THE IDENTIFICATION OF PHONOLOGICAL UNITS

JOHN L. M. TRIM

No-one listening to the papers and ensuing discussions during this congress can fail to have been struck by the clarity with which certain trends have emerged.

There has been, on the one hand, an increasing confidence in methods of specifying the acoustic and articulatory stages of the speech event, and evidence of considerable progress in our theoretical insights into their relation. Interpredictability has seemed to be a matter of time and work rather than of any insuperable theoretical barrier.

On the other hand, the papers of Fant and Cooper, Denes and Ingermann, Fry and Ladefoged, Gill and Shearne and Holmes, all indicate what seems an increasing reserve with regard to the relation of these articulatory and acoustic events to linguistic structure, a reserve which seems to have become stronger, as the events concerned have become better observed and better understood.

Fant pointed out that information contributing to the identification of a particular unit at one place in a linguistic sequence would normally be found distributed over a number of phonetic segments.

Cooper pointed to the necessity of estimating the success of speech synthesis by the proportion of recognitions by listeners, and to the variability of possible stimulus and the dominating role of context.

Ingermann, however, had found it difficult to employ listener identification as a criterion of success in synthesis by rule because of the great variation in scores according to context.

Ladefoged, who has expended very considerable effort in attempts to discover the phonetic basis of the syllable, announced that he had come to the conclusion that the syllable should be regarded as a linguistic unit with no single phonetic correlate.

Shearne and Holmes showed that formant values obtained from vowels pronounced in stressed monosyllables did not obtain for vowels recognized as the same phoneme from samples of running speech.

Fry and Denes both called attention to the dominant role played in speech recognition by the listener's knowledge of the linguistic system concerned, expressing their conviction that progress in this field was to be expected primarily from providing a machine with suitably organized linguistic information and looking forward to the employment of digital computers for this purpose.

There seem to be two principal reactions to this state of affairs. One, stated explicitly by Fant and by Jakobson and implied by a number of speakers, holds that unambiguous evidence for the differentiation of linguistic units is to be found in the records of articulatory and acoustic events, but that our techniques are not yet sufficiently refined for us to say, beyond doubt, what these are.

This view holds to the famous prediction of Bloomfield that the definition of the phoneme would come out of the laboratory in 30 years. Indeed, the most powerful school of linguists, deriving from Bloomfield, has nailed this colour firmly to its masthead. Given a long enough sample of recorded speech, the linguist claims, by the employment of substitution techniques, to deduce the linguistic system underlying it; or, if you like, with different emphasis, the task of the linguist is simply to take a certain corpus of phonic material, and then invent a system that will give a complete and economic account of this material.

The task of the experimental phonetician should surely then be a very simple one; apart from his duty to provide the linguist with acoustic and articulatory records as complete as possible (taking care to include "irrelevant" data for the linguist to eliminate) he has merely to work backwards along the analytic path of the linguist and discover in detail what it was that the linguist picked out by aural impression.

Yet it is precisely this simple task that is proving so refractory, and concerning which so many of the most experienced and sophisticated experimental phoneticians are increasingly sceptical. May one therefore not consider the possiblity that far from being a temporary embarrassment which we may hope to overcome with more refined analytical tools, the indeterminate relation between articulatory and acoustic events on the one hand and phonological units on the other is theoretically necessary, essentially rooted in the very nature of speech communication itself?

That this is in fact the case, is surely apparent as soon as we try to base our work on a comprehensive view of the speech-event – and is it not the overall object of the phonetic sciences to develop a comprehensive theory of the speech event, and to apply this theory to the understanding of individual speech events?

In general, of course, the speech event is too well known to require glossing here. It is generally accepted that we have to bring into relation successive stages comprising events different in kind, each initiated by its predecessor, and determined in detail partly by its own inherent characteristics, partly by characteristics of its predecessor, partly by features of the context.

On the other hand, while we know a great deal about the articulatory and acoustic stages, we know very little about the activity of the speaker preceding speech and that of the listener upon receiving it. Of course, these events are difficult of access, and it was this that led Bloomfield to urge us to forget them and concentrate on the observable and accessible. But it is now clear that we cannot understand articulation or sound waves in isolation from the speech event in which they are set. Phonetics – and linguistics – as behavioural sciences must take as their material the whole range of human language behaviour, and apply to it the principles of scientific method – observation, induction, deduction, validation – as outlined by Peterson yesterday. Linguistic, including phonological, analysis which concentrates upon the purity of its procedures for inventing a system to give a complete and economic description of a closed corpus of observed utterances is not scientific, but scholastic. We are interested in the results of a linguistic analysis in so far as it enables us to handle a continuing human activity, to understand and to some extent predict new events as they come along. It is the adequacy of the system to account for and predict the behaviour of listener and speaker, rather than the process by which it is arrived at, which is the real criterion of its validity. An analysis of a living language which lacked this predictive power, however elegant and refined, would be trivial. The linguistic system which matters, at any rate to the experimental phonetician and linguist, is that which is operative in the speech events he observes, that in terms of which the speaker formulates and the listener identifies his utterance.

The various processes involved in the listener's reaction are of particular importance, and seem more complex than is often assumed. A neural input (concerning which Prof. Mol has reminded us of our ignorance), gives rise, under the powerful influence of contextual factors, to primary perceptions of sound quality, which are then used as evidence for the central task of recognizing the linguistic text formulated by the speaker. I do not think we should talk of the "perception" of linguistic units, because the perception of sound quality is not subsumed in the linguistic identification. We identify Delattre's synthetic syllables as a succession of [b]s followed by some [d]s without failing to perceive that a succession of differences is involved. As Fry insisted, information from one stage utilized at a later stage is not thrown away. We respond simultaneously at a number of different levels. Each contributes to the others but retains its identity. The linguistic identification itself involves reference to linguistic knowledge organized at different levels – phonology, lexicon, grammar – and integrates them to reproduce a text. In the process of integration, inter-level correction may occur – but after, not before consciousness. It is the correction and non-correction together which make the "infantry" of linguistic signs – and their "fissures" – funny. Because the identification of phonological units is normally integrated with that of lexical and grammatical units, we are in danger of overlooking the separate levels of organization. It is in abnormal language usage, such as verbal play, especially nonsense, and various kinds of dysphasia that their autonomy is most clearly seen. The study of these kinds of speech is therefore of considerable importance for the understanding of the speech event.

Since one may reasonably posit an autonomous phonological level of linguistic organization, I think that it is proper to conduct experiments requiring people to identify particular phonological units. (I prefer to use the word "identify" rather than "recognize", in order to emphasize that it is more a case of using a given input as evidence in selecting one from a range of pre-set categories than of recognizing a category inherent in that input.)

Of course, it must be a potential text which the listener is asked to identify (including perhaps certain attested types of nonsense, e.g.). It is part of the experimenter's skill to frame choices from which unambiguous conclusions can be drawn.

Experiments of this kind, particularly those carried out at Haskins laboratories and elsewhere employing synthetic speech with its possibilities of controlling the input, have shown beyond doubt that successful identifications are possible on the basis of a multiplicity of acoustic clues. One may cite, for example, the work of Denes on the contribution of relative durations and intensities to the identification of [s] or [z], that of Fry on duration, intensity and fundamental frequency to the identification of stress-differentiated word pairs. Both sets of results were predictable from the results of aural phonetic observation, and others spring to mind – vowel quality in the case of stress.

Intonation patterns are recognized on the basis of changes in fundamental frequency, intensity and spectrum. Under particular circumstances, one clue may dominate and contribute more to identification than others. We may test this as Prof. Delattre showed in the Voback experiments, by setting one clue against another: fundamental frequency v. residual structure and seeing which wins – in this case fundamental frequency. But in breathed or whispered speech, where no fundamental frequency is found, and in pseudo-voice, where it may not be similarly controlled, identification of intonation pattern still occurs, though the number of recognitions (matched identifications) is reduced in a manner dependent on various factors (strength of clues, acuity, intelligence and experience of listener, contextual factors). Some learning occurs, but recognition is high without it.

In general, we may say (1) that no single phonetic clue is indispensable to the identification of a phonological unit, (2) that clues to the identification of any phonological unit may be found in any number of acoustic dimensions over an indefinite stretch of the utterance. The characterization of a linguistic text by a linear sequence of phonemes, for instance, does not imply a strictly successive presentation of clues, much less a one to one relation with a physically definable utterance-segment. (3) that a single utterance-segment in a particular acoustic dimension may contain information relevant to a variety of linguistic judgements. The length of a vocoid may be relevant to the identification of the vowel phoneme concerned, the following consonant, the stress of the syllable concerned, the foot to which it belongs, the emphasis attaching to the word containing it, the point of word or syllable division, the kind of tonal nucleus involved, the style of speech involved and probably more. Or, viewed from the point of view of the speaker, the length may be seen as the resultant of a number of forces represented by these factors.

Beyond this, of course, the contribution to identifications made by perceptual clues based on such acoustic features varies widely according to the contextual constraints which operate in any speech event.

Most individual speakers maintain a range of styles – or rather a continuum of stylistic variation, arising from the law of minimal effort operating on linguistic redundancy in context, restrained by certain social constraints (e.g. politeness).

The most formal style is precisely that in which linguistic units have their most elaborate form, and in which articulatory-acoustic-perceptual differentiations are constant and maximal.

This style of utterance is generally that chosen for institutionalization and formal public utterance. It may also be used for linguistic initiation, since it has fewest presuppositions–though, in fact, children hear the full range of styles and soon learn some inter-style predictions. It is used to foreigners, and being "edited" and stable is suitable for analysis by linguists too.

Once learnt, the fully developed system is latent, rarely explicit and in familiar situations, where situational information is high and social constraint on behaviour low, redundancy is exploited in various ways by speaker and listener alike, the sole criterion being successful recognition.

There are various possible ways of dealing with this stylistic variation. It may be categorized into a certain number of separate styles, constituting a series of related, but separate languages, each to be separately analyzed. This can be done, but has disadvantages: there is something rather arbitrary about the categorization; the presupposition of relative stability is shaky, and in very familiar style the analysis gets very complicated – the phonemic inventory becomes very large, and morphophonology a nightmare. It comes to look as though for convenience's sake people have invented an unmanageably complex language!

Alternatively, the fully developed system may be considered as still operative, but with reduced clues to its recognition; or again, one may distribute the reduction between the text to be recognized and the clues to its recognition, stating rules of derivation for each, in terms of morphophonological and acoustical transforms.

Although such procedures may reduce the apparent indeterminacy of relation between acoustic events and linguistic units, the difference of level involved is absolute. It is essential to clarity of thought in this field that these levels should be most clearly distinguished and that a terminology should be developed which does not permit – or at least not encourage – the confusion of features of the articulatory or acoustic stages of particular speech events (the *phonetic* level) with features of the centrally organized linguistic structure (the *phonological* or more generally *linguistic* level). "Utterance", for instance, should be restricted to the phonetic level, its counterpart on the linguistic level being the "text". Only a text, and never an utterance, can be said to contain phonemic or prosodic units. Only an atterunce, and never a text, can be viewed as a sequence of phonatory and articulatory movements, or a train of sound-waves. The use of "utterance" indiscriminately to cover text and utterance is a continuing source of confusion.

Following out this principle, we should evolve sets of distinct terms for correlated features at different stages and levels, distinguishing at least the articulatory, acoustic, perceptual and linguistic. It is clearly increasingly common practice to distinguish in this way between duration (physical), length (perceived) and quantity (phonological), or between frequency, pitch and tone. To extend such conceptual and terminological distinctions to cover the entire field and then to think and work consistently in these terms, is surely among our most important immediate tasks.

University of Cambridge

DISCUSSION

Mr. Haas agreed with Mr. Trim that, in identifying phonological units, we do not rely merely or necessarily on recognition of those few distinctive features by which we define the units. (Trubetzkoy spoke of the "subsidiary diacritical power" of non-distinctive features.) At the same time, the features by which we define phonological units, seem to be selected as *capable of carrying the maximum burden* in the identification of those units. More especially they seem to be more resistant to varying contextual conditions.

Mr. Trim replied that attempts to define phonemes as clusters of phonetically identifiable distinctive features involve the selection of certain clues to recognition as primary and the relegation of others to a subsidiary position. The criterion employed might be either a perceptual one, i.e. which features contributes most to recognition, or a distributional one, i.e. which features are present in all allophones of a given phoneme. There is no reason, a priori, to equate the two criteria, though this is often done. The invariant feature, selected on a Highest Common Factor basis, is a logical necessity only as a consequence of the requirement of a minimal phonetic similarity between allophones and its contribution to recognition in any particular context may well be very slight. While power of the concept of correlation to structure the phonemic inventry should not be belittled, "non-distinctive" [i.e. contextually bound] features, too, characterize parallel series. If phonemes are resolved into bundles of "features", these too must be seen as categories recognized on the basis of a multiplicity of clues.