

# 1

## *Setting the Stage*

Most people speaking their native language do not notice either the sounds that they produce or the sounds that they hear. They focus directly on the meaning of the input and output: the sounds serve as a channel for the information, but not as a focus in themselves (cf. Brown 1977: 4–5) This is obviously the most efficient way to communicate. If we were to allow a preoccupation with sounds to get in the way of understanding, we would seriously handicap our interactions. One consequence of this opacity of the sound medium is that our notion of how we pronounce words and longer utterances can be very different from what we actually say.

Take a sentence like ‘And the suspicious cases were excluded.’ Whereas a speaker of English might well think they are saying:

(a) ændðəsəs'pɪʃskeɪsɪz wəɪɛks'kludɪd

what they may be producing is

(b) nə:s'pɪʃskeɪsɪsə'xs'kludɪt

This book will look how you get from (a) to (b). It deals with pronunciation as found in everyday speech – i.e. normal pronunciation.

Years of listening closely to English as spoken by people from a great variety of groups (age, sex, status, geographic origin, education) leads me to believe that there are some phonological differences

## 2 *Setting the Stage*

from citation form which occur in many types of spoken English. Further, these differences are very common within these varieties of English and fall into easily recognizable types which can be described using a small number of phonological processes, most of which can be seen to operate in English under other circumstances.

I call these differences ‘reductions’ (though this term is a loose one: sometimes characteristics are added or simply changed rather than lost). A citation form is the most formal pronunciation used by a particular person. It can be different for different people: for example, the most formal form of the word ‘celery’ has three syllables for some people and two syllables for others. For the former group, the pronunciation [ˈsɛl.ɪ] involves a reduction, for the latter group, it does not.

[ˈsɛl.ɪ] could, however, have been a reduced form in the history of the language of the two-syllable group, even if not within the lifetime of current speakers. That it is no longer a reduced form attests to its ‘promotion’: the word is pronounced in its reduced form so often that the reduced form becomes standard. I speak as if promotion occurs to individual lexical items rather than classes of items, because it can be shown that not all words which have a given structure will undergo reduction and promotion: ‘raillery’, for example, will presumably remain a three-syllable word for those who have only two in ‘celery’, perhaps because the former is an unusual word, perhaps because it has more internal structure than ‘celery’ perhaps for other reasons. In general, the more common an item is, the more likely it is to reduce, given that it contains elements which are reduction-prone (see chapter 2).

The idea of lexeme-specific phonology is not a new one: many phonologists and sociolinguists have worked under the assumption that phonological change over time occurs first in a single word or small set of words, then spreads to a larger set – what is known as ‘lexical diffusion’. (For an early treatment, see Wang, 1977.)

The citation form is therefore not the same as a phonological underlying form: it must be pronounceable and will appear as such in a pronouncing dictionary. Words like ‘celery’ generally appear with both pronunciations cited above.

Deciding what is a reduced form can hence be difficult, but there are few debatable cases in the material I present here: nearly every

native speaker of English will agree that the word 'first' has a /t/ at the end in citation form, but virtually none of them will pronounce it under certain conditions.

The material which I cover in this treatise overlaps the boundaries of several areas of study: sociolinguistics, for example, is interested in which reductions are used most frequently by given groups and what social forces spark them off. Lexicography may be interested in reduced variants, but only in so far as they are found in words in isolation, whereas this work looks at reductions very much in terms of the stream of speech in which they occur. Rhetoricians or singing teachers may regard reductions as dangerous deviations from maximal intelligibility, and a similar attitude may be found in speech scientists attempting to do automatic speech recognition. This book recognizes reductions as a normal part of speech and further suggests that the forces which cause them in English are the same forces which result in most-favoured output in others of the world's languages.

### 1.1 Phonetics or Phonology?

It has been demonstrated (Lieberman, 1970; Fowler and Housum, 1987; Fowler, 1988) that there is *phonetic* reduction in connected speech, especially in words which have once been focal but have since passed to a lower information status: the first time a word is used, its articulation is more precise and the resulting acoustic signal more distinct than in subsequent tokens of the same word. By 'phonetic' I mean that the effect can be described in terms of vocal tract inertia: since the topic is known, it is not necessary to make the effort to achieve a maximal pronunciation after the first token. We expect the same to happen in all languages, though there may be differences of degree.

Phonetic effects are not the only ones which one finds in relaxed, connected speech: there are also language-specific reductions which occur in predictable environments and which appear to be controlled by cognitive mechanisms rather than by physical ones. These we term phonological reductions because they are part of the linguistic plan of a particular language. Sotillo (1997) has shown that

#### 4 *Setting the Stage*

these behave quite differently from the phonetic effects described above: whereas phonetic effects are sensitive to previous mention, phonological reductions are not.

We speak here as if phonetics and phonology were distinct disciplines, and some feel confident in assigning a given ‘phenomenon’ to one or the other (Keating, 1988; Farnetani and Recasens, 1996). Both comprise the study of sounds, but can this study be divided into two neat sections?

‘Phonology’ has meant different things to different people over the course of the history of linguistics. Looking at it logically, what are possible meanings for the term, given that it has to mean ‘something more abstract than phonetics’?

(1) One could take the stance that phonology deals only with the relationship between sound units in a language (segmental and suprasegmental) and meaning (provided you are referring to lexical rather than indexical meaning). Truly phonological events would then involve exchanges of sound units which made a difference in meaning, either:

(a) from meaning 1 to meaning 2 (e.g. pin/pan) or

(b) from meaning 1 to non-meaning or vice versa (e.g. pan/pon).

Phonetics would be everything else and would deal with how these units are realized: all variation, conditioned or unconditioned would then be phonetics. As far as I know, this does not correspond to a position ever taken by a real school of phonology, but is a logical possibility.

(2) Phonology could be seen as the study of meaning-changing sound units and their representatives in different environments, regardless of whether they change the meaning, and with no constraints on the relationship between the abstract phoneme and its representatives in speech: anything can change to anything else, as long as the change is regular/predictable, that is, as long as the linkage to the underlying phonemic identity of each item is discoverable. This will allow one-to-one, many-to-one, and one-to-many mappings between underlying components and surface components, as well as no mapping (in which an underlying component has no phonetic realization).

This type of phonology would look at the sound system of a language as an abstract code in which the identity of each element is determined entirely by its own original description and by its relationship to other elements. Fudge (1967) provides an early example of introducing phonological primes with no implicit phonetic content.

Foley's point of view (1977) is not unlike this: his thesis is that phonological elements can be identified only through their participation in phonological rules:

As, for example, the elements of a psychological theory must be established without reduction to neurology or physiology, so too the elements of a phonological theory must be established by consideration of phonological processes, without reduction to the phonetic characteristics of the superficial elements. (p. 27)

and 'Only when phonology frees itself from phonetic reductionism will it attain scientific status.'

Kelly and Local (1989) also take a position of this sort: 'We draw a strict distinction between phonology and phonetics. Phonology is formal and to be treated in the algebraic domain; phonetics is physical and in the temporal domain.'

Any school which determines membership of a phonological class by distribution alone might be said to take a similar stance: de Saussure's analogy between phonological units and pieces in the game of chess could be interpreted this way.

(3) Phonology could be seen as the study of meaning-bearing sound units and their representatives in different environments, regardless of whether they change the meaning, with the addition of constraints as to what sorts of substitutions are likely or even possible.

If constraints are specified, phonology offers some insight into why changes take place, based on the articulatory and perceptual properties of the input and output. A congruous assumption is that since vocal tracts, ears, and brains are essentially the same in all humans, some aspects of phonology are universal.

## 6 *Setting the Stage*

Most currently-favoured phonological theories are like this: in Chomsky's terminology, they attempt to achieve explanatory as well as descriptive adequacy. Generative grammar opted to incorporate links between abstract phonology and the vocal tract through (1) a choice of features which reflect normal human articulatory possibilities and (2) 'parsimony' (the rule using the fewest features is best, hence rules involve small changes which are easily executed by the vocal tract). Linked to this are the 'natural classes': sounds which are articulated similarly are very likely to undergo similar phonological changes. Autosegmental phonology achieves a link with the vocal tract through structuring of feature lattices, gestural phonology through encoding phonological elements in terms of the articulators themselves. (These themes will be taken up in chapter 3.)

It is, of course, generally understood that articulatory involvement cannot always be presupposed by a theory because in some cases the physical motivation for a phonological event has become inadequate (Anderson, 1981). For example, the *f/v* alternation in singular/plural words (shelf/shelves, roof/rooves, loaf/loaves) is not currently productive (\*Smurf/Smurves), though variation owing to this process is still part of the language. These remains of decommissioned processes are often called fossils. Or the alternation could be the result of an interaction with another linguistic level (cf. Kaisse, 1985) rather than having an articulatory origin. For example, in the utterance 'I have to wear what I have to wear', (meaning 'I must wear clothing which I own') the first 'have' can be pronounced [hæf] while the second cannot, for lexical/syntactic reasons.

These cases aside, when we look at motivated alternations, we begin to consider the relationship between abstract categories and human architecture: this could be seen as a small subset of the mind/body problem so beloved of philosophers.

Most theories of phonology assume that spoken language involves categories which exist only in the minds of the speakers and for which there is thought to be a set of templates: some for segmental categories, some for tones, intonation, and voice quality. Another assumption which is usually not overt is that in speech

production, our goal is to articulate strings of perfect tokens of these categories, but are held back from doing so by either communicative or physical demands.

Again musing on logical possibilities, we can imagine several variations on mind–body interaction.

### 1.1.1 *More mind than body (fossils again)*

Some sequences take more attention than others, and some even take more attention than they are worth, because they do not contribute substantially to the understanding of the utterance. Over time, it becomes customary to simplify these forms through a kind of unspoken treaty amongst native speakers of a language. This leads to our not pronouncing, say the ‘t’ in ‘Christmas’, the ‘b’ in ‘bomb’, or the ‘gh’ in ‘knight’. Eventually, the base form starts to be learned as a whole, so that younger speakers of the language do not even know that, for example, ‘bomb’ has a potential ‘b’ at the end and find out only by learning to spell.

These changes, as mentioned above, are primarily matters of convention and history.

### 1.1.2 *A 50/50 mixture*

Articulatory ease is more evidently a cause for change in cases such as word-final devoicing, which occurs very often with English oral obstruents: one rarely encounters a fully voiced final fricative or stop, even in careful speech. This change from the base form has a different psychological status from the previous one, however: native speakers do not know they are devoicing, and new generations are not led to believe that final obstruents are voiceless, though they pick up the habit of devoicing, as they must in order to sound like native speakers. It is easy to find languages where this feature is an overt convention (e.g. the Slavic languages, German, Turkish). It seems that here we have a peaceful settlement between what the vocal tract wants and what the brain decides to do.

Many characteristics of spoken English seem to fall into this intermediate category. For example, in vowel + nasal sequences, it

is not unusual to nasalize the vowel and to not execute the closure for the nasal consonant. This means that words like ‘can’t’ can be realized as [kãt]. At the phonetic level, then, there can be a contrast between plain and nasalized vowels in words like ‘cart’ and ‘can’t’. While this is a full-fledged phonological process in languages like French and Portuguese, it is merely a tendency in English and Japanese: a habit which is picked up by native speakers and used subconsciously.

### 1.1.3 *More body than mind*

In other cases, vocal tract influences seem clear and inevitable, as in the fronting of velar consonants before front vowels. This is called ‘coarticulation’ and is a function of the fact that the vocal tract has to execute sequences in which commands can conflict (‘front’ for [i], ‘back’ for [k], and a compromise is reached. This seems to me a clear case of a phonetic process, but it also seems quite clear that it can have phonological consequences, as in Swedish, where the sequence (which was historically and which is still spelled) [ki] is pronounced [çi], or as in English alternations such as act/action.

Bladon and Al-Bamerni (1976) have also pointed out that *resistance* to coarticulation can occur as a result of other demands of a language. In English, [k] and [i] can coarticulate freely, since a fronted [k] is not likely to be misinterpreted. In languages with a [ç], [k] has less freedom to move about. This indicates that even process which are largely controlled by the vocal tract can be moderated by cognitive processes.

Resistance to coarticulation can also develop for no obvious reason: in Catalan, there is virtually no nasalization of vowels before nasal consonants, though it is found in the other Romance languages. (Stampe (1979: 17) cites denasalization as a natural process, and we can see this at work elsewhere in Catalan: whereas Spanish has [mano] and Portuguese [mãõ] for ‘hand,’ Catalan has [ma], with a plain vowel.)

If we accept that our third definition of phonology is a reasonable one, how can we distinguish phonology from phonetics? What is the difference between saying that changes have to have an



articulatory or perception explanation and saying that the vocal tract is *responsible* for the changes? What is the interaction between the physical demands of the vocal tract and the desire on the part of the speaker to (a) be intelligible and (b) sound like a native speaker?

The answer seems obvious: as long as constraints determined by the shape and movement of the vocal tract are included in one's phonology, there is in principle no way to draw a boundary between phonetics and phonology. Processes which are essentially phonetic (such as nasalization of vowels before nasal consonants) are prerequisites for certain phonological changes (lack of closure for the nasal consonant, leading to distinctiveness of the nasalized vowel). Distinctions which are essentially phonological (such as the word-final voicing contrast in English obstruents) are signalled by largely phonetic features such as duration of the preceding vowel (though, granted, this process is exaggerated in English beyond the purely phonetic). Language features which are said to be phonological are constantly in the process of becoming non-distinctive, while features said to be phonetic are in the process of becoming distinctive. There are obvious cases of truly phonological processes and truly phonetic ones, but between them there is a continuum rather than a definable cutoff point.

#### 1.1.4 *Functional phonology and perception*

The discourse above has been largely couched in terms of the generation of variants. If we are to think of phonology as not just an output device, but also as a facility which allows us to use the sound system of our native language, we must also think of it in terms of perception. In this framework, we can ask how knowledge of variability in a sound system is acquired and used and we can explore the relationship of this knowledge to phonological theory: are the sound units used for perception the units we posit in a phonological analysis? These questions, while normally thought of as psycholinguistic ones, are clearly important for an understanding of casual speech phonology. We will go into this more deeply in the second half of chapter 3.

### 1.1.5 *Have we captured the meaning of 'phonology'?*

We have, rather, shown that there are many ways to define phonology. I propose a further one:

(4) Phonology is the systematic study of the pronunciation/perception targets and processes used by native speakers of a language in everyday life. It presupposes articulatory control of not only the contrasts used meaningfully in a language, but also of other dynamic features which lead to variation in speech sounds, such as tension of the vocal tract walls (cf. Keating, 1988: 286). It therefore includes all articulatory choices which make a native speaker sound native, including sociolinguistic variables such as register and style. It does *not* include simple coarticulation but can place limits on degree of coarticulation (Farnetani and Recasens, 1995; Manuel, 1990; Whalen, 1990).

Note that here again, the boundary between phonetics and phonology is hard to define, though it is clear that version 4 phonology includes a great deal of what is normally thought of as phonetics.

### 1.1.6 *Influence of phonology on phonetics*

We have suggested that phonetics 'works its way up' into phonology. It must also be recognized that phonology 'works its way down' into phonetics. We think of speech sounds as being representatives of abstract categories despite there being a very large number of ways that one realization of a phonological unit can differ from another realization of the same phonological unit. When we do phonetic transcription, we use essentially the same symbol to represent quite different variants because phonology guides our choice of symbols. We can avoid this to some extent when listening to a language we do not know, but once the basics of the new language are assimilated, phonological categorization again takes over. This process has been useful in helping us derive new spelling systems for previously unwritten languages, but stands in the way of our experiencing phonetic events phonetically. The very notion that connected speech can be divided up into segments and represented

with discrete symbols is a phonological one, reinforced by our alphabetic writing system.

### **1.1.7 Back to basics**

Let us now return to the question of whether this book is about phonetics or phonology. In the light of what was said above, it is not clear that this question needs to be answered, or even that it is a meaningful question. By definitions 1 and 2, most of the material covered here will have to be thought of as phonetics. By definitions 3 and 4, it is mainly phonology. Suffice it to say that it deals with systematic behaviour by native speakers (of English in this case, though not in principle) using fluent speech in everyday communicative situations.

## **1.2 Fast Speech?**

Casual speech processes are often referred to as ‘fast speech rules’. Results are not yet conclusive about whether increase in speech rate increases the amount of phonological reduction: it seems clear that phonetic undershoot takes place as less time is available for each linguistic unit, but evidence cited below suggests that cognitive factors are more important than inertia, despite the fact that connected speech processes are often called ‘fast speech rules’.

A commonsense view of connected speech has it that the vocal tract is like any other machine: as you run it faster, it has to cut corners, so the gestures get less and less extreme. Say, for example, you are tracing circles in the air with your index finger. At a rate of one a second, you can draw enormous circles but if you’re asked to do 6 per second, you have to draw much smaller circles, and a rate of 15 per second is impossible, no matter how small they are. So if you try to do 15, you might get only 10 – effectively, 5 have dropped out.

The same reasoning is applied to the vocal tract: as you execute targets faster and faster, the gestures become smaller and smaller, and sometimes they have to drop out entirely, which is why you get deletions in so-called ‘fast speech’.

A moment's thought will convince you that the analogy here is not very good: the vocal tract is a very complicated device, and different parts of it can move simultaneously. The elements which comprise the vocal tract are of different sizes and shapes and have different degrees of mobility. The speech units which are being produced are very different from each other. And, most importantly, speech is not just an activity, it is a means of communication. This means that different messages will be transmitted nearly each time a person speaks, different units will be executed in sequence, and different conditions will be in effect to constrain articulation. For example, one can speak to a person who is very close or very far away, to a skilled or unskilled user of the language, with or without background noise.

The 'finger circle' analogy also does not take into account the relationship between the higher centres of the brain and articulation. Speech is a skill which we practise from infancy and one over which we have great control: does it seem likely that anyone would run their vocal tract so fast that not all of the sounds in a message could be executed? One might imagine singing a song so fast that not all of the notes/words could be included: the difference here is that we are executing a pre-established set of targets with a fixed internal rhythm *intended* for performance at a certain speed. But presumably, in real speech, our output is tailored to the situation in which it is uttered and has no such constraints.

Another argument against our very simplistic view of 'fast speech deletion' is that there are very distinct patterns of reduction in connected speech, related to type of sound and place of occurrence. If one were simply speaking too fast to include all the segments in a message, would not the last few simply drop out, as with our 'finger circles'? Rather, we find specific types of sounds being under-executed, in predictable locations. And these 'shortcuts' are different from language to language as well. Surely the importance of cognitive control of these mechanisms cannot be underrated.

Lindblom (1990) follows this line of reasoning in his 'H&H theory' of speech, which essentially says that in any given situation, the vocal tract will move as little as possible, provided that (situationally-determined) intelligibility can be maintained. This theory thus predicts a limit to the degree of undershoot based on the communicative demands of the moment.

While this point of view has a lot to be said for it, it cannot be considered a phonetic or phonological theory exclusively: it embraces all areas of linguistics, because they all contribute to the ‘communicative demands of the moment’. Take an example from one of my recorded interviews: the speaker said [sof s'kɜ̃ri] ‘social security’. The underarticulation of this phrase is allowed because of discourse features (the topic is ‘welfare mothers’) and other pragmatic features (social security has been mentioned previously) as well as because of the syllable shapes and stress patterns involved. While the interests of the articulators are served by the apparent disappearance of certain sounds, the articulators cannot be said to have caused the underarticulation.

Finally, it is obvious that the types of reduction which we have been looking at also occur in slow speech: if you say ‘eggs and bacon’ slowly, you will probably still pronounce ‘and’ as [m], because it is conventional – that is, your output is being determined by habit rather than by speed or inertia. This brings us back full circle to the question ‘phonetics or phonology?’ Habit and convention are language-specific and are part of the underlying language plan rather than part of moment-to-moment movement of the articulators. Habits of pronunciation are systematic and predictable and can be linked only indirectly to articulator inertia.

### 1.3 Summary

This book is about the differences from citation form pronunciation which occur in conversational English and their perceptual consequences. We call these changes ‘phonological’ because they systematically occur only to certain sounds and in certain parts of words and syllables and because they are different from connected speech processes in other languages. Hence, they form part of the abstract pattern of pronunciation which is the competence of the native speaker. While they reflect constraints in the vocal tract, they are not purely phonetic: the boundary between phonetic and phonological processes is indistinct and probably undiscoverable given present-day notions of phonology. The reductions found in unselfconscious speech cannot legitimately be called ‘fast speech’ processes.