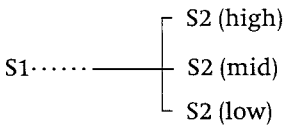
1

A preliminary sketch of the tone unit

Intonation is traditionally equated with variations in the perceived pitch of the speaking voice. We shall begin by examining this characterisation.

Pitch varies continuously from the moment anyone begins speaking to the moment they end. An accurate description of the variation would be exceedingly complex and would reveal nothing of the significantly patterned phenomenon that we take intonation to be. This is because not all the variation has the same kind of communicative significance. The only meaning that the description presented here purports to deal with is of the kind that can be represented as the result of a speaker having made an either/or choice. We are concerned, that is to say, with identifying a set of oppositions* that reside in the language system, knowledge of which we must assume the speaker shares with his/her hearer.

Suppose there is a language system that includes an opposition we can describe in the following terms: having pitched a particular syllable at a certain level, the speaker can, at some subsequent syllable, make a meaningful choice between the same pitch, a higher one or a lower one. Graphically, we might represent the choices at the second syllable thus:



Two important conditions are inherent in the kind of situation we have postulated at S2:

(i) if the speaker produces S2 at all he/she cannot avoid choosing one of the three possibilities specified (there is no fourth choice).

(ii) the communicative value of any one choice is defined negatively by reference to the other options available (the value of 'high' is whatever is *not* meant by 'mid' and 'low' taken together).

This account misses out one very obvious thing we might say about what our speaker does. Having decided, let us say, on a higher choice, he/she must then make a further decision, as to *how much* higher to pitch the syllable. To deal with this kind of decision, any descriptive enterprise would have to concern itself with variation on a continuous scale, and would demand a quite different analytical method from that used here. My aim will be to show that a small set of either/or choices can be identified and related to a set of meaning oppositions that together constitute a distinctive sub-component of the meaning-potential of English.

In accepting this limitation, we are not of course suggesting that variation of the continuous kind has no communicative importance. Individual speakers may well have a predilection for greater or lesser pitch intervals, so that the overall pitch treatment of an utterance will carry a significant component of indexical information; or an increase in anxiety, anger or physical discomfort might result in a gradual rise in the general pitch of the voice: a hearer might, in a given situation, derive a great deal of information of this kind from variation that is superimposed upon the systemically significant changes. Impersonation and similar devices can result in a speaker's deliberately adopting a pitch level or pitch range that is not his/her usual one and the effect is part of the total 'meaning' of his/her present behaviour. None of these effects, or others like them, can be said to be without interest to anyone who wishes to understand how verbal interaction works. But in so far as, say, anger can be attributed to a general rise in pitch, there are theoretically unlimited gradations of 'how high', and so presumably of 'how angry', the speaker may be judged to be. Any attempt to capture the significance of intonation must aim to disentangle, at an early stage, those variations that hearers respond to as pitch placement on a continuous scale and those that they hear and interpret as manifestations of an either/or option. To say this is to adopt a well-established position in linguistics and to assume that it is only phenomena of the latter kind that linguistic techniques are designed to handle. If we re-label the distinction we have made and speak of our object of study as linguistic intonation (as opposed to the other which is non-linguistic) we shall be ascribing to 'linguistic' a definition which probably few people would dispute.

The hypothetical case on which we have based our discussion so far is probably too simple to apply to any real language in at least two ways.

Firstly, we have associated the opposition with a particular syllable. Reality would seem to require that we associate a number of variables globally with a unit that extends over a number of syllables. We shall refer to the stretch of language that carries the systemically-opposed features of intonation as the tone unit*. Our starting assumption will be that hearers are able, on the basis of the overall presentation of the tone unit, to recognise that it carries a number of distinctive intonation characteristics and to differentiate it from otherwise similar tone units having other characteristics. To state what these are, we must make a rough-and-ready reference to what happens at particular syllables, but this practice has to be followed in the awareness that anything affecting physical realisation at one point in the tone unit is likely to affect what happens elsewhere. The several physical features that we find we have to separate out in order to make a usable description occur, in reality, as a complex contour: we may speculate, with some plausibility, that the speaker 'plans' the tone unit and the hearer 'decodes' it *as a whole*. When we say – as expediency compels us to say – that certain speaker-choices are associated with certain syllables, we shall be simplifying to the extent of overlooking this kind of prosodic effect.

The other source of over-simplification is our adherence to the traditional practice of speaking of intonation in terms of pitch variation. It seems inherently improbable that a human being can make systematic variations on one physical parameter without its affecting others. Changes in loudness and in speed result from intimately connected adjustments to the same speech mechanism as that which determines pitch. Isolating any one parameter for attention might well give a quite false notion of what, in the course of ordinary language use, people actually 'hear' when they hear intonation. Instrumental analysis suggests, in fact, that each of the meaningful oppositions our description recognises *can* be identified on the basis of pitch treatment alone. This provides us with a convenient method of characterising the oppositions and enables us to make use of traditional vocabulary like 'high', 'low', 'rise' and 'fall'; we must leave open the possibility that variations other than those in pitch may be of greater operational importance to the language-user in some or even in many cases.

Everything we have said so far in this chapter underlines the need for extreme tentativeness in providing phonetic descriptions of the meaningful choices that make up the intonation system. We have recognised

four possible causes of variation between one occurrence of a systemically-meaningful feature and another occurrence of the same feature: it may be overlaid by variation of the continuous, non-systemic, kind; the variables that a description will want to separate out and identify individually may merge into a unified envelope co-terminous with the tone unit; variation on the dimensions of pitch, loudness and time may interact in complex ways as modifications of the segmental composition of the utterance; and idiolectal and dialectal variation may result in a given formal choice having a whole range of different phonetic realisations. To describe a certain intonation feature in such phonetic terms as, for instance, 'high fall' is to take no account of the phonetically distinct events that hearers might classify together as non-significant variants. More seriously, perhaps, it disposes us prematurely to assume that all occurrences of the feature will fit this strictly arbitrary description: in theory we should start with no firm expectation that a particular realisation will necessarily be either a 'fall' or 'high'. A procedure which would seem unlikely to result in an adequate functional description of intonation would be to observe pitch patterns that appear to recur and then to try to ascribe significance to what has been observed. The use of instruments to refine and sophisticate observational techniques would not, in itself, do anything to reverse the directionality of such an exercise. If the same principles are to be applied as have been applied in establishing the set of segmental elements the language employs, then we must start with a hypothesis about meaning: before proceeding to detailed phonetic specification we need to know how many meaningful oppositions there are and how they are deployed with respect to each other within the area of meaning potential that intonation realises.

It is fair to say, of course, that the only research procedure available is to make tentative phonetic observations and try to associate them with generalisable meaning categories. And what applies to the discovery of the meaning system applies equally to its presentation in such a book as this. We have no alternative but to refer to the variables in pseudo-phonetic terms; but we have to stress that such a characterisation is no more than an approximation and a convenience. It is necessary always to preserve a certain distance from the phonetic fact. When we say that there is a 'fall' at a particular point in the utterance, it is not the fact that the pitch falls that we wish to make focal; it is rather the function of the language item which carries it, a function which, as far as we

can tell from the examination of a lot of data, is typically realised by a falling pitch.

In practice, the possible discrepancy between meaningful choice and phonetic realisation creates fewer problems than we may have seemed to suggest it will. It must, however, be taken seriously in any description if misunderstanding and misapplication of the categories are to be avoided. For this reason, the sketch of the tone unit that forms the central part of this chapter is presented in such a way as to try to avoid inappropriate precision in describing phonetic realisations. This may seem to the reader to be rather tiresome prevarication. We hope, however, that the prevaricating quotation marks we consistently use with words like 'high' and 'fall' will be accepted as the unavoidable consequences of adopting a proper theoretical position. Only when the set of meaningful oppositions has been described shall we suggest a way of approaching the realisation problem.

Our tone units have some superficial similarities to the entities other descriptions have referred to variously as 'sense groups', 'breath groups' and 'tone groups'. One unresolved issue that these terms reveal is uncertainty about how the observable fact that continuous speech is broken up into perceptible blocks or units is to be interpreted. The labels seem severally to suggest semantic, physiological and formal considerations. There is also uncertainty about how different meaning choices are to be attributed to the analytical unit once it has been identified. Once the boundaries of the unit have been determined, on whatever criteria, the descriptive task has generally been thought to involve identifying sub-components and associating speaker options with each of them. So, for instance, one tradition has categories like 'nucleus', 'pre-head', 'head' and 'tail' and distinguishes different types of each, like 'low pre-head' and 'high pre-head'. A fundamental difference between descriptions is the number of sub-components they recognise and the number of opportunities for speaker-choice they associate with each. The significance of this difference should not be underestimated. By specifying how many possibilities of choice there are in the unit, where they are made, and what relations of interdependence exist among them, the phonologist is making a powerful prediction about what the meaning system is like. If he makes one wrong prediction, he will never achieve a mapping of the descriptive categories into the meaning system.

As we have said, the only procedure that will lead to a model of the

meaning system is a lengthy examination of data, during which both ends of the sound-sense continuum are kept in view, and which works gradually towards a matching of intuitively satisfying analyses at both ends. Little purpose would be served by attempting to reconstruct the various stages through which the present description went before it reached the form in which it is presented here. The reader may very reasonably demand that this particular way of analysing the tone unit should be justified: why this way rather than one of the many other ways that the phonetic data would doubtless support? The impossibility of justifying it in advance of the presentation and exemplification of the rest of the description will be apparent from everything we have said. The categories and choices that are here represented are those which are required by a particular view of the meaning system. Their justification is in the extended exposition of that view that constitutes the main part of this book.

The categories of tone unit

It is not difficult to find samples of recorded data in which pauses of some kind segment the stream of speech into units that hearers can readily agree about. At this point we need not enquire further into what, intuitively, is identified for this purpose as a 'pause'. By taking this empirical observation as our starting point, we are emphatically *not* committing ourselves to regarding the pause as criterial. We are simply taking units that do lend themselves to easy identification in this way as specimens of the descriptive category tone unit. Undoubtedly, there are cases where the phonetic evidence for segmentation is less straightforward, and problems of other kinds arise once we move outside a carefully selected corpus. We can, however, reasonably put off confronting such problems until after we have examined some data which, in this particular respect, is easy to deal with:

(1) // i think on the whole // that these officials // do a remarkably good job // we have to remember // that // they're required // by administrative practice // to take these decisions // on paper // and // most often // when they get these decisions wrong // it's because // they haven't had the opportunity // of talking // face to face // with the claimant // and really finding // the facts //

If we now examine and compare all the segments that this simple transcription procedure recognises, we shall be able to make some preliminary generalisations about the internal organisation of the tone unit.

Firstly, each tone unit has either one or two syllables that a hearer can recognise as being in some sense more emphatic than the others. We shall say that these syllables have prominence*, a feature which distinguishes them from all other syllables. There are two prominent syllables, identified by the use of upper case characters, in each tone unit of the original recorded version of:

> // I think on the WHOLE // that THESE of FICials // do a reMARKably good JOB //

It may seem at first sight that the distribution of prominent syllables is determined by the lexis and grammar of the utterance. For instance, the fact that *-mark-* rather than any other syllable of *remarkably* is prominent is clearly connected with the way 'stress' or 'accent' is normally associated with this particular syllable in this particular word. There are evidently constraints upon what a speaker may do, but our present concern is to recognise that – such constraints notwithstanding – speakers do have the option of producing alternative versions of some units. The broadcaster might, for instance, have said:

(2) // i THINK on the WHOLE //

(3) // do a remarkably GOOD JOB //

(4) // do a reMARKably GOOD job //

We shall investigate in detail the significance of these variations in the next chapter. Meanwhile we must emphasise that seeing prominence as a feature which speakers can vary voluntarily – seeing that its occurrence is not merely an automatic reflex of other decisions they may have to make of a grammatical and lexical kind – is an essential first step towards grasping its significance as an exponent of part of the meaning system. Of course, we cannot rule out *ab initio* the possibility that otherwise similar utterances that are differentiated only by the incidence of prominence are sometimes in free distribution: that is to say the fact that certain syllables in a tone unit are prominent while others are not might sometimes have no communicative significance. As a matter of

principle, however, we shall assume that when a speaker does one thing when he/she could have done another, the decision is in some sense a meaningful one. We shall eventually want to consider how the speaker's freedom of action is limited by constraints of the kind that govern his/her treatment of *remarkably*, but our primary concern is with prominence as an independent variable.

There may not seem to be much scope for plausible redistribution of prominence in these tone units:

// that THESE of FICials //

// MOST OFten //

It is not difficult, however, to conceive of utterances otherwise similar to these having one prominent syllable instead of two:

(5) // that these of FICials //

(6) // that THESE officials //

(7) // most OFten //

(8) // MOST often //

The speaker has freedom to vary, not the location, but the number of prominent syllables. Our sample suggests that the latter freedom is limited to assigning *either one or two*. A transcription of the rest of it, indicating all prominent syllables, is as follows:

// we HAVE to reMEMber // THAT // they're reQUIRED // by adMINistrative PRACtice // to TAKE these deCISions // on PAper // AND // MOST OFten // when they GET these decisions WRONG // it's beCAUSE // they HAVEn't had the opporTUNity // of TALKing // FACE to FACE // WITH the CLAIMant // and REALly FINDing // the FACTS //

We can continue to rely upon pre-analytical intuitions to recognise that the syllables we have so far described as prominent are, in reality, of two kinds. To characterise the difference between them, however, we need to go beyond what will be immediately apparent to most hearers. It is best regarded as a difference in the range of meaningful choices associated with the place the syllable occupies. We have already seen that the allocation of prominence is the consequence of a speaker's decision

with respect to a binary prominent/non-prominent choice. Additionally, associated with the last prominent syllable in the tone unit but not with any other one, there is a further choice from a set of significant pitch movements or tones*. The syllable, the point of operation of the tone system*, will be referred to as the tonic syllable*. Thus tonic syllables are to be understood as constituting a sub-set of prominent syllables.

The tonic syllables in the first three tone units of our data sample exemplify in succession the tones that are described conveniently in quasi-phonetic terms as the 'fall-rise', the 'rise' and the 'fall':

(9) WHOLE of FICials JOB

In transcriptions, the tonic syllable will be underlined and a symbol will be placed at the beginning of the tone unit to indicate which tone is selected. So in

> // ⌄ I think on the WHOLE // ↗ that THESE of FICials // ↘ do a reMARKably good JOB //

the symbols are to be interpreted as meaning 'At the next underlined syllable there is an occurrence of the tone we associate for recognition purposes with a phonetic "fall-rise", "rise" and "fall" '.

These additions to our transcription conventions call for two observations. Firstly, since the tonic syllable is, by definition, the last prominent syllable in the tone unit, underlining it is strictly unnecessary in this piece of data. The need for the convention will become apparent when we come to discuss the question of boundary recognition in later pages. Secondly, the symbols used to represent tones make direct reference to their physical representation. Although it is convenient to use such symbols at this stage, we shall wish to replace them later with others that have functional rather than phonetic connotations, so making it easier to maintain a mental separation between the meaningful decisions the speaker makes and the physical events whereby his decisions are made manifest.

Later tone units in the same sample exemplify the use of a 'level' tone:

(10) // → THAT // , // → AND //

It happens that there are no examples of the fifth option, the 'rise-fall'. The complete set of possibilities comprises:

fall; rise; fall-rise; rise-fall; level

In addition to choosing one of the five tones at the tonic syllable (a choice we have associated with the physical phenomenon, pitch *movement*), the speaker has also to select in a three-term system that we will associate, in a similar provisional and undogmatic way, with the physical fact of pitch level. Retaining our unavoidable concern with phonetic realisation, we can best describe the pitch-level choice by assuming simultaneous choice of a 'falling' tone. While keeping the same pitch movement in // do a reMARKably good JOB // a speaker may vary the height at which the fall begins. What actually happens in our sample will be represented as follows:

(11) // do a reMARKably good $\underline{\text{JOB}}$ //

This is to be taken to mean that the level at which the falling tone begins is 'low' compared with the level of the preceding prominent syllable *-mark-*. Alternative versions of an otherwise identical tone unit would be:

(12) // do a reMARKably good $\overline{\text{JOB}}$ // (job falling from 'high' level)

(13) // do a reMARKably good $\underline{\text{JOB}}$ // (job falling from 'mid' level)

It will be apparent that this set of examples is particularly useful for expository purposes because the tone unit has two prominent syllables, the first of which can be said to be pitched at 'mid' level. In discussing the pitch-level treatment of *job* we have had to be concerned only with relative pitch heights within the tone unit. To generalise the description we have now to consider both how the 'mid' value of *-mark-* is determined and what happens when there is only one prominent syllable in the tone unit.

We shall refer to any prominent syllable like *-mark-* which precedes the tonic syllable in its tone unit as an onset syllable*. (Onset syllables thus constitute another sub-set of prominent syllables and are complementary to tonic syllables.) We have already said that onset syllables, thus defined, differ from tonic syllables in having no pitch-movement choice associated with them: tone is a once-for-all choice for each tone unit. The onset syllable is, however, the place of operation of a separate pitch-level choice, a choice whose communicative significance we shall show to be quite different from the phonetically similar one at the tonic syllable.