1

The Functional Typology of Sound Substitution

Chapter Outline

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 reply '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 wins.

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 to give you a new word. In spoken language, it is the replacement of one *sound* with another that serves this same function. Now, while letters are intended to represent sounds, we all know that the English writing system is far from perfect in this respect and would obviously serve as a faulty guide in our exploration of the details of sound patterning; indeed, we'll soon see that the English writing system actually serves to *confound* our understanding of the genuine 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. We'll be calling these *meaning-changing sound substitutions*. 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* the phonetic cues that would otherwise provide evidence for a distinction in word meaning that existed before the substitution (thus resulting in homophony). These are *meaning-merging sound substitutions*. Third and finally, some sound substitutions take place *without* changing a word's meaning (thus maintaining heterophony). These are *meaning-maintaining sound substitutions*. Succinctly stated, sound substitutions are functionally relevant because they have meaning-related consequences; meaning is either (1) changed, (2) merged, or (3) maintained.

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 colours. Your secret code is any combination of four pegs, say, yellow – black – red – green (but you can use colours 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 rather less interesting game conceptually, because there are no constraints on what sorts of colour sequences might be used. Every guess could minimally alter the previous one by replacing one colour 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 hypotheses. Since your tests are constrained by English spelling, you can minimally alter your next guess only if the result is also an actual 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; indeed, English 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 rather different game – and, for my money, a more interesting 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 mentally active 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'? No. A far simpler and much more straightforward reason why I would never use 'bnick' as a guess in Jotto is that I simply 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'. Rather, it's simply that 'there's no English word spelled B-N-I-C-K': 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, in Mastermind, sequences of coloured pegs are

totally arbitrary to me. One sequence is as good or as bad as the next. Since sequences of coloured 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 span that can be paired with a particular *meaning* that is shared by speakers and their speech communities. For example, in English, the difference in sound between, say, 'brick' and 'trick' resides predominantly towards their beginning spans, roughly during the temporal span we take 'r' to represent (yes, you read that right: 'r', not 't' or 'b'!). Both 'brick' and 'trick' are words of English; they mean two different things for a speaker of English. Indeed, 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. We never turn 'brick' into 'bnick' simply because we never learned to pair that sound span – represented by 'bnick' – with a particular meaning, and so it's not English; it serves no linguistic function. English speakers have no experience with 'bnick', and thus such non-existent forms do not fall within the purview of phonological analysis.

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 – emphatically now – this doesn't mean that this characterization genuinely reflects the cognitive organization of the sound system itself. Indeed, individual speech sounds that we switch out might lack genuine structural status as individual elements, but they make the job of discussing phonological patterns much easier. We've just seen an example of this: whereas the spelling difference – 'brick'–'trick' – involves the switch-out of the first letter, I asserted (correctly, mind you) that the actual sound differences resides primarily during the temporal span indicated by 'r', which does *not* switch out.

Now, among non-existent sound spans in English, some certainly *sound* or *feel* better than others. For example, 'blick' seems better than the aforementioned 'bnick', even though neither is a word. So if some non-words sound or feel better or worse than others, how can I say that the only relevant distinction to be made is whether the sound span is an actual word or not? Isn't the 'word' – 'non-word' distinction a purely binary dichotomy? The answer, I maintain, is 'yes'.

When I speak English, every sound substitution results in one word or another. There are no relevant *feelings* on my part about whether the sound substitution is good or bad. 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 don't think so. I maintain that *our feelings about the words we use cannot be used as evidence for any of the structural properties of linguistic sound systems*. 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 ls 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 spans 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 a spoken instance of 'blick', which we transcribe phonetically as [b̥lɪk], with those of 'bnick', which we transcribe in the IPA (International Phonetic Alphabet) [b̥nɪk], to the inventory of English words that we have in our heads. (The hollow circle [̥] indicates that the vocal

folds are spread apart and thus does not involve vocal fold vibration.) Let's be systematic about it. Let's take each major subpart 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'. In keeping with standard linguistic practice, beginnings and ends of words are indicated by cross-hatching, '#'. Spans that are found in English words are accompanied by examples.

For b̥lɪk, there are many English words that possess each and every sub-span, and so every subpart of the form sounds familiar to an English speaker. The two exceptions are (1) the complete form b̥lɪk, which we already know isn't a word at all, and (2) words that begin with b̥lɪk although we may encounter this span in other languages, for example, the Blix (b̥lɪks).

The case of b̥nɪk is quite different. Look at the blanks in the list for b̥nɪk. These gaps indicate that in English (1) there is an overall absence of words starting with b̥n. and (2) the entire span, b̥nɪk, is nowhere to be found. So 'blick' feels okay, because every partial span of the entire form is fine in English, as is the whole form when embedded in a longer word. But 'bnick' sounds terrible, because several of these spans are never found in English (b̥n, b̥nɪ, b̥nɪk). 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 the point is that any feelings we might have about what is a good word, a possible word, or an impossible word, merely reveals the limits of our linguistic experience, and nothing more.

Table 1.1 Sub-spans of b̥lɪk and b̥nɪk

b̥lɪk	**Examples**	b̥nɪk	**Examples**
#b̥	bean, birth	#b̥	bean, birth
#b̥l	blend, blue	#b̥n	
#b̥lɪ	blimp, blister	#b̥nɪ	
#b̥lɪk		#b̥nɪk	
b̥lɪk	publication	b̥nɪk	
#b̥lɪk#		#b̥nɪk#	
b̥lɪk#	public	b̥nɪk#	
lɪ	flip, slit	nɪ	snicker, turnip
lɪk#	slick, lick	nɪk#	panic, picnic
ɪk#	sick, kick	ɪk#	sick, kick
k#	folk, lock	k#	folk, lock

Tellingly, there might be a few words that really *do* feel funny. For example, every New Yorker knows that a knish (kn̥ɪʃ) is a savoury potato or kasha pastry, but kn̥ really does feel a bit off, even to us New Yorkers, given that it's a loanword from Yiddish (which, of course, has its own sound-sequencing patterns, one of which is #kn̥). In fact, English used to have spans like #kn̥, and also the somewhat similar #g̊n, as indicated in the spelling of 'knee' and 'gnat', but they fell out of the language about 300 years ago. So nowadays, since we have so little experience with such words, they sound odd.

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 – the lips and the tongue – that function independently of each other: there's no reason that the transition from the b̥ to the n should pose any pronunciation difficulties whatsoever. I can even think of a few examples in English that have b̥n (or something very similar to it) in the middle of the word: 'Ab̥ner', 'ob̥noxious', 'ab̥normal', 'hob̥nob'. So b̥n is not more difficult to say than b̥l. Indeed, if you think 'bnick' is difficult to pronounce, it's probably because you have practically no experience in making the b̥n span at the beginning of a word, and not because of any intrinsic difficulty.

As for spans like b̥n, kn̥, and g̊n at the beginning of words, it turns out that it is hard to clearly *hear* the first consonants in these contexts, that is, without a vowel immediately preceding. Such *auditory* facts might explain why some sound spans survive and flourish over generations of speakers, while others are readily extinguished or never arise. The rarity or absence of some sound spans, and the prevalence of others, are vitally important for phonologists to take note of, and have to do with a very complex interaction, over generations of speakers, among (1) the physical properties of speech, (2) sound perception, and (3) speech production (sound, mind, and body). Indeed, much of the discussion that follows is devoted to motivating the prevalence of some sounds and sound spans in comparison to others. But the important point for now is that *our effortless mastery of English sound substitutions derives from our familiarity and experience with the English words we use.*

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 the three functional types we have already inventoried. (1) In *meaning-changing* sound substitutions, the replacement of one sound with another changes word meaning, as in the case of 'brick'–'trick'. (2) In *meaning-merging* sound substitutions, the replacement 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', the 'phone' component often comes out sounding exactly like 'foam'. (3) In *meaning-maintaining* sound substitution, the substitution of one sound with another does *not* change the meaning of the word: the meaning remains the same. For example, as discussed in detail below, 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, nor, again, are they relevant to phonological analysis in general.)

These two additional types of sound substitution – meaning-merging and meaning-maintaining – inevitably sabotage a one-to-one correspondence between sound and meaning. In other words, there exist *many-to-many correspondences between form and function*: there are *one-to-many* relations between sound and meaning (due to meaning-merging substitutions), and there are *many-to-one* relations between sound and meaning (due to meaning-maintaining substitutions). In fact, no language possesses a strict one-to-one correspondence between sound and meaning such that each sound span uniquely pairs with a single meaning, and each meaning uniquely pairs with a single sound span.

The inevitable existence of these many-to-many relationship between sound and meaning induces a remarkable complexity in sound–meaning relations that has often stymied linguists in their understanding of the relevant structural properties of language, but it never stymies children as they are learning their language.

So let's consider each of the three types of sound substitution in turn – thus establishing a *functional typology of sound substitution* – to try to get a handle on their basic attributes.

Meaning-changing sound substitution

A sound substitution can change the meaning of a word. Our Jotto example has shown this quite clearly and intuitively. We can substitute the relevant span in 'brick', to produce 'trick'. The result of this sound substitution is a change in word meaning, though 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 t̥ɹɪk can be a 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. Such sounds are thus *contrastive*, in that they and they alone may be involved in a sound substitution that changes meaning. 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. (Like χ̥ʰ also indicates that the vocal folds are not vibrating such that the immediately following sound is partially devoiced. It is used when a vowel as opposed to a consonant immediately follows, and consist of a puff of air, or *aspiration*.) 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 tʰ never precedes l at the beginning of an English word (#t̥l is surely found in certain other languages, though not too many, it turns out).

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' (that's Jotto, in a nutshell). 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, these linguists suspect that our intuitions

about speech sounds, and our ability to consciously manipulate these speech sounds, provide 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, also, our ability to consciously manipulate speech sounds provides us with no insight whatsoever about this 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 *ex*plicit 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, as already noted, quite imperfectly – represents a very brief span of the speech stream. As we master our writing system, we know that switching a letter typically results in a change of sound, and – most important – typically results in a change of 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, though, is that our ability to substitute these brief sound spans 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? Well, 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 typically stems from three broad sources: (1) the absence of a writing system for a particular

language, (2) a lack of formal education, and/or (3) some sort of learning disability. In the case of our Portuguese fisherman, 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 xwa; 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 of the 'brick'–'trick' sort, 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 supposed 'building block' – like sounds in these sorts of language tasks cannot be taken as evidence for their linguistic relevance.

Allow me to make this point in a more concrete way. In the mid-1980s 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 檸 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 simple 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 study this alphabet to jump-start their learning of Chinese characters, which take years to master, though 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 looked up a word in a Chinese-to-English 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 Roman 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 a bit 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 certainly not that I was more intelligent than they were – indeed, they were among China's best and brightest – and I obviously 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 Chinese dictionary, and that's why you can so readily understand the concept of meaning-changing 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.

Meaning-merging sound substitution

The two other types of sound substitution – meaning-merging and meaning-maintaining – are not nearly as intuitively obvious as meaning-changing substitution, exactly because they do not involve a *change* in meaning, and relatedly, they are not usually reflected in our writing system. Moreover, they are qualitatively distinct from meaning-changing sound substitutions in that

they involve *alternation*: switching out one sound with another results in a *different phonetic shape*, but one that may correspond to the *same meaning* as before the switch-out.

In a meaning-merging sound substitution, the replacement of one sound with another eliminates the phonetic distinction between or among words, resulting in homophony: two (or more) words end up being pronounced the same, and so there is no phonetic evidence for their distinction in meaning. That is, the new form indeed has the same meaning as before, but now also corresponds to another meaning; homophones are thus created. But the counter-functional effects of meaning-merging alternations 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 alternation-related homophony, consider again the compound word 'phone book', mentioned earlier. In the pre-digital age, when you were to say this in a natural way – at a natural speech rate, and in a natural conversational context like 'Where's the phone book?', it's quite possible that 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 fõn̪ – with an n (which involves the tongue-tip rising to the alveolar ridge at the upper gum; an *alveolar* consonant) – is somehow a more 'authentic' or a more 'privileged' realization of the word 'phone' than is the alternative pronunciation fõm̪, with an m (which involves the lips; a *labial* consonant). In fact, sometimes the word is pronounced fõn̪, sometimes fõm̪, and sometimes, as in 'phone call', it may be pronounced fõŋ̪ (in which the ŋ symbol indicates the 'ng' pronunciation, which involves the tongue body and the soft palate, or velum; a *velar* consonant). For fõm̪ and fõŋ̪, the final nasal sound matches the following sound (b̪ and k, respectively) in terms of tongue and lip position, but neither of these pronunciations of 'phone' is any less legitimate than fõn̪ (with n). Our intuitions might tell us that fõn̪ 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 "phoŋe", I say "phone", as in "phoŋe call, or phome book". I always say it with an "n"!' This person – quite naturally and expectedly – doesn't realize that the same word can have different pronunciations depending on its phonetic context, and so mistakenly believes that the word is always pronounced in just one way. The reason our intuitions tell us that fõn̪ is somehow more real or authentic

than fõʊ̯m or fõʊ̯ŋ is that our notion of the correct pronunciation of an English word is usually based on its pronunciation in isolation, and, if we enjoy literacy, we are influenced by how the word is spelled.

It might help if I use set diagrams to illustrate meaning-merging sound substitution. 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. (In Chapter 5 we'll investigate some of the actual complexities involved.) If different pronunciations have the same meaning, they are grouped into a set. The ambiguity of fõʊ̯m is indicated by the intersection of the two sets in Figure 1.1.

The example of 'phone' and 'foam' shows us that words which are distinct when standing alone may become homophonous when other words of certain phonetic shapes are added: the form fõʊ̯m corresponds to more than one meaning when b̥ immediately follows.

You might now grant that 'phone' has a number of perfectly acceptable pronunciations, and that one of these – fõʊ̯m – is ambiguous between 'phone' and 'foam'. But what on earth is a foam book? You might imagine that you're about to give the kids their bath, and they're crying for their favourite 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 fõʊ̯m in this phrase will be unambiguously interpreted by listeners as 'phone'. Even when the phonetic distinction among words is eliminated, the real-world and/or grammatical context of the form usually serves a disambiguating function. As I said, the functional consequences of meaning-merging sound substitutions are never very dire. Indeed, as we discuss in Chapter 3, languages seem to passively resist their pervasion.

So we've now seen that words which are distinct when standing alone may become homophones when other words of certain phonetic shapes are added. But this doesn't mean that the spelling pronunciation, or the

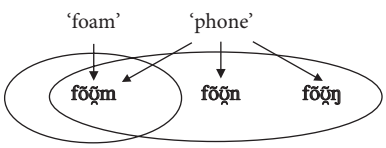

Figure 1.1 Sets for 'foam' and 'phone'.

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 homophony: in some languages, certain words in isolation are homophonous, and they only express their full distinctive status when certain sounds are added. In an extremely influential 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 **dìníh** 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 **í**, 'the one who …', this phantom 't' finally made itself heard: **dìníht^{h}í**. Suffixing **í** to the other **dìníh** ('this one') did not induce the presence of a **t^{h}**. Adding other suffixes to the words also resulted in phonetic distinctions between them. So the two words phonetically manifested their contrastive status only when suffixed but were homophonous when the suffixes were absent. This source of meaning-merging sound substitution is quite different from the English example just discussed, because in English the homophony occurs upon the *presence* of certain additional sounds, whereas in Sarcee it occurs upon the *absence* of additional sounds, such that isolated forms end up homophonous. 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 exact sense of its meaning, predates Sapir's work by a number of decades. In Chapter 7 we consider the profound influence that Sapir's proposals have had on twentieth- and twenty-first-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 psychological organization of the *sound* structure of Sarcee.

Twaddell:

> [T]he ascription of mental reality to the phoneme … fail[s] to meet the requirement of methodological feasibility, i.e. [it identifies] an entity which is inaccessible to scientific methods within the frame of linguistic study … [W]e have no right to guess about the linguistic workings of an inaccessible 'mind', and … we can secure no advantage from such guesses. The linguistic processes of the 'mind' as such are quite simply unobservable; and introspection about linguistic processes is notoriously a fire in a wooden stove. Our only information about the 'mind' is derived from the behavior of the individual whom it inhabits. To interpret that behavior in terms of 'mind' is to commit the logical fallacy of 'explaining' a fact of unknown cause by giving that unknown cause a name, and then citing the name x as the cause of the fact. 'Mind' is indeed a summation of such x's, unknown causes of human behavior […] Any correlation of phenomena which can be established on the basis of mental entities or events can also, and more economically, be established on the basis of the phenomena themselves.

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 dìníhs 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 different

dìnɨ́hs 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, and 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 dìnɨ́h with one meaning and the other dìnɨ́h with the other, rather like when we say 'phome book' without giving the ambiguity of fõ̰m 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 Sarcee sound system.

Was Twaddell right? Sapir wrote that he and Whitney were searching for a way to phonetically distinguish between the two dìnɨ́hs. Although they had to give up this endeavour 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^h 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^h popped up for only one of the meanings, and so he could tell Sapir that he 'felt a "t" '. In this scenario, then, just as Twaddell argued, Whitney was really responding to the difference in *meaning* between the two dìnɨ́hs, not to a difference in *sound*.

Whitney was mistaken in a way similar to someone who says, 'I don't say "phoŋe", I say "phone", like in "phoŋe call", or "phome book" '. Where our English speaker mistakenly thinks that all these 'phone's sound the same because they have the *same* meaning, Whitney was mistakenly setting apart phonetically identical forms that have *different* meanings. In both cases, the confused responses are a consequence of mis-pairing

sound and meaning – the result of the many-to-many relations between form and function, between sound and meaning, that are inherently characteristic of the linguistic system – and not a consequence of the supposed 'psychological reality' of proposed elements of the sound system. But Whitney wasn't the one who made the real mistake. 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 *meaning* in terms of differences in *sound*. 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 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*. Sapir, I would claim, was confusing the knowledge that he possessed as a Sarcee *linguist* with the knowledge that Whitney possessed as a Sarcee *speaker*. 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 meaning-changing 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 important, this sort of substitution, of course, 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 so-called 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 meaning-merging alternations do not play the functionally important role of fully switching one meaning for another, and further, grammatical and real-world contexts typically serve a disambiguating function, such that a speaker's intended meaning is clearly conveyed to a listener.

Meaning-maintaining sound substitution

We've now discussed how fõm̰ may be ambiguous between 'phone' and 'foam'. But what about fõŋ̰? This form doesn't mean anything in isolation, and it can only mean 'phone' in contexts like 'phone call', or 'phone card',

where the following consonant is velar. In this case, the substitution of ŋ for n neither changes word meaning nor results in homophony with any other word. Since it is neither meaning-changing nor meaning-merging, this sound substitution must be of a third (and final) type: a meaning-maintaining sound substitution.

Let's consider another example. Think again about 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 ʔt̚, which is the way I often produce it. The superscripted question mark-like symbol (' ʔ ') indicates a glottal stop, in which the vocal folds suddenly and completely shut tight, prohibiting any air from leaving the lungs to enter the vocal tract. You may have trouble recognizing the glottal stop because our orthography does not use a symbol for it, but (at least if you are a colloquial English speaker) 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 (ʌʔʌ), whereas the positive form has h here (ʌhʌ). 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 may briefly stay up after making contact with the alveolar ridge. We call this an inaudibly released 't', or more often an unreleased 't', because you don't necessarily let the tongue immediately drop from its contact position; you don't immediately release the 't', and most relevantly, you can't hear any acoustic consequences of this (potentially absent) release. Not all English speakers make their word-final 't's 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 alveolar ridge. We transcribe this tap ɾ. So now we have 'invi ʔt̚e someone' and 'inviɾe anyone': two different realizations of 'invite'. The sounds are not the same, but the meaning of the word is, and further, neither of these phonetic forms pairs with any other meanings. So these two sounds in 'invi ʔt̚e someone' versus 'inviɾe anyone' exemplify a sound substitution in which word meaning is

maintained, since both correspond solely with the meaning 'invite'. This is a meaning-maintaining 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 6, the ʔt̚ – ɾ substitution is just one example of a fully regular phonological pattern in American English. Basically, words that have ʔt̚ in final position when a consonant immediately follows, instead have ɾ when a vowel immediately follows. Now, the way I expressed the generalization about glottalization/unrelease (ʔt̚) versus tapping (ɾ) in English certainly seems like a rule or constraint that determines 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 rules or constraints? Actually, 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 rules or constraints on the sorts of sound substitutions and sound spans that are found in the language. As we'll see in later chapters, the generalizations that phonologists make about sound patterning are oftentimes breathtaking in their complexity, in their scope, and indeed, in 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. When I learned English as a toddler 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, even for those privileged few who grow up to be 'big linguists', like me.

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 span of their ambient language, in comparison to when they hear a statistically prevalent sound span of this language. They can use these rarities and prevalences to help predict the next sound, or, as they get a bit 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 5, we'll see that even lower animals have comparable 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. Still, 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 meaning-changing, meaning-merging, or meaning-maintaining 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 meaning-maintaining sound substitutions of the English ʔt̚ – ɾ sort, that is, ones that do not affect word meaning. Although it may seem curious that sound substitutions may regularly occur without inducing functional changes, in Chapter 6 I argue that, despite their superficial functional inertness, meaning-maintaining sound substitutions often evolve exactly to stave off the counter-functional consequences of meaning-merging sound substitutions.

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

spans of sound – 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 spans. At the bottom of this hierarchy are those spans that are clustered across word boundaries. Some of these sound combinations might be encountered in relatively low numbers, because at word boundaries there are few limits on what sound may abut another, and so unusual and rare spans may be found. But *within* words, there are spans 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. (*How* and *why* this limited set has evolved are questions we investigate in subsequent chapters.)

Sound combinations found across word-internal *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 span 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 span ksθs, which is never found within the confines of a single English morpheme. Indeed, this span is limited to this word and this word only! The almost-too-clever-for-his-own-good indie-rocker Stephin Merritt exploited this tongue-twisting sound span in his intentionally annoying band name 'The 6ths'; 'Every lisper's nightmare' he has said. The band's two albums were 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 spans that are internal to the morpheme tend to be more stable, because these sounds least often recombine such that they end up next to other sounds – they are typically 'trapped' in their 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 spans 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 spans with particular *meanings*. It's exactly because certain spans 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 adult speakers. As they learn to associate particular sound spans with particular meanings, learners are beginning to parse – to separate into functional units (morphemes, words, phrases) – the relevant spans of the speech signal.

The structuring of speech into morphemes, words, and phrases emerges as a consequence of those sound patterns that are heard again and again by language learners, 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 span is encountered due to its semantic usefulness, the more readily such a sound-meaning correspondence will be established. And the more often these sound spans combine and recombine with other sound spans, indeed, the more likely that learners will take note of these spans' tendency to combine and recombine in various ways, and so emerge as independent functional units of the language, that is, morphemes and words. 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' proposal. Learners are bombarded with speech at spectacularly punishing levels. The constant repetition of particular sound spans 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 spans which coalesce into morphemes and words, punctuated by less familiar spans, which help to cue word and morpheme boundaries. *During the language-learning process, form and function would seem intertwined to an extent that genuinely precludes their unravelling.*

As amazing as this accomplishment is, it becomes even more astonishing when recalling that there does not exist a one-to-one relationship between sound and meaning. Due to homophone-creating and heterophone-maintaining sound substitutions, sounds *alternate* with each other. That is, 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. Again, one-to-many relations exist between sound and meaning in the form of meaning-merging sound substitutions (thus inducing homophony); many-to-one relations exist between sound and meaning in the form of meaning-maintaining sound substitutions (thus maintaining heterophony).

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:

'Will someone answer the phone?'
'Good lord, look at this insane phome bill!'
'I need to refill my phoŋe card!'
'You missed a phoŋe call from your brother.'
'Someone answer the damn phone!'
'Will someone pick up the goddamn phone already!!'

During this typical day, a child would hear fõ̠n at the end of a sentence twice and with a following vowel once, fõ̠m once, and fõ̠ŋ twice. 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 semantically non-distinct – 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 of English.)

As children begin to make this many-to-one association between sound and meaning, they are learning that several different sound spans play a single functional role. That is, fõ̰m, fõ̰n, and fõ̰ŋ all mean 'phone'. Now the children can use these different sound spans in their own emerging speech. They'll begin to say fõ̰m, fõ̰n, and fõ̰ŋ as appropriate, as a consequence of the speech patterns with which they have become familiar. At this point then, only sounds that actually *alternate* with each other – roughly, the m, n, and ŋ of the various 'phone's, along with the latter portion of õ̰; -õ̰m, -õ̰n, -õ̰ŋ – might emerge from the otherwise stable phonetic background: fõ̰m, fõ̰n, and fõ̰ŋ. By contrast, the remainder of the form 'phone' – roughly, fõ- – 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 fo- patterns as a single, unanalysable 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: fõ̰m, fõ̰n, fõ̰ŋ. So alternations set some elements of the word into high relief against the stable phonetic background, and learners quite naturally and wholly expectedly master their patterning.

A moment ago I remarked that alternations add an astonishing element of complexity to the language-learning task, as they culminate in many-to-many relationships between form (phonetics, sound) and function (semantics, meaning). 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, and acquisition proceeds apace*.

It's vitally important to remember that none of this morpheme- and word-internal structuring is possible without learners assigning *meanings* to the sound spans that they hear. So it's only because fõ̰n and fõ̰ŋ have the same meaning that -õ̰n and -õ̰ŋ 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 fõ̰n in a variety of contexts and fõ̰ŋ in others, but they would have no evidence of the semantic relationship between fõ̰n and fõ̰ŋ, and so would have no evidence for the *phonological* relationship between these particular ̰ns and ̰ŋs. 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.

But learners also must contend with one-to-many mappings, whereby one sound shape corresponds to more than one meaning. This, of course, is meaning-merging sound substitution. And the same principles apply: only when learners are able to pair particular sound spans with particular meanings will it emerge that the -̰m in fõ̰m may bear a functional relationship to the -̰n of fõ̰n and the -̰ŋ of fõ̰ŋ – all meaning 'phone' – 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 emerge only 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 solely on our experience with the words that we hear, how is it that we can both understand and 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 ə̰z̰ – without ever having heard the word 'flanges' before. We clearly have the ability to group morphemes together into novel combinations, and know which version of the morpheme – 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 likely points 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 əz̬, and sometimes it's something else. Through constant exposure to the same words in the plural form, and constant exposure to many other 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: pockets, potato chips, bridgeəz̬, bushəz̬. Once again, through repetition and variation of the plural marker, it is foregrounded from the sound-and-meaning background.

In the case of 'wickets' and 'flanges', what specific generalizations might children be making? Linguists have determined, for example, that whenever a noun ends in certain stop consonants (ˀp, ˀt, ˀk) the s form of the plural marker follows in English. (Stop consonants 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 sounds that precede the plural marker, or by some other means. Indeed, there can be many routes to this generalization. One possibility is that children 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 əz̬, in the plural, for example 'tickets', 'planets', 'buckets'. The problem with this hypothesis is 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.

Another route to the so-called productivity of novel forms that is worthy of consideration involves 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 entrenched action to the next as appropriate to the task, even as novel driving conditions arise. It's probably similar with speech. With the constant repetition of sound spans that is characteristic of all languages, we 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 əz̥, since we have never engaged in that motor activity when producing a plural noun, because we have never *heard* a plural like that before.

Summary and conclusion

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. Meaning-changing substitutions are non-alternating; they change word meaning. The two other sorts of sound substitutions induce alternations. These alternating sound substitutions come in two varieties. Meaning-merging sound substitutions eliminate the phonetic evidence for distinctions in meaning between or among words and/or morphemes, resulting in homophony. Meaning-maintaining sound substitutions maintain the heterophonic status of the alternants.

Homophone-inducing and heterophone-maintaining sound substitutions inevitably sabotage one-to-one sound-meaning correspondences in phonology, but it is this very property of the system that may foreground the alternating subcomponents of words and morphemes from their stable phonetic backgrounds, 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 as they begin to understand and produce the language around them. It's no wonder that speakers make

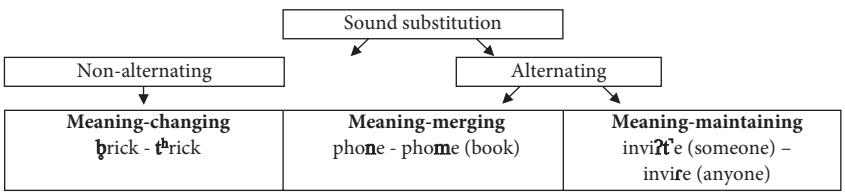

Figure 1.2 The functional typology of sound substitution.

mistakes about their feelings, or intuitions, about language: sometimes the single sound spans correspond to more than one meaning, and sometimes multiple sound spans correspond to a single meaning. 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 system, as evidenced by our ability to produce novel forms like 'wickets' and 'flanges', but such abilities might only emerge after thorough knowledge of words, their various phonetic shapes, and their meanings, is well in place.

While I certainly don't deny the existence of internal mental states (as a strict Skinnerian behaviourist 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 *behaviour* across communities of speakers, and across generations of speakers, that phonology is best characterized as a self-organized system of substantive social conventions that evolves *passively* over generations of speakers. The regularities we observe in phonological systems are due to a complex interaction of phonetic (formal, sound-based) and semantic (functional, meaning-based) 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'.

Doing phonology: Dutch

When analysing phonological patterns, two questions immediately present themselves, one involving phonetics (form, sound), and one involving semantics (function, meaning):

1. What were the historical phonetic circumstances that *triggered*, or *conditioned*, the sounds substitution that we are investigating?
2. What are the semantic consequences of the sound substitution that we are investigating?

Regarding the first question, the answer is remarkably open-ended, as there is, in theory, a virtual infinity of contingent phonetic factors that might

come to hold sway in a sound system that contribute to both its present-day (synchronic) and historical (diachronic) behaviour. Fortunately, as we'll see, most sound patterns (though certainly not all of them) lend themselves to phonetically *plausible* accounts, ones that are consistent with cross-linguistic tendencies *and*, relatedly, with facts we know to be true of the outside world (often referencing the four 'A's of articulation, acoustics, aerodynamics, and audition).

Regarding the second question, recall the sound substitution to be investigated may involve (1) a meaning-changing sound substitution, (2) a meaning-maintaining sound substitution, (3) a meaning-merging sound substitution. Also, of course, (4) the relevant forms may bear no phonological relationship to one another. It is, indeed, vitally important to consider the *meaning* that corresponds the sound patterns under scrutiny. Without knowing the semantic correlates of the forms in question, phonetic regularities in the speech stream may be detected, but there is no way for the investigator (or the language learner, of course) to establish the linguistic relevance of the sound spans under consideration; that is, whether any component of any allomorph might be present as a consequence of sound substitution, thus changing, merging, or maintaining meaning. After all, we now know that single phonetic values sometimes correspond to multiple semantic ones (as a result of meaning-merging sound substitutions), and sometimes single semantic values correspond to multiple phonetic ones (as a result of meaning-maintaining sound substitutions).

As we embark on some basic and (mostly) well-known phonology problem sets, do keep these at the forefront of your mind. I also *very strongly* recommend, at this juncture, that you read Part I of the Appendix, which focuses on both the IPA and the articulatory correlates of its symbols; in the body of the book I discuss new symbols in detail as they are introduced, whereas when we do phonology problems, I do not necessarily do the same; nor do I discuss in detail the IPA chart itself. Again, it is strongly recommended that you read Part I of the Appendix. As for these symbols' acoustic correlates, we address them most fully in Part II of the Appendix.

So let's turn to some data. Consider the two distinct underlined symbols in the forms from Dutch in (1a). We have both unsuffixed roots and their suffixed past tense counterparts.

Let's now consider the first question that any phonological analysis requires addressing: What is the phonetic component that has *triggered*, or *conditioned*, the presence of t or d during the past tense suffix's alveolar closure?

Unsuffixed:	Past tense:	Meaning:
klɑp	klɑptə	to applaud
maf	maftə	to sleep
vɪs	vistə	to fish
lax	laxtə	to laugh
krɑb	krɑbdə	to scratch
klov	klovdə	to split
raz	razdə	to rage
eɣ	leɣdə	to lay
rum	rumdə	to praise
zun	zundə	to kiss
mɛŋ	mɛŋdə	to mix
rur	rurdə	to stir
rɔl	rɔldə	to roll

Figure (1a)

First, note that these values' phonetic differences reside in their distinct voicing: for **d**, the vocal folds are vibrating, whereas for **t** they are not. Now, it turns out that, quite often, the very phonetic property that is in alternation (here, voicing) is *also* the phonetic property that has locally influenced, or conditioned, this alternation. In the case at hand, the alternation in stop voicing suggests that we begin our search for its conditioning factor in some neighbouring difference in voicing. A reliable method of investigating this proposal is to systematically plot both the sounds in alternation and their immediate contexts. Arranged in this fashion, a clear phonetic pattern emerges, especially if we consider the sounds under scrutiny in terms of their location in the IPA chart. Voiceless sounds (without vocal fold vibration) that precede the past tense marker t are indicated with circles, voiced ones (with vocal fold vibration) with triangles.

CONSONANTS (PULMONIC) © 2005 IPA

	Bilabial	Labiodental	Dental	Alveolar	Postalveolar	Retroflex	Palatal	Velar	Uvular	Pharyngeal	Glottal
Plosive	p b			t d		ʈ ɖ	c ɟ	k ɡ	q G		ʔ
Nasal	m̥ m	ɱ		n̥ n		ɳ	ɲ	ŋ̊ ŋ	N		
Trill	B			r					R		
Tap or Flap		ⱱ		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.

Figure (1b)

It appears that (voiceless) t is always preceded by a voiceless sound, and (voiced) d is always preceded by a voiced sound.

To make the suffix allomorphs' distribution perfectly clear, we 'plot the environments' in which the relevant sounds components are found. In (1b) are plotted both the preceding sounds organized by their place of oral constriction, and the consonant alternant of the tense marker itself.

	Biliabials		Labiodentals		Alveolars		Velars	
	voiceless	voiced	voiceless	voiced	voiceless	voiced	voiceless	voiced
Plosive	**pt**	**bd**					**kt**	**gd**
Nasal		**md**				**nd**		**ŋd**
fricative			**ft**	**vd**	**st**	**zd**	**xt**	**ɣd**
Approximant						**rd**		
Lateral Approximant						**ld**		

Figure (1c)

The proper generalization now emerges quite clearly: the voicing alternation observed during the suffix's oral closure indeed correlates *perfectly* with the voicing of the *immediately preceding sound*: (1) when immediately preceded by a voiced sound, voicing is also realized on the suffix-initial sound; (2) when immediately preceded by a voiceless sound, this voicelessness is also realized on the suffix-initial sound. In short, the voiceless suffix-initial stop is found only whenever a root-final voiceless sound immediately precedes, and the voiced suffix stop is found only whenever a root-final voiced sound immediately precedes: the suffix consonant agrees in voicing with the root-final consonant. We now have our answer to the aforementioned question one.

Now for the second question that every phonological investigation must address. Recall: what, if any, is the semantic (functional) relationship between the underlined symbols in tə and də? Is it (1) a meaning-changing sound substitution, (2) a meaning-maintaining sound substitution, (3) a potentially meaning-merging sound substitution, or (4) the underlined forms bear no phonological relationship to each other?

In the precompiled data we have here, the semantic correlates of the various phonetic forms are very clear: the presence of *either* t̲ə or d̲ə marks past tense. We may thus immediately eliminate possibilities (1) and (4) above. Regarding (1), whether t or d is present in this context, the meaning

remains stable, and so we are not dealing with a meaning-changing sound substitution, at least in this context. This fact automatically eliminates possibility (4) as well, since the two sounds are clearly phonologically related to one another: the sound substitution does not change meaning. This leaves us with (2) and (3): this is either a potentially meaning-maintaining sound substitution, or a meaning-merging one.

Here comes the anticlimax: based on the very limited data we are provided, we cannot establish conclusively whether the sound substitution is (2) meaning-maintaining or (3) meaning-merging. I've just ruined your day, I know. But seriously, in the data provided, whether t̲ə or d̲ə is present, past tense is reliably conveyed to listeners, and so we might think that, indeed, meaning is always maintained; end of story. But this would be a premature conclusion, since we simply don't know the forms and behaviours of *all other* morphemes in Dutch. For example, it is certainly possible that Dutch has another suffix that *never* alternates such that it is *always* realized tə. If this were the case, then, for example, klɑp+tə would mean 'applauded' on some occasions, but would mean something else related to 'applaud' on others. The alternation under scrutiny would be meaning-merging in this case, as there would be no word-internal phonetic evidence for one meaning or the other. Alternatively, klɑptə may also be a realization of another *single* morpheme, and so, again, homophony is the result. But again, we just lack the relevant data to make a solid determination here.

In sum, if any Dutch suffix displays this sort of stable behaviour, then we are dealing with (3) a meaning-merging alternation in these cases. If not, then we are dealing with (2) a meaning-maintaining alternation in all cases. We might *tentatively* stick with (2) since there are no data contradicting this proposal. Indeed, in the absence of any contradictory evidence, we are entitled to make the maximal generalization that the data allow. This isn't to say that this will turn out to be the correct conclusion, but in the context of a phonology exercise, which may be viewed as a puzzle or a game that offers a mere incomplete sketch of one small corner of a language's sound pattern, we are usually entitled to draw such conclusions, even when we know that additional data may render our conclusions incorrect.

And how do Dutch learners master this pattern? It could be as straightforward as this: perhaps early Dutch learners must *memorise* the distinct forms of each morpheme they encounter. Once a sufficient number of forms are in memory, a generalization may be made over the motor routines that subsequently recur: for the past tense marker, the tə allomorph

is used when the preceding sound is also voiceless, and the də allomorph is employed when the preceding sound is voiced. As learners encounter new forms, the grammatical and real-world information (including their memory of stored motor routines) are likely to inform them that the relevant structures involve the past tense marker. Moreover, when producing novel forms, once again, Dutch speakers need only dip in their stored inventory of motor routines associated with the past tense marker and effortlessly produce a new word.

Problem solved? Well, not quite, since, as noted in this chapter, we can only make *guesses* about the content and organization of our stored linguistic knowledge, and guessing games about mental states typically strand researchers out at sea in a notoriously leaky dinghy. Still, the proposal just sketched is a fairly straightforward and simple account of the pattern, and after all, despite *complexity of patterning*, *simplicity of explanation* usually takes priority in our guesses (educated or otherwise) about natural patterning in both phonology and elsewhere.

In sum, we may tentatively conclude that the Dutch past tense marker alternates between two forms, neither of which seems to be implicated in a meaning-merging sound substitution. For now then, let's then conclude that t̲ə and d̲ə substitute in a meaning-maintaining fashion, the alternant determined by the voicing of the immediately preceding vocal fold posture.

Further reading

Rule-based and constraint-based approaches to phonology:

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

Prince, Alan and Smolensky, Paul (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 Trager, George L. (1942). *Outline of Linguistic Analysis*. Baltimore, MD: Waverly Press, 40, 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, Yun-fei, Zhang, Hong-yin, Nie and Bao-qing, Ding (1986). '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.

The Twaddell quote ('In so far as this incident …', '[T]he ascription of mental reality …'):

Twaddell, William Freeman (1935 [1957]). 'On defining the phoneme', in Martin Joos, editor, *Readings in Linguistics I*. Chicago, IL: University of Chicago Press, 59.

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

Saffran, Jenny, Aslin, Richard and Newport, Elissa (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.

On learning functional relations between speech sounds:

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

On Dutch:

Booij, Geert (1995). *The Phonology of Dutch*. Oxford: Clarendon Press.