

1. Phonetic and linguistic markers in speech

JOHN LAVER and PETER TRUDGILL

Many disciplines are professionally interested in the analysis of markers of identity in speech, and in the nature of the listener's performance in attributing particular characteristics to the speaker on the basis of such markers. A necessary prerequisite for the development of research in this area, from the standpoint of any discipline, is the provision of an adequate phonetic and linguistic account of the marking phenomena themselves. The aim of this chapter is to indicate some ways in which concepts already available in phonetic and linguistic theory can help to refine the analysis of markers in speech.

The structure of the chapter begins with a brief discussion of the semiotic basis of how speech can serve a marking function as well as fulfilling its role as a vehicle for linguistic communication. Then the phonetic nature of markers in speech is discussed, in terms of voice features, features of tone of voice, and features of linguistic articulations. This is followed by an account of the way in which phonological differences between accents can mark social aspects of speakers' identities. Syntactic markers are then considered. Recent work in the study of discourse and conversational interaction is touched on, and some aspects of lexical markers are examined. The chapter concludes with some tentative hypotheses about the listener's process of attribution which emerge from the phonetic and linguistic discussion.

1. The semiotic basis of marking

Given that the concept of marking deals essentially with the production and perception of communicative and informative signs, it forms part of a general theory of semiotics (cf. Giles, Scherer & Taylor, this volume: ch. 9, 3.1). In recent years, there has been a surge of interest in semiotic theory (cf. Eco 1976), but the history of semiotics as a discipline can be

2 John Laver and Peter Trudgill

traced back to Greek medicine. Morris (1946: 285–7) says that it was first used to refer to the theory of medical symptoms used as signs in the diagnosis and prognosis of disease. The Stoic philosophers then used ‘semiotic’ to mean the general theory of signs. Morris traces the history of semiotics into medieval Europe, through the works of Augustine and Boethius. There, the subject (known as ‘scientia sermocinalis’) developed in the work of figures such as Petrus Hispanus, Abelard, Roger Bacon, Thomas of Erfurt, Sigur of Courtrai and William of Occam. From them, two divergent traditions developed: one led to the work of the British empiricist philosophers like Francis Bacon, Hobbes, Locke, Berkeley, Hume and Bentham; the other, through the work of Leibniz, led to that of modern symbolic logicians such as Boole, Frege, Peano, Russell, Whitehead, Carnap and Tarski. The philosophical basis of an interest in the notion of marking stands on a very long-established foundation, therefore, and there is much of value to be gained from these earlier writings on the subject.

One of the most accessible of the semiotic philosophers, and probably the most relevant for the purposes of this volume, is Charles Saunders Peirce, the American pragmatist philosopher of the late nineteenth century. Six volumes of his collected papers have been edited by Hartshorne & Weiss (1931–5), and Feibleman (1946) provides a useful condensed version of his writings.

Peirce’s definition of ‘semiotic’ was ‘the formal doctrine of signs (where) a sign is something which stands to somebody for something in some respect or capacity’ (Hartshorne & Weiss 1931–5: II, 227–8). Peirce divided the different sorts of signs into three mutually intersecting trichotomies. The second trichotomy is the most relevant here. Feibleman (1946: 90) gives a compressed quotation of this as follows:

‘The second trichotomy of signs consists of the *icon*, a sign which refers to an object by virtue of characters of its own which it possesses whether the object exists or not (2. 247); the *index*, a sign which refers to the object that it denotes by virtue of being really affected by that object (2. 248); and the *symbol*, a sign which refers to the object that it denotes by virtue of a law, usually an association of general ideas, which operates to cause the symbol to be interpreted as referring to that object (2. 249).’

This concept of a symbolic relationship holding between a sign and its referent, where the relationship is conventional and arbitrary, lies at the heart of the linguistic code. But it is Peirce’s notion of an index that is the most interesting in any consideration of how speech identifies the

speaker. He used the term in a number of rather different senses, but the one most useful for marking purposes is his evidential sense illustrated by the quotation above, where an index was said to refer to its object 'by virtue of being really affected by that object'. The orientation of a weathervane would in this usage be evidence for, or an index of, wind-direction; the height of a column of mercury in a thermometer would be an index of heat. 'Index' can thus be equated with 'marker', and terms derived from 'index', such as 'to indicate', 'indicative' and 'indication', can be used as technical terms in association with 'to mark' and 'marking'. One modification to Peirce's basic concept needs to be accepted, however. For Peirce, the connection between an index and its object was nonarbitrary; we shall find it useful, nevertheless, to allow an arbitrary connection between some markers in speech and the personal characteristics indicated by them, or taken by the listener to be so indicated.

One writer in particular who has developed a quite explicit typology of markers of identity in speech, based partly on some of Peirce's ideas, is Abercrombie (1967). He used Peirce's term 'index', and describes three classes of indices in speech that reveal personal characteristics of the speaker:

- (a) those that indicate membership of a group (e.g. a regional or a social group);
- (b) those that characterize the individual;
- (c) those that reveal changing states of the speaker (e.g. changing affective states (Abercrombie 1967: 7–9)).

We shall refer to these three types of indices as *group markers*, *individuating markers* and *affective markers* respectively. The typology that Abercrombie suggests will be adopted here, in conjunction with a cross-cutting classification of markers into three other categories:

- (a) those that mark social characteristics, such as regional affiliation, social status, educational status, occupation and social role;
- (b) those that mark physical characteristics, such as age, sex, physique and state of health;
- (c) those that mark psychological characteristics of personality and affective state.

We shall call these three types of markers *social markers*, *physical markers* and *psychological markers* respectively. The two typologies differ chiefly in the way that psychological attributes are handled.

Another linguist who has taken an explicit interest in indexical information in speech is Lyons. His position on the distinction between 'com-

4 John Laver and Peter Trudgill

‘communicative’ and ‘informative’ signs (which is a not uncontroversial distinction) will be followed in this chapter. Lyons suggests that:

‘a signal is communicative . . . if it is intended by the sender to make the receiver aware of something of which he was not previously aware. Whether a signal is communicative or not rests, then, upon the possibility of choice, or selection, on the part of the sender. If the sender cannot but behave in a certain way (i.e. if he cannot choose between alternative kinds of behaviour), then he obviously cannot communicate anything by behaving in that way . . . ‘Communicative’ means “meaningful for the sender”.’ (Lyons 1977: 33)

This is contrasted with ‘informative’ in the sense that:

‘a signal is informative if (regardless of the intentions of the sender) it makes the receiver aware of something of which he was not previously aware. ‘Informative’ therefore means “meaningful to the receiver”.’ (Lyons 1977: 33)

What a speaker says is thus communicative, but his accent, unless he has the possibility of speaking with more than one accent, is informative.

Lyons adopts Abercrombie’s classification of indices, but proposes one further category, based on Abercrombie’s third type: ‘those that reveal changing states of the speaker’. Lyons calls this proposed category a ‘symptom’, recalling the diagnostic use of signs in medicine, but widens the scope of a symptom beyond the indication of affective information:

‘any information in a . . . signal which indicates to the receiver that the sender is in a particular state, whether this be an emotional state (fear, anger, sexual arousal or readiness, etc.), a state of health (suffering from laryngitis, etc.), a state of intoxication, or whatever, [can be] described as symptomatic of that state’ (Lyons 1977: 108).

2. Phonetic and phonological markers

In order to be able to discuss details of the phonetic and phonological phenomena that act as markers in speech, it will be helpful briefly to consider the physical and auditory variables involved in speech production and perception. Many good introductions to general phonetic theory are available (e.g. Pike 1943; Abercrombie 1967; Ladefoged 1971, 1976; O’Connor 1973). Ladefoged (1962) offers a lucid, short introduction to the acoustic basis of speech, and Hardcastle (1976) gives a schematic, readable account of the anatomy and physiology of speech. The reader is referred to these sources for a more detailed exposition, and the com-

ments on speech production and perception that are offered here are intended only as an orientation.

We can initially consider speech production from the point of view of the different muscle systems which make up all the vocal apparatus. The muscle systems exploited in speaking are almost all anatomically interconnected (Laver 1975), so that no muscular action takes place without affecting the activity of many other parts of the vocal apparatus. Each muscular action has to be cooperatively facilitated by all the muscle systems that could potentially counteract the desired effect of its execution. Speaking thus requires the most complex and skilful collaboration between the different muscle systems, whose cooperative actions all have to be precisely and intricately coordinated in time. It is not at all surprising, therefore, that in learning to control such a complex apparatus sufficiently to be able to produce auditorily acceptable imitations of speech patterns heard in one's social environment, speakers should nevertheless develop idiosyncracies of pronunciation that serve to individuate them within their own social group.

The notion of an isolable muscle system is itself something of a fiction. But if we accept the fiction as analytically convenient, then there are seven basic muscle systems whose contributions to speech can be distinguished. These are: the *respiratory system*, which supplies the driving force that pushes air out of the lungs, up to the larynx and the rest of the vocal tract, for speech purposes; the *phonatory system*, which controls the actions of the vocal folds in producing phonation; the *pharyngeal system*, which controls articulatory activity at the bottom end of the vocal tract; the *velopharyngeal system*, which controls the production of nasality; the *lingual system*, which is responsible for the oral articulations underlying most consonant and vowel segments; the *labial system*, which controls the actions of the lip structures; and the *mandibular system*, which controls movements of the jaw.

From an articulatory and aerodynamic point of view, speech is the joint product of the collaborative interaction of all these muscle systems, as reflected in the continually changing configurations of the vocal tract and larynx, and the pattern of airflow from the lungs to the outside air. These articulatory and aerodynamic changes give rise to associated acoustic changes, which can be described in terms of a number of acoustic parameters. The acoustic parameters in turn have a storable relationship to perceptual, auditory parameters. The auditory variables in speech are basically of two sorts: quality features and dynamic features.

The acoustic correlates of features of auditory quality are essentially

6 John Laver and Peter Trudgill

spectral in nature, and include such aspects as formant frequencies and amplitudes, and the frequency and amplitude of aperiodic noise in the spectrum. The acoustic correlates of dynamic auditory features include fundamental frequency as the correlate of *pitch*, intensity as the correlate of *loudness*, and duration as the correlate of *length*. It should be noted, however, that the allocation of fundamental frequency and intensity to the acoustic realization of dynamic auditory features is not always completely valid: pitch 'jitter' and loudness 'shimmer' (that is, aperiodic cycle-to-cycle variability of fundamental frequency or intensity around the mean value) are both heard as contributing to auditory quality, giving a 'rough', 'harsh' auditory texture.

Auditorily, all speech is made up of sounds describable in terms of quality, pitch, loudness and length. All markers in speech thus depend on these variables for their phonetic realization, and the discussion that follows is an attempt to explain the phonetic basis of different types of speaker-characteristics.

There are three different facets of vocal performance to be considered. Each of these facets is subject to a different time-perspective. Firstly, there is the facet of vocal performance that represents the speaker's permanent or quasi-permanent *voice*, by which he is recognizable even when his consonants and vowels are unintelligible, for example, when heard speaking on the other side of a closed door. The other two facets are *tone of voice* and the *phonetic realizations of linguistic units*. The time-perspective of tone of voice is usually medium-term, and that of linguistic articulations very short-term.

Because voice features are by definition long-term, they lie quite outside any possibility of signalling linguistic meaning, so it is appropriate to refer to such voice features as *extralinguistic*. Since they are not normally consciously manipulated by the speaker, voice features are informative but not communicative. The medium-term features that make up tone of voice, and which have the function of signalling affective information, have a rather closer resemblance in some ways to the short-term use of the vocal apparatus for signalling linguistic meaning, and such features are therefore often referred to as *paralinguistic*. They are 'para'linguistic in the sense that they form a communicative code subject to cultural convention for its interpretation; paralinguistic features are not fully linguistic in the sense that they lack the possibility of signalling meaning through sequential arrangement into structures, which is a criterial property of linguistic communication.

Neither extralinguistic nor paralinguistic features are irrelevant to

directly linguistic interests, since they constitute a background against which the linguistic articulations can achieve their perceptual prominence. Strictly, each of the three types of vocal feature, extralinguistic, paralinguistic and linguistic, acts as a perceptual ground for the figures of the other two types of figure.

Each of these categories of vocal behaviour will now be discussed in more phonetic detail. A summary of the relationship between these vocal variables and their marking functions is given in table 1.

2.1. Extralinguistic voice features

Long-term speaker-characterizing voice features are of two different sorts. One type of voice feature arises from anatomical differences between speakers. The second type is the product of the way in which the individual speaker habitually 'sets' his vocal apparatus for speaking. Unlike this second type, which will be discussed in a moment, the first type of feature is by definition outside any possibility of control by the speaker. It includes anatomical influences on aspects of voice quality and of voice dynamics.

Anatomical influences on voice quality are due to factors such as basic vocal tract length, dimensions of lips, tongue, nasal cavity, pharynx and jaw, dental characteristics, and geometry of laryngeal structures (Abercrombie 1967: 92). These anatomical factors impose limits on the range of spectral effects (in terms of formant frequency and amplitude ranges, and on the distribution of aperiodic noise through the spectrum) that the speaker can potentially control acoustically.

Anatomical influences on voice dynamics are due to factors such as the dimensions and mass of the vocal folds, and respiratory volume. These influence pitch and loudness ranges, by imposing limits on the ranges of fundamental frequency and amplitude that the speaker can produce.

Listeners' judgements of physical attributes, based on the product of such anatomically derived features, are amongst the most accurate conclusions drawn. This is precisely because they are based on invariant, involuntary aspects of a speaker's vocal performance. Physique, age and sex are all judged with a fair accuracy, and interesting information about a speaker's medical condition is also sometimes accurately inferred.

Physique and height are probably judged accurately because of the good correlation that seems to exist between these factors and the dimensions of the speaker's vocal apparatus. A tall, well-built man will tend to

Table 1. The relationship between vocal variables and their marking functions

Signalling function	informative		informative and communicative	
	extralinguistic voice characteristics	paralinguistic 'tone of voice'	phonetic realizations of linguistic units	
Temporal perspective	permanent	quasi-permanent	short-term	
Vocal variables	vocal features deriving from anatomical differences between individuals influencing both quality and dynamic aspects	voice settings, i.e. habitual muscular adjustments of the vocal apparatus, including voice quality settings and voice dynamic settings	'tone of voice' achieved by temporary use of voice settings, including paralinguistic quality settings and paralinguistic dynamic settings	momentary articulatory realizations of phonological units, including short-term manipulations of phonetic quality features and short-term manipulations of phonetic dynamic features
Marking function	physical markers	social and psychological markers		
Potential controllability	uncontrollable, therefore unlearnable	under potential muscular control, therefore learnable and imitable		

have a long vocal tract and large vocal folds. His voice quality will reflect the length of his vocal tract by having correspondingly low ranges of formant frequencies, and his voice dynamic features will indicate the dimensions and mass of his vocal folds by a correspondingly low range of fundamental frequency. His large respiratory volume will be reflected in a powerful loudness range. If we then hear such a voice over the telephone, we normally have a confident expectation that the speaker will turn out to be a large, strong male. In general, our expectations are fulfilled, within a reasonable margin of error. Bonaventura (1935) gave subjects pictures and voices to match, and found that fair accuracy was achieved: in terms of Kretschmerian body-types (Kretschmer 1925), judgements of *pyknic* types were most accurate, accuracy was less for *leptosomic* types, and least for *athletic* types. Moses (1940, 1941) gives general support to this, and Fay & Middleton (1940a) report a more detailed finding: they found that in judging body-types from voices transmitted over a public address system, the results were 22 per cent above chance for *pyknic* types, 20 per cent for *leptosomes*, but only 1 per cent above chance for *athletic* types. Lass, Beverly, Nicosia & Simpson (1978) report that listeners typically judge weight to within 3–4 lbs (though overestimating the weight of males and underestimating that of females), and that they judge height to within 1.5 inches (though underestimating the height of both males and females). There is one class of voices where the general correlation does not apply, but where listeners nevertheless seem to be able to reach successful conclusions about the physical attributes. That is where the formant ranges of the voice are radically discrepant with the fundamental frequency, as in particular types of dwarfism (Vuorenkoski, Tjernlund & Perheentupa 1972; Weinberg & Zlatin 1970). In these cases, the dimensions of the vocal folds are smaller than their general correlation with vocal tract length would lead one to expect.

Exceptions to the general rule of our ability as listeners to attach a particular size and physique to a given voice are sufficiently rare to take us aback when they occur.

Age is judged accurately (Dordain, Chevrie-Muller & Grémy 1967; Hollien & Shipp 1972; Mysak 1959; Ptacek, Sander, Maloney & Roe Jackson 1966; Shipp & Hollien 1969). Voice quality features probably play their part in marking this characteristic, but voice dynamic features are likely to be the more primary cues. Age is marked by pitch in both males and females: Hollien & Shipp (1972) show a progressive lowering of mean pitch with age for males from 20 up to 40, then a rise from age 60 through the 80s. Mysak (1959) also showed this rise in mean pitch from the 50s

10 John Laver and Peter Trudgill

upwards. Dordain *et al.* (1967) report a drop in mean pitch for older women, but a rise with extreme age. Ptacek *et al.* (1966) also report a reduced pitch range with extreme age.

Features of auditory quality can signal aspects of the age of a speaker. These include the quality associated with the 'breaking' voice of puberty, and the quality of extreme old age. Vocal indications of puberty, referred to in clinical literature as 'vocal mutation', often include whispery voice. Luchsinger & Arnold (1965: 132) write that 'In addition to the lowering of the average speaking pitch, the voice is frequently husky during mutation, or it may sound weak.' The senescent voice of extreme old age derives from a complex of endocrinal, anatomical and physiological changes. The mucal fluid supply often becomes disturbed, either greatly increasing or decreasing, tissues become increasingly less elastic, and cartilages become calcified and ossified (Fyfe & Naylor 1958; Luchsinger & Arnold 1965; Meader & Muyskens 1962; Terracol & Azémar 1949). Meader & Muyskens (1962: 77) comment that 'Since the rigidity of tissue is one determination of its resonating qualities, the gradual deposition of lime in . . . cartilages (replacing them by bone) helps to explain the shrill voice and thin voice (deficient in harmonics) of age.' Because muscles atrophy, the glottis of old speakers often has a bowed appearance (Luchsinger & Arnold 1965: 136; Tarneaud 1941); this means that, to achieve phonation, greater effort has to be exerted to bring the vocal folds together, and a rather harsh voice is often the result. When this is combined with inefficient phonation because of an excess of mucus, the type of voice that results is a harsh whispery voice, as suggested by the following comment from Luchsinger & Arnold (1965: 136): 'Tracheal and laryngeal mucous secretions are increased, sometimes on an allergic basis. Together with a tendency to chronic bronchitis, this over-secretion of mucus produces the hacking, coughing, throat-clearing, or "moist" hoarseness of the old man.' In old age, fatty tissue can build up in the ventricles in the sides of the upper larynx (Ferreri 1959), and the ventricular folds above the ventricles can shrink towards the sides of the larynx, giving a wider entrance to the ventricles (Luchsinger & Arnold 1965: 136). All these factors can contribute significantly to the fine detail of the auditory quality of the phonation being produced. Luchsinger & Arnold also mention work by Braus (1924), who 'pointed out that the larynx and the entire respiratory and digestive tract are in a lowered position with senility' (1965: 137), because of the loss of elasticity of the muscular and ligamental structures from which these organs are suspended. Any such elongation of the vocal tract tends to lower the ranges of the formant