, at the phrase and utterance level. This grouping and hierarchical organization together constitute the prosodic organization.

The structure of segments, how segments are combined, and how syllables, metrical feet, and prosodic words are organized, are integral parts of phonology. In this section, we examine syllable structure as an example of prosodic organization above the level of the segment.

3.1 Syllable Structure

Syllable structure influences the ways segments are organized, sometime leading to the insertion or deletion of a segment. Consider an example from Korean, shown in (5) where we observe that sometimes a cluster of consonants occurs and sometimes one of the members of the cluster is deleted. This is an example of what we call an **alternation** where the same morpheme varies in its realization, conditioned by some aspect of the sound system (in this case the allowable syllable structure).

(5) Consonant ~ zero alternations in Korean clusters

root		+ vowel-initial suffix		+ consonant-initial suffix	
		-a nominalizing suffix		**-t'a** infinitive	
/palp/	'tread on'	palp + a	'treading on'	pap + t'a	'to tread on'
/salm/	'boil'	salm + a	'boiling'	sam + t'a	'to boil'

The basic idea is that in some cases, Korean syllable structure can't accommodate all the segments in the underlying representation, so one is deleted in the surface form. The underlying clusters (/lp/ and /lm/) are allowed to surface before a vowel-initial suffix, since the second member of the cluster can be syllabified as the beginning of the second syllable, producing [palpa] and [salma]. But when the root occurs before a consonant-initial suffix, the first consonant of the cluster (here /l/), is deleted, producing [papt'a] and [samt'a]. (In other cases, the second consonant is deleted instead.) The syllabification of forms with vowel-initial and consonant-initial suffixes respectively is shown in (6) for /palp/ (where < > indicates a segment not incorporated into the syllabic structure):

(6)

```
 σ   σ            σ    σ
/|\  /|          /|\   /|
p a l p a        p a p t' a
                  <l>
```

This deletion is directly driven by the allowable syllable structure.

As noted in Section 1, restrictions also exist on possible sequences of sounds. For example in English, *[bn] can't occur at the beginning (7a) or end (7b) of a word.

(7) a. *bnick
b. *kibn
c. lab-network
d. drabness

e. Abner

It is not the case that [bn] is always bad in English. In (7c), this sequence is fine, however, the word is a compound and we might argue that it consists of two prosodic words grouped together ([[lab]$_{\mathrm{PWd}}$[network]$_{\mathrm{PWd}}$]$_{\mathrm{PWd}}$) and therefore it is not held to the same restrictions. The fact that (7d) is allowable might be attributed to the sounds belonging to different morphemes (*drab* and *-ness*). But in (7e) there aren't distinct words or morphemes. What then is the difference between [bn] in (7a & b) and (7e)? In the latter case, the [b] and [n] are in different syllables, while in the former they are in the same syllable. The restriction holds of a sequence within a syllable and seems to be due to the fact that [b] and [n] are too similar in terms of **sonority**. Sonority can be defined loosely as the degree of constriction in the mouth during the production of a particular sound. Most important for our purposes is the observation that there is a hierarchy of how **sonorous** sounds are. Vowels are more sonorous than consonants; and within the consonants, further divisions can be made. Nasals, liquids, and glides, together known as **sonorants**, are more sonorous than fricatives, with constriction creating frication or noise, and stops, with complete closure. Stops and fricatives (together with affricates) are known as **obstruents**, since there is a significant obstruction. Thus we find the following widely observed cross-linguistic pattern:

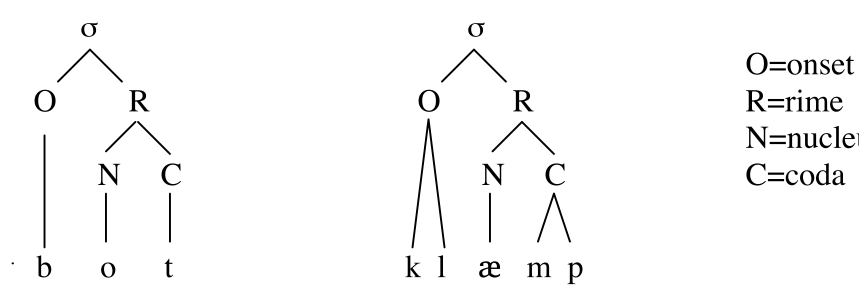
(8) Sonority hierarchy

more sonorous				less sonorous
vowels	>	sonorants	>	obstruents

The sonority hierarchy characterizes the behavior of sounds in syllable structure and many other aspects of phonological patterning. Whether finer-grained distinctions of the sonority hierarchy are required is a question open to much debate. (See Zec 2007 for recent discussion of this question and for an overview of issues in syllable structure more generally.)

Strong evidence exists for making reference to the syllable as part of the hierarchical structure of the phonological system to account for observed alternations and also to capture consonant sequencing restrictions. In addition, the syllable is argued to be divided into subparts. Evidence for this comes from the fact that cooccurrence restrictions hold on the consonants preceding, as well as following, the core of a syllable, but not generally across the subparts of the syllable. One approach to the internal organization of the syllable is shown in (9), where the substructures of *boat* and *clamp* are illustrated:

(9)



Based on a wide range of evidence, there is argued to be a division in the syllable between the **onset** and the **rime** constituents. The division into onset and rime allows us to capture various consonant sequencing restrictions and is also relevant for other aspects of the phonology, as well as language games and poetry. The rime corresponds to the unit that rhymes, e.g. ***oat**, b**oat**, bl**oat***; and the onset is the unit shared in poetic patterns of alliteration, e.g. ***bl**ue, **bl**ow, **bl**ithe, **bl**oat*. Following some views, the rime is further divided into the **nucleus**, the core of the syllable that contains the vowel or vocalic element(s), and the **coda**, which contains any following consonant(s).

In English, the only required element of the syllable is the nucleus (e.g. ***oh*** [o], ***I*** [aj]), although in many languages the onset is also an obligatory part of the syllable.

Let's consider some of the cooccurrence restrictions within onsets and codas in English. All of the examples in (10) are well-formed English syllables (and in these cases independent words too).

(10) Examples of possible syllables in English

onset: \ coda:	Ø	C	CC	CCC
Ø	oh [o]	ode [od]	old [old] amp [æmp]	amps [æmps]
C	bow [bo]	boat [bot]	bolt [bolt]	bolts [bolts]
CC	blow [blo]	bloat [blot] clam [klæm]	clamp [klæmp]	clamps [klæmps]
CCC	spree [spri]	split [splɪt]	splint [splɪnt]	splints [splɪnts]

In English, anything from no consonants to up to three consonants preceding and four following a vowel may constitute a well-formed syllable.[3] Many restrictions hold, however, on possible combinations of consonants preceding or following the vowel and only a small subset of the logically possible combinations occur. For example, in English triconsonantal onsets ($C_1C_2C_3$), the first sound (C_1) must be [s], followed by a voiceless stop ([p, t, k]), followed by a liquid ([r, l]) or glide ([j, w]). Many of the occurring patterns can be characterized with reference to the sonority hierarchy (8), though other factors also come into play. Thus in CCC onset clusters the pattern of C_2 and C_3 follows the sonority hierarchy, with onsets showing a rise in sonority going from C_2 to C_3: stops followed by the more sonorous liquids and glides. But the occurrence of [s] preceding such clusters is not predicted even with modification to the sonority hierarchy, since [s] is not less sonorous than the stops, and therefore requires a distinct explanation. Similarly, in characterizing what coda clusters can occur in English, sonority also plays an important role. In general, the first member of a two-member coda cluster must be of the same or greater sonority than the second member (e.g. *le**nt**, be**lt**, li**ft**, mi**st**, a**pt***). Such patterns can be characterized straightforwardly by making reference to the subparts of the syllable, but are much harder to characterize if we only refer to the string of segments.

How much explicit or formal internal structure to the syllable is warranted and how it should be encoded are much-debated questions; however, reference to some degree of substructure of the syllable allows for insightful generalizations about sequencing restrictions and other aspects of sound distribution. Indeed in English, we can capture the pattern presented in (7) by observing that the sequence [bn] cannot occur together as part of an onset or coda. In addition, reference to syllable subconstituency enables us to capture the broader distribution of sounds in many cases. For example, as noted in (1), the distribution of /h/ in English is limited: it can occur only in the

[3]Four consonants following the vowel is not included in (10); an example is *texts* [tɛksts]. In English, most monosyllabic forms with more than two consonants in the coda are morphologically complex, usually involving the [s] or [z] of the plural or third person singular, or the [t] or [d] of the past tense.

onset of a syllable (and if it is not word-initial, only if the syllable is stressed, e.g. *vehicle* [víıkl̩] vs. *vehicular* [vi**h**íkjəlɚ]).

While it is relatively straightforward to count the number of syllables in a word, it is often trickier to decide where the syllable break falls. Cross-linguistically in forms of the shape (C)VCV, the division generally falls before the medial C, (C)V$CV ($ indicates a syllable break). However, in English, the situation is additionally complicated by the stress pattern. In words such as those in (11a), it is widely agreed that the syllable divisions are as shown, characteristic of the strong cross-linguistic tendency.

(11)	a.	attáck	[ə$tǽk]
		belów	[bə$ló]
	b.	áttic	[ǽtık]
		béllow	[bɛ́lo]

However, many researchers have argued that in the cases such as (11b), the medial consonant belongs either to the first syllable or is shared by the two syllables in order to account for otherwise systematic observations about the relationship between syllable structure and stress in English. (Even though the middle consonants in some of the forms in (11) are written with a doubled letter (tt, ll), they are just single consonants.)

In English, in the case of (C)VCCV(C), the syllabification depends on the specific sequence of consonants. If the CC is an allowable word onset (and therefore an allowable syllable onset), the syllable division is before both consonants (12a), but otherwise it is between the two consonants (12b).

(12)	a.	apply	[ə$**pl**aj]	cf. plea [pli]
		abrupt	[ə$**br**ʌpt]	cf. brush [brʌʃ]
	b.	Adler	[æ**d**$**l**ɚ]	* [dli]
		Abner	[æ**b**$**n**ɚ]	* [bni]
		ardent	[a**r**$**d**n̩t]	* [rdi]

Other languages show much greater restrictions on syllable structure than English does. Consider some examples from Japanese in (13).

(13) Allowable syllables in Japanese: CV, V, CVN, CVC

a. CV, V

[ki]	'tree'
[kokoro]	'heart'
[mado]	'window'
[tegami]	'letter'
[ito]	'string'
[origami]	'paper folding'

b. N$C

[tombo]	‘dragonfly’
[hantai]	‘opposite’
[neŋkin]	‘pension’

c. C$C

[kitte]	‘stamp’
[onna]	‘woman’
[hakka]	‘peppermint’
[kaʃʃa]	‘pulley’

As illustrated in (13), only (C)V(:) and (C)V(:)C syllables occur in Japanese. CV syllables can occur in any position in the word (13a). But CVCs are allowed only if the coda consonant is a nasal (13b), or part of a geminate (long consonant) (13c), and in these cases usually followed by another syllable. Thus, [tom] is a well-formed syllable when followed by [bo], but it would not be allowed if it occurred on its own or as the final syllable in a word. A final alveolar nasal (as in [neŋki**n**] above in (13b)) is well-formed, but other nasals and other consonants in word-final position are not allowed.

Additional evidence for the allowable patterns can be seen by looking at the ways foreign words are modified when they are borrowed into Japanese. Let's consider what happens to some words borrowed from English, as shown in (14).

(14) Borrowings from English into Japanese:

	word	English	Japanese
a.	pin	[pɪn]	[pin]
	pie	[paj]	[paj]
	Chicago	[ʃɪkago]	[ʃikago]
b.	million	[mɪljən]	[mirion]
	avocado	[avəkado]	[abokado]
	rally	[ræli]	[rari:]

Some words are borrowed basically as is (14a), or with modifications to any non-occurring segments, with these being substituted by a similar sound that does occur in Japanese (14b). (The lax vowels of English [ɪ, ɛ, ʊ, ɔ] are realized as short [i, e, u, o] and the tense vowels [i, e, o, u]) are realized as being long [i:, e:, u:, o:] in Japanese.)

Of particular interest are cases where non-allowable consonant clusters occur; in such cases, Japanese uses the strategy of adding extra vowels (epenthesis), as illustrated in (15):

(15) More borrowings from English into Japanese

	word	English	Japanese
a.	free	[fri]	[fUri:]
	spray	[spre]	[sUpUre:]

b.	peak	[pik]	[pi:kU]
	kiss	[kɪs]	[kisU]
	Bill	[bɪl]	[birU]
	beat	[bit]	[bi:tO]
c.	speed	[spid]	[sUpi:dO]
	cross	[krɔs]	[kUrosU]
	test	[tɛst]	[tesUtO]
	street	[strit]	[sUtOri:tO]
	contrast	[kantræst]	[kontOrasUtO]
	baseball	[besbɔl]	[basUbarU]

Consider first cases with onset clusters in (15a). The inserted vowels are indicated in upper case symbols. (The vowel that is inserted in these cases is usually [u] (U), except after alveolar stops, where an [o] (O) is inserted.) (15b) shows borrowings of monosyllables of the shape CVC. Here final vowel epenthesis occurs since a consonant can occur in coda position only if it is followed by an appropriate consonant in the next syllable in the same word. Finally cases with both an onset cluster, and final consonant or consonant cluster are shown in (15c). All of these clusters are broken up into many more syllables in Japanese than found in the original English source, with the exception of [nt] in *contrast* which is well-formed in Japanese (cf. 13b).

In the case of non-allowable clusters in borrowed words, some other languages delete segments. Consider what happens to final consonant clusters in Indonesian in words borrowed from English (or Dutch). In Indonesian, the allowable syllable structure is (C)V(C), so final clusters in borrowed words pose a problem. As shown in (16), the final clusters are simplified by deleting the final consonant (similar to the pattern seen for Korean above in (5), although in those examples, the first member of the cluster was deleted).

(16)	<u>word</u>	<u>English</u>	<u>Indonesian</u>
	sport	[spɔrt]	spor
	aqueduct	[ækwədəkt]	akuaduk
	tolerant	[talɚrn̩t]	toleran
	test	[tɛst]	tes

To account for such systematic syllable patterns, phonologists have proposed various devices including rules, templates, well-formedness conditions, and constraints.

3.2 A Constraint-Based Account

A current approach, **Optimality Theory**, involves the idea of competing phonological constraints, which can be ranked in importance with respect to each other. Due to such ranking, a less important constraint can be violated in order to obey a more important constraint that it conflicts with. Languages differ in how they rank particular constraints. If we have correctly identified the relevant constraints (a major research agenda in itself), then the set of logically possible rankings of those constraints should match up with the range of sound patterns seen across languages. (See Kager 1999 and McCarthy 2008 for an introduction.) Optimality Theory offers an insightful account of syllable patterns and makes strong predictions about allowable syllable types cross-linguistically, and it also accounts for certain implicational universals such as the fact that if a language allows CVC syllables it will also allow CV syllables and if it allows V syllables, again it will also allow CV ones.

As widely discussed, the ideal syllable is CV. Syllables minimally consist of a vowel; onsets are preferred; and codas are dispreferred. To account for the preference for CV syllables as well as the range of cross-linguistic variation observed in syllable structure, two general sorts of constraints interact. First there are markedness constraints—constraints that capture systematic cross-linguistic biases. The preference for CV syllables has been argued to emerge from three constraints, stated here informally:

(17) Syllable structure markedness constraints:

	constraint	informal definition
a.	NUC	Syllables must have a nucleus
b.	ONSET	Syllables must have an onset
c.	NOCODA	Codas are not allowed

If this were the whole story, all languages would have only CV syllables, but this is clearly not the case (look at English!). There are also constraints that mediate between the underlying representation (the input to the constraints) and the actual realization of the form (the output of the constraints). The two constraints relevant for our purposes, again stated informally, limit how different the input and output can be. (* = Don't)

(18) Input/output constraints

	constraint	informal definition
a.	*ADD	Only the material of the input should appear in the output; don't add material to the input [standardly called DEPIO]
b.	*DELETE	Underlying material should be incorporated in the output; don't delete material from the input [standardly called MAXIO]

Other constraints can also affect syllable structure, but these five constraints suffice for our discussion. To test constraint rankings, we compare the input of a form and a list of possible (expected) outputs (placed in the leftmost column of what is termed a **tableau**) with respect to a particular ranking of the relevant constraints (placed in columns, going from higher to lower ranking as we go from left to right). No matter what the relative ranking of these five constraints in a particular language, if we have an input or underlying form of the shape CV(CV)(CV), then all of the above constraints—those affecting syllable structure and those affecting input/output relations—can be satisfied. This is true in both English and Japanese, as shown in (19a) for English *banana* [bənænə] and (19b) for Japanese [kokoro] 'heart'. In these tableaux, the constraints are all unranked, indicated by the dashed vertical lines, in contrast to solid vertical lines that we'll see in the tableaux below.

(19) a. English *banana* [bənænə]

/bənænə/	NUC	ONSET	NOCODA	*ADD	*DELETE
[bə$næ$nə]	√, √, √	√, √, √	√, √, √	√, √, √	√, √, √

b. Japanese [kokoro] 'heart'

/kokoro/	NUC	ONSET	NOCODA	*ADD	*DELETE
[kokoro]	√, √, √	√, √, √	√, √, √	√, √, √	√, √, √

A checkmark in the relevant cell indicates that the constraint is met; there are three checkmarks in each cell referring to each of the three syllables in these cases. ONSET and NOCODA together (no matter what their ranking) ensure that an intervocalic consonant (VCV) will be syllabified as the onset of the second syllable (V$CV).

Let's now consider some cases where the same input results in different outputs in different languages. Consider the English word *test*, which as we saw above is realized as [tesuto] in Japanese and [tes] in Indonesian. (I leave NUC and ONSET out of the following discussion, as they are met by all of the cases we are considering.)

In English, the input [tɛst] matches the output, even though it violates NOCODA twice. This provides evidence that NOCODA is ranked below both *ADD and *DELETE. In other words, in English meeting the requirements of the input/output constraints is more important than adhering to the markedness constraints. We show this by comparing possible outputs (20a – e) and showing that with a given ranking, the optimal or best-formed candidate (20a, [tɛst], indicated by ☛) is the observed output, while the others (20b-e) fail to surface.

(20) English *test* [tɛst]

/tɛst/		*ADD	*DELETE	NOCODA
a. ☛	[tɛst]	√	√	**
b.	[tɛs]	√	*!	*
c.	[tɛ]	√	*!*	√
d.	[tɛstV]	*!	√	*
e.	[tɛsVtV]	*!*	√	√

When a form violates a constraint with respect to the input, we mark that violation with "*". The optimal candidate violates the highest-ranked (leftmost) constraint the least number of times, if at all. In case of a tie at that level, lower-ranked constraints also come into play. An ! indicates an insurmountable violation. This is followed by shading of the successive cells in the same row, indicating that the adherence to these lower ranked constraints isn't relevant to the outcome. (20a) is the optimal candidate in this case, even though this form violates NOCODA twice. This is still preferable to a violation of either *ADD (20d & e) or *DELETE (20b & c), providing evidence that both of these constraints outrank NOCODA (hence NOCODA is positioned to the right, separated by a solid vertical line). Since both *ADD and *DELETE have to be met, we don't have evidence from this example for their relative ranking in English.

The pattern in Japanese is very different. In Japanese, priority is given to the markedness constraints over the input/output constraints. In order to meet the high ranking NOCODA constraint, vowels are inserted, providing evidence that *DELETE outranks *ADD, as shown in the tableau in (21):

(21) Japanese [tesuto] 'test'

/tɛst/		NOCODA	*DELETE	*ADD
a.	[test]	*!*	√	√
b.	[tes]	*!	*	√
c.	[te]	√	*!*	√
d.	[testV]	*!	√	√
e. ☛	[tesVtV]	√	√	**

We see here that (21e) [tesuto], which respects both NOCODA and *DELETE, is the optimal candidate. We use V to represent an inserted vowel. The actual quality of the inserted vowel is assumed to be a language-specific question. As we saw above in (13b & c), Japanese tolerates some limited violations of NOCODA. However, these codas cannot have their own place specification, rather they must share it with the following onset consonant, either as part of a geminate or as part of a nasal-stop cluster agreeing in place of articulation.

Finally in Indonesian, we find a case where deletion is tolerated, indicating the relatively low ranking of *DELETE. This is balanced with a violation of NOCODA, since the optimal form involves one violation of each NOCODA and *DELETE (in contrast to English that violates NOCODA twice and Japanese that violates *ADD twice).

(22) Indonesian [tes] 'test'

/tɛst/		*ADD	NOCODA	*DELETE
a.	[test]	√	*!*	√
b. ☛	[tes]	√	*	*
c.	[te]	√	√	*!*
d.	[testV]	*!	*	√
e.	[tesVtV]	*!*	√	√

The optimal candidate in Indonesian is (22b). Our analysis accounts for the fact that both (22d & e) are eliminated, but more needs to be said about why the optimal outcome is (22b) rather than (22a or c). An additional constraint must be involved; the intuition is that a single consonant in coda position is more acceptable than a cluster and also, there is a limit to how much deletion the system will tolerate. There is also more to the story, since in the case of onset clusters, vowels are inserted rather than consonants being deleted, for example Indonesian *stasiun* [sətasiun] from Dutch *station*, but we leave aside these additional details in our current discussion.

There are clearly additional complexities, since all three languages allow vowel-initial words (hence limited violations of ONSET) and more needs to be said about why, in Japanese, a final syllable such as [kin] is allowed but one such as [tom] is not. Finally, additional constraints are needed to account for the division of medial consonant clusters into codas and onsets, e.g. English *abrupt* [ə$brʌpt] vs. *Abner* [æb$nɚ]. In many languages, VCCV will surface as V$CCV if CC is an allowable onset (again additional constraints are required). If CC is not an allowable onset, the VC$CV syllabification would be the optimal candidate.

While I haven't provided a complete account of these three cases, we can see that the relative ranking of this limited set of constraints allows us to capture these different strategies of syllabification. Other languages are predicted to show different outputs. For example, the output form [testV] would be optimal in a language that had some tolerance of single consonant codas (like Indonesian), but ranked *DELETE over *ADD.

In this section we have seen that grouping of sounds into syllables and subsyllabic constituents offers a more insightful account of sound patterns than one where only reference to the segment is made. In addition we have looked briefly at how a constraint-based approach, with minimal constraint violation, allows us to account for some of the cross-linguistic variation observed in syllable structure. Grouping of sounds into syllables is the lowest level of prosodic organization.

4 Subsegmental Structure

4.1 Features and Segmenthood

Till this point, we have focused on segments and larger units. Good evidence for the psychological reality of segments exists, including speaker intuition, alphabetical writing systems, speech errors, and the fact that phonological alternations involve such units. But there is also good evidence that segments are made up of smaller units and that by making reference to these smaller units, a more insightful discussion of sound patterning is possible. We have an intuition that [p, b] are more similar to each other than [p, l] are. This is because the former share more sound properties than the latter. These sound properties are called **distinctive features**. The notion of distinctive features grows out of the work of Trubetzkoy, Jakobson, and others (see Anderson 1985 for an excellent survey of the history of phonology). While numerous specific systems have been proposed, most current systems have developed from Chomsky and Halle (1968). Most approaches to phonology assume some kind of feature system and take the features to be the smallest building blocks of phonology. Segments thus consist of bundles of features, or feature matrices, as exemplified for *bill* /bɪl/, in (23):

(23) Feature matrices

$$\begin{array}{ccc} \text{b} & \text{ɪ} & \text{l} \\ \begin{bmatrix} \text{+consonantal} \\ \text{-continuant} \\ \text{-sonorant} \\ \text{-nasal} \\ \text{labial} \end{bmatrix} & \begin{bmatrix} \text{-consonantal} \\ \text{+high} \\ \text{-back} \\ \text{-tense} \end{bmatrix} & \begin{bmatrix} \text{+consonantal} \\ \text{-continuant} \\ \text{+sonorant} \\ \text{-nasal} \\ \text{coronal} \\ \text{lateral} \end{bmatrix} \end{array}$$

Evidence for specific feature proposals comes from their adequacy in capturing the recurrent cross-linguistic grouping of sounds, referred to as **natural classes**. Take for example the feature [±sonorant] (where [+sonorant] defines the class of sonorants and [-sonorant] defines the class of obstruents). [+sonorant] is defined as that class of sounds for which spontaneous vocal cord vibration (or voicing) is possible, including nasals, liquids, glides, vowels, sounds for which there is not a close obstruction of the vocal tract. In the typical case, sonorants are voiced and do not show a contrast between voiced and voiceless. For obstruents ([-sonorant]: stops, fricatives, and affricates), voicing involves certain articulatory adjustments to maintain subglottal air pressure and keep the vocal cords vibrating. For obstruents, the least marked category is voiceless, but obstruents often show a contrast between [+voice] and [-voice]. A strong implicational universal is that if a language has a voicing contrast among sonorants (as found, for example, in Burmese), then it also has a voicing contrast among obstruents. The natural class defined by [±sonorant] is also shown by syllabic consonants in English (the nasals and liquids in words like *bottle* [baɾl̩] and *button* [bʌʔn̩]) and the division between the sonorants and obstruents is crucial to the sonority hierarchy discussed above.

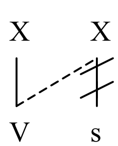
A striking fact is that features themselves are not arbitrary classificatory elements, but rather are closely linked to phonetic structure. Thus we find a convergence of phonetic events and the sounds that are found to pattern together in the phonologies of language after language. I will not provide a systematic discussion of distinctive features, since a number of good overviews are available (see, for example, Clements and Hume 1995). I refer here rather informally to specific features.

Many interesting and important issues surround the status of features. The fundamental question is whether features are universal, with the same set characterizing the phonological patterns in all languages, and if so why. The dominant view (at least till recently) was that features are universal because they are innate (part of the human language endowment). This has led to much debate about a specific set of features that can account for all the occurring sounds in the languages of the world. If the striking patterns of sounds across languages are not due to an innate set of features, the questions to be answered are different, but equally interesting. How are distinctive features learned and to what degree are they similar or identical across languages? Much current work in phonetics, phonology, and language acquisition addresses itself to these questions. (For discussion and review of this literature, see Cohn 2011, Mielke 2008). Here I assume that phonological features are roughly equivalent across languages and leave open the question of how they are acquired.

Often the patterning of sounds is characterized in terms of the specific featural content of segments, but other times the presence or absence of segments themselves accounts for an observed pattern. Thus sometimes it is appropriate to refer to the segment as a unit independent of its featural content. In some cases, a segment is deleted without leaving any evidence behind (such as the Korean consonant deletion case illustrated above in (5)), but in other cases, the timing of a deleted segment "stays behind". This is called **compensatory lengthening**. Consider the widely discussed case from Latin illustrated in (24).

(24)	/ko**s**mis/	[ko:mis]	'courteous'
	/ka**s**nus/	[ka:nus]	'gray'
	/fide**s**lia/	[fide:lia]	'pot'

We see in (24) that an /s/ is deleted before another consonant (/m, n, l/). But the /s/ doesn't completely disappear; rather it leaves behind its timing unit (indicated here by an X), resulting in a lengthening of the preceding vowel. We can capture this change as follows (where **V** and **s** informally represent the relevant bundles of features).

(25)


The feature bundle of /s/ is deleted but its timing unit is reassociated with the preceding vowel. Direct reference to the timing aspect of a segment allows us to capture this straightforwardly.[4]

[4]To incorporate this notion of "segmenthood", some approaches include timing units, and others propose an internal hierarchical grouping of features within the segment, including a "root node", as discussed below. An alternative approach to timing is "moraic theory" where the basic units—morae—characterize the weight properties of segments; see Broselow (1995) for a comparison of these approaches.

4.2 Alternations

With these further refinements of the representation of phonological units—features organized into segments and timing units, in turn grouped into larger units—we are ready to consider one of the central observations in phonology. Oftentimes phonemes are realized in different ways in different contexts as determined by position in the word, neighboring sounds, (un)stressed position, and so forth. Such differences in the realization of a phoneme are the clearest evidence of the effects of phonology. As seen above, alternations can result from aspects of the higher level organization (for example, in the consonant ~ zero alternations in Korean due to syllable structure). But effects are also found due to the quality of neighboring segments. To take a simple example from English, the prefix /ɪn-/ 'not' changes its shape depending on the following consonant:

(26) [ɪn]
inappropriate
intolerant
indecent

[ɪm]
impossible
imbalance

[ɪŋ]
incoherent
inglorious

The nasal becomes more similar to the following consonant by sharing its place of articulation, with a coronal nasal [n] before coronals (and also vowels), a bilabial nasal [m] before bilabial stops, and a velar nasal [ŋ] before velars. The morpheme /-ɪn/ has several allomorphs, including [ɪn-, ɪm-, ɪŋ-]. This is an example of **assimilation**, whereby a sound becomes more similar to its neighbor(s). While such patterns of nasal place assimilation are very common cross-linguistically, this pattern is not as systematic in English as in some other languages, since a nasal consonant doesn't always share the place of articulation of the following consonant. For example, in forms compounded with the particle /ɪn-/, for some speakers, assimilation doesn't take place: cf. *input*, [n-p] *income* [n-k]. The (non-)application of assimilation in this case is argued to be due to a difference in morphological and prosodic structure between these cases.

It is also assimilation, in this case, of voicing, that accounts (in part) for the alternation in the shape of the regular plural marker in English that we saw above in (1). As observed above, what is spelled as *s* or *es* is pronounced as [s], [z], or [ɪz]. The distribution of these three variant shapes of the plural morpheme is not arbitrary. Rather the distribution is systematically determined by the voicing and place of articulation of the final sound of the stem.

(27)

a. [s]		b. [z]		c. [ɪz]	
cap	[p]	cab	[b]	match	[tʃ]
cat	[t]	fad	[d]	judge	[dʒ]
book	[k]	dog	[g]	mess	[s]
		can	[n]	buzz	[z]
		file	[l]	wish	[ʃ]
		bow	[o]	garage	[ʒ]

If the final sound of the stem is voiceless, as shown in (27a), then the shape of the plural marker is [s]. (This holds systematically for the stops, but the situation with voiceless fricatives is more complicated: sometimes the voiceless fricative itself becomes voiced and then takes the voiced allomorph [z], such as *leaf* [f], *leaves* [vz], but sometimes the pattern for the stops is found, *chef* [f] *chefs* [fs].) As shown in (27b), if the final sound of the stem (whether an obstruent, sonorant, or vowel) is voiced, then the shape of the plural marker is [z]. Thus the voicing of the final sound in the stem conditions the shape of the plural marker, which agrees in voicing with that sound. But there is a systematic exception to the pattern seen in (27a & b), as illustrated in (27c). If the final sound is either an affricate [tʃ, dʒ], or an alveolar or palato-alveolar fricative [s, z, ʃ, ʒ], then the shape of the plural marker is [ɪz]. The intuition here is that [s] or [z], added to stems ending in these sounds would be too similar to be perceptually distinct and so a vowel is inserted to break up the cluster. While some limited exceptions exist, such as *mouse-mice, sheep-sheep, child-children*, there is good evidence that speakers intuitively know the "rule" that accounts for the phonetic shape of the plural marker. Such evidence comes from the fact that both children acquiring English and adults when faced with new words added to the language apply these rules in forming the plural, for example *macs* [s] and *PCs* [z] and some people even say *mouses* [ɪz].

Such patterns of assimilation are very common across the languages of the world. This is an area where we see a close parallel between the phonology and phonetics. It is a common property of speech that neighboring sounds are coarticulated, that is, that the articulation of adjacent sounds overlaps. Such phonetic effects can become exaggerated and over time result in phonological assimilation. Let's consider another example, the case of vowel nasalization in Sundanese (a regional language of Indonesia).

(28) Sundanese vowel nasalization

a.	[atur]	'arrange'	[ŋãtur]	'arrange' (active)
	[obah]	'change'	[ŋõbah]	'change' (active)
	[parios]	'examine'	[mãrios]	'examine' (active)
b.	[tiis]	'relax in a cool place'	[nĩĩs]	'relax in a cool place' (active)
	[saur]	'say'	[ɲãũr]	'say' (active)

In Sundanese, an initial vowel or one following an oral consonant is oral, while one following a nasal consonant is nasalized. This alternation between nasalized and oral vowels can be seen in corresponding bare stems and active forms, since the active is formed by adding [ŋ] or [+nasal] to the initial consonant of the root, as shown in (28a). Not only is a single vowel following a nasal consonant affected, but a sequence of vowels becomes nasalized, as shown in (28b). Such examples illustrate the importance of distinctive features for an adequate description of such alternations. If we couldn't make reference to a single feature (e.g. [voice] or [nasal]) or set of features (needed, for example, to account for nasal place assimilation), we would be missing a fundamental insight into what is going on in such cases. Within the generative framework, following the seminal work of Chomsky and Halle (1968), *The Sound Pattern of English* (SPE), such patterns are accounted for by rules of the form: a —> b/ c__ d, "a becomes b in the environment following c and preceding d". The general rule schema offers a formalism for accounting for observed phonological alternations. Rather than just describing the distribution of the differing allophones, this rule formalism incorporates the fundamental idea that one of the variants is basic, or underlying, and that the other variant(s) are derived by rule. Such rules are an attempt to capture the knowledge that a speaker has about the sound patterns of his or her language. Alternatively such patterns can be analyzed as the interaction of violable constraints (as discussed for syllable structure in Section 3.2).

Following the SPE rule-based approach, the pattern of nasalization in Sundanese can be represented as shown in (29a), with an example of the application of the rule or "derivation" in (29b).

(29) Sundanese Vowel Nasalization:

a. V —> [+nasal] / [+nasal] ____
Condition: applies iteratively
"A vowel becomes [+nasal] when it is in the environment following a sound that is [+nasal]"

b.

Underlying representation	/tiis/	/[+nasal] + tiis/
Vowel Nasalization	—	nĩis
iterative		nĩĩs
Surface representation	[tiis]	[nĩĩs]

Further developments suggest that a more accurate account follows from the idea of assimilation as "feature spreading", rather than the changing of feature values. This is part of a more general approach termed **autosegmental phonology**, where specific features can function independently of segments. Following this approach, we could characterize vowel nasalization in Sundanese as follows:

(30) a.
X V
[+nasal]

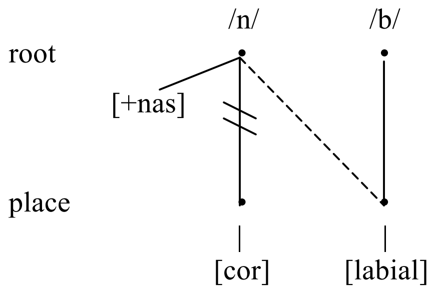
b.
n i i s
[+nasal] [-nasal]

The autosegmental rule in (30a) indicates that the [+nasal] feature specification spreads to the right to a following vowel, resulting in structures such as that illustrated in (30b). Here the pattern of assimilation is captured directly through the sharing of a single feature specification. This has the added advantage of allowing us a straightforward account of the iterative nature of this process.

We also saw an example of spreading in our characterization of compensatory lengthening above in (25), where the whole feature matrix specifying the vowel is shared between the vowel's timing unit and the following timing unit, freed up by the loss of the feature matrix of the /s/. Viewed in this way, this too can be seen as a sort of assimilation, in this case, total assimilation.

In addition to assimilation of a single feature (e.g. vowel nasalization) and total assimilation (e.g. compensatory lengthening), there are cases where two or more features systematically pattern together. Such is the case of nasal place assimilation, exemplified above for the English prefix /ɪn-/. Cases where a particular set of features pattern together in assimilation and other phonological processes provide strong evidence for the grouping of features (see Clements and Hume 1995 and work cited therein). This general approach, termed **feature geometry**, not only captures the notion of the segment as a unit independent from its featural content (represented by a **root node**), but it also offers an explicit proposal of hierarchical structure or subgrouping of features, making direct reference to elements such as the **place node**. An account of nasal place assimilation following this approach is schematized in (31).

(31) Nasal Place Assimilation



Segments can influence each other in a wide variety of ways. There is a rich array of patterns of assimilation, including cases where the segments affecting each other are not adjacent, such as **vowel harmony** where vowels agree in a certain property (e.g. height, backness or rounding) irrespective of the quality of the intervening consonants. For example, suffixes in Turkish vary in their shape depending on the backness of the vowel(s) in the root. Thus *balta* 'axe', *balta**lar*** 'axes', but *kedi* 'cat', *ked**iler*** 'cats'.

The contrast between segments can be lost in a particular environment. This is known as **neutralization**. Feature changes may be brought about due to the segmental context (that is, influence of neighboring segments), but it is also the case that prosodic structure can drive such effects. It is quite common that the range of contrasts is more restricted in syllable codas than in syllable onsets. One very common pattern of neutralization is what is known as **Final Devoicing**. Consider the following example from Polish:

(32) Polish voicing alternations

a.	klu**p**	'club' sg.	klu**b**i	'club' pl.
	tru**t**	'labor' sg.	tru**d**i	'labor' pl.
b.	tru**p**	'corpse' sg.	tru**p**i	'corpse' pl.
	ko**t**	'cat' sg.	ko**t**i	'cat' pl.

We see an alternation in the voicing of the final consonant of the stem. Just looking at the forms in (32a), we might think that either the voiceless stops are underlying and become voiced between vowels, or that the voiced stops are underlying and become voiceless at the end of the word. But looking at the data in (32b), we see that not all cases show the same alternation; here a voiceless stop surfaces in both forms. This makes it clear that the voiced stops are becoming voiceless. We also note that this pattern seems to apply to a natural class of sounds, in this case, the stop consonants. As shown in (33), this pattern also applies to velar stops (33a), and to fricatives (33b), that is, the class of obstruents (characterized as [-sonorant]). We can capture this pattern by positing underlying forms as shown in (33c) and applying a rule of Final Devoicing, which can be characterized in SPE terms as shown in (33d). Or we can account for such patterns in an autosegmental notation with the delinking of the relevant feature specification, in this case [+voice] (33e). In either case, we can see that the rule works by looking at sample derivations in (33f).

(33) Polish Final Devoicing

a.	wu**k**	'lye' sg.	wu**g**i	'lye' pl.

b. gru**s** ‘rubble’ sg. gru**z**i ‘rubble’ pl.

c. underlying forms:
/klub/ ‘club’
/trud/ ‘labor’
/trup/ ‘corpse’
/kot/ ‘cat’
/wug/ ‘lye’
/gruz/ ‘rubble’
/-Ø/ singular
/-i/ plural

d. [-sonorant] —> [-voice] / ____ # (# = word boundary)
“A member of the class of [-sonorant] becomes voiceless in word final position.”

e. root [-son] #

laryngeal [+voice]

f. Underlying representation /klub + Ø/ /klub + i/

Final Devoicing klup —

Surface representation [klup] [klubi]

We can also capture this pattern in Optimality Theoretic terms, as constraint interaction. There is a tension between a markedness constraint avoiding voiced codas, and an input-output constraint requiring faithfulness to the underlying [+voice] specification. Following Kager (1999, pp. 14-16), the relevant constraints are *VOICED-CODA “Obstruents must not be voiced in coda position”, which is in conflict with IDENT-IO(voice) “The specification for the feature [voice] of an input segment must be preserved in its output correspondent.” In Polish, *VOICED-CODA outranks IDENT-IO(voice), resulting in the output form [klup] as illustrated in the tableau in (34).

(34) Polish, /klub/ [klup] ‘club’ sg.

/klub/	*VOICED-CODA	IDENT-IO(voice)
a. [klub]	*!	
b. ☛ [klup]		*

In English where voicing is maintained in codas, IDENT-IO(voice) outranks *VOICED-CODA, resulting in a form faithful to the input voicing specification, as shown in (35).

(35) English, *club* /klʌb/ [klʌb]

/klʌb/	IDENT-IO(voice)	*VOICED-CODA
b. ☛ [klʌb]		*
a. [klʌp]	*!	

Before concluding this section, let's return to the question raised above about the status of [ŋ] in English. While we included /ŋ/ in the chart of the sound inventory in English presented in (3) above, we also noted in (1) that [ŋ] has a defective distribution. One approach to this would be to say that /ŋ/ just has a defective distribution similar to /h/, end of story. Yet this would leave a number of distributional observations unaccounted for. Consider the distributions of the three nasals of English, [m, n, ŋ] in (36):

(36)

	initial	medial	final	N-Vstop	N+V stop	N-Vstop	N+Vstop
m	map	dimmer	dim, bomb	camper	amber	camp	—
n	nap	sinner	sin	canter	candor	can't	land
ŋ	—	singer[5] [ŋ]	sing [ŋ]	canker [ŋk]	anger, finger [ŋg]	bank [ŋk]	—

[m, n] can occur in word-initial position, as well as medially and finally. They can also occur before an oral stop, either medially or finally (except that [mb] doesn't occur as a cluster within a syllable coda, hence *bomb* [bam], but *bombardment* [bəmbardmn̩t]). [ŋ], on the other hand, doesn't occur in word-initial position. Basically [ŋ] only occurs in the syllable coda, not in the onset. This generalization accounts for its absence word initially and accounts for all the cases except *singer*. Notably *singer* consists of the root *sing* plus the suffix *-er* and so the [ŋ] is, in effect, in syllable-final position until the suffix is added.

This generalization accounts for the distribution, but doesn't explain why it should be so. As noted above, sometimes sounds are limited in their distribution, but cross-linguistically we usually find if a consonant is limited, the greater restriction holds in the coda, not the onset. In other words, neutralization (such as Final Devoicing) tends to occur in codas, not onsets. If we take the spelling as a cue, a solution presents itself. We might argue that [ŋ] is not part of the underlying inventory of English, but rather that it is derived from /ng/ or /nk/ sequences. Briefly the analysis would work as follows. The underlying nasal consonants in English are /m, n/. As noted above, English has a rule of Nasal Place Assimilation whereby a nasal assimilates to a following stop (schematized above in (31)). Based on the evidence from the lack of word-final [mb] clusters, we might also posit a rule of Voiced Stop Deletion that applies to non-coronals, whereby a voiced stop following a nasal consonant is deleted word finally (37a). Given the underlying representations in (37b), the rules of Nasal Place Assimilation and Voiced Stop Deletion together (and an analysis for *singer* that we won't develop here) account for the observed patterns, as shown in the derivations in (37c).

---

[5] For some speakers, the author included, this is pronounced [sɪŋgɚ], rhyming with *finger* [fɪŋgɚ].

(37) a. Voiced Stop Deletion

$$\begin{bmatrix}\text{-sonorant}\\ \text{-continuant}\\ \text{+voice}\\ \text{-coronal}\end{bmatrix} \longrightarrow \text{Ø} \;/\; \begin{bmatrix}\text{+consonantal}\\ \text{+nasal}\end{bmatrix} \;\_\_\_\_ \#$$

"A voiced non-coronal stop is deleted word finally following a nasal consonant."

b. Underlying representation:

/dɪm/

/banb/[6]

/bænk/

/sɪng/

/fɪngɚ/

c.

Underlying representation	/banb/	/bænk/	/sɪng/	/fɪngɚ/
Nasal Place Assimilation	bamb	bæŋk	sɪŋg	fɪŋgɚ
Voiced Stop Deletion	bam	—	sɪŋ	—
Surface representation	[bam]	[bæŋk]	[sɪŋ]	[fɪŋgɚ]

When /n/ occurs before a velar consonant it assimilates in place of articulation and if the stop is voiced, it is then deleted. These rules are crucially ordered, as shown by the derivation for *bomb* and *sing* in (38). If Voiced Stop Deletion applies first, it would remove the trigger for Nasal Place Assimilation.

(38) Derivation – wrong order

(38) Derivation – wrong order

Underlying representation	/banb/	/sɪng/
Voiced Stop Deletion	ban	sɪn
Nasal Place Assimilation	—	—
Surface representation	*[ban]	*[sɪn]

[6]For *bomb*, we might assume an underlying /n/ or /m/ or even a nasal consonant that is unspecified for place of articulation.

Thus Nasal Place Assimilation must apply first. The result of Nasal Place Assimilation applying first is that it avoids Voiced Stop Deletion taking away or "bleeding" inputs to Nasal Place Assimilation. This particular type of rule ordering is called "counter-bleeding".

Some additional evidence for this analysis of [ŋ] comes from spoonerisms (speech errors where segments are transposed). In a corpus collected by Fromkin (1971) the following three examples suggest that [ŋ] is indeed composed of /ng/, since the properties of nasality and velarity can be split:

(39) Spoonerisms involving [ŋ]:

target	observed realization
Bing Crosby	bi**g** cro**n**sby
swing and sway	swi**n** n sway**g**
Springtime for Hitler	spri**g**time for Hi**n**tler

The restricted distribution of [ŋ] in English follows directly from this approach without our having to posit an underlying phoneme /ŋ/ with a defective distribution.

We have modeled the interactions of these two processes as ordered rules, but what about following a constraint-based approach? One of the central claims in most versions of Optimality Theory is that all constraints apply in parallel. The application of all constraints, while potentially highly complex (since *a lot* of constraints are active in any given phonology), should be surface true, but certain patterns of interaction, such as that seen for English nasals, are **opaque**, that is, the result of a sequence of rules is not transparent or surface-true. Since cases of opacity are predicted not to exist under the simplest set of assumptions within Optimality Theory and yet are well attested across languages, the issue of opacity is of central concern in Optimality Theory (see McCarthy 2007 for a review of these issues and some possible solutions).

In this section, we have seen a number of ways in which segments might affect each other, and evidence for reference to distinctive features, as well as their grouping. We have also seen that the division we made between structure above the level of the segment and subsegmental structure is somewhat artificial, since syllable structure can affect feature specification.

5 Phonology in a Broader Context

In this final section, I summarize the view of phonology sketched here (Section 5.1) and then put the discussion in a broader context (Section 5.2), briefly addressing alternative views to phonology and highlighting some of the current questions that are likely to propel the field forward. Section 5.3 lists journals and societies specific to phonology, as well as suggestions for further reading.

5.1 Phonology as a System

In concluding this introduction to phonology, it is useful to step back and consider how all the aspects we have discussed fit together. Most basically, a phonology consists of a set of representations—an inventory of sounds, and lexical entries defined by distinctive feature matrices—and a system of rules or constraints that act on the representations (Anderson 1985). Fundamental to the generative phonology approach is the idea that the idiosyncratic and predictable information are treated separately: the former is part of underlying representations and the predictable patterns arise through the systematic manipulation of these sounds through rules or constraints. Consider the following schematic figure:

(38) underlying representations: input

constraints/rules

surface representations: output

The underlying representation includes the phonemes for each morpheme in the language and the surface representation incorporates the phonetic variations or allophones, seen in the systematic alternations of the language, introduced as a result of the applications of a system of rules or constraints. The phonological representation includes not only the sequence of sounds, made up of timing units and featural content, but also the hierarchical grouping of these sounds into syllables and higher level prosodic units. The phonology of a language consists of the whole system taken together. Only by studying the whole phonology of a language, comprising dozens and dozens of rules (or constraints), can we understand its full complexity.

5.2 Emerging Trends and Research Questions

The view presented so far is the general view of generative phonology. While there are interesting and important differences between rule based and constraint based views, both of these views take a similar approach by understanding phonology to be systematically governed behavior which results from the phonological grammar, that is, what a speaker/hearer knows about his/her language—the linguistic competence. There are aspects of phonology that this approach captures elegantly and other aspects that are less well accounted for. In particular there is a tension about the relationship between competence and performance. The dominant view here is that the primary object of interest is competence, which can only be studied through performance (in this case, a language's surface forms).

An alternative view is that competence and performance are inextricably intertwined. In phonological terms, we can consider this in terms of the role of variation. Speech is characterized by variation of all sorts. Variation in pronunciation may be due to differences across across styles or registers used by individual speakers or across speakers, dialects, rate of speech, as well as phonological context (the main "stuff" of phonology). A traditional generative approach privileges two kinds of variation: the idiosyncratic differences that capture meaningful distinctions defining sound-meaning relationships of the lexicon, and the systematic or predictable variations defined by phonological contexts (adjacent segments, word position, syllable position, relative prominence and so forth). Work focusing on other aspects of this variation, especially **variation in the lexicon** and **sociolinguistic variation**, pushes us to see this way of looking at variation as an oversimplification (albeit, I would argue, a very useful oversimplification, see Cohn 2011 in this regard). See Chapter 13 The Lexicon and Chapter 25 Language Variation of Aronoff and Rees-Miller (eds. to appear) for fuller consideration.

In this section, we pursue these issues a bit further by considering the following question: What is the relationship between a speaker/hearer's knowledge of their lexicon and knowledge of phonological patterns? Is it the case that the phonology is purely generalizations across the lexicon, or do we "know" the rules or constraint interactions of our language independently?

Pierrehumbert (1994) asks how we can account for the distribution of medial clusters, that is, the fact that certain English consonant sequences are well formed but others are not, e.g. /mpr/, /ndr/ but not */rpm/ or */rdn/. A generative phonology approach, such as that discussed above in Section 3.1, predicts that medial clusters consist of possible codas + possible onsets. On the other hand, if such phonotactic patterns are purely statistical generalizations across the lexicon: "the likelihood of medial clusters [is] derived from the independent likelihoods of the component codas and onsets" (p. 174).

In a systematic dictionary analysis, Pierrehumbert found roughly 50 monomorphemic medial clusters. The same dictionary listed 147 possible codas and 129 possible onsets. If these could combine freely, we would find 18,963 medial clusters. With some expected restrictions, Pierrehumbert concludes that we could still expect approximately 8,708. She observes "It turned out that almost all the occurring triconsonantal clusters were among the 200 most likely combinations, and that a stochastic interpretation of syllable grammar effectively ruled out a huge number of possible clusters, eliminating the need for many idiosyncratic constraints in the grammar" (p. 169). Pierrehumbert then discusses systematic restrictions that play a role in determining the particular 50 or so medial combinations that are attested among the 200 most likely. She concludes that a probability-based syllable grammar understood in the context of certain more traditional sorts of phonological constraints accounts for the observed patterns.

These results highlight the complex interactions between competence and performance, suggesting that statistical generalizations across the lexicon may be part of how we learn phonology, but once learned indeed constitute patterns of knowledge that are independent from these generalizations. Pierrehumbert (2003) argues that some phonotactic knowledge is not tied to frequency and indeed is true abstraction across the lexicon, that is, there is phonological knowledge independent of statistical generalizations across the lexicon. "In light of such results, I will assume, following mainstream thought in linguistics, that an abstract phonological level is to be distinguished from the lexicon proper" (p. 191).

Such integrated approaches to modeling phonological patterns and knowledge fit into a view of **phonology as cognitive science** and have led to the inclusion of new methods, redefining **phonology as an experimental field** –what is termed **Laboratory Phonology** (see Cohn, Fougeron, Huffman 2012 for a comprehensive introduction). This includes more extensive work on the multiple facets of variation in sound patterning, as mentioned above. Increasing attention has also been paid to the issue of **the acquisition of phonology**, both in first and second language learning (see Chapter 19 First Language Acquisition and Chapter 28 Second Language Acquisition of Aronoff and Rees-Miller (eds. to appear)).

These new approaches and broader perspectives together help advance an understanding of the fundamental question of what speakers/hearers know when they "know" the phonology of their language.

5.3 Suggestions for Further Reading and Relevant Journals and Societies

Suggestions for further reading:

Historical perspective:

Anderson, S. (1985). *Phonology in the Twentieth Century*. Chicago: University of Chicago Press.

Textbooks:

Hayes, B. (2009). *Introductory phonology*. Malden, MA: Wiley-Blackwell.

McCarthy, J. (2008) *Doing Optimality Theory: Applying Theory to Data*. Oxford: Blackwell Publishing.

Odden, D. (2005). *Introducing phonology*. Cambridge, UK: Cambridge University Press.

Zsiga, E. (2102). *The Sounds of Language: An Introduction to Phonetics and Phonology*. Malden, MA: John Wiley and Sons.

Handbooks:

Goldsmith, J., J. Riggles, A. Yu (2011) *The Handbook of Phonological Theory, second edition*. Malden, MA: Wiley-Blackwell.

de Lacy, P. (2007). *The Cambridge Handbook of Phonology*. Cambridge: Cambridge University Press.

van Oostendorp, M., C. Ewen, E. Hume and K. Rice (2011). *The Blackwell Companion to Phonology*. Oxford: Wiley-Blackwell.

Cohn, A., C. Fougeron, and M. Huffman (2012). *Oxford Handbook of Laboratory Phonology*. Oxford: OUP.

Journals and societies specific to phonology:

Phonology, http://journals.cambridge.org/action/displayJournal?jid=PHO

Laboratory Phonology, http://www.labphon.org/home/journal

The Association for Laboratory Phonology, http://www.labphon.org/

Manchester Phonology Meeting, http://www.lel.ed.ac.uk/mfm/mfm.html

References

Anderson, Stephen. 1985. *Phonology in the Twentieth Century*. Chicago: University of Chicago Press.
Aronoff and Rees-Miller (eds. To appear) *Handbook of Linguistics, 2nd edition*. Oxford: Blackwell.
Broselow, Ellen. 1995. Skeletal positions and moras. In *The Handbook of Phonological Theory*, ed. J. Goldsmith, 175–205. Oxford: Blackwell.
Chomsky, Noam and Morris Halle. 1968. *The Sound Pattern of English.* (SPE) New York: Harper and Row.
Clements, G. N. 2003. Feature economy in sound systems. *Phonology* 20:287–333.
Clements, G. N. and Elizabeth Hume. 1995. The internal organization of speech sounds. In *The Handbook of Phonological Theory*, ed. J. Goldsmith. Cambridge: Blackwell, 245–306.
Cohn, Abigail C. 2011. Features, segments, and the sources of phonological primitives. In *Where Do Phonological Features Come From?*, ed. G. N. Clements and R. Ridouane. Amsterdam: John Benjamins, 15–41.
Cohn, A, C. Fougeron, and M. Huffman. 2012. *Oxford Handbook of Laboratory Phonology*. Oxford: OUP.
de Lacy, P. 2007. *The Cambridge Handbook of Phonology*. Cambridge: Cambridge University Press.
Fromkin, Victoria. 1971. The non-anomalous nature of anomalous utterances. *Language* 47:27–52.
Goldsmith, J., J. Riggle, A. Yu. 2011. *The Handbook of Phonological Theory, second edition*. Malden, MA: Wiley-Blackwell.
Hayes, Bruce. 1995. *Metrical Stress Theory: Principles and Case Studies*. Chicago: University of Chicago Press.
Hayes, Bruce. 2009. *Introductory Phonology*. Malden, MA: Wiley-Blackwell.
Hyman, Larry. 2008. Universals in phonology. *The Linguistic Review* 25:83–137.
Kager, René. 1999. *Optimality Theory*. Cambridge: Cambridge University Press.
Kenstowicz, Michael and Charles Kisseberth. 1979. *Generative Phonology: Description and Theory*. New York: Academic Press.

Ladefoged, Peter and Keith Johnson. 2010. *A Course in Phonetics, 7th Edition*. Boston: Wadsworth.
Ladefoged, Peter and Sandra Disner. 2012. *Vowels and Consonants, 3rd Edition*. Oxford: Wiley-Blackwell.

Maddieson, Ian. 1984. *Patterns of Sounds*. Cambridge: Cambridge University Press.

Maddieson, Ian. and Kristin Precoda. 1990. Updating UPSID. In *UCLA Working Papers in Phonetics* 74:104–111.

McCarthy, John. 2007. Derivations and levels of representation. In *The Cambridge handbook of phonology*, ed. P. de Lacy. Cambridge: Cambridge University Press, 99–117.

McCarthy, John. 2008. *Doing Optimality Theory: Applying Theory to Data*. Oxford: Blackwell Publishing.

Mielke, Jeff. 2008. *The emergence of distinctive features*. Oxford: Oxford University Press.

Odden, David. 2005. *Introducing phonology*. Cambridge, UK: Cambridge University Press.

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. P. Keating. Cambridge: CUP, 168–188.

Pierrehumbert, Janet. 2003. Probabilistic Phonology: Discrimination and Robustness. In *Probabilistic Linguistics*, ed. R. Bod, J. Hay and S. Jannedy. Cambridge, MA: The MIT Press, 177–228.

van Oostendorp, M., C. Ewen, E. Hume and K. Rice. 2011. *The Blackwell Companion to Phonology*. Oxford: Wiley-Blackwell.

Zec, Draga. 2007. The syllable. In *The Cambridge Handbook of Phonology*, ed. P. de Lacy. Cambridge: Cambridge University Press, 161.

Department of Linguistics
Cornell University
203 Morrill Hall
159 Central Ave.
Ithaca, New York 14853
acc4@cornell.edu