CHAPTER 3

ABCS: SEGMENTS, TRANSCRIPTION, AND THE BUILDING BLOCKS OF INVENTORIES

3.1 THE DELIMITATION OF UNITS

What are the basic units, the "primitives" (Cohn 2011: 17) of human speech? When we speak, when we plan an utterance, what are the building blocks we reach for to build it out of? When languages choose an inventory of contrasts from which to build a vocabulary, what is the stock from which they draw?

Saussure, as he was defining modern linguistics at the turn of the twentieth century, wrote that language was, at bottom, a system of "signs," pairings of a chunk of sound with a chunk of meaning. Before this pairing takes place, he wrote, both thought and sound are chaotic, indistinct, and undifferentiated. It is the creation of the linguistic sign that "subdivides" and "orders" each plane, by pairing some distinct "phonic material" with some distinct meaning, thus bringing about "the reciprocal delimitation of units" (1959: 112). He illustrates the idea with the diagram in Figure 3.1.

Language is neither thought nor sound, but the system that links them: "language works out its units while taking shape between two shapeless masses" (1959: 112).
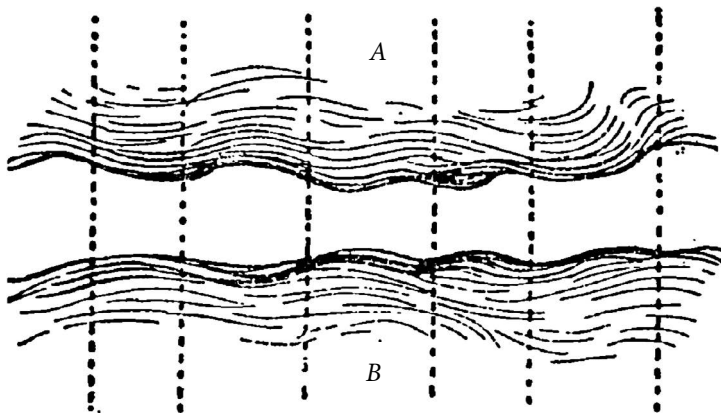

Figure 3.1   Language is "a series of contiguous subdivisions marked off on both the indefinite plane of jumbled ideas (A) and the equally vague plane of sounds (B)"

Source: Saussure (1959: 112).

Phonology and phonetics must deal with the question of how the "phonic material" associated with meanings can be characterized.

All phonology works on the premise that words are not "vocal wholes" that have no subcomponents. Hockett (1960) defined the pairing of sound and meaning with the term "duality of patterning": any utterance can be analyzed simultaneously as a string of meaningful morphemes and as a string of meaningless sounds. "The meaningful elements in any language—'words' in everyday parlance, 'morphemes' to the linguist—constitute an enormous stock. Yet they are represented by small arrangements of a relatively very small stock of distinguishable sounds which are themselves wholly meaningless" (Hockett 1960: 90). Hockett took duality of patterning to be a "basic design feature" of human language, along with arbitrariness, discreteness, productivity, and prevarication, among others.

What, then, are the units that languages use to pair sound with meaning? Of what does this "very small stock" consist? Hockett describes the stock as being composed of strings of "distinguishable sounds." Like Hockett, those of us who were raised on alphabetic writing systems take the segmental nature of speech for granted: just as written words and sentences are made up of sequences of letters, so spoken utterances are assumed to be made up of sequences of sounds to which the letters more or less correspond. Phonologists and phoneticians alike rely on segment-based descriptions of language sound systems, and represent spoken language using letter-based systems of phonetic transcription. Statements like "Hawai'ian has five vowels [i, e, a, o, u] and eight consonants [p, k, s, h, m, n, w, ʔ]" or "the English word [sprɛd] begins with three consonants" are ubiquitous. Hockett took the existence of "strings of sounds" as obvious and axiomatic. Yet phonetic study quickly shows that "segmenting" the acoustic record is not obvious. Saussure's diagram notwithstanding, it is impossible to draw exact lines in a waveform indicating where one segment ends and another begins. In both articulation and acoustics, characteristics of segment sequences overlap and blend. Given this difficultly, can we be so sure that the segment is the basic unit of both phonology and phonetics? Could it be something larger (mora, syllable, or word), smaller (distinctive feature), or different (articulatory gesture)? This chapter considers the evidence.

We begin (Section 3.2) with a discussion of the segment as a phonetic and phonological unit, which leads to a discussion of orthography and alphabetic writing in general in Section 3.3. Writing systems are not always considered to be in the domain of either phonology or phonetics, and may be more or less closely based on spoken language, but writing systems are symbolic representations of inventories and are thus implicit theories of what the basic units of a language are. Section 3.4 turns to orthographies developed for phonetic transcription, including Alexander Melville Bell's "Visible Speech" and the extremely useful but theoretically problematic International Phonetic Alphabet. Does the usefulness of segmental transcription prove that the segment is the basic unit of Language, or do the problems with the IPA prove that the segment is not basic? Section 3.5 changes the focus somewhat, looking not at linear strings, but at the problem of selecting the set of segments for any given language inventory, briefly discussing Dispersion Theory and Quantal Theory. Section 3.6 then begins the consideration of subsegmental "parameters" or

features as basic units, a discussion continued in Chapter 4. Concluding the chapter, Section 3.7 considers the question of how all of the previously-described principles might apply to non-spoken signed languages, which also show duality of patterning. What insight do signed languages give us about what the basic units of Language might be?

3.2 SEGMENTATION

As has often been pointed out, due to coarticulation and gestural overlap, it is impossible to perfectly "segment" the stream of speech by indicating exact timepoints where one segment begins and the previous one ends. Thus segmentation is one of the first and most basic problems confronting the phonology/phonetics interface. We assume discrete and categorical segments, but how can we identify them in the continuous acoustic record?

Ladd (2014, in press) emphasizes that reliance on segmental transcription has been a constant through a century of change in phonological theory: theories from the Prague School through SPE to various contemporary versions of Optimality Theory "[take] for granted the scientific validity of a segment-based idealized phonetic representation of speech" (in press: 7) and "all assume that the primary phonetic data can be expressed in terms of a segmented symbolic representation" (in press: 8). Bloch (1948: 13), for example, in laying out his "postulates for phonemic analysis" states in Postulate 12.4 that "Every utterance consists wholly of segments." Ladd (2014: 31–7) further points out that the structuralist ideas of phoneme and allophone presuppose the "phone" whose distribution can be analyzed (e.g., in Bloch 1948 and Hockett 1958), and that most generative analyses, whether rule-based or constraint-based, assume a phonological "output" that consists of a segmental string.

The problem with assuming that segments are obvious, however, can be illustrated with the simplest of utterances, such as [apa], as shown in Figure 3.2. The figure shows the waveform, spectrogram, and a proposed segmentation. The lines are not drawn randomly, and the acoustic representation is not amorphous. Discontinuities can be marked. The vowels show high amplitude, periodicity, and complex resonance structure. The consonant shows low amplitude, lack of periodicity, and no resonance structure. But are the lines in exactly the right place? The discontinuities do not all exactly line up. The voicing of the vowel dies out gradually into the voiceless closure. At exactly what point does vowel end and consonant begin? The voicelessness of the aspiration is a characteristic of [p], but the open vocal tract is a characteristic of [a]. To which segment does the period of aspiration belong? Should the utterance be transcribed as [apʰa], [apha], or perhaps [apa̯]? The answer cannot be simply read off the acoustic record, but requires a phonological analysis to determine the units, as sample problems presented for students in both Hockett (1958) and Kenstowicz and Kisseberth (1977) attest.

The problem only gets worse if one considers perception rather than acoustics. During the voiceless closure for the [p], which is where we might want to say the real consonant segment resides, the hearer perceives only silence. The labiality of the
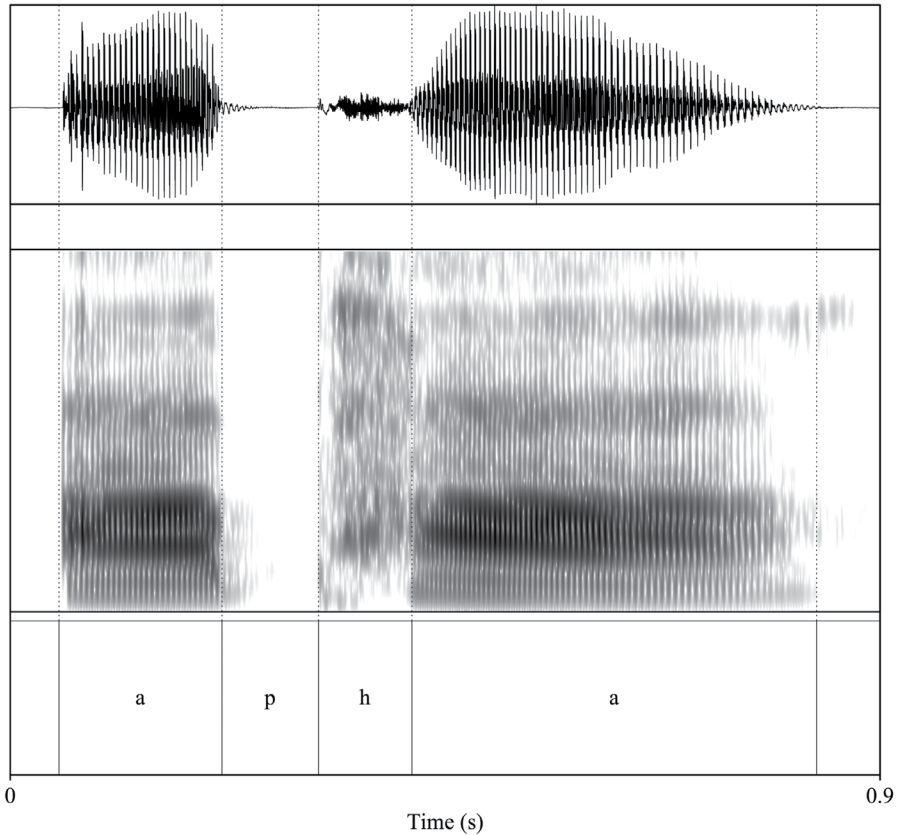
Figure 3.2   Waveform, spectrogram, and segmentation for the utterance "apa"

consonant is actually perceived in the formant transitions into and out of the closure, caused by the gradual closing of the lips, during the periods Figure 3.2 delineates as belonging to the vowel.

Yet, even if the edges are not crisp, phonologists have assumed that evidence for segments resides in perception of the (relatively) steady states. Bloch (1948: 12) writes in Postulate 11 that discrete segments can still be perceived in speech even if they can't be marked off in a spectrogram.

> Phoneticians have long known that the movements of the vocal organs from one "position" to another proceed by continuous, uninterrupted flux . . . Such instrumental data, however, need not be taken as evidence that speech AS PERCEIVED cannot be segmented; every phonetician has had the experience of breaking up the smooth flow of speech into perceptibly discrete successive parts. In Postulate 11 we do not imply that the vocal organs assume static positions or move in unidirectional ways at constant acceleration; rather, we imply that a phonetically trained observer can interpret the auditory fractions of an

utterance in terms of articulations that seem (to his perception) to be static or unidirectional. (Emphasis original)

Some more recent models of speech perception, notably Stevens (2002), also assume that steady states, interspersed with rapid transitions, serve as "landmarks" that listeners recognize as signaling segmental units. (See further discussion in Chapter 10.)

Another argument for the existence of segments is that phonology needs them. Ohala (1992) argues that segmental organization is crucial to the creation and maintenance of phonological contrast, and that language must have evolved to have segmental structure. Segmental organization, he argues, provides the necessary temporal coordination of features to one another in order to efficiently build contrastive units, and the alternation of steady states and transitions that segments provide make the different parts of the signal more readily perceptible. Ladd (2014), as quoted above, reminds us that most phonological theories are premised on segmental structure. As Anderson (1974: 6) puts it,

> The only justification [for positing segments in the continuous speech stream] that can be given is a pragmatic one; such a description, based on a segmental structure imposed on the event by the analyst, has been the basis of virtually every result of note that has ever been obtained in the field of linguistic phonetics or phonology.

That is, if segments didn't exist, we would have had to invent them.

Not everyone agrees with Anderson's conclusion, however. A number of researchers have argued that processes and patterns that have traditionally been modeled in terms of segments, including phonological alternations (Port and Leary 2005; Silverman 2006; Golston and Kehrien 2015), speech errors (Mowrey and MacKay 1990; Pouplier 2003; Brown 2004; Pouplier and Hardcastle 2005), and language acquisition and processing (Jusczyk 1992; Cheung et al. 2001; Port and Leary 2005; Golston and Kehrein 2015), can be as well or better modeled in terms of larger units such as onsets or codas, or smaller units such as features or articulatory gestures. (See also Raimy and Cairns 2015a for further examples and arguments both for and against the utility of segmental analyses in phonology.) For example, speech errors in which whole segments seem to move from one word to another, such as "toin coss" for "coin toss" and the famous "tasted the whole worm" for "wasted the whole term," seem to provide clear evidence that words consist of segmental sequences that can be permuted in error (Fromkin 1971). The authors cited above, however, argue that the transposed units could equally well be onset constituents or tongue body gestures.

Throughout this book, we'll see many cases of competing analyses that do and do not assume segmental organization. For example, Chapter 9, Articulatory Phonology, is an extended discussion of a phonological system without segments. In the sections that follow in this chapter, we'll look at some further evidence for the identity of basic units of sound structure: from orthography, from systems of phonetic transcription,

and from the ways that inventories are organized. Notwithstanding the claims of Bloch (1948), Ohala (1992), and Anderson (1974) for the centrality of the segment to phonological analysis, and the assumption of segmental structure by most phonologists (Ladd 2014), some have argued (e.g., Read et al. 1986; Silverman 2006) that our assumption that the segment is a basic unit of phonology comes solely from exposure to an alphabetic writing system.

3.3 ORTHOGRAPHIES

3.3.1 MORPHEME, SYLLABLE, SEGMENT

Orthographies are relevant to the phonology/phonetics interface because every orthography is an inventory: a set of symbols that represent the units of the language, whether those units are words/morphemes (attested in Egyptian and Chinese logograms), syllables/moras (attested in numerous syllabaries from Korean to Cherokee), or segments (in the Semitic, or Phoenician/Roman alphabets). The way that people write reflects, albeit imperfectly, an analysis of the structure of language (Berent 2013). As Aronoff (1985: 28) puts it, "Written language is a product of linguistic awareness." Thus, looking at writing systems provides a window on how languages (and thus Language) "mark[s] off . . . both the indefinite plane of jumbled ideas and the equally vague plane of sounds" (Saussure 1959: 112). Important references on the relation of writing to speech include Gelb (1963), DeFrancis (1989), Coulmas (1989, 1999), Daniels and Bright (1996), Rogers (2004), Sproat (2006), Gnanadesikan (2009), and Sampson (2015). (This section is indebted to these sources, but covers only a few of the major points raised therein.)

   Orthographies differ in how close they are to spoken language. In principle, an orthography could be completely "semasiographic," consisting of symbols like ⚠ and ⊘, which are designed by the International Organization for Standardization (https://www.iso.org) to be comprehensible in any language, and stand for ideas rather than specific words. While the inventory of such symbols is large, no semasiographic system exists that is capable of conveying the full semantic range of human language, including utterances as simple and practical as "Haven't we met before?" (Try it. Or consider the online games that attempt to write passages of famous novels in emojis.)

   In practice, all orthographies are used to represent specific utterances in specific languages, though orthographies differ in how directly they represent pronunciation. Some orthographies are closer to logo-graphic, in which every symbol represents a morpheme; others are more phono-graphic, in which every symbol represents a sound. None is perfectly one or the other, however. Even basically logo-graphic systems, like Chinese, have some phonetic elements, and even the most transparently phono-graphic have some non-phonetic symbols, such as 2 or $. (Exceptions are linguistically-designed phonetic alphabets, discussed below.) English is more toward the phono-graphic end of the spectrum, where most words can be "sounded out" based on spelling, but there are many cases where the pronunciation is unpredictable and thus has to be memorized, and many cases where

the reader has to know what morpheme is represented by the spelled word in order to know how to pronounce it.

The earliest orthographies (Egyptian and Sumerian, about 3000 BCE) began as logo-graphic. Essentially, writing began as a series of pictures, a drawing representing each morpheme. Through repeated use, the drawings became more and more simplified and stylized, until the symbols became shapes only vaguely reminiscent of the meaning of the morpheme (a circle for "mouth" or a wavy line for "water"). Pictograms, however, are cumbersome. A language has tens of thousands of words, and it is difficult to draw pictures of every morpheme, particularly very specific items (such as "donkey" vs "horse") or anything abstract ("yesterday," "loyalty"). Pictograms undergo extension, both semantic extension (such as using the pictogram for "foot" to also mean "walk" or "go") and phonetic extension (as in a rebus puzzle, where ◉ stands for "I"). If pictograms develop into a completely stylized and abstract system, the result is morpho-graphic writing, as in Chinese characters. If the phonetic "rebus" use takes over, the result is a syllabary, where each symbol stands for a particular chunk of sound, not a particular meaning.

Of the different writing systems in use around the world, the majority are syllabaries. The name implies that in these systems there is a one-to-one correspondence between syllables and symbols, though there is some question as to whether the unit represented in these systems is in fact a syllable or a mora. According to Poser (1992; see also Sproat 2006; Buckley 2006), it is very rare for a "syllabary" to actually include a separate symbol for every possible syllable, which would number at least in the hundreds for most languages. Rather, a set of a few dozen "core" CV syllables are symbolized, with extra symbols used to extend the inventory, for example to indicate vowel length.

Regardless of whether "syllabaries" are based on syllables or moras, there is little question as to the utility of either the syllable or mora as a phonological unit (as evidenced by allophony, stress systems, poetry, reduplication, language games, the ability of naïve speakers to count syllables, patterns of articulatory coordination, etc.). If we follow the claims of Aronoff (1985) and Berent (2013) that orthographies reflect linguistic awareness, then the fact that syllabaries have been independently invented numerous times is evidence for the "core" CV syllable as a basic natural division in the stream of speech.

On the other hand, the alphabet as we know it was invented only once. According to scholars of the ancient Near East (beginning with Gardiner 1916), sometime around 1700 BCE Semitic speakers, the ancestors of the Phoenicians, borrowed Egyptian symbols to write their own language. In a crucial development, however, the symbols were used acro-phonically: the symbol stands not for the meaning nor for the syllable, but for the first sound in the word (a strategy that is also used by children's puzzle-makers today). Since all the words in this ancient variety of Phoenician began with consonants, only consonants could be written. This kind of writing system, where only the consonants are represented, is known as an "abjad." Modern Arabic and Hebrew orthographies are fundamentally abjads (though vowel diacritics can be added), direct descendants of this earliest Semitic writing system.

The abjad became a true alphabet when the Greeks borrowed writing from the Phoenicians, around 700 BCE. (A syllabic writing system, "Linear B," had been used on the Greek peninsula by the Mycenaeans centuries earlier, but was lost with the collapse of the Mycenaean civilization.) The Greek sound system differed from Phoenician in several important ways. First, Greek syllables, and thus words, can begin with vowels. In Greek, vowels and consonants are equally important in conveying lexical contrast, while in Semitic the consonants carry lexical meaning and the vowels indicate grammatical distinctions. For both these reasons, a consonant-only writing system would be problematic for Greek. Finally, Greek doesn't use the pharyngeal and laryngeal consonants that Semitic does. At some point, an ancient Greek writer may have made a conscious choice to use the otherwise un-needed pharyngeal symbols for the needed vowels, or it may have been a mistake. Sampson (2015: 101) imagines the scenario:

> A Greek sees a Phoenician using a mysterious system of written marks and asks for an explanation. The Phoenician . . . begins, "This mark is called ʔalp—no, not 'alp'—ʔalp, ʔalp, can't you hear, ʔʔʔalp!", while the bewildered Greek perceives only the [alp], and ends up calling the letter "alpʰ-a" and using it for /a/ since by the acrophonic principle that will now seem to him to be its proper value. This would explain the Greek use of Semitic <ʔ h ħ ʕ> as vowel letters, without the need to attribute any special linguistic sophistication to the first Greek user of the alphabet.

Every truly alphabetic system is descended from the Phoenician/Greek system. The Greek alphabet was borrowed by the Romans, and the Roman alphabet was then adapted, with varying degrees of success (as discussed below), for numerous languages around the world.

One might argue that the one-time development of alphabetic writing argues against the idea of the segment as a basic linguistic unit. If segments are in fact the default basic units of speech, we would expect many different alphabets to have been independently created. Alternatively, one could argue that the popularity of the alphabet, once invented, argues in favor of the segment as a basic unit. Most neutrally, one can assume that orthographies can represent different levels of phonological/morphological structure, which exist independent of writing. Syllabaries exist because syllables are basic units, and alphabets exist because segments are basic units.

3.3.2 ALPHABETS AND PRONUNCIATION

The preceding section has described alphabetic writing as "phono-graphic," which implies that each letter transparently represents a specific sound. That is not necessarily (or usually) the case, however, depending on the structure of the language and the depth of time over which the alphabet has been used. Despite its popularity, the Latin alphabet is not a very good fit for most of the languages of the world.

Because of the high regard for Latin among literate Europeans (who for most of the history of writing would have been a very small minority of the population),

there was a belief that the way Latin was written was "correct," and all other languages ought to correspond to it. English, for example, has thirteen or so vowels (depending on the dialect and how you count diphthongs and vowel-r combinations), written with only five symbols. We get around this by using digraphs "ea" "oi" "ai," doubled letters "ee" "oo," or other spelling cues (final "e" makes the vowel long), though these are far from perfect (e.g., "read" as [rid] or [rɛd], and the famous triplet "though," "through," "trough"). English readers rely a lot on memorization and context. English spelling was closer to pronunciation 500 years ago, but mismatches between sound systems and orthographies are a perennial cross-linguistic problem.

The "First Grammarian" was an anonymous scribe writing in Iceland about 1150 CE (so called because his was the first "Grammatical Treatise" of four that were appended to the Prose Edda (Haugen 1950)). The Grammarian praises the idea of each country writing its own history and laws in its own language, but complains at length about the inadequacy of the Roman alphabet to capture the more numerous contrasts of the Icelandic language (which he calls "Danish" but which we now call Old Norse). Through extensive demonstrations with minimal pairs, he argues that Icelandic contrasts thirty-six different vowels—nine different qualities, any one of which can be contrastively nasalized and/or contrastively lengthened—far beyond the power of "a, e, i, o, u" to capture. Almost 900 years later, Trudell and Adger (2015: 17) discuss similar problems in teaching Maasai children to read in the same Latin script that the First Grammarian complained about: "Maasai-language text is reported to be difficult to read, even for native speakers of the language. The principal problem seems to be that the orthography marks neither tone nor four of the nine Maasai vowels."

The only orthography in use today (other than those explicitly designed by linguists) that at least somewhat systematically represents units smaller than the segment is Korean Hangul. This writing system is said to have been created *c.*1500 CE by King Seychong, but was not widely used until the nineteenth century. Up until that time, the educated elite wrote Korean using Chinese characters. In Hangul, the shape of the character to some extent corresponds to the articulatory characteristics of the sound represented.

Some of the Hangul consonants are shown in Figure 3.3, with their IPA equivalents. The labials are based on a square shape that (plausibly) indicates the lips, while the dentals and velars are based on a wedge that could correspond to raising the front and back of the tongue respectively. The most basic symbol is the nasal (though /ŋ/ is an exception, probably because it never occurs syllable-initially). The plain (voiceless lenis) stop then adds a horizontal stroke, the aspirated stop two horizontal strokes. The fortis consonants are indicated by doubling the symbol for the lenis, straightforwardly indicating their extra length. Other consonant and vowel symbols are similarly systematic. However, as with English spelling, the phoneme/grapheme correspondence has grown opaque over time, so Hangul cannot be read as simple phonetic symbols today.

In summary, then, cross-linguistically, orthographies can represent different levels of language: the morpheme, the syllable, the segment, or in the case of Hangul even

ㅁ	ㅂ	ㅍ	ㅃ
[m]	[p]	[pʰ]	[pˈ]
ㄴ	ㄷ	ㅌ	ㄸ
[n]	[t]	[tʰ]	[tˈ]
○	ㄱ	ㅋ	ㄲ
[ŋ]	[k]	[kʰ]	[kˈ]

Figure 3.3   Some consonants in Hangul (Korean)

a subsegmental articulatory configuration. Orthographies thus can provide evidence that each of these levels is a valid and useful way of dividing the speech stream into discrete re-combinable categories. Yet even the most phono-graphic writing systems are not perfectly so. Because written marks are relatively permanent and because they are strongly influenced by conservative social pressures (like respect for Latin), orthographic conventions change much more slowly than spoken language. Conventionalized (even fossilized) writing systems can serve a purpose of communication across time and across diverse speech communities. But for accurately representing the sounds of actual utterances, a different system, one of explicit phonetic transcription, is needed.

3.4 PHONETIC TRANSCRIPTION

Accurate phonetic transcription is an indispensable tool for recording, describing, and analyzing sound systems. Yet even in a "purely phonetic" transcription, the questions of what the units of representation should be, and what level of detail should be represented, must be addressed. As was noted above and as discussed by Ladd (2014), in the Western tradition both phonologists and phoneticians have assumed that transcription should consist of strings of segments. But segments can be described with more or less detail, and the acoustic and articulatory similarities between segments can be emphasized or ignored. Segments can be represented as simple units (as is for the most part the case with the International Phonetic Alphabet) or as composites of more basic constituents, as was assumed by Alexander Melville Bell's "Visible Speech" (Bell 1867).

Visible Speech was the first orthography that attempted an explicit and universally-applicable phonetic transcription. The work was subtitled *The science of universal alphabetics: Or self-interpreting physiological letters, for the writing of all languages in one alphabet.* Figure 3.4 shows the alphabet as Bell presented
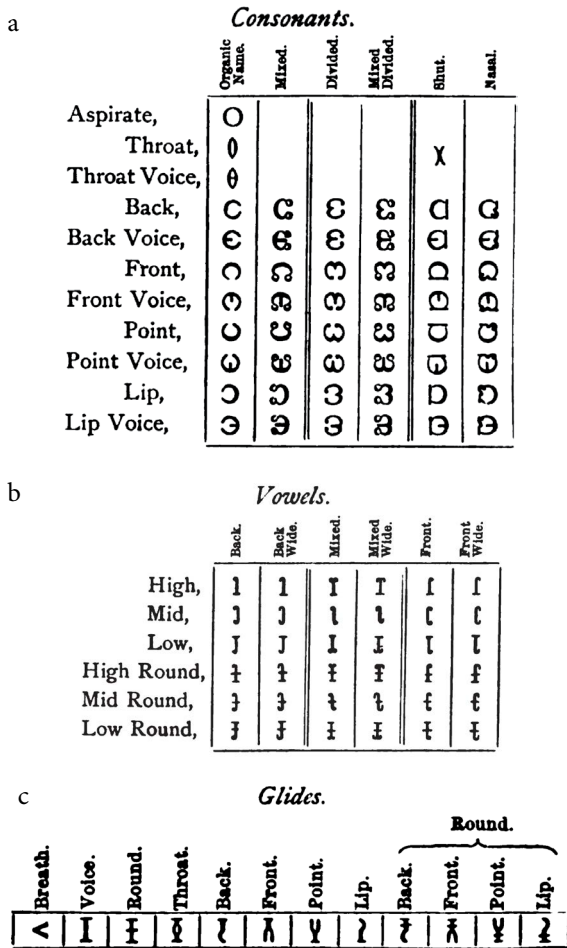
a

Consonants.

	Organic Name.	Mixed.	Divided.	Mixed Divided.	Shut.	Nasal.
Aspirate,	O					
Throat,	Ი				Х	
Throat Voice,	θ					
Back,	C	Ⴚ	Ɛ	Ɛ	ᴅ	Ꮐ
Back Voice,	Ე	Ҕ	Ꮛ	Ꮛ	ᴓ	Ꮐ
Front,	ᗡ	ᘘ	ᙆ	ᙆ	ᗟ	Ꮔ
Front Voice,	ᗝ	ᙈ	ᙈ	ᙈ	ᗠ	Ꮐ
Point,	ᘂ	ᘒ	Ꮗ	Ꮗ	ᘁ	Ꮕ
Point Voice,	ᗯ	ᘓ	Ꮗ	Ꮗ	ᘂ	Ꮗ
Lip,	Ꮐ	Ꮐ	Ꮝ	Ꮝ	ᴆ	Ꮝ
Lip Voice,	Ꮐ	Ꮝ	Ꮝ	Ꮝ	Ꮐ	Ꮝ

b

Vowels.

	Back.	Back Wide.	Mixed.	Mixed Wide.	Front.	Front Wide.
High,	ı	ı	ɪ	ɪ	ſ	ſ
Mid,	]	]	ı	ı	ɾ	ɾ
Low,	]	]	ɪ	ɪ	ɾ	ɾ
High Round,	ɫ	ɫ	ɫ	ɫ	ɟ	ɟ
Mid Round,	}	}	ł	ł	ɟ	ɟ
Low Round,	ɟ	ɟ	ɫ	ɫ	ɫ	ɫ

c

Glides.

									Round.			
Breath.	Voice.	Round.	Throat.	Back.	Front.	Point.	Lip.	Back.	Front.	Point.	Lip.
<	I	ƚ	ᘔ	ᘖ	ᗑ	Ɏ	ᘓ	₹	ᗑ	Ɏ	ᘓ

Figure 3.4   Alexander Melville Bell's "Visible Speech"
Source: Bell (1867: 37).

it; Figure 3.5 shows modern equivalents for the consonants in the International Phonetic Alphabet.

The letters are "self-interpreting" because each symbol contains explicit instruction for making the sound. As can be seen in Figures 3.4 and 3.5, the basic symbol for most consonants is a rounded horseshoe, with the direction of the opening indicating the place of articulation. Open to the right equals "Back" (a velar), open downward equals "Front" (palatal), upward "Point" (tongue tip), and to the left "Lip" (labial). Without any additional embellishment or diacritic, the horseshoe symbol stands for a voiceless continuant. A line inside the ring adds voicing. A line across the opening indicates a stop, and a curly line, reminiscent of a tilde, indicates a nasal stop. The "mixed" or "modified" consonants allow for intermediate places of articulation such

	Primary	Mixed	Divided	Mixed Divided	Shut	Nasal
Lip	ꓛ	ꓛ	3	ꓮ	ꓷ	ꓷ
	φ	ʍ	f	fˠ	p	m̥
Lip voice	ꓱ	ꓱ	3	ꓮ	ꓷ	ꓷ
	β	w	v	vˠ	b	m
Point	ꓴ	ꓴ	ꙍ	ꙍ	ꙮ	ꙮ
	r̥	s	l̥	θ	t	n̥
Point voice	ꙍ	ꙍ	ꙍ	ꙍ	ꙮ	ꙮ
	r	z	l	ð	d	n
Front	ꓵ	ꙮ	m	ꙮ	ꓰ	ꙮ
	ɕ	ʃ	ʎ̥	ɬ	c	ɲ̥
Front voice	ꓵ	ꙮ	m	ꙮ	ꓰ	ꙮ
	ʝ	ʒ	ʎ	ʓ	ɟ	ɲ
Back	ꓚ	ꓚ	ꓰ	ꓰ	ꓯ	ꓯ
	x	xʷ	ʟ̥	jʷ	k	ŋ̊
Back voice	ꓰ	ꓰ	ꓰ	ꓰ	ꓯ	ꓯ
	ɣ	ɣʷ	ʟ	ʟʷ	g	ŋ
Throat	ꓳ				ꓵ	
	breathy				ʔ	
Throat voice	ꓸ					
	hoarse					
Aspirate	ꓳ					
	h					

Figure 3.5   "Visible Speech" consonants, with equivalents in the International Phonetic Alphabet

Source: Adapted from omniglot.com/writing/visiblespeech.htm.

as alveopalatal and dental, and the "divided" consonants indicate a "side channel," thus corresponding to laterals for the tongue consonants and labio-dentals for the lip consonants. (Note the velar laterals. Bell confesses, p. 49, that this "back divided" consonant is "perhaps the most difficult of all articulations to unpracticed organs.") For the vowels, Bell's system indicates three degrees of backness—Front, Back, and Mixed (or central)—and six degrees of height—High, Mid, and Low, each of which

could further modified as Wide or not. Each of these eighteen vowels could then be either round or unround. Glottal stop (which Bell described as a "cough"), [h], and glides each have their own separate symbols. It is particularly impressive that Bell apparently worked out his system primarily through observation and introspection, without any previous systematic description or phonetic alphabet to rely on.

Bell's goals for his alphabet were very practical—he was primarily interested in teaching Deaf people to speak, and his idea was that reading aloud would be easier if each letter explicitly conveyed the articulatory configuration necessary to make the sound. Bell also predicted that, if popularized, his alphabet would make foreign language pronunciation simple, and would "convert the unlettered millions in all countries into readers" (1867: ix). (He wrote a very indignant preface to his book, criticizing the British government for failing to take him up on his offer of making his system freely available to all if only the government would bear the costs of printing.)

Despite these lofty goals, "Visible Speech" never caught on as a practical orthography (though if you're paying attention you can see it featured in several scenes of Henry Higgins's laboratory in the movie *My Fair Lady*). The system was too cumbersome to print, too different from the alphabets Europeans were used to, and too difficult to learn. It also turned out that Deaf people did not find it helpful. In fact, while Bell was praised during his lifetime for his work, the Deaf community today does not hold the Bell family in high regard (to put it mildly), because the insistence of both father and son on exclusively oral communication was not only unsuccessful, it also prevented Deaf students from becoming proficient in manual sign language. (Signed languages are further discussed in Section 3.7 below.)

Bell's work resulted in practical failure and social harm. What good came of it? Several innovations pioneered by Bell became important to phonetic science, which (as was noted in Chapter 2) was developing rapidly at the end of the nineteenth century. The first was simply the demonstration that a "universal alphabet" was an achievable goal. Henry Sweet, who published *Handbook of Phonetics* in 1877, using a phonetic alphabet of his own, which he called "Romic," wrote (p. viii) that Bell's system "is the first which gives a really adequate and comprehensive view of the whole field of possible sounds . . . applicable to all languages."

Second, as Halle (2009) emphasizes, Bell's system was the first (European) script that expressed the idea that "speech sounds are composite entities." Each sound, and thus each symbol, is a combination of place, manner, voicing, rounding, and nasality.

Third, Bell's work emphasized the need to abstract away from the fine-grained details of shades of articulatory difference and concentrate on differences in sound that produce differences in meaning. As was noted in Chapter 2, in earlier attempts at a phonetic alphabet, focusing on English, Bell was frustrated by the endless variety of possible vowel sounds:

the plasticity of the organs is so great, that shades of vowel quality are endless, arising from infinitesimal differences in the relative positions of the lips and the tongue. The number of possible varieties can as little be estimated as the number of possible shades of colour. (1867: 15)

However, by concentrating on systematic articulatory differences between contrastive sounds, "the expectation of ultimate success in the construction of a complete Physiological Alphabet, on the principle of Elementary Relations, was now, however, fully entertained" (p. 15). A few decades later a more successful universal alphabet, incorporating the idea of describing sounds in terms of "elementary relations," but based on the familiar letters of the Roman alphabet, was developed by the International Phonetic Association.

As noted above, phonetics was flourishing at the end of the nineteenth century. Phoneticians such as Henry Sweet (1845–1912) and his student Daniel Jones (1881–1967) at University College London sought to put the study of speech on a firmly scientific footing. Their laboratories made full use of new inventions that permitted sounds and articulations to be imaged and permanently recorded. In addition to this increasing technical innovation, there was increasing contact between Europeans and the languages of Africa and Asia, through trade and conquest, but also through missionaries who were hoping to create orthographies and Bible translations. Sweet (1877: Preface page b) records that now that linguists are becoming interested in describing "savage languages" that need to be written down for the first time, they need to be interested in phonetics. Increasing literacy and education in Europe led to a greater need to teach reading, writing, and foreign languages at home as well.

It has to be said at this point that the intellectual and social arrogance of Bell, Sweet, and their contemporaries leaps off the page. They see themselves as "men of science" using their knowledge to save the ignorant, whether the ignorant like it or not, and it falls to us to weigh that attitude and the harm that it did alongside their scientific achievements and the good that resulted. The effects of this Eurocentrism in linguistics have been long-lasting: see the discussions in, for example, Gilmour (2006), Errington (2007), and Zimmerman and Kellermeier-Rehbein (2015). One example is found in the International Phonetic Alphabet, which, while aspiring to be universal, is in its core largely based on Germanic and Romance languages, especially French.

The International Phonetic Association first met in 1886, and first published their International Phonetic Alphabet in 1888. "The principles of the International Phonetic Association," a set of numbered statements that described the alphabet and how it was intended to be used, was published by the association in 1912. Note that in the published principles there is much discussion of the usefulness of phonetic writing, and which symbols should be used for which sounds, but it is taken for granted that an alphabetic system should be the basis of phonetic transcription.

Like Bell's "Visible Speech," the IPA was meant to be practical, a tool for language teaching. The association's first president was Paul Passy (1859–1940), who was both a phonetician and a French language teacher. The association recognized that for "scientific" description a very detailed system of "narrow transcription" was needed, but that such details were not needed for other practical tasks. For most purposes, including language teaching and creating new orthographies, a "broad transcription," including only enough detail as necessary to distinguish words, was sufficient.

Thus, contra the ideas of Saussure on the need for phoneticians to completely avoid any reference to meaning, actual phoneticians recognized that writing down every shade of detail was impossible, and the emphasis had to be on representing dif-

ferences in sound that created differences in meaning. As stated in the principles of the International Phonetic Association:

> Principle 65: "The general rule for strictly practical phonetic transcription is therefore to leave out everything that is self-evident, and everything that can be explained once for all. In transcribing any given language, it is in general sufficient to represent the distinctive sounds only; for each distinctive sound the typical international [nb: = European] symbol should be chosen; and if necessary, the exact shades of sound used either throughout or in certain positions may be explained (with the use of modifiers) in an introductory note." (1912: 15)

> Principle 70: distinctive sounds are "those which if confused might conceivably alter the meanings of words." There is no need to symbolize "shades of sound which are occasioned by proximity to other sounds, absence of stress, and the like." (1912: 16)

As Anderson (1985a) notes, the more phonetic science improved during this time, the clearer it became that most of the detail that phoneticians were able to measure was not relevant to phonology as they understood it. The goal of "practical" transcription is to abstract away from coarticulation and all predictable aspects of speech. However, Anderson also notes that the IPA was designed to accommodate "scientific" description as well: the systematic differences between languages that are neither contrastive within the language nor predictable across languages. For example, because final stops are always released in Georgian, sometimes released in English, and never released in Korean, for the purpose of language description the IPA needs a way to indicate stop release, even if it is never contrastive. The goal, if all systematic aspects of pronunciation are indicated, is to create a "language-neutral phonetic transcription" (Anderson 1974: 8).

Unlike "Visible Speech," the IPA used, insofar as possible, the familiar letters of the Roman alphabet, forgoing Bell's principle of making every symbol "self-interpreting" in favor of simplicity, a strategy that has been successful. But note that there are still some "featural" aspects, where articulation is indicated by an aspect of the symbol, such as hooks (originally under-dots) for retroflex consonants, or the convention of using small caps for uvulars. This is not consistent however: there is no consistent difference between the symbols for voiced and voiceless consonants, for example, or any common visual design element shared by all the labials. Nonetheless, the IPA follows Bell in presenting consonants as organized into a table by place, manner, and voicing, and in so doing, recognizing that speech sounds are "composite entities."

While the IPA has been extremely successful and widely adopted, the goal of creating a phonetic transcription that indicated all non-predictable aspects of pronunciation was never reached. There is no "language-neutral" way of representing all the systematic aspects of pronunciation, since there is no language-neutral way of deciding a priori what is systematic. Even the tiniest details of quality or quantity may be systematic and language-particular, and most transcription is done in the
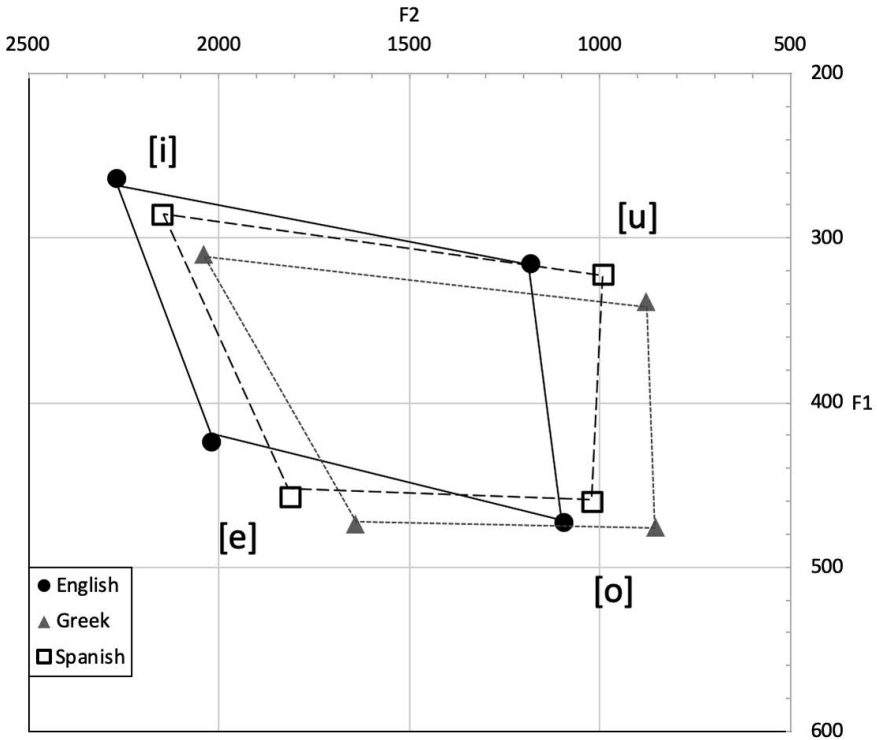
Figure 3.6   The vowel spaces of Greek, Spanish, and English are not identical
Source: Adapted from Bradlow (1993: 57).

absence of phonetic analyses that would reveal these details. (See further discussion of language-specific phonetics in Chapter 5.) In using the IPA, the analyst always had the choice of how much detail to represent, and of how exactly to divide the phonetic space (especially for vowels), and such choices were not consistently made or explained. For example, the actual quality of the front mid vowel of a five vowel-system is quite variable cross-linguistically, and thus the symbol [e] does *not* stand for the exact same sound in different languages. Figure 3.6, adapted from Bradlow (1993: 57), shows that the typical quality of the vowels transcribed /i/, /e/, /o/ and /u/ differ systematically between Greek, Spanish, and English.

To take another example, there is a longstanding problem of deciding whether languages with four vowel heights should be transcribed as [i, ɪ, e, a] or [i, e, ɛ, a]. The transcription of the non-peripheral vowels has to be based on phonological pattern-ing, such as participation in ATR harmony, not phonetic quality (Casali 2008; Rose 2018).

Further, a reader attempting to pronounce an IPA transcription such as [bed] would probably have to guess as to whether there was any dipthongization or not: was [e] chosen over [eɪ] because slight dipthongization is a predictable detail that could be left off, or because dipthongization is absent? How much dipthongization

is enough to be worthy of transcribing? The same is true of a consonant symbol such as [b]. Laryngeal configurations and amount of vocal fold vibration for "voiced" consonants differ widely across languages (see e.g., Kingston and Diehl 1994). If there are two stop consonants in the language, one with some prevoicing and the other with some aspiration, should the consonants be transcribed as [b] and [p] or as [p] and [pʰ]? Honeybone (2005) notes that different analysts of the same language often differ in the choice they make. Because such questions are in general answered according to the particular goals and choices of the transcriber, Pierrehumbert (1990b: 388) found it necessary to note of "fine phonetic transcription" that "the representation it is claimed to provide is not a coherent one." Pierrehumbert et al. (2000: 286) further assert that because "it is impossible to equate phonological inventories across languages," an IPA transcription is therefore not "a technically valid level of representation in a scientific model" but rather "a useful method of note-taking and indexing." The same point is made by Lindau and Ladefoged (1986), Bradlow (1993), Cohn (2011), and Ladd (2014), among others.

Port and Leary (2005: 927) take this line of reasoning a step further, arguing that because the IPA fails in the task of being a systematic, language-neutral, phonetic description, and because no other segment-based transcription system is any better, or is likely to be any better, that therefore discrete, symbolic phonology has no empirical basis. They argue that because "languages differ in far more ways than any learnable alphabet could represent," then "an alphabet cannot capture most of the important structure in speech," although it may be "a very useful technology for reading a language one knows well."

> Much of the evidence about the inadequacy of traditional transcription has been available for a long time, but most of us have ignored it or made excuses for it. Since all formal models of language are built on a foundation of discrete, a priori phonetic symbols, the very idea of a formal model of language is rendered impossible, unless someone figures out a way to provide a genuinely discrete universal phonetics. But there is little hope for that. (2005: 927)

The problem with Port and Leary's critique, however, lies in equating "discrete phonetic symbol" with "IPA symbol." As Pierrehumbert et al. (2000) note, if the IPA is taken to be simply a very useful tool for taking notes on the important of aspects of pronunciation, it works, but it is not in itself a phonological theory. The "atoms" or units of phonological organization are not to be found in the symbols of an orthography, no matter how phonetically clear. That is, the IPA symbol [e] should not be taken as referring to a particular set quality, but as a "cover symbol" or short-hand representation for the unit "mid front vowel," however that might be instantiated in a particular system.

As the problem was posed at the beginning of this chapter, does the usefulness of segmental transcription prove that the segment is the basic unit of Language, or do the problems with the IPA prove that it is not? On balance, the success of the IPA and of segment-based phonetic and phonological analyses show that the segment is indeed a *useful* level of analysis. Whether or not it is the ultimately *correct* analysis

continues to be debated (see Chapters 9 and 11). Even if the segment is accepted as a unit, analyses at levels both suprasegmental (Chapters 7 and 8), and subsegmental (Chapter 4) also provide important insights. Crucially, when considering how segments are organized into systems of contrast, the nature of the "elementary relations" (Bell 1867: 15) between them must be taken into account. It is to these elementary relations that the next sections now turn.

3.5 SELECTING THE INVENTORY

The International Phonetic Alphabet provides symbols for seventy-nine consonants and twenty-nine vowels, not counting diacritics. How do languages choose which to use in creating inventories? According to the database compiled by Maddieson (2013), the smallest consonant inventory is six (Rotokas, Papua New Guinea) and the largest is 122 (!Xóõ, Botswana). The largest inventories include many secondary and complex articulations, in addition to contrasts in voice quality and airstream mechanism, thus many "segments" use a combination of IPA symbols. In fact, for such complex articulations, it can be difficult to tell whether a combination of constrictions should count as one segment or two, perhaps further evidence against the segment as a basic unit. Most languages, however, select a much simpler inventory. According to Maddieson (2013), the median size for a consonant inventory is twenty-one. English has (about) twenty-four, depending on dialect and how you count. The range for contrastive vowel qualities (not counting tone, voice quality, or diphthongs) is from two (Yimas, Papua New Guinea) to fourteen (German, Western Europe). English is on the high end for vowels, with thirteen (British English). The median number of contrastive vowel qualities is five, and according to Maddieson, fully one-third of the languages in the database have five vowels.

While both very large and very small inventories exist, Maddieson (2013) shows that inventory size is close to normally distributed—most languages are average. Communicative needs keep inventories from getting too small—there have to be enough segment combinations to keep tens of thousands of vocabulary items distinct. (Languages with small segment inventories tend to have long words.) At the other end of the spectrum, acoustic and articulatory pressures keep inventories from becoming too big. It has to be possible, in environments sometimes not perfectly conducive to speaking and hearing, to keep too many words from becoming too confusable.

Within an inventory, the set of sounds is not chosen randomly. It is not accidental that the inventory of Rotokas is [p, t, k, b, d, g] and [i, e, a, o, u], not [b, l, q, h, x, ŋ] and [y, ø, a, ʌ, ɨ]. For a given inventory size, how are contrastive segments chosen from among the indefinitely large number of ways that sounds can vary? How do inventories emerge? Linguists have long argued, beginning with Passy (1890: 227) as well as Martinet (1955) and Jakobson (1968), that if phonology is about contrast, then systems should be organized so that contrasts are *maximized*.

"Dispersion Theory" (Liljencrants and Lindblom 1972; Lindblom 1990; Flemming 1996, 2004) quantifies the idea of "maximizing contrast." According to Dispersion Theory, the members of an inventory are selected so that each segment is as acoustically distinct from each other segment as possible, given the number of segments.
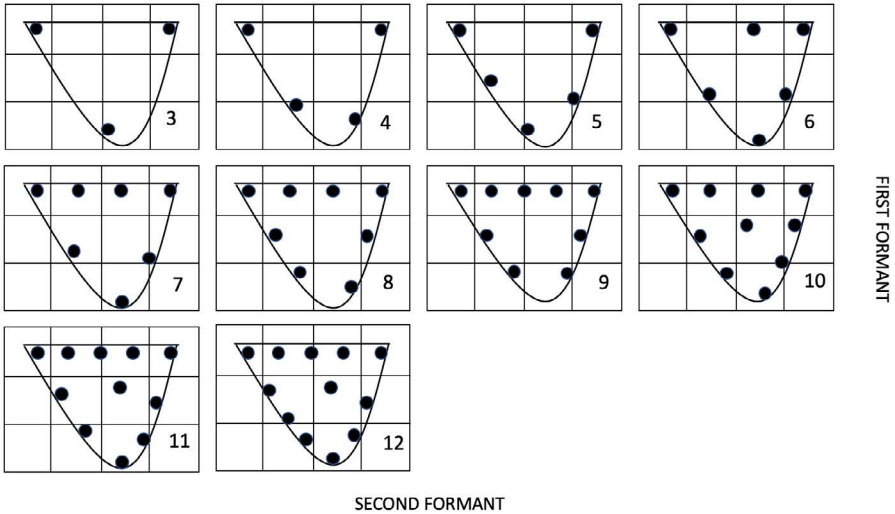
FIRST FORMANT

SECOND FORMANT

Figure 3.7   Dispersion Theory
Source: Adapted from Liljencrants and Lindblom (1972: 843, Figure 2).

Liljencrants and Lindblom suggest an analogy to magnets attached to corks floating in a tub of water: the magnets will repel each other, resulting in a stable configuration in which the distance between the magnets is as large as possible within the confines of the tub. The positions of the corks will depend on their number: the more there are, the closer together they will be forced.

For vowel systems, Liljencrants and Lindblom propose a formula to maximize the Euclidean distance between points in the F1/F2 space (in mel units), within the limits of possible human articulation/perception, based on the number of points (vowels) in the system. The results of their simulations do not do badly, as shown in Figure 3.7, adapted from Liljencrants and Lindblom's Figure 2 (1972: 843). (The graphs have been rotated from the original to match the orientation of a familiar vowel chart, where /i/ is in the top left corner and /a/ at the bottom.) For a three-vowel system, the results of the simulation produce formant values close to those of /i, a, u/. For a five-vowel system, the simulation produces values close to /i, e, a, o, u/. As the number of vowels increases, however, the predictions become less accurate (see Disner 1984 and Vaux and Samuels 2015 for further discussion). One of the biggest problems is that the model predicts that the high central space is utilized more extensively than is actually the case in attested languages. As seen in Figure 3.7, for example, the theory predicts that a nine-vowel system will have five degrees of backness, but only three vowel heights. Instead, actual nine-vowel systems (such as Akan or Maasai) use only front vs back, and five distinct heights: [i/u, ɪ/ʊ, e/o, ɛ/ɔ, a]. Languages also use the center of the acoustic space (schwa-like vowels) more often than Liljencrants and Lindblom predict. More recent versions of Dispersion Theory (Flemming 1996, 2004) refine the theory by incorporating both acoustic and articulatory information into the predictions. (Flemming's work is described in more detail in Chapter 6.)
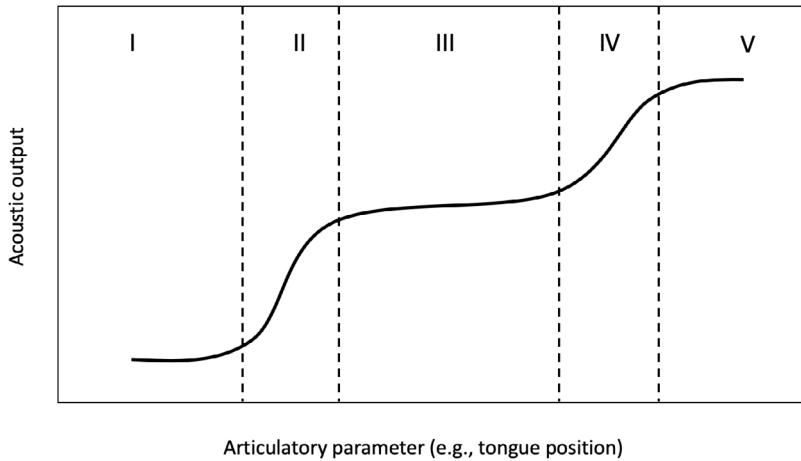
Figure 3.8   *Quantal regions in the relations between acoustic and*
articulatory parameters

Source: Adapted from Stevens (1989: 4, Figure 1).

Stevens's (1989) Quantal Theory was one of the first to propose a quantitative model of the interaction of acoustics and articulation in predicting inventories. Using a schematic model of the vocal tract and calculating predicted resonance frequencies for different configurations, Stevens maps the acoustic output for a range of vocal tract positions. He finds that there are certain regions where a slight change in articulation results in large changes in acoustic output, but other regions where even fairly large changes in articulation result in little change in acoustic output. Stevens argues that languages organize their inventories using these regions of stability.

The idea is graphed in Figure 3.8, a schematic (adapted from Stevens 1989, Figure 1) that could be applied to the relation between any articulatory parameter and acoustic output, such as tongue position and F2, or constriction degree and fricative noise. To take the case of a velar fricative, there would be a large region (I in the diagram) where the tongue approaches the velum, but no fricative noise is produced. There would then be a sudden transition (region II) to fricative noise production, which would in turn not change much as the constriction degree became somewhat tighter (region III), until there was another sudden change (IV) to a region of closure/compression (V).

For vowels, Stevens models the vocal tract as a series of tubes with different resonance frequencies (similar to organ pipes). Changes in tongue position create different tube lengths with higher or lower frequencies. In configurations where the resonance frequencies of two tubes approach one another, the two systems interact in such a way as to produce regions of stability: because of the acoustic linkage between the two tubes, small tongue movements have less of an effect on the overall system (see Stevens 1989 and 2000 for details and formulas). Languages organize their vowel systems, Stevens argues, around such areas of stability, avoiding other areas where

Table 3.1 Relative difficulty of producing a voiced stop based on place of articulation and environment. Values in shaded cells are > 25

Environment	Place of articulation		
	b	d	g
After an obstruent	43	50	52
Word-initial	23	27	35
After a non-nasal sonorant	10	20	30
After a nasal	0	0	0

Source: Adapted from Hayes (1999: 251).

slight changes in tongue position have large acoustic effects. This point that "phonemes are located in well-behaved regions of the articulatory-to-acoustics mapping" is also emphasized by Pierrehumbert (2000: 16). It turns out that the palatal region is an area of instability, where small movements produce large changes. Thus, a great deal of articulatory precision would be required to maintain multiple vowel contrasts in the palatal region, explaining why languages don't use such systems, and why a purely acoustic Dispersion Theory makes the wrong predictions. (Chapter 10 returns to the discussion of the role of perceptual distinctiveness in predicting not only inventories but alternations.)

Pierrehumbert (2000) also notes that overall characteristics of the human vocal tract, such as the fact that the fundamental frequency of vocal fold vibration is lower than the tube resonances of the oral cavity, allow vowel inventories to encode linguistic information. But biology is not enough. Inventories arise, Pierrehumbert argues, from multiple interacting constraints, some of which are physical and biological, and some of which are cognitive.

Implicit in the preceding discussion are two ideas. First is the idea that languages create contrastive units (segments) from acoustic continua. The second is the idea that languages are not choosing segments per se (in which case they might select any random set), but choosing and reusing intersections of phonetic parameters, such as F1 and F2, or place of articulation, nasality, and voicing. Thus, very generally, if a language chooses [p, t, k] for voiceless stops, it will reuse those places with voiced stops [b, d, g] and nasals [m, n, ŋ]. A language will choose three symmetrical series of three, rather than nine randomly related segments, even if those nine are easy to both produce and perceive.

This point was made by Hayes (1999) in his proposal for "Phonetically-driven phonology." (See also the further discussion in Chapter 6.) Hayes argues that systems of contrast emerge through a combination of "phonetic sensibleness" and "good design" (1999: 244). An example is the "difficulty map" in Table 3.1.

The "map" shows the relative difficulty (based on an aerodynamic model of the vocal tract) of producing a voiced stop based on place of articulation and environment, where a higher number means more difficult. In any given environment, a voiced bilabial stop is easier than a voiced coronal which is easier than a voiced velar, because

airflow is easier to maintain when there is a larger space between larynx and closure. For the effect of environment, post-nasal position favors voicing (0 difficulty), and then difficulty increases following other sonorants, in word-initial position, and then finally in post-obstruent position, the most difficult environment in which to induce voicing.

Hayes' point is that if languages organized their phonology based solely on articulatory difficulty, you would expect to find phonologies that allowed any voiced stop with degree of difficulty less than twenty-five: that is, [b, d, g] after nasals, [b, d] after non-nasal sonorants, only [b] in word-initial position, and no voiced stops at all after other obstruents. Instead, languages choose the easier over the harder, but in symmetrical ways. To bring in examples from other languages, Dutch (Gussenhoven 1992) bans [g] and allows [b, d] across environments, even though post-obstruent [b] (forty-three) is more difficult than intervocalic [g] (thirty) would be. In Sakha (formerly known as Yakut; Krueger 1962), on the other hand, all stops are voiced in intersonorant position and voiceless in initial position, even though initial [b] (twenty-three) would be easier than intervocalic [g] (thirty). Hayes concludes that "phonological constraints tend to ban phonetic difficulty in simple, formally symmetrical ways" (1999: 252). Phonetic difficulty matters, but is mediated by a formal phonology.

Further, Hayes argues that "The 'boundary lines' that divide the prohibited cases from the legal ones are characteristically statable in rather simple terms, with a small logical conjunction of feature predicates" (1999: 252). That is, possible inventories and possible sequences, on both phonetic and phonological grounds, are not defined by reference to segments per se. A segment, represented by an IPA symbol, is not an atomic whole, but can be understood as a combination of distinctive features that delimit its function within the linguistic system.

3.6 PHONETIC PARAMETERS AND PHONOLOGICAL FEATURES

Are distinctive features, then, rather than segments, the true building blocks of language? Linguists have been classifying segments according to their articulatory characteristics since antiquity (see further discussion in Chapter 4), and in the twentieth century, with the emphasis of phonological theory on systems of contrasts, characterizing the phonetic parameters that differentiate segments took on even greater importance. Saussure wrote that "[t]he important thing in the word is not the sound alone, but the phonic differences that make it possible to distinguish this word from all others, for differences carry signification" (1959: 118). Trubetzkoy took the emphasis on contrast a step further in arguing that an inventory is not a list of sounds, but a list of distinctive parameters. "Phonemes should not be considered as building blocks out of which individual words are assembled," he writes (1969: 35). "The phonemic inventory of a language is actually only a corollary of the system of distinctive oppositions. It should always be remembered that in phonology the major role is played, not by the phonemes, but by the distinctive oppositions" (1969: 67). The inventory, then, is a by-product. The language doesn't choose a list of sounds, but a list of distinctive dimensions, and the intersections of dimensions result in particular sounds.

It remains a question, however, whether features are actual constituents of segments (as were the components of Bell's "Visible Speech") or labels that describe and

categorize segments (much as place and manner describe and categorize the set of IPA symbols). As Ladd (2014: 2) put the question, are features "particles or attributes"?

Trubetzkoy saw the phoneme as an abstraction that characterizes all and only the features that all of its allophones share. He writes, "The phoneme is not identical with an actual sound, but only with its phonologically relevant properties. One can say that the phoneme is the sum of the phonologically relevant properties of a sound" (1969: 36). At the allophonic level, features characterize a segment. At the phonemic level, they constitute the contrastive unit.

Later phonologists (including Jakobson et al. 1952 and Chomsky and Halle 1968) would argue explicitly that phonetic segments, not just abstract phonemes, are literally nothing more than "bundles" of features. Ladd quotes phonology textbooks by Harms (1968) and Hyman (1975) as explicitly stating that "the fundamental unit of generative phonology is the distinctive feature" (Harms 1968: 1) and "symbols such as *p, t, k, i, a, u* are used as convenient shortcuts for the feature compositions which combine to produce these segments" (Hyman 1975: 24ff.). In a more recent textbook, Zsiga (2013a: 257) states "Sentences are made up of words, words are made up of segments, and segments are made up of features." If segmental representations, including IPA symbols, are just a convenient notation for sets of features, which are the "real" constituents, that removes the problem of the failure to define IPA transcription as "a technically valid level of representation in a scientific model" (Pierrehumbert, et al. 2000: 286). Instead, it puts the burden on feature theory to define the basic units.

Thus for many phonologists, features, not segments, are the basic building blocks of language. Ladd (2014) goes on to argue, however, that considering all features as "particles" autonomous from the segment turned out to be more confusing than helpful, and others such as Ohala (1990) have argued against the "reification" of features as entities rather than descriptors. The question of whether segments are basic units to which featural labels attach, composite entities composed of features, or just useful fictions to which linguists trained to read the alphabet are predisposed remains a matter of debate. Decades of phonological and phonetic research have shown that the feature, the segment, and larger units such as syllables are all useful ways of breaking down continuous speech into component parts. The question of which level is the most important or basic has not been fully answered.

Chapter 4 delves deeper into distinctive feature theory, considering in detail some of the sets of actual features that have been proposed in the last 100 years or so since Saussure and Trubetzkoy, concentrating on the ways that distinctive feature theory illuminates and illustrates the phonology/phonetics interface, including the question of basic units. However, before digging in to the specifics of feature theories for spoken languages, Section 3.7 turns to the question of the basic units in a different modality: manually signed languages.

3.7 THE UNITS OF SIGNED LANGUAGES

Spoken languages are not the only ones that have phonology and phonetics. The same questions of phonological contrast and phonetic implementation apply to signed languages as well. In fact, studying the phonology and phonetics of languages

that use a visual rather than an auditory modality provides a different and important perspective on the general nature of the relationship between the cognitive and physical aspects of language, and allows us to better investigate the ways that the physical modality of expression constrains the structure of language. Overviews of sign language phonology and its relation to spoken phonology include Sandler and Lillo-Martin (2006), Johnson and Liddell (2010), and Brentari (2010). Linguistic studies of signed languages are much more recent than linguistic studies of spoken languages; there is a smaller body of research and arguably less consensus on representation.

What is the "basic unit" of a signed language, the "building blocks" out of which utterances are built? The top-level answer is, of course, that signed languages consist of "signs": gestures of the hands, arms, and face that express a meaning. But just as words are not "vocal wholes" but can be broken down into successive syllables or segments, signs are not "gestural wholes" either. What are the distinctive parameters, to use Trubetzkoy's terminology, that make up the signs?

There are several issues that arise when trying to apply the principles that work for spoken languages to signed languages. First, there are many more degrees of freedom in the visual/spatial domain. Given these many degrees of freedom, there turn out to be few minimal pairs. Thus the top-down principle, of contrasting morphemes/signs to determine the smallest sequential unit, doesn't work well. Third, more "features" can be realized simultaneously in sound than in speech. Every meaningful sign by definition corresponds to a morpheme, and signs definitely combine into phrases within utterances, but it is not at all clear if there are sign correspondents to the spoken segment or syllable (see especially Battison 1980; Sandler and Lillo-Martin 2006). If there are not, we cannot argue that the segment and syllable are necessary and universal components of Language, just convenient ways of organizing *spoken* languages.

Linguists disagree on how signs should be decomposed into more basic units, and there is no generally-accepted transcription system for signed languages, analogous to the IPA for spoken languages. Publications often use pictures or diagrams to reference signs, if they reference the physical form of signs at all, without making any claim or assumption about sign components. The problem faced by all transcription systems for signed languages, as for spoken languages, is choosing what aspects to represent. The greater number of degrees of freedom in signed languages (the hand has many more possibly contrastive shapes than the tongue) makes the problem that much harder. (See Eccarius and Brentari 2008 and Hochgesang 2014 for further examples and discussion.)

Stokoe (1960) offered the first linguistic analysis of a signed language. He proposed that each sign could be specified in terms of three contrastive parameters: handshape, location, and movement. Obviously, each of these parameters had multiple possible values. Eccarius and Brentari (2008) propose a more differentiated system for transcribing handshapes (based on Brentari 1998), specifying the "base" shape of the hand, as well as joint configurations for selected fingers and thumb. The features are arranged in a hierarchical structure similar to the feature geometry proposed for spoken language segments (see the discussion in Chapter 4). Johnson and Liddell (2010) offer a different system (Sign Language Phonetic Annotation, or SLPA), in which handshapes are broken down into simultaneous feature bundles that specify

hand configuration, placement, and orientation. The authors explicitly compare these bundles to the segmental feature bundles of spoken language phonology. They argue that signed languages have two types of segments. "Postural" feature bundles are connected by "transforming" segments that specify path and contact, among other aspects of movement.

Both the Eccarius and Brentari system and the Johnson and Liddell system are cumbersome as methods of transcription, however. As Tkachman et al. (2016) point out, a single handshape requires twenty-four to thirty-four characters to transcribe in SLPA, and many feature combinations are either anatomically impossible or non-occurring. Tkachman et al. suggest a number of modifications to SLPA that would make it easier to use, especially for large-scale corpus studies, modifications that include reducing the degrees of freedom in hand configurations, and creating templates for feature combinations that often re-occur. While different systems are still in development, no general consensus has been reached on transcription of signed languages. There simply has not been enough research into the phonology and phonetics of different signed languages to work out a complete theory of what the "distinctive parameters" are, so representation systems err on the side of more detail rather than less.

At the phonology/phonetics interface, we find that there is the same pressure to adapt the message to the medium in signed and spoken languages. There is the pressure of ease of articulation, resulting in assimilation and reduction. There is the pressure of clarity of communication, which results, for example, in an increase in signing space when the interlocutor is more distant, and the familiar effect of less assimilation and reduction in a more formal context. There are physical constraints: no sign requires the ring finger to move independently, for example. There are "markedness" constraints: in signs that require two hands, for example, the non-dominant hand can either mimic the dominant hand, or take on one of a limited set of simplified default shapes, such as a flat palm or fist (Battison 1978). There are dialects of signed languages, and signer-specific individual variation (Crasborn 2012).

A difference is that "iconicity" seems to play a more important role in signed than in spoken languages. Onomatopoeia plays a real but marginal role in spoken language, such that Saussure's concept of the "arbitrariness" of the sound/meaning mapping has become a central tenet of phonology. There is definitely arbitrariness in sign as well, but many signs bear a physical resemblance to the things or concepts they signify. For example, Figure 3.9 (Klima and Belugi 1979; Hamilton 2018) shows the sign for "tree" in three different signed languages. American Sign Language and Danish Sign Language show the full shape of the tree in different ways, while Chinese Sign Language depicts the trunk.

Work continues on the role of iconicity in signed languages (e.g., Hamilton 2018; Becker 2018). We simply don't have the data yet to determine whether iconicity plays a similar role in spoken and signed languages, just to a greater extent in signed because the visual modality provides a greater opportunity, or whether the role of iconicity in the grammar of a signed language is somehow different and deeper.

So for the major questions of signed language phonology, the state of the art is that work continues, but much more data and analysis is needed before consensus can be reached. Given the overwhelming preponderance of data and discussion in the

(a) American Sign Language   (b) Danish Sign Language   (c) Chinese Sign Language

Figure 3.9   The sign for tree in American, Danish, and Chinese sign languages
Source: Klima and Bellugi (1979: 21).

literature, most of what is discussed in the following chapters of this book is based on spoken language research. Based on the spoken language data, this chapter has argued for the utility of representing the basic building blocks of language as morphemes made up of syllables, syllables made up of segments, and segments made up of features. But it is just not clear if this same breakdown applies to sign, and if it does, exactly how a syllable, segment, or feature in sign should be defined. It is always worth asking, as we go forward, how the questions raised might be applied to sign, and how research into signed languages could help us reach a more truly universal answer.

RECOMMENDED READING

Overviews

Raimy, E. and C. E. Cairns (2015b), "Introduction," in *The Segment in Phonetics and Phonology*. Malden, MA and Oxford: Wiley-Blackwell.

- What are the "contemporary issues concerning the segment" that Raimy and Cairns discuss?

Sampson, G. (2015), *Writing Systems*, 2nd edn. Sheffield: Equinox Publishing.

- According to Sampson, how do writing systems represent linguistic knowledge?

Sandler, W. and D. Lillo-Martin (2006), *Sign Language and Linguistic Universals*. Cambridge, UK and New York: Cambridge University Press.

- Why is studying sign language important for understanding linguistic universals?

Exemplary research

Something old

Liljencrants, J. and B. Lindblom (1972), "Numerical simulation of vowel quality systems: the role of perceptual contrast," *Language, 48*(4): 839–62.

- How do Liljencrants and Lindblom "simulate" perceptual contrast? What are some of the limits of their simulation?

Stevens, K. N. (1989), "On the quantal nature of speech," *Journal of Phonetics, 17*: 3–45.

- Stevens's simulations are a little harder going than those of Liljencrants and Lindblom, but are well worth the effort. How does "Quantal Theory" take advantage of the intersection articulation and acoustics? Make reference to Figure 3.8 in your answer.

Something new

Hayes, B. (1999), "Phonetically-driven phonology: the role of Optimality Theory and inductive grounding," in M. Darnell, E. Moravscik, M. Noonan, F. Newmeyer, and K. Wheatly (eds), *Functionalism and Formalism in Linguistics, Volume I: General Papers*. Amsterdam, The Netherlands: John Benjamins, pp. 243–85.

- According to Hayes, how do systems of contrast emerge through a combination of "phonetic sensibleness" and "good design"?

Hochgesang, J. A. (2014), "Using design principles to consider representation of the hand in some notation systems," *Sign Language Studies*, *14*(4): 488–542.

- What "design principles" does Hochgesang propose? How do they compare to those proposed by Hayes?

Opposing views

Nolan, F. (1992), "The descriptive role of segments: evidence from assimilation," in G. J. Docherty and D. R. Ladd (eds), *Papers in Laboratory Phonology II: Gesture, Segment, Prosody*. Cambridge, UK: Cambridge University Press, pp. 261–79.

- Read Nolan's paper, and then the commentaries by Hayes, Ohala, and Browman that follow. Which of the four accounts of assimilation do you find most convincing?

QUESTIONS FOR FURTHER DISCUSSION

1. Given that orthographic systems only imperfectly represent the structure of language, how do such systems provide evidence concerning units of phonological representation? Investigate a writing system you do not know. Is the system based on the morpheme, syllable, mora, or segment? What evidence did you use to make this determination?

2. Discuss: The IPA is a *useful* level of representation, even if it is not "a technically valid level of representation in a scientific model" (Pierrehumbert et al. 2000: 286). What do Pierrehumbert et al. mean by that phrase, and how can the IPA be useful anyway? What evidence does the IPA provide both for and against the segment as a basic unit of spoken language?

3. How would you redesign the IPA? How might you make it easier to read and remember? How could it be less Eurocentric? What would the IPA look like if it were designed on the principle of "Visible Speech" and Hangul?

4. Consider the vowel inventory of a language other than English. Would Dispersion Theory correctly predict this inventory? If not, would Quantal Theory help? Is the inventory symmetrical? If the answer to all three questions is "no," how do you think this inventory might have come to be?

5. Consider the consonant inventory of a language other than English. Is the inventory symmetrical? What dimensions of contrast are utilized? What tradeoffs do you see between ease of articulation and ease of perception?

6. What insight do signed languages give us about what the basic units of Language might be? Should the basic units of signed and spoken languages be the same? What factors make analyzing the units of signed languages difficult?

7. In what ways did linguists such as Bell and Sweet "see themselves as 'men of science' using their knowledge to save the ignorant, whether the ignorant like it or not"? If you're not sure, read Bell's preface to "Visible Speech." What effects of Eurocentrism in linguistics have you come across? What efforts can we make now to reduce it?