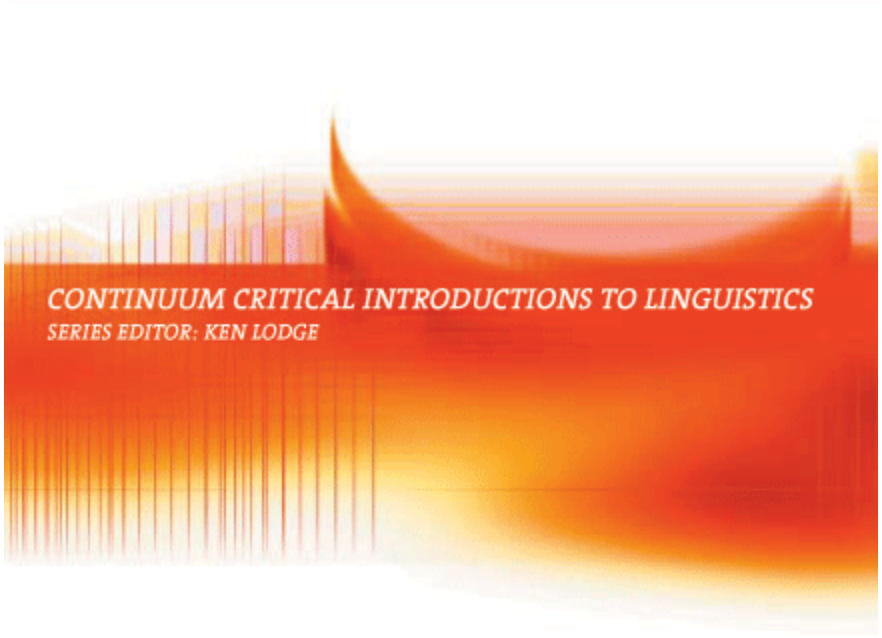


A Critical Introduction to Phonology

Of Sound, Mind, and Body

DANIEL SILVERMAN

ERRATA

"There is hope in honest error; none in the icy perfections of the mere stylist"
-Charles Rennie Macintosh

i, 29, 171, 216:	"parlo ergo es" → "**loquor ergo es**"
viii.14:	"a theory of" → "**an approach to**"
24.10:	"sound-meaning relationships→ "**sound and meaning**"
25.26:	"stimulus" → "**the stimulus**"
27.20:	"manual" → "**stick**"
29.35	"1997" → "**1986**"
44 Figure 2.10:	"300Hz → "**3000Hz**"
48.22	"lowering → "**raising**"
48.24	"raising → "**lowering**"
54 Figure 2.12:	"Back Front" → "**Front Back**"
75 Table 3.2:	"top" → "**tip**"
78.29:	"standard word" → "**standard southern word**"
78.29:	first instance of "[gɑt]" → "**[gɑl]**"
93 Table 4.1:	first column "L" → "**L#**"
165.29:	"voicless" → "**voiceless**"
167.14:	"two" → "**a number of**"
168 Table 6.1:	"tʰak" → "**tʰăk**"
178.11:	"voicelessness" → "**voicelessness**"
179 Figure 6.2.	top circle: "closer to devoiced" → "**closer to aspirated**"
184 Figures 6.4, 6.5:	Eliminate lower left diagonal ellipses.
214.26-27:	Eliminate first instance of "listeners perceive it"
218.24:	"vocal tract" → "**respiratory and ingestive systems**"
234.24:	frɜm → **frɜ̃m**
234.25:	ˈmɑɹ̆ → **ˈmɑ̃ɹ̃**
234.25:	ɔ̬~~ɬ~~inḏˢ → **ɔ̬̃~~ɬ~~inḏ̃ˢ**
240 Figure A.11.3,4:	"half of the tube length" → "**twice the tube length**"
244.1:	"As back cavity size inversely correlates with tongue height" → "**As back cavity size correlates with tongue height**"
259:	"Gould, Stephen Jay 189" → "**Gould, Stephen Jay 138**"

Table of contents

Dedication

To my mom.

Here's one of the many reasons why:

The scene: our living room, early evening in Chautauqua, Summer 2000. My mom is sitting on the couch reading the Times.

Me (entering from outside):
Ok Mom, I have a pronouncement:
Religion is for children!
Philosophy is for adolescents!
But science is for adults!

My mom (looking up from the paper):
Dan, pronouncements are for children.

Acknowledgments

Large portions of this work were conceived and written in 2002, while I was a Fellow at Edinburgh University's Institute for Advanced Studies in the Humanities. In Edinburgh, I am very grateful to Anthea Taylor, Donald Ferguson, and John Frow at the Institute, to my good friend and colleague Mircea Itu for great times at restaurants and the pub, to Russell Stewart Fountain who lived across the Crescent, to Michael Conway who took me up the hill, and especially to lifelong friends Gary, Karen, and Rosie (and now Duncan) Biggar who were always there for me. Thanks also to Arpad Hornyak, Bob Ladd, Benjamin Platts-Mills, Jim Scobbie, and Ivan Yuen, and to Viktor Tron for his astonishingly careful reading of an early draft of this book, and for all the great discussions at the Blind Poet and elsewhere. I am very much indebted to you, Viktor.

The remainder of this work was written in New York in the latter half of 2004 and into 2005, where I was a fellow of the more mundane variety. Back in the states, Naomi Gurevich, Gary Linebaugh, and Margaret Russell provided me with invaluable assistance and genuine friendship. Special thanks to Devin Casenhiser, whose big brain is matched by his big heart.

For additional comments on portions large and small, and/or for general support, I am grateful to Ree Adler, Phil Carr, Dan Everett, Dong-Bo Hsu, Keith Johnson, Peter Ladefoged, Jerry Levinson, Ken Lodge, Jenny Lovel, Paul Mathews, John Ohala, Kevin Ryan, Michal Schnaider Beeri, Steve Scher, Helaine Silverman, Paul Silverman, and Richard Wright.

Endless thanks to my way cool bro, Jeremy Silverman, who provided me with two—count them two—toothcomb edits of this work. Jer, you're an honorary linguist in my book (is that a compliment?).

Of course, thanks to Jie Zhang and Lois Leardi.

There are many other individuals who have influenced my thinking, helped me maintain my enthusiasm, or otherwise aided, abetted, and inspired from afar. Among them are Neil Babitch, Ian Buruma, Frederick C. Crews, Terrence Rafferty, Ira Robbins, Tim Smith, Jeff Starrs, and Hans Vandenburg.

And if anyone wants to know what my life was like after leaving the wonderful Department of Linguistics at UCLA in 1997 until returning to wonderful New York in 2004, just ask Theodore Roszak!

New York
March, 2005

Preface
I speak from my heart

This book on phonology can be read not only by linguists, but also by philosophers, anthropologists, psychologists, evolutionary biologists, and computer scientists. It's been my experience that experts in these fields often harbor a passing interest in linguistics, but quickly hit a brick wall when trying to penetrate the theoretical intricacies of the discipline. This book is written with these scholars very much in mind, because the approach espoused herein does not call upon the reader to make any concessions to the particularism that dominates mainstream linguistic theory. My hope is to establish a dialogue with these scholars by leveling the playing field: phonology, when done right, is not the exclusive domain of linguists, but should be open to all who can make a contribution to our thinking. In turn, experts in other fields might find that we phonologists can make a contribution to theirs.

The book espouses an approach to phonology that is interdisciplinary in scope. Phonetic theory is featured quite prominently, but both evolutionary biology (as metaphor) and cognitive psychology make significant appearances as well. Phonology herein is viewed as a self-organized and self-sustaining system of social conventions that passively evolves as a consequence of language use. Due to the limited variation that is inherent to speech production, phonological systems are at once sufficiently stable to fulfill their communicative function, and sufficiently variable to be under constant—if slow-going—modification. Systemic changes are often the consequence of the communicative success or failure of the word variants that we use. Successful speech propagates; today's spontaneous, unplanned innovation may become tomorrow's new norm. An indebtedness to Darwin's theory of evolution by natural selection should be apparent even to the most casual of readers.

As the book assumes no previous knowledge of either phonology or linguistics in general, it may even be of interest to the general reading public. If you can make it past Chapter Two—the most difficult chapter in the book—then the remainder may prove quite accessible, and dare I say, enjoyable.

For students of phonology, the content of this book may help to hone their analytical skills. It might help solidify their own inclinations on the subject, or better, it might instead help them "liquefy" their thinking, giving them the impetus to ask their professors some challenging questions, or to rethink certain received notions as they embark on writing their dissertations.

For established scholars who teach phonology, this book may serve as a supplement/alternative to mainstream books. Thinking positively, I envision much constructive class discussion emerging. Phonologists will be quick to note my indebtedness to a number of schools of linguistic thought. The Kazan School is featured quite prominently, but strong traces of both The Prague School and American Structuralism are present as well. The influence of the Generative School should be obvious, since rule-based and constraint-based analyses are featured, albeit re-conceptualized as generalizations about sound change. Finally, the phonological theories of John Ohala served as a constant source of inspiration.

So if this book is for established experts, for students of phonology, for experts in other fields, and also for absolute beginners, then I guess this book is for everybody. Ah,

would that it were true! Truth is, phonology isn't easy (and over the years, I've noticed that some people find it a tad esoteric…): it focuses on the most obscure minutiae of the most everyday and natural of topics: spoken language. It requires patience, concentration, and most importantly, an ability to wipe clean one's subjective feelings about the language that one uses, and to place in their stead a rigorous objectivity. But phonology is not the exclusive domain of linguists or academics. It can be understood and appreciated by anyone who is ready to expend a little effort.

A NOTE ON FONTS

All phonetic transcriptions are written in IPA, the International Phonetic Alphabet, and are enclosed in square brackets. As new symbols are introduced in the text, they are noted, and their basic articulatory and acoustic characteristics are usually discussed. Still, for beginners, there will be a lot to internalize. For readers unfamiliar with the IPA, it is discussed in some detail in the Appendix. I also provide the full IPA chart at the end of this section.

Important terms are usually written in italics when they are introduced. Most of these terms also have glossary entries. Experts will be very quick to notice that I occasionally modify standard definitions: I narrow the definition of neutralization; I broaden the definitions of allophony, sound substitution and sound change; I reconfigure the definition of bi-uniqueness, all in service to the theory of phonology that I explore. I also use italics for emphasis, to add some orthographic intonation to the text. Finally, conclusions or assertions that are especially pithy, pointed, or provocative are written in italics as well.

Quotation marks are usually used for terms whose standard definitions, or whose very value as theoretically relevant notions, I call into question, for example, "phonologization". I suppose these might be regarded as scare quotes. I also use quotation marks for the English translation of foreign languages.

THE INTERNATIONAL PHONETIC ALPHABET (revised to 1993, updated 1996)

CONSONANTS (PULMONIC)

	Bilabial	Labiodental	Dental	Alveolar	Postalveolar	Retroflex	Palatal	Velar	Uvular	Pharyngeal	Glottal
Plosive	p b			t d		ʈ ɖ	c ɟ	k g	q ɢ		ʔ
Nasal	m	ɱ		n		ɳ	ɲ	ŋ	ɴ		
Trill	ʙ			r					ʀ		
Tap or Flap				ɾ		ɽ					
Fricative	ɸ β	f v	θ ð	s z	ʃ ʒ	ʂ ʐ	ç ʝ	x ɣ	χ ʁ	ħ ʕ	h ɦ
Lateral fricative				ɬ ɮ							
Approximant		ʋ		ɹ		ɻ	j	ɰ			
Lateral approximant				l		ɭ	ʎ	ʟ			

Where symbols appear in pairs, the one to the right represents a voiced consonant. Shaded areas denote articulations judged impossible.

CONSONANTS (NON-PULMONIC)

Clicks		Voiced implosives		Ejectives	
ʘ	Bilabial	ɓ	Bilabial	’	Examples:
|	Dental	ɗ	Dental/alveolar	pʼ	Bilabial
!	(Post)alveolar	ʄ	Palatal	tʼ	Dental/alveolar
ǂ	Palatoalveolar	ɠ	Velar	kʼ	Velar
‖	Alveolar lateral	ʛ	Uvular	sʼ	Alveolar fricative

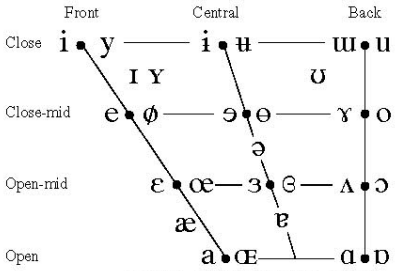
OTHER SYMBOLS

ʍ	Voiceless labial-velar fricative	ɕ ʑ	Alveolo-palatal fricatives
w	Voiced labial-velar approximant	ɺ	Alveolar lateral flap
ɥ	Voiced labial-palatal approximant	ʧ	Simultaneous ʃ and x
ʜ	Voiceless epiglottal fricative		
ʢ	Voiced epiglottal fricative		
ʡ	Epiglottal plosive		

Affricates and double articulations can be represented by two symbols joined by a tie bar if necessary.

k͡p t͡s

VOWELS



Where symbols appear in pairs, the one to the right represents a rounded vowel.

SUPRASEGMENTALS

ˈ	Primary stress
ˌ	Secondary stress ˌfoʊnəˈtɪʃən
ː	Long eː
ˑ	Half-long eˑ
˘	Extra-short ĕ
|	Minor (foot) group
‖	Major (intonation) group
.	Syllable break ɹi.ækt
‿	Linking (absence of a break)

TONES AND WORD ACCENTS

LEVEL		CONTOUR	
é͡e or ˥	Extra high	e͡e or ˩˥	Rising
é ˦	High	ê ˥˩	Falling
ē ˧	Mid	e͡e ˦˥	High rising
è ˨	Low	e͡e ˩˧	Low rising
ȅ ˩	Extra low	e͡e ˥˩˥	Rising-falling
↓	Downstep	↗	Global rise
↑	Upstep	↘	Global fall

DIACRITICS Diacritics may be placed above a symbol with a descender, e.g. n̥ŋ̊

◌̥	Voiceless	n̥ d̥	◌̤	Breathy voiced	b̤ a̤	◌̪	Dental	t̪ d̪
◌̬	Voiced	s̬ t̬	◌̰	Creaky voiced	b̰ a̰	◌̺	Apical	t̺ d̺
◌ʰ	Aspirated	tʰ dʰ	◌̼	Linguolabial	t̼ d̼	◌̻	Laminal	t̻ d̻
◌̹	More rounded	ɔ̹	◌ʷ	Labialized	tʷ dʷ	◌̃	Nasalized	ẽ
◌̜	Less rounded	ɔ̜	◌ʲ	Palatalized	tʲ dʲ	◌ⁿ	Nasal release	dⁿ
◌̟	Advanced	u̟	◌ˠ	Velarized	tˠ dˠ	◌ˡ	Lateral release	dˡ
◌̠	Retracted	e̠	◌ˤ	Pharyngealized	tˤ dˤ	◌̚	No audible release	d̚
◌̈	Centralized	ë	◌̴	Velarized or pharyngealized	ɫ			
◌̽	Mid-centralized	e̽	◌̝	Raised	e̝	(ɹ̝ = voiced alveolar fricative)		
◌̩	Syllabic	n̩	◌̞	Lowered	e̞	(β̞ = voiced bilabial approximant)		
◌̯	Non-syllabic	e̯	◌̙	Advanced Tongue Root	e̙			
◌˞	Rhoticity	ə˞ a˞	◌̘	Retracted Tongue Root	e̘			

Language occupies a completely isolated place in the realm of nature: it is a combination of physiological and acoustic phenomena governed by physical laws, and of unconscious and psychical phenomena governed by laws of an entirely different kind. This fact leads us to a most important question: what is the relation...between the physical principle and the unconscious and psychical principle?

-Mikołaj Kruszewski, 1881

Part 1
I speak with my mouth

Chapter 1
Three types of sound substitution

SETTING THE SCENE

When I was little my father and I sometimes played a game called Jotto. To play Jotto, each of us would secretly write down a five-letter word, and then take turns guessing what the other person wrote. After each guess the other player would report the number of letters that appeared in his original word. For example, my father might write down "sport", S-P-O-R-T. If I guessed "brick", B-R-I-C-K, he'd say "one", because only one of the letters in "brick"—the "R"—appears in his word "sport". Although my next guess could be a word with a whole new batch of letters, a conservative strategy was to find a word with only a single letter different from the previous guess. For example, if I next guessed "trick", T-R-I-C-K, my father would have answered "two", because two of the letters in this word, "T" and "R" appear in his word. At this point, I could safely conclude that "T" is in his word, that "B" is not in his word, and that one of the remaining four letters of "R", "I", "C", and "K" is also in his word. By systematically eliminating certain letters, and systematically determining the presence of others, the alphabet could eventually be whittled down to just a few letters out of which the right word could be spelled. The first person to guess the other's word would win.

This book is about phonology—the study of linguistic sound systems. Broadly speaking, whereas phonetics explores the physical aspects of speech, phonology explores its functional aspects. Both disciplines thus explore speech patterns, but to rather different—if highly interdependent—ends. Phonologists are primarily concerned with documenting sound substitutions—the replacement of one sound with another. In Jotto, you replace one letter with another letter to give you a new word. In spoken language, it is the replacement of one sound with another that serves this same function. While letters are intended to represent sounds, we all know that the English writing system is far from perfect in this respect. In fact, we'll soon see that the English writing system serves to confound our understanding of the true nature of the English sound system.

Investigating sound substitutions is one of the primary tasks of phonologists because of the functional consequences that these substitutions have for word meaning. As in the "brick"-"trick" example, we see that sound substitutions can change the meaning of a word. Substituting one sound for another is a very efficient way to create many words from the sounds that we produce with our vocal tracts, and so it's no accident that phonological systems have evolved this property. But, as we'll soon see, some sound substitutions eliminate a distinction in word meaning that existed before the substitution, and other sound substitutions take place without changing a word's meaning.

In my childhood I also played a game that was quite similar to Jotto, called Mastermind. Instead of five-letter words, Mastermind uses pegs of six colors. Your secret

code is any combination of four pegs, say, Yellow-Black-Red-Green (but you could use colors more than once if you want to). The logical strategy of Jotto applies in the same way with Mastermind: take an initial guess about the code of your opponent, get feedback, and modify your next guess accordingly. The feedback in Mastermind is a bit more detailed than it is in Jotto, because you're told how many pegs are positioned correctly in your row of four, in addition to how many are merely present. However, Mastermind is a much less interesting game conceptually, because there are no constraints on what sorts of color sequences might be used. Every guess could minimally alter the previous one by replacing one color with another, or by minimally altering the sequence: I could follow a guess Red-Blue-White-Green with Black-Blue-White-Green, or maybe Blue-Red-White-Green. By contrast, in Jotto you can't necessarily replace any one letter with another to directly test your hypothesis. Since your tests are constrained by English spelling, you can minimally alter your next guess only if the result is also a well-formed word. So sometimes circuitous trial-and-error routes are required to isolate the correct letters. For example, when I played Jotto, I could never change B-R-I-C-K to, say, B-N-I-C-K, to test for "R" and "N", because "bnick" isn't a word of English; the language has no words that start with "B-N". In fact, you can't make any word out of those five letters, so you have to try a different strategy. This constraint on letter sequences and combinations adds a significant level of sophistication and challenge to Jotto, and makes it a much better game than Mastermind, which has no restrictions on sequences and combinations.

So, whereas Mastermind players are fully unconstrained, Jotto players must operate in accordance with letter-sequential or combinatorial constraints. But what might be the origin of these constraints in Jotto? Did I have unconscious knowledge of some sort of spelling constraint that prohibited words from starting with "B-N", and is this constraint the reason why I would never even think to guess "bnick"? I don't think so. A simpler and more straightforward reason for rejecting "bnick" as a guess in Jotto is that I just didn't know any word spelled B-N-I-C-K. Since I didn't know such a word, it would never even occur to me to use it as my next guess. In Jotto, the constraint isn't a matter of "B-N-I-C-K violates the spelling rules of English". It's simply that "there's no English word spelled that way"; good guesses in Jotto are constrained only by our experience and familiarity with reading and writing English. So there isn't a structural constraint on possible Jotto guesses. Rather, there's an experiential constraint based on my knowledge and use of actual English words. By contrast, sequences of colored pegs are totally arbitrary to me. One sequence is as good or as bad as the next. Since sequences of colored pegs serve no function in my life, I have no greater or lesser experience or familiarity with any particular combination of them.

Phonology is more like Jotto than like Mastermind, but not merely because Jotto deals with actual words of language while Mastermind doesn't. Rather, just as in Jotto, there's an experience-based constraint on what sounds can be substituted for what other sounds. In phonology, just as in Jotto, the constraint is very simple, almost trivial: the replacement of one sound with another always results in a sound sequence that can be paired with a particular meaning that is shared by speakers and their speech communities. For example, in English, the difference between, say, "fit" and "pit" resides in the first sound of each word. Both "fit" and "pit" are words of English; they mean different things for a speaker of English. Sound replacements that change meaning are known as

contrastive sound substitutions. English speakers never engage in a sound substitution if the results aren't meaningful, that is, if the resulting form is not a word of English. The sound substitutions that we employ are strictly constrained, but—contrary to the beliefs of many linguists—I believe that sound substitutions are not governed by a system of rules or constraints on good form. We never substitute an "n" for the "r" in "brick", but not because the result would violate English rules or constraints on sound sequences. Instead, we don't say "bnick" simply because we never learned to pair that sound sequence with a particular meaning, and so it's not English—it serves no linguistic function.

Before continuing, let me clarify something. When I say that speakers engage in sound substitutions, I do not mean this in any procedural sense, like the way someone plays Jotto. Speakers don't start with one word, and then change that word into another by replacing—or adding and/or subtracting—sounds. Rather, when I use the term "sound substitution" I am simply offering a helpful characterization of the patterns that phonologists take note of as they document speech. But these substitutions should not be regarded as a result of a process. Focusing on sound substitutions of the "brick"-"trick" sort helps to reveal the remarkable systematicity that is present in linguistic sound systems, but this doesn't mean that this characterization genuinely reflects the cognitive organization of the sound system itself. Some linguists, including me, suspect that we should treat many of the individual sounds that we might extract from the speech signal as convenient fictions; they might lack genuine structural status as individual elements, but they make the job of discussing phonological patterns much easier.

Now, among non-existent sound sequences in English, some certainly sound better than others. For example, "blick" sounds better than "bnick", even though neither is a word. If some non-words feel better or worse than others, how can I say that the only relevant distinction to be made is whether the sound sequence is an actual word or not? Many phonologists—though not I—think that "blick" is a possible word because it doesn't violate any sound-sequencing constraints of English, except that it just happens to be missing, and so it feels okay. These phonologists propose that "bnick", by contrast, involves a genuine violation of an unconscious sound-sequencing constraint, and so it sounds awful to English ears. Such a constraint might strictly prohibit English words starting with the sound sequence "bn".

But such an approach, flawed when applied to Jotto, is just as flawed when applied to English. When I speak English, every sound substitution is always one word or another. There are no relevant feelings on my part about whether the sound substitution is good or not. They are all good, because they are all English. So if "blick" feels good, and "bnick" doesn't, parallel sorts of feelings are nowhere to be found when we compare real words. Does it even make sense to ask whether the word "brick" feels better than the word "trick?" Even if some people have an intuition on the matter, would their feelings somehow teach us anything about linguistic sound structure? I maintain that we can't determine the structural properties of linguistic sound systems based on how people feel about the sounds they use. This has been stated quite emphatically by the scholars Bernard Bloch and George L. Trager. Writing in 1942, they assert that "The ordinary speaker of English, we are told…'feels' or 'conceives of' the two [l]s in 'little' as 'the same sound.' This may or may not be true; if true, it is an interesting fact, but it can never be used by the linguist as a criterion for his classifications, or even as a proof that he has

classified correctly". Bloch and Trager continue: "The native speaker's feeling about sounds or about anything else is inaccessible to investigation by the techniques of linguistic science, and any appeal to it is a plain evasion of the linguist's proper function. The linguist is concerned solely with the facts of speech. The psychological correlates of these facts are undoubtedly important; but the linguist has no means—as a linguist—of analyzing them".

Just out of curiosity though, what can we say about the feelings engendered by "blick" versus "bnick"? Since we can't assign meanings to these two sound sequences when we speak them, we can only treat these spoken forms as non-linguistic sound signals. What we have to do then, is compare the purely physical properties of "blick", which we transcribe phonetically as [b̥lık], and those of "bnick", which we transcribe [b̥nık], to the inventory of English words that we have in our heads. Let's be systematic about it. Let's take each major subpart of the speech signal of these two words—as represented by the phonetic symbols employed—and see if there are any English words that share these properties. The lists in Table 1.1 include every substring of the component sounds of "blick" and "bnick". Beginnings and ends of words are indicated by cross-hatching, "#". Sequences that are found in English words are accompanied by some examples.

[b̥lık]	Examples:
#[b̥…]	bean, birth
#[b̥l…]	blend, blue
#[b̥lı…]	blimp, blister
#[b̥lık]#	
[...lı…]	clip, slit
[…lık]#	slick, lick
[…ık]#	sick, kick
[…k]#	folk, lock

[b̥nık]	Examples:
#[b̥…]	bean, birth
#[b̥n…]	
#[b̥nı…]	
#[b̥nık]#	
[…nı…]	snicker, catnip
[…nık]#	nick, picnic
[…ık]#	sick, kick
[…k]#	folk, lock

Table 1.1. subsequences of [b̥lık] and [b̥nık].

For [b̥lık], there are many perfectly good English words that also have these sound sequences, and so every subpart of the form sounds familiar to an English speaker. The one exception, of course, is the complete form [b̥lık], which we already know isn't a word at all, although English speakers are indeed familiar with its occurrence as a possible sequence (if not a complete word) in other languages, for example, Hans Blix ([b̥lıks]).

The case of [b̥nık] is quite different. Look at the blanks in the list for [b̥nık]. These gaps have one thing in common: in English, there is an overall absence of words starting with [b̥n]. When I hear [b̥nık], I can't think of any words that start with that sequence of sounds. So "blick" feels okay, because every partial sequence of sounds is fine in English. But "bnick" sounds terrible, because several of these sequences are never found in English ([b̥n], [b̥nı] [b̥nık]). Maybe, just maybe, some pronunciations of "banana" or "benevolence" begin with [b̥n] when speaking fast. If so, [b̥n] is not completely absent,

but instead is just very rare indeed. But does this mean, again—and as many phonologists actually believe—that somewhere in my brain I have an inventory of sound-sequencing constraints, one of which forbids [b̥n] at the beginning of words? No, it doesn't. All it means is that there are no English words pronounced this way (bananas aside), just as in Jotto, where there are no words spelled that way.

So, any feelings we might have about what is a good word, a possible word, or an impossible word, merely reveal the limits of our linguistic experience, and nothing more. Tellingly, there might be a few words that really do feel funny. For example, every New Yorker knows that a knish [kʰniʃ] is a savory potato or kasha pastry, but [kʰn] really does feel a bit off, even to us New Yorkers. Does English have a rule that says "no [kʰn] at the beginning of words"? Obviously not, because we have "knish", which is, albeit, a Yiddish loan. Rather, it's just that [kʰn]-initial words are very rarely encountered. In fact, English used to have sequences like [kʰn] and [g̥n], as indicated in the spelling of "knee" [kʰni] and "gnat" [g̥næt], but they fell out of the language about 300 years ago. So nowadays, since we have so little experience with such words, they sound funny.

Actually, there is another possible objection to my account for our different feelings for "bnick" versus "blick": maybe "bnick" is more difficult to pronounce, and so it sounds strange to us. Well, this is simply not the case. I can very easily close my lips for the [b̥], and then lift my tongue tip while letting air flow out my nose for the [n]. Making these sounds in sequence is no problem at all, especially since the two sounds are made with parts of the mouth that function independently from each other: there's no reason that the transition from the [b̥] to the [n] should pose any articulatory difficulties whatsoever. I can even think of a few examples in English that have [b̥n] in the middle of the word: "Abner", "obnoxious", "abnormal", "hobnob". So [b̥n] is not more difficult to say than [b̥l]. If you do think "bnick" is difficult to pronounce, it's probably because you have practically no experience with making the [b̥n] sequence at the beginning of a word, and not because of any intrinsic difficulty.

As for sequences like [b̥n], [kʰn], and [g̥n] at the beginning of words, it turns out that it is hard to clearly hear sounds like [b̥], [k], or [g̥] in this context. Such auditory facts might explain why some sequences survive and flourish over generations of speakers, while others are extinguished or never arise. The rarity or absence of some sound sequences, and the prevalence of others, are important for phonologists to take note of, and have to do with a very complex interaction, over generations of speakers, among the physical properties of sound, of sound perception, and yes, of speech production (hence the sub-title of this book). Indeed, much of the discussion that follows is devoted to motivating the prevalence of some sounds and sound sequences in comparison to others. But the important point for now is that our effortless mastery of English sound substitutions derives from familiarity and experience with English words themselves, and not from a mentally-compiled list of sound-sequential rules or constraints on what constitutes a good word of English.

THREE TYPES OF SOUND SUBSTITUTION

If we think about the situation logically, the sound substitutions that we observe in English, or in any language, are limited to only three functional types. In contrastive

sound substitution, word meaning is changed (as in the case of "brick"-"trick"). There are, in addition, two types of non-contrastive sound substitution. In neutralizing sound substitution, the substitution of one sound with another results in two words sounding exactly the same, and so phonetic evidence for their distinction in meaning is eliminated. For example, as we discuss in the next section, when you say "phone book", "phone" often comes out sounding exactly like "foam". In allophonic sound substitution, the substitution of one sound with another does not change the meaning of the word: the meaning remains the same even after the sound substitution. For example, as discussed below in detail, the last sound of the word "invite" is pronounced differently when you say "invite someone" compared to when you say "invite anyone", yet both pronunciations have the same meaning. (Recall that language is more like Jotto than like Mastermind; we only engage in sound substitutions that have functional relevance, and so logical possibilities that are not real words are not relevant to our discussion.

These two additional types of sound substitution inevitably sabotage what we might call bi-uniqueness between sound and meaning. If phonological systems had the property of bi-uniqueness, then each sound sequence which makes up a word would uniquely match up with a single meaning, and each meaning would uniquely match up with a single sound sequence. But no language has this property. Instead, there is inevitably a one-to-many relation between sound and meaning (neutralization), and also a many-to-one relation between sound and meaning (allophony) (All languages also have homophones that are not a consequence of sound substitutions, but instead are words with different meanings that simply are pronounced the same way, like English "dear" and "deer", for example.) Non-bi-uniqueness due to sound substitution creates a remarkable complexity in sound-meaning relations that has often stymied linguists in their understanding of the relevant structural properties of language, but non-bi-uniqueness never stymies children as they are learning their language. So let's consider each of the three types of sound substitution in turn, to try to get a handle on their basic attributes.

1. CONTRASTIVE SOUND SUBSTITUTION

First, a sound substitution can change the meaning of a word. Our Jotto example has shown this quite clearly and intuitively. We can substitute the first sound in "brick", which we transcribe [b̥ɹɪk], with the first sound in "trick", which we transcribe [tʰɹɪk]. The result of this sound substitution is a change in word meaning. The specific change in meaning is not important for our purposes. The fact that [b̥ɹɪk] usually refers to a block of stone or concrete, and [tʰɹɪk] can be prank or a ploy, is only incidental. The important point is that each form corresponds to a different meaning, regardless of the particular meanings involved. Given the words "brick" and "trick", we can conclude that [b̥] and [tʰ] are contrastive in the context [__ɹɪk]. That is, substituting [tʰ] for [b̥] in the context [__ɹɪk] results in a change of word meaning. But for now, that's all we can conclude about the relationship between [b̥] and [tʰ]. We can't yet conclude that [b̥] and [tʰ] are contrastive in any other context (although further investigation will very quickly show that they do indeed contrast in many other contexts as well, for example, "back"-"tack", "best"-"test"). In order to determine the extent to which [b̥] and [tʰ] can be substituted for each other, phonologists have to look at many other words, and other sound contexts. For example, In English, [tʰ] is never substituted for [b̥] in "block". Indeed, further

investigation would reveal that [t^h] never precedes [l] at the beginning of an English word.

Interestingly, we can substitute one sound for another rather effortlessly on demand, by consciously manipulating our speech sounds. For example, If I ask you to take the word "brick" and substitute a "t" for the "b", you can perform the task effortlessly, and give me "trick". Some linguists argue that our ability to consciously manipulate speech sounds on demand opens a window onto the genuine structural properties of linguistic sound systems. That is, our intuitions about speech sounds, and our ability to consciously manipulate these speech sounds, provides evidence of these sounds' status as linguistically significant phonological entities. But just as our feelings about language are extremely unreliable with respect to offering insight into linguistic sound structure, our ability to consciously manipulate speech sounds provides us with no insight whatsoever about linguistic sound structure. When we play with our language, there is no reason to assume that the elements we are manipulating are the genuine building blocks of the sound system.

In fact, our ability to consciously manipulate speech sounds does not derive from our implicit knowledge of English phonology at all. Rather, it derives from our explicit knowledge of the orthographic (writing) system we use to visually (or tactually, in the case of Braille) represent language. English uses an alphabetic writing system, in which each symbol loosely—though quite imperfectly—represents a contrastive sound of the language. As we master our writing system, we know that switching a letter typically results in a change of sound, and—most importantly—typically results in a change of word meaning: replacing the "B" in B-R-I-C-K with a "T" gives us another word, T-R-I-C-K. Remember that this is explicit knowledge, not implicit. We are taught the alphabet, and how we can use it as a tool to represent the sounds and words of English. As a consequence of this explicit learning, we can apply the alphabetic principle in order to understand how sound substitutions might change word meaning. The claim then, is that our ability to substitute sounds on demand does not provide evidence for the cognitive organization of language, and so it does not establish a direct link between our intuitions about sound structure and the genuine structural properties of our phonological system.

How have researchers come to this conclusion? It turns out that illiterates do not have the same skills and intuitions about contrastive sound substitutions that come so effortlessly to you and me. For example, illiterate Portuguese fishermen have a great deal of difficulty substituting one sound for another when asked to do so by an experimenter, by swapping a sound, or adding or subtracting a sound. Where you and I can effortlessly chop off the first sound of a word and replace it with another if asked to do so, these illiterate fishermen have difficulty even understanding the task. We might conclude, therefore, that our ability to manipulate speech sounds in this fashion is largely dependent —perhaps solely dependent—on our familiarity with an alphabetic writing system.

But of course, these fishermen were illiterate. Illiteracy may stem from two broad sources: a lack of formal education, and/or some sort of learning disability. It certainly might be the case that they were illiterate not because of a lack of formal education, but instead, because of an inability to learn how to read. Perhaps this inability is also responsible for their failing to learn the sound-substitution task. A number of clever experimentalists have examined this hypothesis, and have devised a rather straightforward method to test it. These researchers have given a similar sound-

substitution task to educated adults who are fully literate in a writing system that is not alphabetic in nature. In Chinese, for example, each symbol, or character, represents a whole word, and contains absolutely no information about the component sounds that combine to form the phonetic quality of the word. (There is, typically, some phonetic information that is encoded in a Chinese character, but it is never of the alphabetic sort. Instead, it encodes information about the phonetic quality of the form as a whole. For example, 花, 哗, and 华 all mean different things, but all sound like [xwɑ] in the Mandarin dialect, as they all have the component "化", a character on its own, which also sounds like [xwɑ]; [x] is the sound at the end of "Bach"). When literate, educated Chinese adults (who have no knowledge of an alphabetic writing system) are asked by an experimenter to perform a sound-substitution task, in fact, just like the illiterate Portuguese fishermen, they can't do it! So, one's ability to manipulate speech sounds is clearly unrelated to general intelligence, and instead is rooted in the explicit learning of an alphabetic writing system. Consequently, the ability to manipulate these "building block"-like sounds in these sorts of language tasks cannot be taken as evidence for their relevance at the linguistic level.

Allow me to make this point in a more concrete way. I lived in China for a few years, teaching conversational English to graduate students at a prestigious university in Shanghai. Sometimes, when meeting with my students, they would need to consult a Chinese-English dictionary to find an English word they didn't know. Since Chinese is not written alphabetically, you might wonder how dictionaries are organized. Most are organized by "stroke"-order, where a stroke is one of the dots or lines that, combined with others, forms a complete character. For example, 止 has four strokes, 怍 has eight, and 樽 has eighteen. By learning a few simple rules of stroke ordering (which all Chinese do when they learn to read and write), looking up words in the dictionary becomes as trivial as it is for you and me. However, some Chinese-English dictionaries are organized by a Roman alphabetic system, called "hanyu pinyin" ("Chinese spell-sound"). In school, young students learn this alphabet to jump-start their learning of Chinese characters, which take years to master. But the hanyu pinyin system quickly falls into disuse during the course of education. (Importantly, the aforementioned literate adults had no training in hanyu pinyin.) Now, when my students used a dictionary organized by the hanyu pinyin system, they would slowly and laboriously mouth each of the component sounds of the word in their effort to determine its spelling, because they had so precious little familiarity with an alphabetic writing system, and so were never called upon to break words down into smaller, sound-based units. I confess that I used to get impatient with them, because I knew I could find the word much more quickly than they could, even though they were using a Chinese-to-English dictionary. It's not that I was more intelligent than they were—indeed, they were among China's best and brightest—and I certainly didn't have better intuitions about Chinese phonology than they did. Instead, unlike them, I was well-trained in an alphabetic writing system. This made my mastering hanyu pinyin almost trivial. That's the reason I could find words faster than they could in a spelling-based dictionary, and that's why you can so readily understand the concept of contrastive sound substitutions, and consciously manipulate speech sounds accordingly.

Bloch and Trager, once again, express the subtleties of this argument very succinctly. When a researcher is confused about the sound-structural properties of some foreign language, "This uncertainty cannot be resolved simply by asking the informant. If

[the informant] is sophisticated enough to understand such finespun questions, he is probably literate in his native language and hence likely to be misled by the way in which words are written, by the tradition of the schools, and by other equally fallible guides; and if he is unspoiled by education, the chances are that questions about the identity of words will only baffle him".

2. NEUTRALIZING SOUND SUBSTITUTION

The two other types of sound substitution—neutralization and allophony—are not nearly as intuitively obvious as contrastive substitution because they do not involve a change in meaning, and so they are not usually reflected in our writing system. In a neutralizing sound substitution, the replacement of one sound with another eliminates the phonetic distinction among words, resulting in homophony. This obviously creates the potential to eliminate meaning distinctions: two (or more) words end up being pronounced the same, and so there is no phonetic evidence for their distinction in meaning. But the counter-functional effects of neutralization are never very devastating, since the real-world context and/or the grammatical context normally makes a speaker's intended meaning clear. As an example of neutralization, consider again the compound word "phone book", mentioned earlier. When you say this in a natural way—at a natural speech rate, and in a natural conversational context like, "Where's the phone book!?" quite possibly, your lips will close during the last sound in "phone", resulting in something that sounds very much like "foam book". So, in the context of "-book", "phone" may end up sounding like "foam".

Nonetheless some might feel that the pronunciation [fõ$^{\breve{o}}$n]—with an [n]—is somehow a more "authentic", or a more "privileged" realization of the word "phone" than is the alternative pronunciation [fõ$^{\breve{o}}$m], with an [m] (for now, you can just concentrate on the [n]-[m] substitution, which is the primary difference in the two forms). In fact, sometimes the word is pronounced [fõ$^{\breve{o}}$n], sometimes [fõ$^{\breve{o}}$m], and sometimes, as in "phone call", it may be pronounced [fõ$^{\breve{o}}$ŋ] (in which the symbol [ŋ] indicates the "ng" pronunciation). For [fõ$^{\breve{o}}$m] and [fõ$^{\breve{o}}$ŋ], the final sound in "phone" matches the following sound ([b̥] and [k], respectively) in terms of tongue and lip position, but neither of these pronunciations of "phone" is any less legitimate than [fõ$^{\breve{o}}$n] (with [n]). Our intuitions might tell us that [fõ$^{\breve{o}}$n] is the real pronunciation of "phone", but I've already emphasized that our feelings about language are of no help in determining its structural properties. In fact, I can just imagine someone protesting, "I don't say 'pho[ŋ]e,' I say 'pho[n]e,' like in 'pho[ŋ]e call, or pho[m]e book.' I always say it with an 'n'!" This person doesn't realize that the same word can have different pronunciations depending on its context, and so mistakenly believes that the word is always pronounced in just one way. The reason our intuitions tell us that [fõ$^{\breve{o}}$n] is somehow more real or authentic than [fõ$^{\breve{o}}$m] or [fõ$^{\breve{o}}$ŋ] is that our notion of the correct pronunciation of an English word is usually based on its pronunciation in isolation; also, it is often influenced by how the word is spelled.

It might help if I use set diagrams to illustrate neutralization. Let's suppose for the moment that the word "foam" has only one pronunciation, whereas "phone" has the three that we have just considered, which are dependent on the context in which the word is found. While this is a simplification, for now, let's just suppose it's true. If different

pronunciations have the same meaning, they are grouped into a set. The ambiguity of [fõʊ̃m] is indicated by the intersection of the two sets in Figure 1.1.

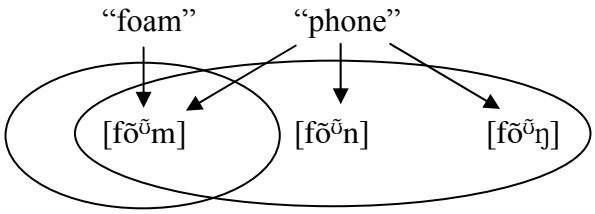


Figure 1.1. Sets for "foam" and "phone"

The example of "phone" and "foam" shows us that words that are distinct when standing alone may neutralize when other words of certain phonetic shapes are added: the form [fõʊ̃m] corresponds to more than one meaning when [b̥] immediately follows. In other words, the [n]-[m] sound substitution in this phonetic context has the potential to eliminate a distinction in word meaning due to the resultant homophony. This is neutralization.

You might now grant that "phone" has a number of perfectly acceptable pronunciations, and that one of these—[fõʊ̃m]—is ambiguous between "phone" and "foam". But what is a foam book? You might imagine that you're about to give the children their bath, and they're crying for their favorite bath-safe book, and you can't find it, and so you mutter under your breath, "Where's the foam book!?" But I think you'll agree that in most cases, the intended meaning of the form [fõʊ̃m] in this phrase will be unambiguously interpreted by listeners as "phone". Even when the phonetic distinction among words is neutralized, the real-world or grammatical context of the neutralized form usually serves a disambiguating function. As I said, the functional consequences of neutralization are never very dire.

We've now seen that words that are distinct when standing alone may neutralize when other words of certain phonetic shapes are added. But this doesn't mean that the spelling pronunciation, or the pronunciation that we use in isolation, has any privileged structural status, or is any more real or authentic than other pronunciations. To show this clearly, let's consider another route to neutralization: in some languages, certain words in isolation are homophonous, and they only express their full contrastive status when certain sounds are added. In a well known paper written in 1933, the linguist and anthropologist Edward Sapir reported some findings on the Sarcee language of Alberta, Canada. His Sarcee language consultant, John Whitney, provided him with two words that sounded exactly the same to Sapir, and yet Whitney insisted that the words were different. What Sapir heard as [dìnɪ́h] corresponded to two meanings for Whitney, "this one", and "it makes a sound". (The grave and acute accents indicate, respectively, lower pitch and higher pitch; Sarcee is a tone language, and so the relative pitch—higher or lower—can change the meaning of a word. Changing the pitch in this way is a sound substitution like any other; we'll come back to tone in later chapters.) Although Whitney himself could not actually hear any difference between the two, and could not even sense

an articulatory difference in his mouth, he nonetheless felt that the words were not identical. After he and Sapir worked at some length to figure out the exact nature of the distinction between the two words, Whitney finally said that he "felt a 't'" at the end of the form for "it makes a sound". Both Sapir and Whitney were rather stumped by this phantom "t" that Whitney could mentally feel, but for which there was no tangible evidence. But as Sapir learned more about the structure of Sarcee, he found that by adding the suffix [ɁÍi], "the one who…", this phantom "t" finally made itself heard: [dìnɁít^{h}Ɂí]. Suffixing [ɁÍi] to the other [dìnɁíh] ("this one") did not induce the presence of a [t]. Adding other suffixes to the words also resulted in phonetic distinctions between them. So the words phonetically manifested their contrastive status only when suffixed, but were homophonous when the suffixes were subtracted. This source of neutralization is quite different from the English example just discussed, because in English the neutralization occurs upon the addition of certain sounds, whereas in Sarcee it occurs upon the subtraction of sounds. The Sarcee case, then, clearly shows us that forms in isolation are not linguistically "privileged" in any sense.

Sapir concluded that Whitney's feelings about the "t" constituted evidence for Whitney's otherwise hidden knowledge of Sarcee's sound system. Despite its physical absence, Whitney could mentally feel its presence, and so the "t" was "psychologically real" in Sapir's parlance. Sapir proposed that there is an abstract phonological value—a "phoneme"—that is psychologically somehow more basic, more prominent, than any of the phonetic ways in which this value is phonetically expressed, and that this abstract phonological value may be psychologically present even in the absence of any physical manifestation. (The term "phoneme", though not necessarily this concept of its meaning, predates Sapir's work by a number of decades. In Chapter Seven we consider the profound influence that Sapir's proposals have had on twentieth century phonological theory.)

Two years after Sapir's paper appeared, another scholar, William Freeman Twaddell, published a work in which he challenged Sapir's interpretation of the Sarcee phantom "t". Twaddell wrote, "In so far as this incident may be interpreted as evidence of any mental reality, it would appear to be rather a morphological class or lexical unit than any phonetic or quasi-phonetic class or unit". In simpler language, Twaddell didn't deny that there was linguistic significance to Whitney's feelings, but these feelings reflect the distinction in meaning between the two words, a distinction for which the phonetic cues are absent when the words are unsuffixed. These feelings, according to Twaddell, don't tell us anything about the supposed psychological reality of the sound structure of Sarcee.

If Twaddell was correct in his interpretation, why did Whitney himself, a native speaker of Sarcee, report his feelings in terms of sound, and not meaning? Let's try to recreate how the difference between the two [dìnɁíh]s might have come to Sapir's attention in the first place. Although Sapir never discussed it, we might imagine that he first encountered the two forms of [dìnɁíh] at different times during his work with Whitney, when the meaning distinction between the two was very clear from the context in which the words were used. So there would have been no confusion on Whitney's part about their pronunciation. Since he was not specifically comparing and contrasting the two forms, Whitney was probably perfectly happy in pairing one [dìnɁíh] with one meaning, the other [dìnɁíh] with the other, sort of like when we say "pho[m]e book"

without giving the ambiguity of [fõ$^{\tilde{o}}$m] a second thought. Maybe later, Sapir checked his notes and realized that these two different words were pronounced the same way. At that point, he might have asked Whitney to compare the two. If Sapir asked Whitney about the word [dìnɨ́h] in complete isolation, without a context that would link the word to one or the other of its meanings, Whitney might have said, "Hmm, that word can mean two different things!" But when Sapir put the word in two different contexts which induced the two different meanings, only now—when Sapir specifically juxtaposed the two words in a way that brought to the fore their distinction in meaning (though not yet their distinction in sound)—would he engender in Whitney the confusing feelings about the difference between the two words. But, according to Twaddell, these feelings might not tell us anything about the mental organization of the sound system. How can Twaddell be right? Sapir wrote that he and Whitney were searching for a way to phonetically distinguish between the two [dìnɨ́h]s. Although they had to give up this endeavor when no phonetic distinction was found, they were still in a sound-based frame of mind; they were still looking for a sound-based explanation for Whitney's feelings. I imagine that Whitney began silently thinking about the two words in different contexts, for example, in a suffixing context where a [t] was actually present for the one meaning, but absent for the other. Now Whitney could give Sapir an answer that they could be satisfied with. He mentally plugged [dìnɨ́h] into different contexts, a [t] popped up for only one of the meanings, and so he could tell Sapir that he "felt a 't'". Under this scenario, then, just as Twaddell argued, Whitney was really responding to the difference in <u>meaning</u> between the two [dìnɨ́h]s, not to a difference in phonological structure.

Whitney was mistaken in a way similar to someone who says, "I don't say 'pho[ŋ]e,' I say 'pho[n]e,' like in 'pho[ŋ]e call, or 'pho[m]e book". Where our English speaker mistakenly thinks that all these "phone"s sound the same because they have the <u>same</u> meaning, Whitney was mistakenly setting apart phonetically identical forms that have <u>different</u> meanings. In both cases, the confused responses are a consequence of mis-pairing sound and meaning—a consequence of non-bi-uniqueness induced by non-contrastive sound substitution—not a consequence of the supposed "psychological reality" of elements of the sound system. But Whitney wasn't the one who made the real mistake, however. It was Sapir who was inducing the confusion that Whitney experienced—a confusion that Whitney would never feel in an everyday language context—by asking him to report his feelings about differences in <u>meaning</u> in terms of differences in <u>sound</u>. But this doesn't make any sense, as Whitney's legitimately confused reaction shows us. Indeed, the task of the phonologist and the task of the language learner/user are very different, and <u>there's no reason to assume that the methods I employ and the generalizations that I make as a phonologist are the methods employed and generalizations that people make when they are actually learning and using their language</u>. Sapir, I would claim, was confusing the knowledge that he possessed as a Sarcee <u>linguist</u> with the knowledge that Whitney possessed as a Sarcee <u>speaker</u>. As a Sarcee linguist, Sapir—a remarkable field linguist by anyone's standards—was busy establishing generalizations about the sounds of Sarcee. As a Sarcee speaker, Whitney was busy extracting meaning from the speech signal.

It was very easy for me to illustrate contrastive sound substitution with words like "brick" and "trick" because we have conscious awareness of this phenomenon. This sort

of substitution is reflected in the writing system which we are taught, and more importantly, this sort of substitution changes word meaning: it is the change in word meaning that truly resonates with language users, because of its functional importance. But I suspect it took a bit more to convince you that the "n"s of "phone" and "phone book" are phonetically distinct from each other, in part because, usually, our writing system only encodes sound changes that produce changes in meaning, but mostly because such neutralizing sound substitutions do not play the functionally important role of switching one meaning for another.

3. ALLOPHONIC SOUND SUBSTITUTION

We've now discussed how [fõ$^{\tilde{o}}$m] may be ambiguous between "phone" and "foam". But what about [fõ$^{\tilde{o}}$ŋ]? This form doesn't mean anything on its own, and it can only mean "phone" in contexts like "phone call". In this case, the substitution of [ŋ] for [n] neither changes word meaning nor induces homophony with any other word. Since it is neither contrastive nor neutralizing, this sound substitution must be of a third type. This third and final type of sound substitution is called allophony, where "allo-" means "same" and "-phon(e)" refers to Sapir's "phoneme". These sounds are different phonetic realizations of the same phoneme. But since we are not embracing the theoretical construct "phoneme" in this book, the term "allophony" is consequently slightly misleading for us. For our purposes, the term "allophony" refers to the fact that the sounds in questions are phonetically distinct, though functionally non-distinct.

Consider again the last sound in the word "invite" when you say "invite someone". There are a few ways in which a speaker of American English might say this word. Each is as good as another to illustrate my point, so let me transcribe this last sound [$^{\text{ʔ}}$t$^{\neg}$], which is the way I often produce it. The superscripted question mark-like symbol indicates a glottal stop, in which the vocal folds suddenly and completely shut tight, prohibiting any air from leaving the lungs. You may have trouble recognizing the glottal stop because our orthography does not use a symbol for it, but you make one every time you answer a question negatively with "uh-uh", as opposed to the positive "uh-huh". In the negative form, the silence between the two vowel sounds is the glottal stop ([ɐ̃ʔɐ̃]), whereas the positive form has [h] here ([ɐ̃ɦɐ̃]). Similarly, when you say "invite someone" you might feel a slight tightening around your larynx, just before the end of "invite". That's the glottal stop. (If you say "invite", then "inside", you might feel a difference in your throat at the very end of the words, since "inside" doesn't have the glottal stop.) The symbol after the "t" indicates that the tongue tip stays up after making contact with the roof of your mouth. We call this an inaudibly released "t", or more often an unreleased "t", because you don't let the tongue immediately drop from its contact position; you don't immediately release the "t". Not all English speakers make their word-final "t" in this way, but it's quite common.

Now let's replace "someone" with "anyone": "invite anyone". In my pronunciation, I no longer make a glottal stop, and since the closure is immediately followed by a vowel, the [t] is immediately released into the first vowel of the next word. In this context, the sound is extremely short in duration, and consists only of a little tap of the tongue tip against the roof of the mouth. We transcribe this tap [ɾ]. Now we have "invi[$^{\text{ʔ}}$t$^{\neg}$]e someone" and "invi[ɾ]e anyone"; two different realizations of "invite". The

sounds are not the same, but the meaning of the word is. So these two sounds in "invi[ʔtˤ]e someone" versus "invi[ɾ]e anyone" exemplify a sound substitution in which word meaning is maintained, since both involve the meaning "invite". This is an allophonic sound substitution.

It turns out that there are very systematic changes that a sound may undergo, depending on the phonetic character of the sounds that are near to it. As we discuss in great detail in Chapter Six, the [ʔtˤ]–[ɾ] substitution is just one example of a fully regular phonological pattern in American English. Basically, words that have [ʔtˤ] in final position when a consonant immediately follows (except [ɹ] under certain circumstances), instead have [ɾ] when a vowel immediately follows (again, under certain circumstances). Now, the way I expressed the generalization about glottalization/unrelease versus tapping in English certainly seems like a rule that constrains English sound sequences. So haven't I just contradicted my earlier claim that sound substitutions are a consequence of experience with actual words, and not a consequence of internalized sound-sequencing constraints? No, I haven't. What I do as a phonologist is quite different from what I do as a learner of a language. As a phonologist, my first task is to document the sorts of sound substitutions I observe. Once I have investigated the phonetic form of many words in many contexts, my next task is to establish the correct generalizations about the patterning of the sounds: what are the systematic properties of the sound substitutions that I have documented? An efficient method of characterizing the observed systematicity is by setting up constraints and general rules on the sorts of sound substitutions and sound sequences that are found in the language. The generalizations that phonologists make about sound patterning are oftentimes breathtaking in their complexity, their scope, and their beauty, and if we are eventually going to have a good understanding of the nature of language, establishing the proper generalizations is absolutely essential. But, again—and I'm repeating this for emphasis—there is no reason to assume that the methods I employ and the generalizations I make as a phonologist are the same methods employed and generalizations made when people are actually learning and using their language. As I said earlier, we can characterize phonological systems in terms of sound substitutions, but that doesn't mean that this characterization genuinely reflects the cognitive organization of language. In contrast to the views of many linguists, I maintain that when I learned English I wasn't a "little linguist" formulating and testing hypotheses about the structural properties of my language. Instead, it is the relation between sound and meaning that is most relevant for learners.

Now, it's certainly true that language learners become aware of the sound-sequencing regularities of their language. For example, even infants have different physiological responses when they hear a rare or absent sound sequence of an ambient language in comparison to when they hear a statistically prevalent sound sequence of this language. They can use these rarities and prevalences to help predict the next sound, or, as they get older, even the next word; we touch on this in the next section. But in fact, the sorts of statistical analyses that infants may engage in are not special to language. Instead, they are the automatic response to any patterned perceptual experience. In Chapter Five we'll see that even lower animals have comparable physiological reactions to rare-versus-common patterns of stimuli, indicating that they too engage in complex statistical calculations over their perceptual experience, which is part of an evolution-derived survival-enhancing mechanism. But there is no reason to assume that the statistical

analyses that young language learners engage in assist them in determining the functional relationships among the sounds of their language, that is, the contrastive, neutralizing or allophonic consequences of specific sound substitutions in specific contexts. Learners must know word meanings in order to establish any functional relationships among sounds. After all, sounds only serve a linguistic function if they contribute to the conveyance of meaning. It is these sound-meaning relationships that have functional consequences for young learners, and it is these that surely emerge to them as they master their language.

Subsequent chapters will thoroughly explore allophonic sound substitutions of the English [ˀt˺] – [ɾ] sort. Two sounds that are related in this way are functionally the same, even though they are not physically the same. Although it may seem curious that a sound substitution may take place that seems to have no functional consequences, in Chapter Six I argue that, despite their superficial functional inertness, allophonic substitutions often evolve exactly to stave off the counter-functional consequences of neutralization.

LEARNING THE ALTERNANTS

When we hear a completely unfamiliar language, we perhaps get some sense of what speech sounds like at the very earliest stages of language learning—an unbroken jumble of sounds that has virtually no discernible structure, neither rhyme nor reason. But because language learners are exposed to a daily barrage of speech, and because speech consists of words that are used over and over again, certain sound patterns are repeated and repeated. These sequences of sounds—exactly because of this repetition—begin to emerge and be recognized amid the chaos.

In any given language, there is actually a hierarchy of frequency among sound sequences. At the bottom of this hierarchy are those sequences that are clustered across word boundaries. Some of these sound combinations might be encountered in relatively low numbers, because there are few limits on what sound may abut another, and so unusual and rare sequences may be found. But within words, there are sequences that are encountered more often, because words are repeated and repeated, and are comprised of the comparatively limited set of word-internal sound combinations that the language has evolved.

Sound combinations found within words but across morpheme boundaries are more often encountered than sound combinations across word boundaries, exactly because words are used again and again by speakers. Morphemes are the "bits of meaning" out of which words are made. So "six" ([sɪks]) has one morpheme—a root—which includes the rare morpheme-internal sequence [ks]; "sixth" ([sɪks+θ]) has two morphemes, the root "six" and the ordinal suffix "-th" (word-internal morpheme boundaries are indicated with "+"). The plural form "sixths" ([sɪks+θ+s]) has the sequence [ksθs], which is never found within the confines of a single English morpheme. Indeed, this sequence is limited to this word and this word only! The almost-too-clever-for-his-own-good indie-rocker Stephin Merritt has exploited this tongue-twisting sound sequence in his intentionally annoying band name "The 6^{ths}". The band's two albums to date are called "Wasps' Nests" and "Hyacinths and Thistles"!

The sound combinations that are encountered most often are those within morphemes. Usually, only sounds at the beginning and the end of morphemes combine in new ways with other sounds, while sound sequences that are internal to the morpheme

are usually fairly stable, because these sounds least often recombine such that they end up next to other sounds—they are typically "trapped" in a morpheme-internal context—and so are the most stable and consistent in terms of their combinatory properties.

Repetition breeds familiarity:

Least often encountered sound combinations:	across word boundaries
More often encountered sound combinations:	across morpheme boundaries
Most often encountered sound combinations:	within morphemes

As mentioned in the previous section, we have experimental evidence showing that children are differentially sensitive to the more common and less common sound sequences they encounter in the speech signal, even at the early pre-linguistic levels of infancy. However, children cannot possibly understand that the speech signal might be structured into words and morphemes until they begin to associate these particular sound sequences with particular meanings. It's exactly because certain chunks of the speech signal are semantically relevant and useful to speakers of the language that they are repeated over and over again in particular real-world situations. Because of this repetition, they are constantly encountered by learners, and eventually emerge to these learners as the functional units that they are for speakers. As they learn to associate particular sound chunks with particular meanings, learners are beginning to parse—or separate out—the functionally relevant chunks of the speech signal.

Structuring the speech signal into sentences, words, and morphemes emerges as a consequence of patterns of sounds that are heard again and again by language learners, with which they come to associate with a particular meaning, due to what we might call the richness of the stimulus. At the earliest stages of vocabulary building, the more often a particular sound sequence is encountered, the more readily that such a sound-meaning correspondence will be established. And the more often these sound sequences combine and recombine with other sound sequences, indeed, the more likely that learners will take note of these sequences' tendency to combine and recombine in various ways, and so emerge as independent functional units of the language. In this sense, learners' knowledge of the form of language is determined to a great extent by the very function that language has for speakers.

The only reasonable explanation for our effortless mastery of the inordinate complexities of the linguistic system is the aforementioned "richness of the stimulus" hypothesis. Learners are bombarded with speech at spectacularly punishing levels. The constant repetition of particular sound sequences in particular real-world contexts will induce their emergence as functional units of the language—words and morphemes. Consequently, languages that have a richer and/or more complex phonological structure should be no more difficult to acquire than languages that have simpler structures, because phonological complexity is necessarily matched by evidence for this complexity. Indeed, no language takes significantly longer to acquire by children—or is significantly more difficult to acquire by children—than any other language.

We can well imagine the early learner beginning to make sense of the speech stream, encountering patches of increasingly familiar sequences which coalesce into words and morphemes, punctuated by less familiar sequences, which help to cue word

and morpheme boundaries. During the learning process, form and function would seem intertwined to an extent that genuinely precludes their unraveling.

As amazing as this accomplishment is, it becomes even more astonishing when recalling that there isn't a one-to-one relationship between sound and meaning. Due to neutralization and allophony, sounds alternate with each other: due to the substitution of one sound with another, the same word or morpheme may possess several context-dependent realizations. The variation in sounds that is a consequence of these alternations can provide evidence to learners that words and morphemes have internal structural properties. Learners come to master all the context-dependent realizations of morphemes, such that they establish the one-to-many, and many-to-one relationships between sound and meaning that exist in the ambient system. One-to-many relations exist between sound and meaning in the form of neutralization. Many-to-one relations exist between sound and meaning in the form of allophony, although we might just as readily call this allomorphy, since we are dealing with different phonetic shapes at the level of the morpheme, not at the level of individual sounds. What I mean is that, for example, at this level of description, the pronunciations [fõ$^{\tilde{U}}$n] and [fõ$^{\tilde{U}}$ŋ]k (for "phone") are phonetically distinct though have the same meaning, and so the forms are allomorphic. As the linguist Jan Baudouin de Courtenay wrote in 1895, "Strictly speaking…alternation concerns not isolated phonemes (sounds), but entire morphemes, or even words."

So let's see how this might work with our "phone" example. (For the present, let's just ignore the complications introduced by the fact that "foam" is also a word in English.) Consider a few sentences that a child might hear in the course of a typical day:

"Someone answer the pho[n]e!"
"Where's the pho[m]e book?"
"The pho[ŋ]e cord is twisted again!"
"You missed a pho[ŋ]e call from your brother".
"Someone answer the damn pho[n]e!"
"Honey, the pho[m]e bill is overdue again".
"Will someone pick up the goddamn pho[n]e already!!"

During this typical day, a child would hear [fõ$^{\tilde{U}}$n] at the end of a sentence two times, [fõ$^{\tilde{U}}$ŋ] twice, [fõ$^{\tilde{U}}$m] twice, and [fõ$^{\tilde{U}}$n] with a following vowel once. Due to the real-world contexts in which these sentences are spoken, the child will quite rapidly come to figure out that these three phonetically distinct forms are allomorphs—that is, they are all associated with the same meaning. (They would also, of course, hear many other words which pattern similarly in this regard, thus increasing their exposure to the sound pattern.)

As children begin to make this many-to-one association between sound and meaning, they are learning that several different (albeit similar) sound sequences play a single functional role. That is, [fõ$^{\tilde{U}}$m], [fõ$^{\tilde{U}}$n], and [fõ$^{\tilde{U}}$ŋ] all mean "phone". Now the children can use these different sound sequences in their own emerging speech. They'll begin to say [fõ$^{\tilde{U}}$m], [fõ$^{\tilde{U}}$n], and [fõ$^{\tilde{U}}$ŋ] as appropriate, as a consequence of the speech patterns that they have become familiar with. At this point then, only sounds that actually

alternate with each other—the [m], [n], and [ŋ] of the various "phone"s—might emerge from the otherwise stable phonetic background: [fõ$^{\tilde{U}}$m], [fõ$^{\tilde{U}}$n], and [fõ$^{\tilde{U}}$ŋ]. By contrast, the remainder of the form "phone"—roughly, [fõ$^{\tilde{U}}$]—does not engage in alternation, and so there is no evidence to learners that these phonetic aspects of the various phonetic realizations of "phone" may be broken down into smaller, reusable bits: if learners encounter no evidence to the contrary, then [fõ$^{\tilde{U}}$] patterns as a single, unanalyzable whole, or Gestalt. We might say that sounds in alternation are foregrounded for the learner, exactly because they behave somewhat independently from the remainder of the morpheme or word with which they are affiliated: [fõ$^{\tilde{U}}$m], [fõ$^{\tilde{U}}$n], [fõ$^{\tilde{U}}$ŋ]. So alternations set some elements of the word into high relief against the stable phonetic background, and learners quite naturally and expectedly master their patterning.

A moment ago I remarked that alternations add an astonishing element of complexity to the language learning task. But now, it turns out that it is this very complexity of the pattern that assists learners in the structuring process itself. The richer the set of alternations, the more frequently learners are exposed to these alternations, the more readily they master these alternations. Here again, it is the richness of the stimulus that reveals the structural complexities to learners. As I mentioned, languages which have a richer and/or more complex set of alternations are no more difficult to acquire, and take no longer to acquire, than languages with fewer and/or simpler alternations: complexity in patterns of alternation is necessarily matched by evidence for this complexity.

It's vitally important to remember that none of this word- and morpheme-internal structuring is possible without learners assigning a meaning to the sounds that they hear. So it's only because [fõ$^{\tilde{U}}$n] and [fõ$^{\tilde{U}}$ŋ] have the same meaning that [n] and [ŋ] may emerge as alternants of each other in this context. If learners did not assign meaning to these two phonetically distinct forms, then learners would hear [fõ$^{\tilde{U}}$n] in a variety of contexts, and [fõ$^{\tilde{U}}$ŋ] in others, but they would have no evidence of the allomorphic relationship between [fõ$^{\tilde{U}}$n] and [fõ$^{\tilde{U}}$ŋ], and so would have no evidence for the allophonic relationship between these particular [n]s, and [ŋ]s. Indeed, experimental evidence suggests that learners begin to establish the functional relationships among sounds in alternation during about the tenth to twelfth month of life, which is, not coincidentally, just about when they also start to establish systematic associations between sound and meaning: learning allophonic relations is dependent upon learning allomorphic relations.

But learners also must contend with one-to-many mappings, whereby one sound shape corresponds to more than one meaning. This, of course, is neutralization, as in "phone book" and "foam book". And the same principles apply as well. Only when learners are able to pair particular sound sequences with particular meanings will it emerge that the [m] in [fõ$^{\tilde{U}}$m] may bear a functional relationship to the [n] of [fõ$^{\tilde{U}}$n] and the [ŋ] of [fõ$^{\tilde{U}}$ŋ]k, or that this [m] may be part of another word entirely, that is, "foam".

Of course, children aren't consciously aware of the generalizations that they make. The point is that it's only through vast linguistic experience—exposure to thousands of words on an everyday basis; the richness of the stimulus again—that learners come to extract the relevant patterns from the speech signal. Generalizations about the pairing of sound and meaning can only emerge through experience with an enormous number of examples.

Although primarily concerned with pairing sound and meaning, listeners' experience with repeated patterns may eventually lead to the passive emergence of statistically-derived generalizations, which may account for their ability to reproduce these patterns in novel contexts. This generalizing ability on the part of language learners might also help explain how they can come to use words that they have never encountered before. Indeed, you may have already been wondering: if our knowledge of phonology is based on our experience with the words that we hear, and not based on structural constraints, how is it that we can produce words that we have never heard before? For example, children might never have heard of a "wicket" before, but they'll know without hesitation that more than one wicket is a set of "wickets", with the plural marker consisting of [s], as in a fictional company called "World Wide Wickets". These children will also know that a fictional company called "Continental Flange" deals in some way with "flanges"—in which the plural marker consists of [ɪz̬]—without ever having heard the word "flanges" before. We clearly have the ability to group morphemes together into novel combinations, and know which allomorph to use, without ever having encountered the word previously. How do we acquire this knowledge, and what does this knowledge consist of?

The issue is far from resolved, but once again, the "richness of the stimulus" hypothesis might point the way to a satisfying answer. Based on the thousands and thousands of examples that children are exposed to, they build an inventory of forms that take a plural marker. Sometimes the plural marker is [s], sometimes it's [ɪz̬], and sometimes it might be something else. Through constant exposure to the same words in the plural form, and constant exposure to many <u>other</u> words taking one or another plural form, the plural marker is eventually set into high relief against the phonetic background of the nouns that it accompanies: bridge[ɪz̬], pocket[s], bush[ɪz̬], potato chip[s]. Once again, through repetition and variation of the plural marker, it is <u>foregrounded</u> from the sound-and-meaning background.

In the case of "wickets" and "flanges", what specific generalizations might children be making? Linguists have determined that whenever a noun ends in certain stop consonants ([ʔp,ʔt,ʔk]) the [s] form of the plural marker follows in English. (<u>Stop consonants</u> are those in which air is completely blocked from exiting the mouth.) The fact is, however, that we really have no way of knowing if children have mastered the plural form by taking note of the consonants which precede the plural marker, or by some other means. Indeed, there can be many routes to this generalization. One route might involve automatized routines of movement. For example, after having mastered a manual transmission, my driving a stick car proceeds unencumbered by reflection. I now effortlessly glide from one automatized action to the next as appropriate to the task. It's probably similar with speech. With the constant repetition of sound sequences that is characteristic of all languages, we probably develop automatized actions. So when we pluralize "wicket", we tap into our inventory of motor routines and employ the one which we have always used before. We move effortlessly from the articulatory posture for [ʔt] to the one for [s] without a moment's reflection. It would never occur to us to move from [ʔt] to [ɪz̬], since we have never engaged in that motor activity when producing a plural noun, because we have never <u>heard</u> a plural like that before.

Alternatively, children might exploit the similarity that "wicket" bears to other nouns, and pluralize "wicket" accordingly. Children know, for example, that words similar to "wicket" take [s], never [ɨz̥] in the plural, for example, "ticket[s]", planet[s]", "bucket[s]". The problem with this hypothesis that no one has ever come up with a compelling and quantifiable determination of similarity. What are the specific qualities and relations that render some objects similar, and others dissimilar? It remains an impressionistic notion for both linguists and psychologists, and so its scientific use remains elusive for the present. This is not to say that children don't exploit some extremely sophisticated diagnostic for similarity, only that similarity has not, to date, been compellingly operationalized by researchers.

SUMMARY

As speakers of a language, our effortless mastery of sound substitutions derives from our vast experience with the speech that we hear. As shown in Figure 1.2, these sound substitutions can be divided into non-alternating and alternating types. Contrastive substitutions are non-alternating; they change word meaning. Non-contrastive sound substitutions induce alternations. These alternating sound substitutions come in two varieties which result in non-bi-uniqueness between sound and meaning. Neutralizing substitutions create homophones, thus eliminating the phonetic evidence for distinctions in meaning. Allophonic substitutions maintain meaning distinctions.

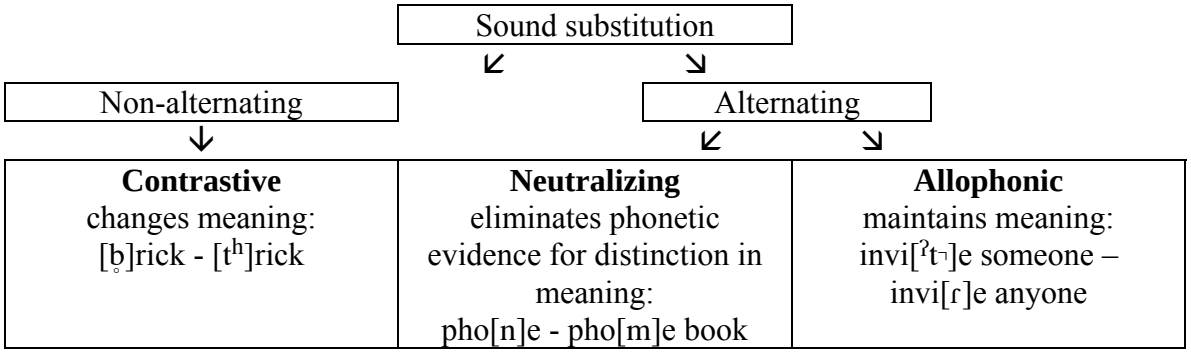


Figure 1.2. Three types of sound substitution.

Neutralizing and allophonic alternations inevitably sabotage bi-uniqueness in phonology, but is the very property of non-bi-uniqueness that may foreground the alternating sub-components of words and morphemes from the stable phonetic background, such that learners may effortlessly recycle them in novel forms.

Although one of the main jobs of phonologists is to document the regularities and systematic properties of sound substitutions, there is little evidence that language learners focus their energies similarly. Instead, learners are busy pairing sound and meaning, so that they can understand and produce the language around them. It's no wonder that speakers make mistakes about their feelings, or intuitions, about language: sometimes the same sound sequences correspond to different meanings (neutralization), and sometimes different sound sequences correspond to a single meaning (allophony). Furthermore, conscious awareness of speech sounds, and one's ability to consciously manipulate their patterning on demand, reveals nothing about the nature of phonological structure. We do

eventually become aware of the regularities of our sound systems, as evidenced by our ability to produce novel forms like “wickets” and “flanges”, but such knowledge might only emerge after thorough knowledge of words and their meanings is well in place. It seems that learning allophonic relations is dependent upon learning allomorphic relations.

While I certainly don’t deny the existence of internal mental states (as a strict Skinnerian behaviorist might), I do believe that we should proceed with extreme caution in our hypotheses about their content. In subsequent chapters I will argue that the nature of linguistic knowledge can only be indirectly ascertained, through direct inspection and documentation of linguistic behavior across communities of speakers, and across generations of speakers; that phonology is best characterized as a self-organized system of substantive social conventions which evolves passively over generations of speakers. The regularities we observe in phonological systems are due to a complex interaction of phonetic and cognitive pressures acting over generations and generations of language use, and can be understood only when considering the communicative function of language itself. Hence, to challenge the inclinations of any would-be solipsists reading this book, loquor ergo es “I speak, therefore you are”.

FURTHER READING:

The Kruszewski quote ("Language occupies…"):

Kruszewski, Mikołaj. (1881 [1995]). "On sound alternation", in Konrad Koerner, editor, Writings in General Linguistics. Amsterdam Classics in Linguistics 11. Amsterdam: John Benjamins. Pages 8-9.

Rule-based and constraint-based approaches to phonology:

Chomsky, Noam, and Morris Halle. (1968). The Sound Pattern of English. New York: Harper and Row.

Prince, Alan, and Paul Smolensky. (2004). Optimality Theory: Constraint Interaction in Generative Grammar. Oxford: Blackwell.

The Bloch and Trager quotes ("The ordinary speaker of English…;" "This uncertainty cannot be resolved…"):

Bloch, Bernard, and George L. Trager. (1942). Outline of Linguistic Analysis. Baltimore: Waverly Press. Page 40; Page 38.

The experiment with illiterate Portuguese fishermen:

Morais, Jose. (1985). "Literacy and awareness of the units of speech: implications for research on the units of perception", Linguistics 23:707-721.

The experiment with educated, literate Chinese adults:

Read, Charles, and Zhang Yun-fei, Nie Hong-yin, and Ding Bao-qing. (1997). "The ability to manipulate speech sounds depends on knowing alphabetic writing", Cognition 24:31-44.

Sapir's report on his work with John Whitney:

Sapir, Edward. (1933 [1949]). "The psychological reality of phonemes", in David G. Mandelbaum, editor, Selected Writings of Edward Sapir in Language, Culture, and Personality. Berkeley: University of California Press.

Twaddell's rejoinder to Sapir ("In so far as this incident…"):

Twaddell, William Freeman. (1935 [1957]). "On defining the phoneme", In Martin Joos, Editor, Readings in Linguistics I. Chicago: University of Chicago Press. Page 59.

Evidence for infants' sensitivity to rare versus common sound sequences:

Saffran, Jenny, Richard Aslin, and Elissa Newport. (1996). "Statistical learning by 8-month-old infants", Science 274:1926–1928.

An alternative to "the richness of the stimulus":

Chomsky, Noam. (1980). "On cognitive structures and their development: a reply to Piaget", in Massimo Piattelli-Palmarini, editor, Language and Learning: The Debate Between Jean Piaget and Noam Chomsky. London: Routledge and Kegan Paul.

Chomsky, Noam. (1980). Rules and Representations. Oxford: Basil Blackwell.

Chomsky, Noam. (1988). Language and Problems of Knowledge: The Managua Lectures. Cambridge, MA: MIT Press.

The Baudouin de Courtenay quote ("Strictly speaking…alternation concerns not isolated phonemes):

Baudouin de Courtenay, Jan. (1895 [1972]). "An attempt at a theory of phonetic alternations", in Edward Stankiewicz, editor, A Baudouin de Courtenay Reader. Bloomington: Indiana University Press. Page 175.

On learning functional relations among speech sounds:

Werker, Janet F., and Chris E. Lalonde. (1988). "Cross-language speech perception: initial capabilities and developmental change", Developmental Psychology 24.5:672-683.