Chapter Two: Pure phonology

The tradition (particularly the American tradition) of scholarship in the 1930s and 1940s from which we have derived most of the content of the notion of the phoneme would not accept the line of exploration that we have considered so far. Let us put our conclusions (limited as they are) on hold for now and listen to part of the reason why this tradition, known as American structuralism, would prefer a different approach to the question of the flap in American English. We are going to explore the structuralist view in this chapter – not for historical reasons, but because the structuralists were trying to do something interesting and quite relevant to us in this day of computers. We will see, I believe, that this project can't be made to work, but figuring out just why is very instructive.

The structuralist position was shaped by the awareness that there is an element of circularity in the procedure that we have been engaged in, as we have already noticed. We relate the phones [*t*?] and [D] because we see that the same word (*let*, *kit*, etc.) is realized sometimes with the one, sometimes with the other, in a regular way; but we wouldn't *know* that there is such a thing as "the word *let*" unless we had already internalized the regular relationship between word-final [*t*?] and [D]. Isn't there another way to discover a regular relationship between these two phones in English, one which focuses only on sounds?

The answer is a conditional Yes. Yes, there are ways of analyzing the sound system of a language which allow us to discover relationships between sounds that do not depend on the kind of implicit knowledge of the language that we know that we have been using. But these ways have a number of weaknesses, and we shall ultimately decide that the weaknesses inherent in these purer approaches are so severe and so limiting that their methodological purity is not worth the price we are forced to pay if we limit ourselves to their use.

Sometimes these methods fail us by allowing us to come up with two or more *different* ways to analyze the phonemes of a language. Sometimes they fail us by not relating sounds that our earlier method (the one that tries to find a single underlying form for each word) could successfully relate. And finally, sometimes they fail because they need to make reference to something called a *phonological word*, and if we need to introduce the notion of the phonological word, we have given up a good deal of the purity of method that the structuralist program was hoping to retain.

The structuralist approach does not close the door completely to the matter of alternations, but it insists that analysis that is based on alternations must not be put in the same category as analysis that does not require any knowledge of alternations. This latter analysis would be phonology in the strict sense, while analysis that requires knowledge of alternations, or of words more generally, would be *morpho*phonology (the root *morpho-*, meaning *form* or *shape*, referring to the form of words as such).

The reader may detect -- rightly -- a sense in this approach that privileges the phonological over the morphophonological, in this sense: it's certainly possible that one and the same phenomenon might come to light by either looking at alternations, or through the structuralist means that does not require knowledge of alternations -- a distributionalist approach, as we will say. How then should the phenomenon be treated? -- should its treatment depend on how we discovered it? The answer to the last question is No; and the general point of view (which privileges phonology over morphophonology) is that if a phenomenon *can* be treated as purely phonological, then it *ought* to be so treated. We will come back to this contention later on in this book.

These notions will become clearer as we proceed. Let's step back and look at sounds in a different way.

Look at sounds in a purer way: the structuralist approach to distribution

What we must renounce now, if we are to consider this purer way to analyze a language's phonology, is knowledge of *words*. We have to imagine that we are machines of some sort, or using machines that only have access to recordings of the language's utterances, with no knowledge of words or usage.

But we *can* break utterances down into sequences of sounds, and we *can* identify various sounds as being instances of the same sound when we spot it in different utterances.[1] These finely defined classes of sounds we call *phones*. This (almost) amounts to saying that when we begin to analyze a given language or dialect, we can set up an inventory of the sounds of the language, and we can transcribe on paper any utterance with a finite set of symbols. I say "almost," because our assumption doesn't really guarantee us that there is some particular length of time after which we can be sure that we won't hear a sound that we will have to classify as a new phone -- just as biologists can't know for certain that there isn't a new species or variety waiting to be discovered in the Brazilian jungle or a deep oceanic hot spot. But in reality we encounter almost all of the sounds very soon. And soon we have a broad enough coverage of utterances in our language that we can do more than just list the phones that are there: we can start to observe logical dependencies between the appearances of the sounds. We say that we are studying the *distribution* of the various phones of the language.

Distribution

As an exercise to limber up our powers of perspective -- to allow us to take a step back from the phonological system -- it can be quite instructive to think about language as if it were a sort of random system -- a "stochastic" system, as these sorts of systems are often called. It is not our purpose to believe that this is how language works: far from it. The reason we might want to

[1] Yes, there is some cheating here in my allowing that. How can it be that the data tell you how to categorize sounds? suppose there's just a range of vowel sounds for which your computer can't see any "clustering", that is, any distribution which would lead you to see subgroupings (since no two pronunciations are every really *exactly* identical). I don't mean to minimize the importance of this question, but it's a different sort of problem from the one that we are trying to come to grips with in the text, and we have to stick to one deep problem at a time.

look at language this way is to use this view as a point of comparison against which the reality of phonology can be measured.

Let's be a bit more specific. Let's assume we have established a set of perhaps a hundred different phones in English (the number doesn't much matter; it could as easily be a thousand as a hundred, or as little as twenty-five). Let's imagine that we give a code-number to each sound, since we may have too many to conveniently give names to. We'll have S-1 ("sound 1") , S-2, and so on, up to S-100 or S-1000. We are accustomed to thinking of an utterance as consisting of a sequence of these sounds, so any utterance of English can be converted into a series of numbers: S-83 + S-44 + S-23, etc.

And if we really knew nothing at all about the language in question, other than what the whole inventory of sounds were, what would we do to decrease our ignorance and increase our knowledge? Assuming that we have access to a lot of data -- a lot of utterances or recordings of utterances -- one natural thing to do would be to calculate the relative proportions of each of the sounds. Over a total set of recordings, we would find that some sounds were vastly more common than others: *t* is much more common than *g*, for example, and ε (the vowel of *get*) is more common than *u*. And this calculation would be extremely simple: we just count up the number of occurrences of each sound in a particular recording (or our whole set of recordings), and divide by the number of sounds there are in total.

sample table:

We might see how much we'd learned about English by then using these relative proportions that we had computed to create a random pattern that might look a bit more like English than totally random strings are. That is, suppose we write a computer program to produce a sequence of sounds, subject to the condition that the probability of producing a particular sound is determined by the relative frequency of the sound as we have just computed it. We would get some strings like this:

leistsoa...

(Looking at that, we might even decide to make it look more like English by adding a "space" character, one that indicates a boundary between words, based on the actual probability of finding a word boundary at any particular moment. -- so that the example might look more like

lei st oa...)

This looks very little like English, and the main reason is because our little model so far involves no connections between one sound and the next -- though in reality there are very rich and very real connections between adjacent sounds. Some sequences of sounds are very common (t@, stI, tr, etc.), others are rare, while a very large number are rare to the point of being simply unobserved.

Let's continue imagining how a statistically-minded observer might deal with this system. In order to deal with the fact that there are *probabilistic serial dependencies* (that is, the probability that a particular sound will appear depends on the sound(s) that preceded it), we could ask the following set of questions: for each sound S, what is the probability that any particular sound will follow? That is, for each sound S_i, what is the probability that sound Sj will immediately follow? (there might be as many as 100 i's and a 100 j's, so that makes 10,000 probabilities to determine; obviously none of these probabilities will be very large in absolute terms, and most will be extremely small).

If we do some counting like this, we will establish a list much like that in Table 2. Traditionally, however, linguists have tried to look past these list of probabilities (to which we shall return in Chapter 4) to find patterns of an all-or-none character. Or to put it another way, to minimize the difference from the probabilistic approach: phonologists have traditionally focused on phenomena involving frequencies hovering right around zero. Are there certain environments, that is, where a given sound S will *never* appear, no matter how long and how hard we look?

This brings us to our first encounter with the procedures that establish phonemes on purely distributional grounds, involving phones with limited distribution.

phones with distinctly limited distribution: part 1

To begin with, if there were a sound with no limitations on its distribution, it could appear anywhere -- by definition. To put this another way, and ignoring one or two really irrelevant niceties, if an *s* could appear anywhere and no distributional limitations, then anytime we selected a string of phones in English that really was found, then we ought to be able to stick an *s* in it anywhere at all, and still have a permissible sequence of English sounds. In fact, with *s* we're as close as we're going to get to a sound with no distributional restrictions. If we take the word *cat*, and add an *s* at each position in the word (on separate blackboards, so to speak), we get *scat* (or *skat*?), *csat* (or *ksat*), *cast*, and *cats*. Of these four, three are words that exist, though the fourth, *csat*, does not, and I daresay *could* not be a word of English (though that is a strong enough statement to require further elaboration, which will have to be postponed to a later chapter). Other sounds would not do as well: consider what an *n* would produce. "*ncat"* as a single syllable is just about unpronounceable, *cnat* or *knat* is pronounceable no doubt, and English once had words starting with *k-n* sequences, such as *knee* and *knight*, but English no longer does; *cant*, of course, is a fine word, but *catn* (again, a single syllable, not a disyllabic word like *cotton*) does not conform to the pattern of sounds in English words. So of the four possible places to try out an *n* in *cat*, only one is a fine sequence of sounds (though another used to be fine). If we had tried an *m* instead of an *n*, none of the forms would have been fine (*mcat, cmat, camt, catm*).

The sound represented by the letter *h*, as in *hat* or *happy,* has rather severe restrictions on its occurrence. It usually appears only when a stressed vowel immediately follows, as in *hat* or *happy*; it never appears before another consonant, or at the end of a word (regardless of whether the next word begins with a vowel or not), and it can only appear in front of an unstressed vowel when it appears at the beginning of a word. That is, we find words like *homogenize*, with no stress on the first syllable and a leading *h*, but in words where we might expect an h before an unstressed vowels word-internally, we don't find it, as in *vehicle.*[2] The vast majority of its

[2] Article in Phonology on prefixes. Other cases to worry about: inhalator; inhibition?; Allahabad; maharajah; but good: Pan-Hellenic; rehabilitate vs. rehabituate;

occurrences are word-initial [document that!], but it certainly does appear word-internally if the following vowel is stressed, as in *vehicular*. The related word *vehicle* has two pronunciations in the dictionary, one with no stress on the second syllable (this writer's pronunciation), in which case the *h* in the orthography has no counterpart in the pronunciation, and the other with a secondary stress on the second syllable, in which case an *h* is pronounced.

Another simple example of a sound with limited distribution in English is the sound represented by the letters *ng* in *sing,* a sound called the *velar nasal* [N]. We will look at its distribution in greater detail later in this chapter, but roughly speaking, the velar nasal can appear only word-finally (as in *song*), or before a *g* or a *k;* it never appears immediately before a vowel.

We will look at a third example below, drawn from German, where voiced obstruents -- the sounds *b, d, g* --- cannot appear at the end of a word; otherwise, they have the same distribution as other consonants. While words can be spelled with these letters at the end (*Bund, ruhig, starb*), the spelling does not reflect pronunciation, for the sounds that appear word-finally in these (and all other) words are voiceless (*t*, *x, p*, resp.; explain)

As these two examples from English and another from German, suggest, we must bear in mind that when we look for sounds with limited distribution, we mean something that we may not be able to rigorously define, at least at this point: when we say that a sound S has limitations on its distribution, we normally mean limitations compared to what we would expect, given the patterns that we have already seen for sounds that are in the same class as S -- but we have not yet made explicit what "being in the same class as S" would mean rigorously. It is overly crude to say that the classes we have in mind are vowel and consonants, but we would not be too far off the mark with that, and that rough statement will have to do for now. A consonant C has limitations on its distribution if it fails to appear in places where (most, or all) other consonants appear. *h* has a limitation on its distribution, in that it cannot appear at the end of a word, although all the *other* consonants of English *can* appear there.

The simplest case is one where sound #1 only appears next to sound #2, though sound #2 can appear regardless of whether sound #1 appears next to it or not. In Japanese, for example, we find both the sounds *s* and *sh* (as we write them in English!). But the sound *sh* only appears when the vowel *i* follows, as in *sushi,* or *Hashimoto* or *watashi*. The vowel *i*, on the other hand, can perfectly well appear with many, many other sounds preceding or following it.

Phones with limited distribution: the special case of Dr. Jeckyll and Mr. Hyde

But very often this is *not* the whole story, for many cases that start off this way continue to a second chapter. For instance, if we learn that *sh* only appears in Japanese before *i,* we find out next that of all the sounds that *do* appear to the left of *i,* one sound is missing and never shows up: the sound *s*. There is a *gap,* we will say, in *s*'s distribution, and that gap coincides with the location in which *sh* appears, the *only* location in which *sh* appears. If Mr. Hyde appears only at night, and only on nights with a full moon, though nights with full moons continue to occur with the same rhythm that they have been doing for millennia; and if Dr. Jeckyll never shows up on those nights when Mr. Hyde appears -- then the gap in Dr. Jeckyll's appearance may well mean that there is a deeper relationship between Dr. Jeckyll and Mr. Hyde.

In the case of Dr. Jeckyll and Mr. Hyde, we may want to say that they are, somehow, the same *person*, and in the case of the Japanese *s* and *sh*, we may want to say that these two phones are the same *phoneme*. For this analysis to make sense (in the Jeckyll/Hyde case, just as with the Japanese *s/sh* case), we must find three conditions met:

◆there is a nearly perfect agreement between the cases where one appears and the other fails to appear;

◆the two items we are considering are reasonably similar to each other: Jeckyll and Hyde are both people (adult males, in fact), after all, just as *s* and *sh* are both consonants, and voiceless, and so forth;

◆we are not able to find any one stage on which both appear simultaneously. Here the Jeckyll-Hide-phoneme analogy is harder to state simply, but the logic remains parallel. If Jeckyll and Hyde were to present themselves in the same room at the same time, we would have to give up our theory that they were really the same person. Similarly, if both *s* and *sh* were to appear in front of the same vowel to form two different words, we would have to give up our theory that *s* and *sh* were two phones of the same phoneme -- crudely put.

That last point was too crudely put, though, and we need to be more careful at this point. What phonologists do when trying to decide if two phones (phone X and phone Y) are phones of the same phoneme is this: they look for pairs of words that are exactly identical, as far as sounds are concerned, except that one of the words has phone X where the other has phone Y. If they find such pairs (called *minimal pairs*), then the phones are *not* phones of the same phoneme. In the Japanese case, this would means finding two words, like *sin* and *shin*: but pairs like that are never found in Japanese, because *s* and *sh* are in fact phones of the same phoneme! If we *did* find such a pair, a minimal pair, we would be able to say that these two sounds *contrast* in the language in the context that the minimal pair embodies. We will return to the notion of *contrast* below.

Let's step back and reflect on what we have done. We've done something that at first seems radically different from our earlier analysis involving *let* and the flap [D]. Now we look only at phones, and we look at their distributions in our recordings. This is fine, up to a point. But there was a moment when even this pure study of distribution slips back into using the (impure!) knowledge of words of the language that we banned when we started off this section. We slipped back when we say that finding -- or rather, *not* finding -- minimal pairs could be part of the way we decide whether two phones are part of the same phoneme. That's precisely the kind of knowledge, the higher-level knowledge of vocabulary, that we decided we wanted to avoid. In our earlier strategy, we were looking for an explanation of how it was that the same word shows up in two different phonetic forms (let?] and [leD]); now we're saying that if we hear two phonetic forms, *sin* and *shin*, we know that these are (or aren't) two different words in the language in question.

The purist could respond this way: OK, you're right. If you want to remain pure, just use the first two purely distributional criteria. The minimal pair test is just something you can use on the side if you want to as a way to help you formulate your hunches and your guesses.

And that's not a bad answer. It does show that it's very tricky to set up a set of basic principles for phonology that are pure, so to speak, in avoiding antecedent knowledge of the vocabulary.

The pure study of sounds, then, could begin by concentrating on the distribution of various sounds, and gaps in the distribution of various sounds; when there was a close fit between the distribution of one sound and a gap in the distribution of another sound, we would have a strong candidate for a phoneme. If Sound #1 is found only in Context A, and Sound #2 is widely found *except* in Context A, we could consider setting up a phoneme P, where P is realized as Sound #1 in Context A, and as Sound #2 elsewhere.

Some more examples

Buy, bide, and bite

In these three words -- *buy*, *bide*, and *bite* -- the vowel (though spelled differently in the first word) seems to be much the same, a vowel that was called by our elementary school teachers a "long i". But if we listen carefully, we will notice that the vowel in *bite* is actually different from the vowel in *buy* and *bide*. It is a good deal shorter, as far as duration is concerned, but in addition, its actual vowel quality is somewhat different. Both the vowels in *bide* and *bite* are *diphthongs*, that is, sounds that glide from one vowel quality to another. But the beginning of the glide is different in the two cases. In *buy* and *bide*, the gliding is from a vowel much like that of *ma* or *cod* (written [a]) to a vowel-like sound much like that of *see* (written [i]); a naive transcriber might write this as "aah -- ee", but the phonologist writes this [ay], using the symbol *y* to indicate a sound much like [i], but used as the final part of a diphthong (used as an *offglide*, we say).

Thus *bide* and *buy* have the diphthong [ay], but *bite* has a different diphthong, one which begins with the vowel that we might naively write as "uh" (though phonologists write it: ^), the vowel of *cut*. In *bite*, the vowel is a diphthong that might again naively be transcribed as "uh - ee", but which phonologists write [^y].

I daresay it comes as a surprise to most speakers that the vowels of *bide* and *bite* are as different as they are (though I have heard them as different for so long that I have only a faint recollection of the surprise of learning that they are different -- or even that they are composed of diphthongs, gliding combination of two vowelish sounds).

With some perseverance and practice, the reader should be able to take the word *bite,* and pronounce it without the final *t*, without turning it into *buy*. That is, with a bit of practice, you ought to be able to say [b^y], and to hear the striking difference between that sound and the sound of the real word *buy*, pronounced [bay]. Of course [b^y] isn't an English word, nor even a *possible* English word, and to say it, you must wrap your mouth around sequences of sounds in a way that it has never had to before. (But that's the life of the phonologist.)

If we explore various combinations of consonants standing on either side of the vowels [ay] and [^y], to see which combinations of sounds do make English words and which do not, we get results much as in (3).

(3) [ay] and [^y]

[ay] bide, guide, tribe, aisle, smile, mime, mine, buy, try

[^y]	bite, kite, tripe, type, like, Mike, knife, nice, slice

The generalization that emerges out of the data sketched in (3) is twofold. First of all, it is the sound to the right of the diphthong that is critical, and second, the consonants that follow [^y] are all limited to a very specific set of sounds, voiceless ones. [more on what a voiceless sound is]. On the other hand, the diphthong [ay] may be followed either by a voiced sound -- or by nothing at all. In a situation like this, the phonologist is strongly inclined to see a qualitative difference between these two statements of the right "context" (as it is called). To the right of the [^uy] will be found a voiceless consonant: fine. But that the right-hand context for the [ay] should be *either* a voiced sound *or* nothing strikes the phonologist as odd, or rather, as a context better understood as really meaning "all contexts other than the one described for that other sound just mentioned". (Actually, the situation is sharper than I have described it, because the set of sounds I have referred to as "voiced sounds" includes three different groups of sounds: vowels; voiced obstruents like *b, d*, or *g*; and also sonorants like *l,r, m*, and *n*. While it is true that the three can be summarized neatly as "voiced sounds" -- for they are all voiced -- many phonologists, myself included, would suspect that these sounds do *not* form a natural class in this case. ?more on this in a later chapter?)

We can conclude that in a dialect of English such as the one we have just looked at, [ay] and [^y] are phones pertaining to a single phoneme, with the [^y] used before voiceless sounds, and [ay] used elsewhere.

Front low vowels in American English

Let us take a look at another case of distribution in American English. Most dialects of American English a peculiar vowel in the word *can't* , a vowel that strikes many people (when they sit back and reflect upon it) as a rather nasal and whiney vowel. We're not here to pass judgment on the aesthetics of the vowel; we want to observe it, identify it, and see what its distribution is. There is enormous variation, however, among American dialects regarding its distribution. For many Chicago speakers, for example, the vowel can appear virtually anywhere where a vowel can appear, while for many speakers from New York (like this writer) the restrictions are rather tight. I cannot offer the reader a range of data for each major region in the United States; let us restrict our attention to the data as it appears in my speech.

First, a symbol: we will refer to the vowel in *can't* with the symbol [A]. We find it also in the word *tan*; thus [tAn]. But if we look at some other words that are quite similar, like *tap*, *tat*, and *tack*, we find a different vowel, one which is a bit shorter, and "lower" (as phonologists say). The reader should do his or her best to attend to the difference in sound between the vowels of *tat* or *tap*, on the one hand, and *tan* or *can't* on the other. The former vowel is represented [ae], while the latter is the [A] we have spoken of. In (1) I have given a list of monosyllabic words divided in two groups, those with the vowel [ae] and those with [A], in my speech. As I noted, some speakers may find that they have the vowel [A] in all of these cases, but most Americans will agree with most of these judgments.

(1) a.	ae	cap	cat	pack
		nap	Nat	knack

	gap	bat	back
	bang	sang	gang
b. A	Sam	tan	
	Nam	Nan	
	dam	Dan	

As we look at more and more monosyllabic words, we find that it is the consonant that follows the vowel that influences whether [ae] or [A] may appear, and never the preceding consonant. If an *m* or *n* follows, then we never find [ae], though we can find [A], as the examples in (1b) illustrate. Now, *m* and *n* are consonants that are called *nasals*, because they are characteristically produced by closing the passage of air through the mouth but opening the velum so that air can flow freely out the lungs and through the nose. The other nasal consonant in English is the angma [N], the velar nasal, spelled usually *ng* (or *n* before *k*) in English. Yet before an angma, we don't find what we find before *m* and *n*, oddly enough: before *ng,* we do not find [A], but we do find [ae].

(1a) shows some examples of [ae] before angma, and also [ae] before *p, t*, and *k*, the sounds that we call *voiceless stops*. They are stops because the flow of air out the lungs is completely blocked off for a brief period of time when we produced a stop; *p*, *t*, and *k* are considered *voiceless* because their production tends to inhibit "voicing," that is, the rapid pulsing of the vocal chords that gives rise to what is virtually the essence of speaking out loud. A voice that has no voicing is a voice that is whispering; if you whisper out loud for a moment, and then stop whispering, the loud robustness that reemerges in your voice is the vibration or voicing of the vocal chords. We will return to this in the next chapter when we discuss vowels and syllables, but for now we may content ourselves with the knowledge that *p*, *t*, and *k* are considered voiceless not only because there is no voicing during their production, but because (if a vowel immediately follows) voicing is inhibited even after these voiceless stops for a relatively long period of time -- about 90 milliseconds, almost a tenth of a second. They have a major influence on the length of the vowel that precedes them, too: vowels that precede a voiceless consonant are much, much shorter than consonants that precede a voiced consonant; the difference in the length of the vowel in *tap* and *tab* is easily perceptible if one pays attention. And before these voiceless consonants, we do not find [A], though we do find [ae].

On these distributional grounds, therefore, we are led to suggest that there is a relationship between [A] and [ae]; we may hypothesize that these two phones, or sounds, are realizations of the same phoneme, because [ae] is not found before [m] and [n] -- in monosyllables, which is all we've considered so far -- while [A] is found there; and conversely, in front of voiceless stops, [ae] is found while [A] is not.

As we extend our observations further, looking at monosyllables, and checking for the presence of [ae] or [A], we can report that when a voiced stop follows, [ae] is possible and [A] is not, as the data in (2) report.

(2) [ae], not [A]	nab	cab
	bad	mad
	bag	tag

When I attend to my speech fairly carefully, though, I find that my behavior is not clear-cut as (2) suggests; in words like *cab* and *mad*, the vowel [A] does sometimes creep in; and when I look at my pronunciation in front of the fricatives in (3), words like jazz, calf, ass, the choice of [ae] or [A] seems to float freely back and forth. This takes us to the matter of free variation.

Free Variation

The second part of the traditional method for establishing phonemes in what we have called a pure way involves some very sticky and messy methodological points. This involves the decision as to when two sounds are close enough, as far as the language is concerned, to be treated the same way.

Now, it is certainly a fact that each language we look at permits a range of possible ways to produce each of its sounds, and more particularly, it is frequently the case that in one and the same context, two or more versions of a sound will be permitted. I remarked just above that in cab and mad, either [ae] or [A] could appear, apparently freely varying.

Looking back at the case we started with -- the production of a *t* as a flap [D] or some other sort of *t* -- we do find such variation in one particular context. In my speech, a t that is found between unstressed vowels inside a word may be pronounced either as a flap [D], or as a "normal" [t]: not glottalized like the [t?], but instead much the same as the word-initial *t* of *tea* or *telephone*. Words that end in -ity are good examples of this: I can equally well produce the pronunciation *sani[D]y* or *sani[t]y*. In fact, if we look across a wide sample of my speech, or that of many other Americans, we will find that [t] and [D] both appear in large numbers in the context:

(1) [t] and [D] in free variation in this context: unstressed __ unstressed V

[explanation of that notation]

To take another example: in virtually all dialects of Spanish, the sound represented by the letters *b* or *v*, when found in-between vowels, is what is called a voiced bilabial fricative, and symbolized with a [β]. This is a sound that we don't have in English; it is produced with the two lips nearly touching, though otherwise it is much like a *v* in English. Most dialects of Spanish don't have the sound *v* at all. But in the speech of many Chileans, [v] does indeed appear, in free variation with the [b].

When a language does not care about the difference between two particular sounds in a particular environment, we ultimately want to be able to analyze these sounds as being reflections of a single phoneme: that is a big part of the motivation of the notion of the phoneme in the first place (indeed, the notion of the phoneme is strongly rooted in the tradition of work meant to help in the design of spelling systems for languages that had never been written down before. How many different letters does a given language need? We have to check it out, and see which of all the sounds are really different, and which are just minor and unimportant differences as far as the language is concerned.)

We could try to summarize this second step in our modeling of phonemes like this: If two phones are used indifferently in a given phonological context -- indifferently in the sense that their different pronunciations does not and cannot be used to mark two different words -- then in that context, we will say that they are the realizations of the same phoneme.

There are two major problems that we have just stumbled onto, however. The first is this: how can we allow ourselves to decide whether two pronunciations are "the same word" -- like *sanity* with a flap [D] and *sanity* with a true released [t] -- if we thought we were engaged in some kind of pure phonological enterprise that was not going to engage in any of the slippery "same word" business that we were involved in Chapter One? We began this chapter by saying that the structuralists would not allow us to do phonology by picking a single word and then trying to figure out how that word's pronunciation changed in different contexts, since that enterprise presupposed that you already knew how to identify a word as such, as that particular word -- an assumption that begs the question. Yet here we need to make much the same assumption in order to decide whether two pronunciations (both of which are found in our recordings of a particular language) represent the same word (and hence the sounds are variants of the same phoneme) or are really two different words (hence the phones are representatives of distinct phonemes).

That's a serious problem, but there's another one that's every bit as serious. It's this: if we are going to give statements like the one above in (1), we will frequently have to put a condition on them to the effect that they hold only *inside of words*. The principle in (1) in fact only holds inside of words. If a t is found between two unstressed vowels, with all of those sounds inside a single word, then the *t* may in fact freely be pronounced as a flap or as a released [t]. But we've already seen that if the first unstressed vowel and the *t* form the end of a word, and the following unstressed vowel starts a new word -- as in examples like *Planet of the Apes* or xx -- then the normal American pronunciation is as a flap; or, if a pause of some sort is introduced after *Planet*, the *t* can be glottalized; but a true released [t] is unnatural in standard American pronunciation here. On the other hand, if the *t* and the following unstressed vowel marks the beginning of a word -- as in a phrase like *the tomato* -- the flap is quite unusual, to the point of being nearly impossible (recall the discussion in the last chapter). We shall see in the next section that this is just the tip of the iceberg: our dependence on finding ends of words will turn out to be considerable: we might even say overwhelming.

Now, this second problem doesn't rest upon needing to be able to identify a certain set of sounds as "the same word" in various utterances, and so in a sense it seems to be a less serious problem for the purity of the phonemicist's strategy that we are exploring in this chapter. It seems that we only need to be able to break up an utterance into words, though we don't need to be able to say *which* word is *which*.

We might formulate the task (or the problem, to be pessimistic about it) like this: we need to be able to define, or at least recognize, the *phonological words* of any given utterance. If we could do that, we'd undoubtedly expect that in most cases, the stretch of sounds that was a phonological word in this sense was usually a word in the normal sense, the sense we began discussing in Chapter One. The two notions might not *always* coincide, but we'd expect that most of the time they would.

Summarizing

Let us recapitulate. A pure study of the distribution of sounds allows us to find cases where two sounds behave in a complementary fashion, like Dr. Jeckyll and Mr. Hyde (or Superman and

Clark Kent, if you prefer): each sound has a set of environments in which it can be found, and these two environments do not overlap; there is no context in which both sounds could occur. In that case, we may assign these two sounds to the same phoneme.

The second step we took was to say that if two sounds have a set of contexts in which both occur, and yet in occurring they are merely different pronunciations of the same word, then we will again permit ourselves to say that the two sounds are realizations of the same phoneme. Unfortunately, much of our methodological purism is endangered with this step, however.

And we have (though only in sketchy fashion) seen that the notion of "context" may require reference to where words begin and end -- so that when we define a particular phonological context we can be clear about where words begin and end.

Let us look at that point in greater detail, because it is important, and our case of the American English flap is a bit complex, and does not provide the sharpest illustration of this point.

Limited Distribution, Part 2

Earlier in this chapter we began our discussion of a distribution-based account of phonology with a consideration of sounds which have a limited distribution, but which seem to have a Mr. Jeckyll to their Dr. Hyde -- that is, that have a cousin whose limited distribution is complementary to their own. We are able to put these sounds together and see them as forming a whole phoneme with a less limited distribution.

But limited distribution is by no means always like this. German provides a very straightforward example of a sound system in which there is a limitation on where certain consonants can appear, without the possibility of finding alternate sounds with complementary distributions.

The obstruents in German are: p, t, k, ts, b, d, g, pf, h... All of these obstruents can appear before a vowel, but word-finally (and in the speech of many Northern Germans, syllable-finally) only the voiceless ones (p, t, k, ts, though not h) can appear. Despite what the standard spelling suggests, we do not find words ending with a [b]: the word "died" is spelled "starb", but it is pronounced [shtarp]. It is not hard to find words whose spelling ends with any of these "forbidden" voiced sounds (*Bund, starb, lebendig*, etc.), but in modern German these sounds are not voiced (though at an earlier time in the language, these sounds were indeed voiced); they are all voiceless.

But there are no other sounds that appear only where voiceless sounds do *not* appear -- no sounds in complementary distribution with the voiceless ones. So there is nothing to be done about collapsing these sounds together with others to make phonemes.

We can conclude, then, that German has the characteristic that we might call word-final *voicelessness,* and this feature holds with considerable generality and exceptionlessness. [FN, Rich J's counterexamples] But this very exceptionlessness does push us to ask the nagging methodological question: is knowing where a word ends knowledge that is purely phonological, or does it require knowledge of words -- the kind of knowledge that the purity of distributional analysis was trying to root out of purely phonological analyses (leaving the rest of the less pure analysis to that related field to be known as morphophonemics)?

We face the horns of a trilemma, and we have to choose one of three paths:

a. We can conclude that the desire to have a "pure" methodology for phonology, one that obliges us to do phonology with no knowledge of word-structure, was a mistake in the first place, and there is no distinction between "pure" phonology, on the one hand, and morphophonology on the other.

b. The distinction still holds, but examples like word-final voicelessness in German are not phonological (alas); pure phonology may just be smaller (in terms of the range of phenomena that it ought to handle) than we probably expected it to be.

c. Our effort to purify the methodology of phonology should not forbid us to make reference to notions like "word-final", despite the fact that they seem to require knowledge of words, an affront to our sense of methodological purity. There must be a way to distinguish between crude knowledge of where word-boundaries are, on the one hand, and knowledge of particular lexemes or lexical words, on the other. Perhaps (i) we can finesse this problem -- overcome it somehow -- by observing that most phenomena associated with the end of a word also take place at the end of an utterance (since utterances typically end with completed words!). Perhaps German voicelessness was originally a fact about sounds at the ends of *utterances* (they're always voiceless in German), and by some sort of analogy, when the same words are pronounced when they're *not* utterance-final, that voicelessness carries over. [FN: Or perhaps (ii) there really is an indication in the signal which we have not transcribed in German, one which marks word-final syllables as different from others -- perhaps in the measurable duration, which we have ignored in our transcriptions].

Fifty years ago, when this trilemma was first posed, no one was willing or interested in choosing alternative a, and hence abandoning phonemic analysis; phonologists had to choose between alternatives (b) and (c), and of course they offered arguments as to why the one or the other was the better way to go. By the early 1960s, the voice of generative phonology (proposed by Noam Chomsky and Morris Halle) became widely influential, and it argued that neither (b) nor (c) were acceptable, and the only recourse was (a): abandon the program of methodological purity, and at its crudest, that is the view held by most phonologists in the anglophone world today.

But that's not the end of the story -- if it were, there would not have been much point in retracing these steps in the detail that we have. Phonologists are still looking for ways to classify the degree of phonological purity of each part of the analysis that they offer of a language, and this remains an extremely important part of virtually all work done today, though often this remains only implicit in the work. What this means in practice is that while phonologists reject the notion that knowledge of words is inappropriate for phonological analysis, they try to divide up their analysis into separate components. Phenomena that require absolutely no knowledge of words may fall into one large heading; phenomena that require knowledge of where words begin and end, but no knowledge of which *particular* word is involved, will fall into another heading; and phenomena that are indeed sensitive to knowledge of the particularities of lexical items may well fall into a third category.

The case of German word-final voicelessness is a very simple example, but one that reveals a lot about the consequences of the assumptions that we make. We must choose one of the three choices:

a. We abandon methodological purity, and so we allow ourselves to make reference to what word is involved when we do a phonological analysis. When we analyze the word [shtarp] in German, which means " s/he died", we can bring into the picture the fact that the infinitive of this verb is *sterben* [shterb@n], with a [b], and we can use that fact to argue for an underlying representation in which the consonant in question really is a *b*, adding to this a principle that says that all word-final obstruents in German are voiceless, thus forcing the underlying *b* to become a [p] -- more about what this sort of "forcing" might be to come, of course. This has been the modern position for several decades.

b. If we were to stick to the purest position, then we would be able to say no more than that there was indeed limited distribution of consonants in word-final position. This has led some phonologists to assign a special status to the sounds that *do* appear in these positions. [add more, perhaps; perhaps not]

c. The third and final possibility would be to say that a phonological analysis will not need to provide a *unique* result. The methodologically purest component of a phonology may make reference to where words begin and end, but will not have access to which particular lexeme is being analyzed at a particular moment. So the analysis of a German word like [starp] will recognize that the final [p] appears in a position in which no voiced obstruents ever appear, despite the fact that the language possesses a healthy supply of voiced obstruents; the phonological analysis will then allow for a non-unique solution to the analysis: the phonetic [shtarp] would have *two* phonological analyses, /shtarp/ and /shtarb/, and only a knowledge of the inventory of lexemes of German, and perhaps a knowledge of the subject of the utterance being analyzed, will allow us to refine the analysis any further.

As I noted above, the first approach (a) is the view of most phonologists; (b) is at this point the view of few or no phonologists; and (c) is the approach taken in effect by many computational approaches to language. Indeed, as has long been noted, there are two quite different styles to doing phonology (and these styles become formulated as theories). In one style, we begin with a stock of observations, and we apply our methods to those observations, gradually building up a set of conclusions, so that after a while, we have produced a set of generalizations that allows us to neatly and cleanly analyze the next piece of data that comes along from the same language. This is the perspective taken by the structuralists, and it contrasts with the point of view taken generativists, whose goal in effect is to figure out what is going on the head of a speaker, looking at data as a source of relevant information in two ways: the data can be looked at from the point of the naive language learner, who is soaking in the data just as a linguist might be, or as the structuralists' methods might be doing. But the generativist also looks at data as the telltale product of a mind (so to speak) that already has a phonology of the language installed in the brain. If I know that *you* know English (or Swahili), and if I am interested in what *you* know, and I know that *you* know which word you're saying when you say it, then what is the point in my producing a whole analysis based on the principle that phonology must absolutely not make reference to knowledge of particular words?

Of course that seems very plausible, and the argument goes back and forth; the reply is, That's fine as a model of the *speaker*, but phonology must deal equally well with the hearer, and the hearer doesn't know what words have been spoken until he has done *all* the analysis of the sounds -- so there must be a way to do phonology *before* you've figured out what word it is that you are listening to. And that is pretty plausible too, as a response, and it is one that offers considerable plausibility to the third position that I mentioned above, the one implicit in much computational work: phonology is a "device" -- a chunk of systematic knowledge -- that allows a hearer to decode a phonetic signal, though not without leaving considerable ambiguity in many cases.

This will be the final subject of this chapter: the "ambiguity" that may arise in a phonological analysis when it is viewed as a model of what the hearer is doing. In a model of phonology that allows the analyst to make reference to word boundaries -....

Section to be moved to Chapter Five (Word structure)?

If there are such things as words, we should be able to unambiguously count them when we find them. "The bell rang." That's three words. What about "The doorbell rang." Is that three, four, or five words: how do we count "doorbell"? "Doorbell" is in the dictionary, and its meaning is not completely specifiable from the meaning of its parts: surely *doorbell* is a word. But *door* is a word, as is *bell,* and a *doorbell* has something to do with *doors* and *bells*. Should we say that we're looking at three words, then? After all, it's not just a funny accident that *doorbell* has subparts; it's not like, say, *chocolate,* which has inside it the word *chalk* (at least from the point of view of the sounds involved); or *satellite,* which no adult would think is made up of *saddle* + *light*. But what about words like *human* and *woman*? Do they contain the word *man* inside them? Does the word *female* contain the word *male* inside it? If we check the dictionary, we find that *woman* was indeed a compound formed from *man,* but *human* was not (it's ultimately from Latin), and *male* and *female* are unrelated historically (as a knowledge of French will confirm, which gave us these words: *male* and *femelle*, respectively).

So *doorbell* is what we call a compound; it is a word, yes, but a word made up of two words. And, as a handful of examples show, there is a terribly fuzzy line between compound words and words that are not compounds. We will come back to this, and worry about very tough cases like *Bloomington* and *Hingham*: compounds or not? But for now, it would be best if we reached a simple conclusion: in *some* cases, at least, two words can be combined to form a third, which we call a compound:

(6) doorbell

door bell

So when we see the word *doorbell*, we would have to say that there are three words there: *door*, *bell*, and *doorbell*. {FN It might be helpful to compare this state of affairs with grammatical matters. If we consider the sentence *Mars is rising and Saturn is setting*, , how many sentences are there? Surely there are at least two (*Mars is rising* and *Saturn is setting*), and there's no doubt a third: the whole sentence.)

Are the reasons for this decision phonological, however? And does this answer -- that there are three words in *doorbell* -- tell us that there are three *phonological* words there?

Our reasons so far for deciding whether something is a word have largely *not* been phonological; they involve some crude judgment regarding what is the smallest meaningful unit in a few cases. If there is to be a notion of a phonological word, it should have two characteristics: it must roughly coincide with other notions of word (like this roughly "semiotic" notion of word that we have just been considering) and it must involve phonological motivation. What might that be?

Here is a list of some of the reasons we might be able to offer to justify setting up such a thing as a phonological word:

(i) There could be restrictions on what phones appear at the beginning or at the end of a phonological word.

(ii) There could be possibilities of phones, or combinations of phones, that may appear at the beginning or the end of a phonological word, but not internal to it.

(iii) There might be restrictions on what sequences can occur inside of words (though these restrictions would not affect sequences of phones that stretch across from one word to another)

(iv) There might be restrictions -- a lower limit, typically -- on the number of syllables in a phonological word.

All of these are found, and found in most languages, and several other related effects arise, testifying to the utility of the notion of a phonological word. Let's consider a few simple examples from English.

(a) No identical consonants in sequence inside a word

While we frequently spell words using two identical, adjacent letters (as in *spell* or *letters*), we do not pronounce double letters. Many languages do: Italian may be the most familiar (if you know how a word like *bello* or *fatto* sounds like in Italian -- the doubled consonant is stretched out and long); Japanese likewise. English permits such pairings of identical consonants just so long as they come from separate words. *Nick can never write ten new words so fast*. may not make much sense, but it is a perfectly fine English sentence, and if you listen closely, you'll hear doubled consonants at virtually every transition from one word to another. Or listen to the difference between *May king* and *make king*. The latter has a doubled k in its pronunciation. But we won't find any such sound inside a word in English -- and that's why the doubling of a consonant can be used (as it is) to indicate that a preceding vowel is short, or occasionally, to mark that it's stressed.

(b) No schwa-vowel sequences inside a word

We have in English a very important little vowel, a "reduced" vowel that is called *schwa*. This vowel has no special spelling associated with it: many vowel-ish letters are used to represent this colorless vowel. It is the vowel of the second syllable of *sofa* or *human*; it is the vowel of the unstressed articles *a* and *the* when words that begin with consonants follow (*a pill, the book*). This vowel can be followed by any vowel in the language provided we place two appropriate words together. For example, schwa can be followed by the vowel of *is* (denoted by phonologists as I) if we construct a sentence such as *The sofa is getting musty*. But this sequence never appears inside a word.

Now, there are some apparent exceptions to the two generalizations I have just made, regarding doubled consonants and schwa-vowel sequences word-internally. I don't want to give them away quite yet, but a lucky or observant reader may have thought of one. Good for you.

But I don't have compounds in mind. Compounds do indeed provide an opportunity for finding sequences that would not normally be found in a single word, like doubled consonants, as in *hard-drive, hat-trick*, or . And we can find sequences of the schwa+vowel sort that wouldn't be found word-internally, as in *an idea-inventor* (a compound that I have just made up), or XX.

Let's go back to our list of ways in which it can be useful to define such a thing as a phonological word in English (or any other language). There can be restrictions on what phones appear at the beginning or end of a word, or conversely, there may be segments that only appear at the beginning or end of a word. The sound represented by the letters *ng* in a word like *song* or *ring* is an example of such a sound. This sound (called angma, and represented by phonologists thusly:) is a nasal consonant produced with the back portion of the tongue: it is not composed of an *n* and a *g*, as the spelling might suggest (just as th is not composed of a *t* plus an *h*, normally) The angma is restricted in English to appearing in only two places: first, it can appear immediately before *k* or *g*, as in *finger* or *sink*.. Notice that in *finger*, I have said that the angma -- a sound -- appears before a *g*-sound: bear in mind that I am talking about pronunciation, not spelling, and *finger* is pronounced [fINg@r]. The second place where an angma can appear in English is at the end of a word, as in *sing* or *strong*. Some analyses of English propose that this second case is really a special case of the first, by saying that words like *strong* really have a *g* at the end, but that the g is not pronounced at the end of a word after an angma. On this analysis, *strong*'s underlying representation would be [stroNg]. This may sound like a bit of circularity: we postulate a *g* underlyingly (that is, for the abstract underlying representation) just in order to simplify the statement as to where angma can appear in English, and then we wave our hands and say that the *g* doesn't need to be pronounced to prove our case. But that proposal is not as outlandish as such a description would make it seem, for at least two reasons. First and most importantly, there are no words in English that actually end in ...*Ng*, though there are lots of words that end in ...*Nk*, and even trying to pronounce an English word ending in *Ng* isn't particularly easy for an English speaker: this means pronouncing a word something like *song*, say, but followed by a pronounced *g*. So we may well feel obliged to give some kind of a systematic account for the *absence* of a possible sound sequence in English words. The phonologist has two quite different ways of providing an account for a missing combination of sounds in the pronunciations. In general, as we have seen, phonetic pronunciations are conceived of as being the realizations of underlying combinations of sounds, so the phonologist can either (a) approach the problem directly, and give an account of why a certain combination of sounds is not permitted for the underlying representation, or (b) approach the problem indirectly, and allow that the appropriate combination of sounds *does* occur underlyingly, but the principles that determine the phonetic realizations of the sounds in some way clouds the matter, by masking the realization of these segments phonetically (typically by imposing a "null" realization of one of the segments, or as phonologists say, "deleting" that segment: in short, making sure that the segment is not pronounced). In this case, the phonologist pursuing this approach would say that a word-final g's realization after an angma is zero, or in slightly different language, a g is deleted after an angma at the end of a word. (Recall we have to add the condition on the g being word-final because the *g* doesn't delete in words like *finger* or *shingle*.) [FN: There are two exceptions to the generalization that I have given here: the name of the town of Hingham [hIN@m], Massachusetts, and the name of the fabric *gingham* [gIN@m]. There are other apparent exceptions, like singer [sIN@r] or singing [sININ] -- but we will return to these exceptions, which are apparent rather than real.]

To draw these points together

The appearance of angma tells us something about phonological word in English, regardless of which account of angma we choose. On the first account, word-final position is special because it is the only position where an angma can appear without being followed by a *g* or a *k*. On the second account, word-final position is special because a *g* is not realized there when it is preceded by an angma.

Compounds again show that they are composed of two separate phonological words, since angma can appear (without a following g or k) at the end of either word of the compound, as in *songbird, stringsort,* or *singsong.*

The considerations we have seen so far exemplify why it might be useful to employ the notion of a *phonological word*. As we noted above, the notion of the phonological word ought to coincide in a large range of cases with the other notion (or notions) of *word* that arise, such as the "semiotic" sense which corresponds to our more down-home sense of what a word is. And the two do, largely, coincide: a single word from the semiotic point of view is typically a word from a phonological point of view (*dog, jump, book, yellow,* etc.) But there can be deviations as well; what we call *contractions* are typically of this sort. The two (semiotic) words *do not* in English typically contract to *don't* in spoken English, a single phonological word. We know *don't* is a single phonological word, because in English as in all spoken languages, a word must be composed of at least one syllable, and *don't* is composed of only one syllable: hence it can't be two words.